Journal of Applied Research in Memory and Cognition xxx (2019) xxx–xxx

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Identification Performance from Multiple Lineups: Should Eyewitnesses Who Pick Fillers Be Burned?☆

Laura Smalarz*
Williams College, USA

Nate Kornell
Williams College, USA

Kalif E. Vaughn
Northern Kentucky University, USA

Matthew A. Palmer
University of Tasmania, Australia

Over the course of a criminal investigation, eyewitnesses are sometimes shown multiple lineups in an attempt to identify the culprit, yet little research has examined eyewitness identification performance from multiple lineups. In two experiments, we examined eyewitness identification accuracy among witnesses who made an inaccurate identification from an initial lineup, correctly rejected an initial lineup, or saw no initial lineup. Consistent with the legal practice of "burning" eyewitnesses who pick fillers, witnesses who made an inaccurate identification from an initial lineup provided subsequent identification evidence that had little diagnostic value and reflected poor memory performance. Critically, these eyewitnesses' initial-identification confidence did not predict their subsequent identification accuracy, thus precluding the identification of witnesses who could provide diagnostic evidence in a subsequent lineup. Eyewitnesses who correctly rejected the initial lineup performed similarly to eyewitnesses who saw only one lineup, and initial-rejection confidence was associated with subsequent identification accuracy under some conditions.

General Audience Summary

In as many as 74% of real-world lineups, eyewitnesses fail to identify the police suspect (Wells, Steblay, & Dysart, 2015). Eyewitnesses in these cases may be asked to view a second lineup, yet little research has examined eyewitness identification performance from multiple lineups. In two experiments, we examined identification performance on a second lineup following exposure to an initial lineup. Participant-eyewitnesses watched a video of a mock-crime and then either viewed an initial lineup that did not contain the culprit or viewed no initial lineup. Using this procedure, we sorted eyewitnesses into three groups: Eyewitnesses who made an inaccurate identification from the initial lineup; eyewitnesses who correctly indicated that the culprit was not in the initial lineup; and eyewitnesses who saw no initial lineup. All eyewitnesses then attempted an identification from a

Author Note.
Nate Kornell, Williams College, Williamstown, MA, USA.
Kalif E. Vaughn, Northern Kentucky University, Highland Heights, Kentucky, USA.
Matthew A. Palmer, University of Tasmania, Hobart, Tasmania, Australia.

Materials and data for this research are available at osf.io/k529y.

* Correspondence concerning this article should be addressed to Laura Smalarz, Department of Psychology, Williams College, 25 Stetson Court, Williamstown, MA, 01267, USA. Contact: laura.smalarz@williams.edu.

Please cite this article in press as: Smalarz, L., et al. Identification Performance from Multiple Lineups: Should Eyewitnesses Who Pick Fillers Be Burned? *Journal of Applied Research in Memory and Cognition* (2019), https://doi.org/10.1016/j.jarmac.2019.03.001

final lineup that contained the culprit (Experiment 1) or manipulated the presence of the culprit (Experiment 2). Eyewitnesses' confidence was collected following all lineup decisions. In both experiments, eyewitnesses who incorrectly identified someone from the initial lineup provided less reliable identification evidence from a subsequent lineup compared to witnesses who saw no initial lineup, whereas the subsequent identification performance of witnesses who rejected the initial lineup did not differ from that of witnesses who saw no initial lineup. Moreover, the confidence with which eyewitnesses made their initial inaccurate identification did not predict their subsequent identification accuracy, thereby precluding the possibility of identifying which eyewitnesses would be able to provide reliable identification evidence from a subsequent lineup. These findings are consistent with the legal practice of "burning" (i.e., ceasing to rely on) eyewitnesses who pick fillers from lineups and highlight the importance of good record-keeping at the time of an initial identification procedure.

Keywords: Eyewitness identification, Eyewitness memory, Eyewitness confidence, Multiple lineups, Diagnosticity, Discriminability

In as many as 74 percent of real-world police lineups, eyewitnesses do not pick the person under suspicion and instead identify a known-innocent lineup filler or reject the lineup (Wells, Steblay, & Dysart, 2015). It is unlikely that police simply abandon their investigations in these cases. If the police pursue a new potential suspect, they may wish to present this person to the eyewitness in a subsequent identification procedure. Yet little is known about eyewitness identification performance from multiple lineups. The goal of the current research was to investigate this issue by examining identification performance following exposure to an initial lineup that does not contain the culprit. We compared identification performance across three groups: witnesses who made an inaccurate identification from an initial lineup, witnesses who correctly rejected an initial lineup, and witnesses who were not exposed to an initial lineup.

In the legal system, eyewitnesses who make an initial inaccurate identification are considered to be "burned" for purposes of attempting a future identification (Steblay, Dysart, & Wells, 2011; Wells et al., 2015). This can happen when the eyewitness picks a police suspect who is later discovered to be innocent, but it is more common for the eyewitness to pick a known-innocent lineup filler. The logic underlying the practice of "burning" eyewitnesses who pick fillers is that these eyewitnesses have impugned their reliability by committing a known error. By contrast, witnesses who reject an initial lineup are still considered reliable and can be used in a subsequent identification procedure. To date, however, surprisingly little research has tested the presumption that witnesses who pick fillers are less reliable in subsequent identification attempts than are witnesses who pick no one.

Early on in the eyewitness-identification literature, Wells (1984) proposed that eyewitnesses who make an inaccurate identification from an initial lineup are more prone to error on a subsequent lineup than are eyewitnesses who reject an initial lineup. This rationale was the primary justification for Wells' development of the blank lineup procedure, in which police show eyewitnesses an initial lineup that does not contain the culprit in order to "weed out" unreliable eyewitnesses who are prone to making mistaken identifications. In an initial test of the blank lineup procedure, Wells found that witnesses who made identifications from a blank lineup were less accurate in their identifications from a subsequent lineup than were witnesses who rejected the blank lineup, a finding later replicated by Palmer, Brewer, and Weber (2012).

Although the findings of Wells (1984) and Palmer et al. (2012) are consistent with the idea of "burning" witnesses who make inaccurate identifications from an initial lineup, this issue warrants closer investigation for two reasons. First, neither of these studies used conditions that realistically map onto what is likely to happen in actual police investigations. Perhaps most importantly, neither Wells nor Palmer et al. told witnesses anything after they made an initial identification from the blank lineup. In other words, witnesses who picked from the first lineup were not informed that their initial identification was inaccurate—a methodological feature that might have led eyewitnesses who chose from the initial lineup to remain committed to their choice and behave differently in the subsequent lineup (e.g., Gorenstein & Ellsworth, 1980; Hinz & Pezdek, 2001; Steblay, Tix, & Benson, 2013). In real cases involving multiple lineups, witnesses who pick someone from an initial lineup will virtually always learn that their initial identification was inaccurate; after all, why else would the police be showing them a new lineup? In the current research, therefore, we informed witnesses who made an initial inaccurate identification that their identification was inaccurate.

Second, some empirical evidence suggests that witnesses who make an inaccurate identification from an initial lineup might have better memories of the culprit than has typically been assumed. In Palmer et al.'s (2012) second experiment investigating the blank lineup procedure, witnesses chose from or rejected an initial blank lineup and then were forced to choose from a second culprit-present lineup. Although accuracy was better on the second lineup for initial non-choosers than for initial choosers, this difference was modest rather than marked: Witnesses who rejected the blank lineup were only 1.25 times more likely than were witnesses who made an identification from the blank lineup to accurately identify the culprit in the subsequent lineup. Moreover, diagnosticity ratios from both Wells (1984) and Palmer et al. indicated that even witnesses who made an initial inaccurate identification provided subsequent identification evidence that had diagnostic value (suspect identification diagnosticity ratios of 3.60 and 4.15, respectively). Hence, extant research suggests that the

Please cite this article in press as: Smalarz, L., et al. Identification Performance from Multiple Lineups: Should Eyewitnesses Who Pick Fillers Be Burned? *Journal of Applied Research in Memory and Cognition* (2019), https://doi.org/10.1016/j.jarmac.2019.03.001

practice of indiscriminately burning all eyewitnesses who pick fillers might lead to a loss of diagnostic subsequent identification evidence.

One potential strategy for identifying which eyewitnesses are most likely to provide diagnostic evidence from a subsequent identification procedure is to examine the confidence with which the eyewitnesses make their initial identification response. It is widely recommended that police collect confidence statements from eyewitnesses immediately at the time of an identification, and it is increasingly being suggested that confidence is collected not only from identifying eyewitnesses but also from eyewitnesses who reject the lineup (Wells et al., in prep). In cases involving multiple lineups, eyewitnesses' confidence in their initial lineup decision might prove informative regarding the likely reliability of identification evidence obtained in a subsequent lineup; for example, eyewitnesses who made an initial inaccurate identification with high confidence might be especially inaccurate on a second lineup. Somewhat consistent with this idea, Palmer et al. (2012) found evidence of differential diagnosticity of high- and low-confidence decisions for eyewitnesses who rejected an initial blank lineup but not for eyewitnesses who made an initial inaccurate identification. Recall, however, that witnesses in Palmer et al. who made an initial inaccurate identification were not informed prior to viewing the second lineup that they were inaccurate, which could have influenced subsequent identification performance. The current research examined the extent to which confidence in an initial decision predicts subsequent identification performance under more realistic conditions in which identifying witnesses are no longer committed to their initial choice upon viewing the subsequent lineup.

Research Overview

We conducted two experiments that compared subsequent identification performance among witnesses who made an inaccurate identification from an initial lineup, witnesses who correctly rejected an initial lineup, and witnesses who saw no initial lineup. In both experiments, witnesses were randomly assigned to either view an initial lineup that did not contain the culprit or to view no initial lineup. In Experiment 1, subsequent identification performance was measured using a forced-choice culprit-present lineup task. In Experiment 2, subsequent identification performance was measured using a free-choice culprit-present orculprit-absent lineup.

Experiment 1

Method

Participants. Participants were 210 Amazon Mechanical Turk workers who received $1.50 for participating. We restricted recruitment to participants in the United States who had at least a 95% approval rate (Peer, Vosgerau, & Acquisti, 2013). We excluded participants who failed to finish the study ($n = 21$), participants who restarted the study midway through ($n = 2$), participants who failed to respond to the initial lineup ($n = 2$), and participants who reported experiencing serious technical problems during the study ($n = 3$). A total of 182 participants remained, all of whom spoke English fluently. Participants' ages ranged from 19 to 69 ($M = 34.5$, $SD = 10.6$) and 49.5% identified as female, 50.5% as male. Participants' highest education levels were reported as follows: 0.5% completed some high school, 13.2% graduated from high school, 28.0% completed some college, 10.4% obtained an associate's degree, 35.2% obtained a bachelor's degree, and 11.5% obtained a graduate degree. Two participants did not report their education level.

Design. Experiment 1 used a two-cell, between-subjects design in which participants viewed an initial culprit-absent lineup or no initial lineup before the final memory test. The final memory test consisted of a culprit-present lineup that included new fillers who had not appeared in the initial culprit-absent lineup. We collected data from approximately twice as many participants in the initial culprit-absent lineup condition ($n = 122$) as in the no-initial lineup condition ($n = 60$) in order to ensure adequate sample sizes of witnesses who made an identification from the initial lineup and those who rejected the initial lineup.

Materials and measures. The materials were adapted from Smalarz and Wells (2014). The stimulus video depicted a degraded version of an airport scene in which a suspicious individual switched his suitcase with another passenger's suitcase at the check-in line. After switching the suitcase, the culprit exited the airport. The video lasted 1 min 28 s and showed multiple views of the culprit's face. The initial culprit-absent lineup included six individuals who fit the general description of the culprit; identification responses to the initial lineup are shown in the Results. The final culprit-present lineup included a photo of the culprit along with five individuals who fit the general description of the culprit and who had not appeared in the initial culprit-absent lineup. Both lineups are available at the following URL: osf.io/k529y.

Our dependent measures included participants' responses to the initial lineup (inaccurate identification, lineup rejection) and final lineup (culprit identification, filler identification, lineup rejection) and their confidence in each of those decisions, reported on a scale from 0 (*no confidence*) to 100 (*total confidence*). We also collected forced-choice responses from witnesses who rejected the final culprit-present lineup. After their rejection decision, they were asked, "If you were forced to choose someone, which person would you pick?" We collected demographic information pertaining to age, gender, education level, fluency in English, and country of residence.

Procedure. Participants were recruited online to participate in a study on the impressions people form of others. They began by providing demographic information and informed consent. They were then instructed that they would view a short video and were asked to pay attention and watch for any suspicious behavior. Participants were informed that the video quality may be degraded and that this was normal. After watching the video, participants learned that the suitcase that the man switched had a bomb in it and they were told that they were now eyewitnesses to this crime. All participants then completed a five-minute filler task (the game of Tetris).

Following the filler task, participants in the initial-lineup condition were shown the initial culprit-absent lineup. Participants were informed that the actual culprit may or may not be in the

Please cite this article in press as: Smalarz, L., et al. Identification Performance from Multiple Lineups: Should Eyewitnesses Who Pick Fillers Be Burned? *Journal of Applied Research in Memory and Cognition* (2019), https://doi.org/10.1016/j.jarmac.2019.03.001

lineup and they were given the option to indicate that the man who switched the bags was not present in the lineup. After making an identification decision, participants indicated on a scale from 0 to 100 how confident they were in their decision. In order to prevent commitment effects from influencing subsequent responding, participants who picked someone from the initial lineup were told, "Sorry, the person you chose was not the person who switched the bags." Participants who rejected the lineup were not told anything about their identification decision. Participants in the initial-lineup condition then completed a second five-minute filler task. Participants in the no-initial lineup condition completed the first and second five-minute filler tasks consecutively.

All participants then viewed the final culprit-present lineup. Participants were informed that the actual culprit may or may not be in the lineup and they were given the option to indicate that the man who switched the bags was not present in the lineup. Following their identification decision, participants indicated how confident they were in their decision. Participants who provided a "Not there" response were then asked to provide a forced-choice response. Finally, participants indicated whether there were any problems in the experiment.

Results

Identification behavior in the initial culprit-absent lineup. Recall that we collected data from twice as many participants in the initial lineup condition ($n = 122$) compared to the no initial lineup condition ($n = 60$) in an effort to obtain sufficient samples of identifying and non-identifying eyewitnesses. Of the 122 participants who saw the initial lineup, 65 made an identification and 57 rejected the lineup (see Table 1 for a full breakdown). Hence, participants' responses were well-distributed across the two choices.

Identification behavior in the final culprit-present lineup. We compared final culprit-present identification performance among witnesses who made an identification from the initial lineup, witnesses who rejected the initial lineup, and witnesses who saw no initial lineup using a Chi-square test of independence on witnesses' identification behavior from the final culprit-present lineup (culprit identification, filler identification, lineup rejection; see Table 2). Witnesses' initial lineup group significantly predicted their identification behavior in the final culprit-present lineup, $\chi^2(4) = 12.49$, $p = .014$.

Table 1
Response Frequencies for the Initial Culprit-Absent Lineup

Lineup choice	Percent	Frequency
1	0.8	1
2	24.6	30
3	5.7	7
4	9.8	12
5	4.9	6
6	7.4	9
Not there	46.7	57
Total	100.0	122

Note. A calculation of Tredoux' (1998) E based on the distribution of affirmative responses indicated that the effective size of the initial culprit-absent lineup was 3.49, 95% CI [2.73, 4.84].

Table 2
Final Lineup Performance

Initial lineup group	Culprit	Filler	Reject
No initial lineup	40.0 (24)$_a$	31.7 (19)$_a$	28.3 (17)$_{ab}$
Identified from initial lineup	29.2 (19)$_a$	55.4 (36)$_b$	15.4 (10)$_a$
Rejected initial lineup	31.6 (18)$_a$	31.6 (18)$_a$	36.8 (21)$_b$

Note. Percent (frequency). Proportions with differing subscripts within columns are significantly different at $p < .05$.

Follow-up analyses indicated that the final lineup behavior of witnesses who rejected the initial lineup did not differ significantly from that of witnesses who did not view the initial lineup, $\chi^2(2) = 1.23$, $p = .541$. However, the final lineup behavior of witnesses who made an identification from the initial lineup differed significantly from that of witnesses who did not view the initial lineup, $\chi^2(2) = 7.46$, $p = .024$, and from witnesses who rejected the initial lineup, $\chi^2(2) = 9.45$, $p = .009$. Specifically, witnesses who made an identification from the initial lineup were significantly more likely to identify a filler from the final culprit-present lineup (55.4%) than were witnesses who saw no initial lineup (31.7%), $\chi^2(1) = 7.12$, $p = .008$, OR = 2.68, 95% CI [1.29, 5.57], or witnesses who rejected the initial lineup (31.6%), $\chi^2(1) = 6.98$, $p = .008$, OR = 2.69, 95% CI [1.28, 5.65]. Witnesses who made a identification from the initial lineup were also significantly less likely to reject the final culprit-present lineup (15.4%) than were witnesses who rejected the initial lineup (36.8%), $\chi^2(1) = 7.38$, $p = .007$, OR = 0.31, 95% CI [0.13, 0.74]. Rejection rates did not differ significantly between witnesses who made an identification from the initial lineup and those who saw no initial lineup, $\chi^2(1) = 3.09$, $p = .079$. The accurate culprit identification rate did not differ significantly between witnesses who made an identification from the initial lineup and that of the other two groups, χ^2s$(1) \leq 1.60$, ps $\geq .205$.

We also examined the ratio of correct identifications to filler identifications, which provides a measure of discriminability for one important aspect of eyewitness identification decisions: The ability to recognize the culprit, given that a positive identification is made (see Palmer, Brewer, & Weber, 2010). This ratio was significantly lower for witnesses who identified someone from the initial lineup (.35) than for witnesses who saw no initial lineup (.56), $\chi^2(1) = 4.43$, $p = .035$, but not significantly lower than that of witnesses who rejected the initial lineup (.50), $\chi^2(1) = 2.15$, $p = .142$. There was no significant difference between witnesses who rejected the initial lineup and witnesses who did not view an initial lineup, $\chi^2(1) = 0.27$, $p = .606$.

The comparisons of identification responses reported above suggest that there was no difference in identification performance on the final lineup between witnesses who rejected the initial lineup and those who saw no initial lineup. We conducted Bayesian contingency table analyses using JASP (Jamil et al., 2017; JASP Team, 2018) to explore whether these results are best interpreted as a true null effect or an inconclusive result, potentially due to lack of power. The results of a 2 (Initial Lineup

Please cite this article in press as: Smalarz, L., et al. Identification Performance from Multiple Lineups: Should Eyewitnesses Who Pick Fillers Be Burned? *Journal of Applied Research in Memory and Cognition* (2019), https://doi.org/10.1016/j.jarmac.2019.03.001

Group: rejected initial lineup, no initial lineup) × 3 (Identification Response: suspect identification, filler identification, lineup rejection) analysis supported the former interpretation: The observed data are seven times more likely under the null hypothesis (no difference between the groups) than the alternative hypothesis (a difference in identification responses between the groups), $BF_{01} = 7.04$, $N = 117$. Hence, the results are best interpreted as a true null difference in identification responses between these groups.

Witnesses who rejected the final culprit-present lineup were asked to indicate which lineup member they would choose if they were forced to make a choice. A total of 48 witnesses rejected the final culprit-present lineup. We assessed whether the three initial-lineup groups differed in their *overall* ability to identify the culprit from the final lineup by combining the identification responses of witnesses who chose from the final lineup voluntarily and those who were forced to choose from the final lineup to create a measure of overall culprit identification accuracy. The results yielded a significant effect of initial lineup group on overall culprit identification accuracy, $\chi^2(2) = 6.74$, $p = .034$. Specifically, witnesses who made an identification from the initial lineup were significantly less likely to accurately identify the culprit from the final lineup (30.8%) than were witnesses who saw no initial lineup (53.3%), $\chi^2(1) = 6.54$, $p = .011$, OR = 0.39, 95% CI [0.19, 0.81]. Overall culprit identification accuracy did not differ significantly between witnesses who rejected the initial lineup (45.6%) and those who saw no initial lineup or between witnesses who rejected the initial lineup and those who made an identification from the initial lineup, χ^2s$(1) \leq 2.85$, $ps \geq .091$.

Finally, we examined the relation between confidence and accuracy in the final lineup separately for each of the three groups of witnesses. Final identification decision confidence was significantly correlated with accuracy from the final lineup for witnesses who made an identification from the initial lineup, $r(62) = .32$, $p = .010$, witnesses who rejected the initial lineup, $r(55) = .33$, $p = .013$, and witnesses who saw no initial lineup, $r(58) = .30$, $p = .022$.

Initial lineup decision confidence as a predictor of subsequent performance. We examined whether the confidence with which witnesses made their decision from the initial culprit-absent lineup predicted their performance in the subsequent culprit-present lineup. We predicted that witnesses who make an initial inaccurate identification with high confidence perform more poorly on a subsequent lineup than do witnesses who make an initial inaccurate identification with low confidence. Conversely, we predicted that witnesses who correctly reject an initial lineup with high confidence perform better on the subsequent lineup than do witnesses who correctly reject an initial lineup with low confidence (Palmer et al., 2012).

We used the combined identification accuracy measure for voluntary and forced choosers, such that all witnesses were coded as having made either an accurate culprit identification or an inaccurate filler identification from the final lineup. We first conducted correlation analyses between witnesses' initial confidence and their subsequent culprit identification accuracy separately for witnesses who made an identification from the initial lineup and witnesses who rejected the initial lineup. Among witnesses who made an identification from the initial lineup, confidence in the initial identification did not significantly predict subsequent culprit identification accuracy, $r(63) = .20$, $p = .102$. Among witnesses who rejected the initial lineup, however, confidence in the initial rejection decision was significantly positively associated with subsequent culprit identification accuracy, such that higher confidence in the initial lineup rejection corresponded to an increased likelihood of making a subsequent accurate culprit identification, $r(55) = .27$, $p = .042$.

Given that witnesses' confidence in their rejection of the initial culprit-absent lineup was significantly correlated with subsequent culprit identification accuracy, we examined whether subsequent accuracy varied significantly as a function of whether witnesses expressed high or low confidence in their initial rejection decision. We median-split the data by witness confidence, with witnesses who rejected the initial lineup with confidence of 75% or greater representing the upper half ($n = 28$) and witnesses who rejected the initial lineup with confidence of less than 75% representing the lower half ($n = 29$). Although high-confidence rejectors (57.1%) tended to perform better on the final lineup than did low-confidence rejectors (34.5%), a Chi-square analysis comparing these two groups was not significant, $\chi^2(1) = 2.95$, $p = .086$. An additional analysis that compared culprit identification accuracy rates across witnesses who rejected the lineup with confidence of 75% or greater and witnesses who saw no initial lineup (53.3%) likewise did not produce a significant difference, $\chi^2(1) = 0.11$, $p = .738$.

A critic might argue that our median-split analysis did not provide a good test of whether eyewitnesses who reject an initial lineup with high versus low confidence differ in their subsequent identification performance because the low-confidence group could include witnesses with confidence levels as high as 74%. Accordingly, we conducted another analysis in which we compared culprit identification accuracy rates from the final lineup among eyewitnesses who rejected the initial lineup with confidence levels in the top and bottom quartiles. The top quartile included witnesses who rejected the initial lineup with confidence of 90% or greater ($n = 14$) and the bottom quartile included witnesses who rejected the initial lineup with confidence of 55% or lower ($n = 14$). A Chi-square analysis indicated that eyewitnesses who rejected the initial lineup with confidence of 90% or greater performed significantly better on the final lineup (64.3%) than did eyewitnesses who rejected the initial lineup with confidence of 55% or lower (14.3%), $\chi^2(1) = 7.34$, $p = .007$, OR = 10.80, 95% CI [1.69, 68.94]. However, eyewitnesses who rejected the initial lineup with confidence of 90% or greater did not perform significantly better on the final lineup compared to eyewitnesses who saw no initial lineup (53.3%), $\chi^2(1) = .55$, $p = .458$.

Discussion

Experiment 1 examined subsequent identification performance among witnesses who saw no initial lineup and witnesses who made an inaccurate identification or a rejection decision from an initial lineup. The subsequent identification performance of witnesses who rejected the initial lineup did not differ

Please cite this article in press as: Smalarz, L., et al. Identification Performance from Multiple Lineups: Should Eyewitnesses Who Pick Fillers Be Burned? *Journal of Applied Research in Memory and Cognition* (2019), https://doi.org/10.1016/j.jarmac.2019.03.001

significantly from that of witnesses who saw no initial lineup on any measure of performance; Bayesian analyses indicated that these results are best interpreted as null effects rather than inconclusive (or potentially underpowered) results. By contrast, witnesses who made an identification from the initial lineup had a lower ratio of correct identifications to filler identifications and were significantly less able to accurately identify the culprit (i.e., for combined free and forced-choice responses) compared to witnesses who saw no initial lineup. Differences in subsequent identification performance among witnesses who made an initial identification and witnesses who rejected the initial lineup were somewhat less pronounced, emerging only in terms of relative rates of rejection decisions versus filler identifications. Hence, the primary difference between these two groups of witnesses may simply be in their willingness to choose from a subsequent lineup.

Contrary to our hypothesis, eyewitnesses' confidence in their inaccurate identification from the initial lineup did not reliably predict identification accuracy in the subsequent lineup. And although the confidence expressed by witnesses who rejected the initial lineup was positively correlated with subsequent identification accuracy, it was only when we compared subsequent identification performance among witnesses who rejected the initial lineup with confidence levels in the upper and lower quartiles that we observed superior performance among high-confidence rejectors than among low-confidence rejectors. Results of a median-split comparison did not yield a significant difference in subsequent identification accuracy, perhaps because the effect size for that comparison was smaller. In Experiment 2, we increased our sample size to achieve sufficient power to detect a significant difference of the size observed in the median-split analysis with power of at least 80%.

In sum, the findings from Experiment 1 suggest that witnesses who make an identification from an initial culprit-absent lineup provide less reliable identification evidence from a subsequent lineup than do witnesses who see no initial lineup. By contrast, witnesses who reject an initial lineup retain an ability to provide reliable identification evidence from a subsequent lineup. Because we did not include a culprit-absent lineup condition in the final lineup test, however, the results from Experiment 1 are somewhat ambiguous with respect to the relative contribution of response bias effects and discriminability effects on subsequent lineup performance across these three groups of witnesses. Accordingly, we next conducted a more robust test of the effects of exposure to an initial lineup on subsequent identification performance by including both a culprit-present and a culprit-absent condition in the final lineup test.

Experiment 2

Method

Power analysis. In Experiment 1, eyewitness confidence in an initial rejection decision was significantly correlated with identification performance from a subsequent lineup, but a median-split analysis comparing the final identification performance of eyewitnesses who rejected the initial lineup with high and low confidence did not yield a significant effect. To address the possibility that this non-significant result was because of a lack of statistical power, we used G*Power (Erdfelder, Faul, & Buchner, 1996) to calculate the necessary sample size to detect a difference in final identification accuracy of the size observed in Experiment 1 with at least 80% power. The analysis indicated that 76 participants were required in each condition. We aimed to collect 100 participants per condition in Experiment 2 to achieve a sufficient sample size after excluding participants who did not follow instructions, failed to complete the study, or experienced technical problems.

Participants. Participants were 601 Amazon Mechanical Turk workers who received $1.80 for participating. We restricted recruitment to participants in the United States who had at least a 95% approval rate (Peer et al., 2013). We excluded participants who failed to finish the study ($n = 70$), participants who restarted the study midway through ($n = 1$), and participants who reported experiencing serious technical problems during the study ($n = 9$). A total of 521 participants remained (approximately 87 per cell), all but one of whom reported speaking English fluently. Participants' ages ranged from 18 to 72 ($M = 35.2$, $SD = 10.7$) and 58.3% identified as female, 41.7% as male. One participant did not report her age. Participants' highest education levels were reported as follows: 0.4% completed some high school, 9.8% graduated from high school, 27.3% completed some college, 12.3% obtained an associate's degree, 36.1% obtained a bachelor's degree, and 13.1% obtained a graduate degree. Six participants did not report their education level.

Design. The experiment used a 2 (Initial Lineup: no initial lineup vs. initial culprit-absent lineup) × 2 (Final Lineup: culprit-present lineup vs. culprit-absent lineup) between-subjects design. As in Experiment 1, we collected data from twice as many participants in the initial culprit-absent lineup condition ($n = 352$) than in the no initial lineup condition ($n = 169$) in order to ensure adequate sample sizes of witnesses who made an identification from the initial lineup and those who rejected the initial lineup.

Materials and procedure. The stimulus video and initial culprit-absent lineup were the same as in Experiment 1. We created two six-person final lineups for Experiment 2, one of which contained the culprit and the other of which contained a designated innocent suspect. We chose as the innocent suspect a lineup member from the culprit-present lineup in Experiment 1 who competed with the culprit for identifications (16.4% of identifying witnesses chose him in Experiment 1). For the new culprit-present lineup, this lineup member's photo was replaced with the photo of a lineup member who was a weaker competitor for identifications, as established by similarity-rating data from Smalarz and Wells (2014). Hence, the final culprit-present lineup in Experiment 2 was a slightly easier task than it was in Experiment 1 (because a strong competitor had been replaced by a weak competitor), and the final culprit-absent lineup included an innocent suspect who was a plausible response option. A calculation of Tredoux' (1998) E based on the distribution of affirmative responses in the initial and the final culprit-absent lineups indicated that their effective sizes were 4.19 (95% CI [3.66, 4.91]) and 3.36 (95% CI [2.99, 3.84]) respectively. All lineups are available at osf.io/k529y. The procedure was the same as that of

Please cite this article in press as: Smalarz, L., et al. Identification Performance from Multiple Lineups: Should Eyewitnesses Who Pick Fillers Be Burned? *Journal of Applied Research in Memory and Cognition* (2019), https://doi.org/10.1016/j.jarmac.2019.03.001

Table 3
Final Lineup Performance Among Witnesses Who Viewed No Initial Lineup, Witnesses Who Made An Identification from the Initial Lineup, and Witnesses Who Rejected The Initial Lineup

Initial lineup	Final lineup	Suspect	Filler	Reject
No initial lineup	Culprit-present	59.8 (52)	18.4 (16)	21.8 (19)
	Culprit-absent	18.3 (15)	37.8 (31)	43.9 (36)
Made identification from initial lineup	Culprit-present	36.4 (32)	40.9 (36)	22.7 (20)
	Culprit-absent	23.5 (19)	49.4 (40)	27.2 (22)
Rejected initial lineup	Culprit-present	47.2 (42)	29.2 (26)	23.6 (21)
	Culprit-absent	16.0 (15)	36.2 (34)	47.9 (45)

Note. Percent (frequency).

Experiment 1 except that participants were randomly assigned to view either the culprit-present or the culprit-absent final lineup.

Results

Identification behavior from the initial culprit-absent lineup. As in Experiment 1, we collected data from twice as many participants in the initial lineup condition ($n = 352$) compared to the no initial lineup condition ($n = 169$). Of the 352 participants who saw the initial lineup, 169 made an identification and 183 rejected the lineup. Hence, responses were again well-distributed across identifying and non-identifying witnesses.

Identification behavior from the final lineups. We examined witnesses' final lineup behavior separately for the culprit-present and culprit-absent lineups using two sets of Chi-square tests of independence. Table 3 displays witnesses' identification behavior from the final culprit-present and culprit-absent lineups.

Culprit-present final lineup. For witnesses who viewed the final culprit-present lineup, initial lineup group was significantly associated with final lineup behavior, $\chi^2(4) = 12.60$, $p = .013$. Follow-up analyses indicated that the final lineup behavior of witnesses who rejected the initial lineup did not differ significantly from that of the other two witness groups, χ^2s$(2) \leq 3.52$, ps $\geq .172$. The final lineup behavior of witnesses who made an identification from the initial lineup, however, differed significantly from that of witnesses who did not view the initial lineup, $\chi^2(2) = 12.48$, $p = .002$. Specifically, witnesses who made an identification from the initial lineup were significantly less likely to identify the culprit (36.4%) and significantly more likely to identify a filler (40.9%) from the final culprit-present lineup than were witnesses who saw no initial lineup (59.8% culprit identifications; 18.4% filler identifications), culprit identifications $\chi^2(1) = 9.60$, $p = .002$, OR = 0.39, 95% CI [0.21, 0.71]; filler identifications $\chi^2(1) = 10.62$, $p = .001$, OR = 3.07, 95% CI [1.54, 6.12]. These two groups did not differ significantly in their rates of rejecting the culprit-present lineup, $\chi^2(1) = 0.02$, $p = .888$.

As in Experiment 1, witnesses who rejected the final culprit-present lineup were asked which lineup member they would choose if they were forced to make a choice. A total of 60 witnesses rejected the final culprit-present lineup. We again combined the identification responses of witnesses who were forced to choose and witnesses who chose voluntarily to create a measure of overall final culprit identification accuracy. Although the effect of prior lineup group on the rate of accurate culprit identifications did not reach significance, the pattern was consistent with the Experiment 1 results, such that overall culprit identification accuracy was highest among witnesses who saw no initial lineup (65.5%) compared to witnesses who rejected (58.4%) or made an identification from (48.9%) the initial lineup, $\chi^2(2) = 5.01$, $p = .082$.

Culprit-absent final lineup. A Chi-square analysis of the choosing rate from the final culprit-absent lineup yielded a significant effect of prior lineup group, $\chi^2(2) = 8.51$, $p = .014$. Follow-up analyses indicated that the inaccurate identification rate did not differ significantly between witnesses who rejected the initial lineup and those who did not view the initial lineup, $\chi^2(1) = 0.278$, $p = .598$. Witnesses who made an identification from the initial lineup, however, were significantly more likely to make an inaccurate identification from the final culprit-absent lineup (72.8%) than were witnesses who did not view the initial lineup (56.1%), $\chi^2(1) = 4.98$, $p = .026$, OR = 2.10, 95% CI [1.09, 4.04] or witnesses who rejected the initial lineup (52.1%), $\chi^2(1) = 7.90$, $p = .005$, OR = 2.46, 95% CI [1.31, 4.65].

Final lineup confidence-accuracy relation. We examined the relation between confidence and accuracy in the final lineup separately for each of the three groups of witnesses. Final lineup decision confidence was significantly correlated with accuracy from the final lineup for witnesses who saw no initial lineup, $r(167) = .28$, $p < .001$, and witnesses who rejected the initial lineup, $r(181) = .22$, $p = .003$, but not for witnesses who made an identification from the initial lineup, $r(167) = .08$, $p = .334$. These patterns held for all three groups of witnesses when analyses were restricted to choosers; when analyses were restricted to non-choosers, the patterns held for witnesses who saw no initial lineup and those who made an identification from the initial lineup. For witnesses who rejected the initial lineup, final lineup accuracy was not significantly correlated with confidence among non-choosers, $r(64) = .16$, $p = .214$.

Diagnosticity of subsequent lineup behavior. We computed diagnosticity ratios to measure the incriminating value of suspect identifications and the exonerating value of filler identifications and lineup rejections from the final lineup for each of the three prior lineup groups (Table 4). We used ratios of relative risk (RRR) to compare diagnosticity ratios across the three groups of witnesses (Altman & Bland, 2003). Note that 95% confidence intervals that do not contain 1.00 indicate a significant difference at $p = .05$.

Suspect identifications from the final lineup were significantly less diagnostic of guilt for witnesses who made an identification from the initial lineup than for witnesses who rejected the initial lineup or saw no initial lineup, RRR = .52 [.45, .61] and RRR = .47 [.40, .56], respectively. The diagnosticity of suspect identifications did not differ significantly for witnesses who rejected the initial lineup and witnesses who saw no initial lineup, RRR = 1.10 [.92, 1.32].

Filler identifications from the final lineup were significantly less diagnostic of the suspect's innocence for witnesses who chose from the initial lineup than for witnesses who saw no initial

Please cite this article in press as: Smalarz, L., et al. Identification Performance from Multiple Lineups: Should Eyewitnesses Who Pick Fillers Be Burned? *Journal of Applied Research in Memory and Cognition* (2019), https://doi.org/10.1016/j.jarmac.2019.03.001

Table 4
Diagnosticity Ratios for Subsequent Lineup Performance

	Suspect identification			Filler identification			Lineup rejection		
	DR	LL	UL	DR	LL	UL	DR	LL	UL
No initial lineup	3.27_a	2.84	3.76	2.05_a	1.86	2.27	2.01_a	1.80	2.25
Made identification from initial lineup	1.55_b	1.40	1.72	1.21_b	1.05	1.39	1.20_b	1.09	1.31
Rejected initial lineup	2.96_a	2.65	3.30	1.24_b	1.12	1.37	2.03_a	1.81	2.28

Note. DR = diagnosticity ratio. LL = lower limit 95% confidence interval. UL = upper limit 95% confidence interval. Diagnosticity of suspect identifications was computed as the ratio of culprit-present/culprit-absent responses. Diagnosticity of filler identifications and lineup rejections was computed as the ratio of culprit-absent/culprit-present responses. Diagnosticity ratios were computed as relative risks because the sampling distributions are known and can be used to compute confidence intervals. Diagnosticity ratios with differing subscripts within columns are significantly different at $p < .05$.

lineup, RRR = .59 [.49, .70]. Filler identifications from the final lineup were also significantly less diagnostic of the suspect's innocence for witnesses who rejected the initial lineup than for witnesses who saw no initial lineup, RRR = .60 [.52, .70]. The diagnosticity of filler identifications did not significantly differ among witnesses who made an identification from the initial lineup and witnesses who rejected the initial lineup, RRR = .97 [.82, 1.16].

Rejections of the final lineup were significantly less diagnostic of the suspect's innocence for witnesses who made an identification from the initial lineup than for witnesses who rejected the initial lineup or saw no initial lineup, RRR = .59 [.51, .68] and RRR = .59 [.51, .69], respectively. The diagnosticity of lineup rejections did not differ significantly for witnesses who rejected the initial lineup and witnesses who saw no initial lineup, RRR = .99 [.84, 1.16].

Compound-decision signal detection theory analysis. The analyses reported in the previous section yielded a number of findings regarding witnesses' final lineup performance. Compared to witnesses who saw no initial lineup, witnesses who made an identification from the initial lineup were significantly less likely to identify the culprit and significantly more likely to identify a filler from the final culprit-present lineup. Witnesses who made an identification from the initial lineup were also significantly more likely to make an inaccurate identification from the final culprit-absent lineup than were either other two groups of witnesses. Diagnosticity-ratio analyses indicated that suspect identifications and lineup rejections from the final lineups were significantly less diagnostic of the guilt and innocence of the suspect, respectively, for witnesses who made an identification from the initial lineup than for the other two groups of witnesses. Filler identifications were also less diagnostic of innocence for witnesses who made an initial identification than for witnesses who saw no initial lineup. These findings converge with the results of Experiment 1 to suggest that witnesses who make an identification from an initial lineup provide less reliable evidence from a subsequent lineup than do witnesses who reject an initial lineup or see no initial lineup. To investigate whether differences in response bias or discriminability underlie these differences, we conducted a compound-decision signal-detection analysis.

Compound-decision signal-detection analysis (SDT-CD) is a statistical method for estimating discriminability and response bias in tasks involving compound decisions. Unlike simple-detection tasks in which respondents provide an old/new or present/not present response under conditions in which the signal is either present or absent, responses to eyewitness lineups involve compound decisions. The detection component of the decision involves determining whether a perpetrator is present; the identification component involves identifying which person that is (Macmillan & Creelman, 2005). Detection performance in an eyewitness lineup is assessed by measuring the proportion of identifications that occur when the culprit is present versus absent, whereas identification performance is assessed by measuring the proportion of identifications in the culprit-present lineup that are of the culprit as opposed to fillers. Compound-decision signal-detection assumes that a single discriminability parameter is common to both the detection and identification components of lineup procedures (Duncan, 2006).

We used an *independent observation* version of SDT-CD, which assumes that the detection component of the decision rests on an assessment of the strength of evidence for each lineup member individually; if the strength of evidence for the best lineup member exceeds the decision criterion, the witness will choose that person. Most applications of SDT-CD to eyewitness identification decisions have used an *integration* model of SDT-CD, which assumes that the detection component of the decision rests on an assessment of the combined strength of evidence for all lineup members (e.g., Palmer & Brewer, 2012; Smith, Wells, Smalarz, & Lampinen, 2018). However, recent evidence suggests that independent observation versions of signal detection models better describe eyewitness decisions than do integration models (Wixted, Vul, Mickes, & Wilson, 2018). Although we present analyses using the independent observation model, we also conducted analyses with the integration model. These produced slightly different values for parameter estimates but yielded the same conclusions.

Table 5 presents estimates of discriminability, decision criterion, and response bias for the three groups of witnesses. Decision criterion refers to the amount of evidence that is required to make an affirmative response, whereas response bias refers to the tendency to make an affirmative response relative to an optimal criterion (i.e., where the target and lure distributions overlap). Non-overlapping 95% confidence intervals indicate a significant difference at $p = .05$ (Tryon, 2001). The model fit statistics indicated that the model fit the data adequately for each condition (all total G statistics <4.5, p values >.21).

As is clear from Table 5, witnesses who made an identification from the initial lineup had lower discriminability and set a more

Please cite this article in press as: Smalarz, L., et al. Identification Performance from Multiple Lineups: Should Eyewitnesses Who Pick Fillers Be Burned? *Journal of Applied Research in Memory and Cognition* (2019), https://doi.org/10.1016/j.jarmac.2019.03.001

Table 5
Estimates of Discriminability (d'), Response Criterion, and Response Bias (c) and 95% Inferential Confidence Intervals for the Three Groups of Witnesses

Group	*d'*	Criterion	*c*
No initial lineup	1.77, 95% CI [1.56, 1.98]	1.23, 95% CI [1.13, 1.33]	0.34, 95% CI [0.22, 0.47]
Identified from initial lineup	0.97, 95% CI [0.75, 1.19]	0.94, 95% CI [0.84, 1.03]	0.46, 95% CI [0.33, 0.58]
Rejected initial lineup	1.47, 95% CI [1.26, 1.68]	1.22, 95% CI [1.12, 1.32]	0.48, 95% CI [0.37, 0.60]

lenient decision criterion than did witnesses in the other two groups. Response bias did not vary significantly across the three groups. Witnesses who rejected the initial lineup did not differ significantly from witnesses who saw no initial lineup on any of the three measures. Together, these results suggest that witnesses who made an inaccurate identification from the initial lineup performed worse on the final lineup than did the other groups of witnesses, and—in accord with classic principles from signal-detection theory (e.g., Macmillan & Creelman, 2005)—set a more lenient decision criterion from the final lineup in order to maximize overall accuracy.

Bayesian contingency table analysis. The identification response data and signal detection parameter estimates suggest that there was no difference in identification performance on the final lineup between witnesses who rejected the initial lineup and those who saw no initial lineup. As in Experiment 1, we conducted Bayesian contingency table analyses comparing these two groups to explore whether these results are best interpreted as true null effects rather than inconclusive results. For culprit-present lineups, the data are more than three times as likely under the null hypothesis (no difference between groups) than the alternative hypothesis, $BF_{01} = 3.75$, $N = 176$. For culprit-absent lineups, the data are 18 times as likely under the null hypothesis than the alternative hypothesis, $BF_{01} = 18.09$, $N = 176$. Thus, the patterns of identification responses are best interpreted as reflecting a true lack of difference in identification responding between these groups.

Initial lineup decision confidence as a predictor of subsequent performance. As in Experiment 1, we examined whether witnesses' confidence in their initial lineup decision predicted their performance from the subsequent lineup. We first conducted correlation analyses between witnesses' initial confidence and their subsequent identification accuracy separately for the culprit-present and culprit-absent final lineups and for witnesses who made an identification from the initial lineup and witnesses who rejected the initial lineup. For the culprit-present final lineup, we used the combined identification accuracy measure for voluntary and forced choosers, such that all witnesses who saw the culprit-present lineup were coded as having made either an accurate culprit identification or an inaccurate filler identification.

Among witnesses who made an inaccurate identification from the initial lineup, witnesses' confidence in their initial identification was not significantly correlated with subsequent culprit-present or culprit-absent lineup performance, $ps \geq .900$. Among witnesses who rejected the initial lineup, confidence in the initial rejection decision was not significantly correlated with performance from the final culprit-present lineup, $p = .326$. However, initial rejection confidence was positively associated with performance from the final culprit-absent lineup, such that higher confidence in the initial lineup rejection corresponded to an increased likelihood of making an accurate rejection decision from the culprit-absent lineup; this effect was only marginally significant, $r(92) = .20$, $p = .057$.

We examined whether accuracy from the final culprit-absent lineup varied significantly as a function of whether witnesses expressed high or low confidence in their initial rejection decision. We median-split the data for witnesses who rejected the initial lineup, with witnesses who rejected the initial lineup with confidence of 80% or greater representing the upper half ($n = 51$) and witnesses who rejected the initial lineup with confidence of less than 80% representing the lower half ($n = 43$). Results of a Chi-square analysis comparing subsequent identification accuracy for these high-confidence (52.9%) and low-confidence (41.9%) rejectors did not yield a significant effect of initial confidence on accuracy from the culprit-absent lineup, $\chi^2(1) = 1.15$, $p = .284$. An additional analysis that compared subsequent identification accuracy for high-confidence rejectors and witnesses who saw no initial lineup (43.9%) likewise did not yield a significant difference, $\chi^2(1) = 1.03$, $p = .310$.

As in Experiment 1, we conducted another analysis in which we compared identification accuracy rates from the final culprit-absent lineup among eyewitnesses who rejected the initial lineup with confidence in the top and bottom quartiles. The top quartile included witnesses who rejected the initial lineup with confidence of 95% or greater ($n = 25$) and the bottom quartile included witnesses who rejected the initial lineup with confidence of 65% or lower ($n = 24$). A Chi-square analysis indicated that eyewitnesses who rejected the initial lineup with confidence of 95% or greater performed significantly better on the final culprit-absent lineup (72.0%) than did eyewitnesses who rejected the initial lineup with confidence of 65% or lower (41.7%), $\chi^2(1) = 4.60$, $p = .032$, OR = 3.60, 95% CI [1.09, 11.86]. Moreover, eyewitnesses who rejected the initial lineup with confidence of 95% or greater also performed significantly better on the final culprit-absent lineup compared to eyewitnesses who saw no initial lineup (43.9%), $\chi^2(1) = 6.05$, $p = .014$, OR = 3.29, 95% CI [1.24, 8.72].

Discussion

Experiment 2 examined identification performance from multiple lineups using both a culprit-present and culprit-absent final lineup. The subsequent identification performance of witnesses who rejected the initial lineup did not differ significantly from that of witnesses who saw no initial lineup on either the culprit-present or the culprit-absent final lineup. By contrast, witnesses who made an identification from the initial lineup were

Please cite this article in press as: Smalarz, L., et al. Identification Performance from Multiple Lineups: Should Eyewitnesses Who Pick Fillers Be Burned? *Journal of Applied Research in Memory and Cognition* (2019), https://doi.org/10.1016/j.jarmac.2019.03.001

less likely to identify the culprit and more likely to identify a filler from the final culprit-present lineup than were witnesses who saw no initial lineup and were more likely to make an inaccurate identification from the final culprit-absent lineup than were either of the other two groups of witnesses. Moreover, diagnosticity-ratio analyses indicated that witnesses who made an identification from the initial lineup provided subsequent identification evidence that had little diagnostic value, with modest diagnosticity ratios of 1.55, 1.21, and 1.20 for suspect, filler, and rejection decisions, respectively. In interpreting these values, however, it is important to keep in mind that intervening lineups may be most detrimental when memory for the target is weak (Pezdek & Blandon-Gitlin, 2005). In the current research, accuracy rates in the no-initial lineup condition were 60% and 44% for the culprit-present and culprit-absent lineups, respectively, suggesting the possibility of potent effects of the initial lineup on subsequent identification performance.

Our compound-decision signal detection analysis further indicated that witnesses who made an identification from the initial lineup had lower final lineup discriminability (*d*') and set a lower criterion for making an identification than did witnesses who rejected the initial lineup or witnesses who saw no initial lineup. Witnesses who rejected the initial lineup did not significantly differ from witnesses who saw no initial lineup in terms of discriminability or response criterion. Critically, the confidence with which eyewitnesses made an initial inaccurate identification did not reliably predict their identification performance from the subsequent lineup, suggesting that confidence is not a useful method for identifying which inaccurate witnesses are likely to provide diagnostic identification evidence in a subsequent procedure.

General Discussion

We examined subsequent identification performance among witnesses who made an inaccurate identification from an initial lineup, correctly rejected an initial lineup, or saw no initial lineup. In both experiments, the performance of witnesses who made an identification from the initial lineup was worse compared to witnesses who saw no initial lineup, whereas the performance of witnesses who rejected the initial lineup did not differ significantly from that of witnesses who saw no initial lineup. These findings suggest that the legal practice of "burning" witnesses who pick a filler from an initial lineup may be justified (Steblay et al., 2011; Wells et al., 2015).

The findings also raise an interesting theoretical question: Do witnesses who make an initial inaccurate identification perform more poorly on a subsequent lineup because making an inaccurate identification impairs memory, or do witnesses who make an initial inaccurate identification have weaker memories going into the initial lineup? We view both possibilities as plausible accounts of our results. Perhaps witnesses who make an initial inaccurate identification have weaker memories compared to witnesses who correctly reject an initial lineup and thus have greater difficulty detecting the absence of the culprit in the lineup (e.g., Wells, 1984). Alternatively, perhaps witnesses who make an initial inaccurate identification simply have a lower criterion for making an identification compared to witnesses who correctly reject an initial lineup, but making an inaccurate identification itself impairs memory. Although it has been theorized that retrieval of information can impair subsequent retrieval of that same information (e.g., Estes, 1997), the notion that making an inaccurate identification itself impairs memory has not, to our knowledge, been explored. Hence, differentiating between these two theoretical accounts remains an open question. This issue is less consequential from an applied perspective, however, because the legal system is likely to dismiss the testimony of witnesses who make inaccurate identifications regardless of whether the witnesses were unreliable prior to or only after participating in the initial identification procedure.

Importantly, this poorer performance among witnesses who made an identification from the initial lineup occurred even though these witnesses had been informed prior to viewing the final lineup that their initial identification was inaccurate. In some ways, this is a surprising finding. One might expect that informing these witnesses of their error following the initial lineup would have made them particularly cautious during the subsequent lineup. Instead, this group of witnesses maintained a high willingness to choose. This willingness to choose may be attributable to the fact that their memory of the culprit was relatively poor; research has shown that witnesses who have a weaker recognition memory experience set a more lenient decision criterion for making an identification (e.g., Lampinen, Erickson, Moore, & Hittson, 2014; Smith, Wells, Lindsay, & Myerson, 2018). Alternatively, this result could simply reflect individual differences in response criterion, which tend to be stable across multiple recognition tests (Kantner & Lindsay, 2012), perhaps regardless of whether witnesses are informed of their initial inaccuracy.

A novel and somewhat surprising finding in our research was that the identification performance of witnesses who rejected the initial lineup was no better than that of witnesses who saw no initial lineup. Indeed, the basic premise of the blank lineup procedure is that it helps to identify or "weed out" unreliable eyewitnesses, such that only reliable witnesses remain (Wells, 1984). Our data do not align with this premise: Analyses of accuracy rates, signal detection estimates of discriminability, and Bayesian analysis of identification responses all point to the conclusion that witnesses who rejected the initial lineup performed comparably to those who did not view an initial lineup.

One interpretation of the lack of superior performance among witnesses who rejected the initial lineup is that some witnesses who rejected the initial lineup did so because they were unsure rather than because they were certain that the culprit was not present (Weber & Perfect, 2012). We tested this possibility by examining whether witnesses who rejected the initial lineup with high confidence performed better on a final lineup compared to witnesses who viewed no initial lineup. Witnesses who rejected the initial lineup outperformed eyewitnesses who were not exposed to an initial lineup only when the initial rejection decision was made with very high confidence and only for the final culprit-absent lineup (Experiment 2), not for the final culprit-present lineup (Experiments 1 and 2). Hence, our findings yielded weak evidence at best for the notion that witnesses

Please cite this article in press as: Smalarz, L., et al. Identification Performance from Multiple Lineups: Should Eyewitnesses Who Pick Fillers Be Burned? *Journal of Applied Research in Memory and Cognition* (2019), https://doi.org/10.1016/j.jarmac.2019.03.001

who reject an initial lineup with high confidence provide more reliable evidence from a subsequent lineup. Another potential interpretation for the lack of superior performance among witnesses who rejected the initial lineup is that being exposed to an initial culprit-absent lineup itself impairs memory. Although the existing data on the effects of exposing witnesses to potential photos of the culprit have shown that mere exposure does not impair memory (Deffenbacher, Bornstein, & Penrod, 2006), those studies involved mugshot searches rather than lineup tests. It is possible that the unique nature of a lineup test might have different effects on eyewitness memory, an interesting possibility that we believe should be explored in future research.

A potential limitation of this work is that we used a compound manipulation across our witness groups by giving disconfirming feedback to witnesses who made an identification from the initial lineup ("Sorry, the person you chose was not the person who switched the bags.") but not to those who rejected the initial lineup. There are two reasons why we judged this to be the best methodological approach for the current work. First, we were interested in examining subsequent identification performance independent of potential commitment effects (e.g., Gorenstein & Ellsworth, 1980; Hinz & Pezdek, 2001; Steblay et al., 2013). Past research on the blank lineup procedure has not informed witnesses that their initial identification was inaccurate, thereby leaving witnesses committed to their initial response and hence potentially less likely to attempt a new identification from the subsequent lineup (e.g., Wells, 1984). Second, we sought to maximize the ecological validity of our manipulation of initial lineup groups. Real eyewitnesses who make an initial identification and are later asked to view another lineup will virtually always be told—or at the very least infer—that their initial identification was not of the culprit, whereas witnesses who reject an initial lineup will not be given such feedback. Nevertheless, this compound manipulation introduces the possibility that the performance decrement among witnesses who made an inaccurate identification from the initial lineup may have been produced by the disconfirming feedback. Indeed, research has shown that feedback can affect memory independently of the choice preceding the feedback (Palmer et al., 2010; Smalarz & Wells, 2014). Most important for applied purposes, however, is the fact that such feedback is likely always present in real cases in which eyewitnesses make an initial mistaken identification, and our findings demonstrate a clear memory detriment among those eyewitnesses on a subsequent identification test.

Conclusion

There are a number of circumstances in which an eyewitness might view multiple lineups for the same culprit. The current research demonstrated that witnesses who made an inaccurate identification from an initial lineup tended to perform poorly on a subsequent lineup, both compared to other witnesses and in an absolute sense. These findings underscore the importance of good record-keeping at the time of any eyewitness identification procedure. Archival research using eyewitness case files indicates that police agencies often do not differentiate between filler identifications and lineup rejections (e.g., Behrman & Davey, 2001; Tollestrup, Turtle, & Yuille, 1994). Faithfully documenting witnesses' responses to identification procedures is critical, as their responses to a prior lineup can be used to assess their likelihood of providing reliable evidence in a subsequent identification procedure.

Author Contributions

L. Smalarz and N. Kornell conceived of and designed the studies; K.E. Vaughn programmed the experiments and collected and processed the data; L. Smalarz and M.A. Palmer analyzed and interpreted the data; L. Smalarz wrote the manuscript and all authors provided critical revisions.

Conflict of interest

The authors declare that they have no conflict of interest

Acknowledgements

M. Palmer was supported by an Australian Research Council Discovery Grant (DP140103746) funded by the Australian Government. N. Kornell was supported by Scholar Award 220020371 from the James S. McDonnell Foundation.

References

Altman, D. G., & Bland, J. M. (2003). Interaction revisited: The difference between two estimates. *BMJ*, *326*, 219. http://dx.doi.org/10.1136/bmj.326.7382.219

Behrman, B. W., & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior*, *25*(5), 475–491. http://dx.doi.org/10.1023/A:1012840831846

Deffenbacher, K. A., Bornstein, B. H., & Penrod, S. D. (2006). Mugshot exposure effects: Retroactive interference, mugshot commitment, source confusion, and unconscious transference. *Law and Human Behavior*, *30*, 287–307. http://dx.doi.org/10.1007/s10979-006-9008-1

Duncan, M. J. (2006). *A signal detection model of compound decision tasks (Tech. Rep. Np. No. TR2006-256)*. Toronto, ON, Canada: Defence Research and Development Canada.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, *28*, 1–11. http://dx.doi.org/10.3758/BF03203630

Estes, W. K. (1997). Processes of memory loss, recovery, and distortion. *Psychological Review*, *104*, 148–169. http://dx.doi.org/10.1037/0033-295X.104.1.148

Gorenstein, G. W., & Ellsworth, P. C. (1980). Effect of choosing an incorrect photograph on later identification by an eyewitness. *Journal of Applied Psychology*, *65*, 616–622. http://dx.doi.org/10.1037/0021-9010.65.5.616

Hinz, T., & Pezdek, K. (2001). The effect of exposure to multiple lineups on face identification accuracy. *Law and Human Behavior*, *25*, 185–198. http://dx.doi.org/10.1023/A:1005697431830

Jamil, T., Ly, A., Morey, R. D., Love, J., Marsman, M., & Wagenmakers, E. J. (2017). Default "Gunel and Dickey" Bayes factors for contingency tables. *Behavior Research Methods*, *49*, 638–652. http://dx.doi.org/10.3758/s13428-016-0739-8

JASP Team. (2018). *JASP (Version 0.9)* [Computer software].

Kantner, J., & Lindsay, S. D. (2012). Response bias in recognition memory as a cognitive trait. *Memory & Cognition*, *40*(8), 1163–1177. http://dx.doi.org/10.3758/s13421-012-0226-0

Lampinen, J. M., Erickson, W. B., Moore, K. N., & Hittson, A. (2014). Effects of distance on face recognition: Implications for eyewitness identification. *Psychonomic Bulletin & Review*, *21*(6), 1489–1494. http://dx.doi.org/10.3758/s13423-014-0641-2

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law and Human Behavior*, *36*, 247–255. http://dx.doi.org/10.1037/h0093923

Palmer, M. A., Brewer, N., & Weber, N. (2010). Postidentification feedback affects subsequent eyewitness identification performance. *Journal of Experimental Psychology: Applied*, *16*, 387–398. http://dx.doi.org/10.1037/a0021034

Palmer, M. A., Brewer, N., & Weber, N. (2012). The information gained for witnesses' responses to an initial "blank" lineup. *Law and Human Behavior*, *36*, 439–447. http://dx.doi.org/10.1037/h0093939

Peer, E., Vosgerau, J., & Acquisti, A. (2013). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavioral Research Methods*, *46*(4), 1023–1031. http://dx.doi.org/10.3758/s13428-013-0434-y

Pezdek, K., & Blandon-Gitlin, I. (2005). When is an intervening line-up most likely to affect eyewitness identification accuracy? *Legal and Criminological Psychology*, *10*, 247–263. http://dx.doi.org/10.1348/135532505X49846

Smalarz, L., & Wells, G. L. (2014). Confirming feedback following a mistaken identification impairs memory for the culprit. *Law and Human Behavior*, *38*(3), 283–292. http://dx.doi.org/10.1037/lhb0000078

Smith, A. M., Wells, G. L., Lindsay, R. C. L., & Myerson, T. (2018). Eyewitness identification performance on showups improves with an additional-opportunities instruction: Evidence for present–absent criteria discrepancy. *Law and Human Behavior*, *42*(3), 215–226. http://dx.doi.org/10.1037/lhb0000284

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

Steblay, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law*, *17*(1), 99–139. http://dx.doi.org/10.1037/a0021650

Steblay, N. K., Tix, R. W., & Benson, S. L. (2013). Double exposure: The effects of repeated identification attempts on eyewitness accuracy. *Applied Cognitive Psychology*, *27*, 644–654. http://dx.doi.org/10.1002/acp.2944

Tollestrup, P. A., Turtle, J. W., & Yuille, J. W. (1994). Actual victims and witnesses to robbery and fraud: An archival analysis. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 144–160). New York, NY: Cambridge University Press.

Tredoux, C. G. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior*, *22*(2), 217–237. http://dx.doi.org/10.1023/A:1025746220886

Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical tests. *Psychological Methods*, *6*, 371–386. http://dx.doi.org/10.1037/1082-989X.6.4.371

Weber, N., & Perfect, T. (2012). Improving eyewitness identification accuracy by screening out those who say they don't know. *Law and Human Behavior*, *36*, 28–36. http://dx.doi.org/10.1037/h0093976

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

Wells, G. L., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C., & Wixted, J. T. (in preparation). Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *American Psychology-Law Society Scientific Review Paper*, Initial draft, July 30, 2018.

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2015). Double-blind photo lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior*, *39*, 1–14. http://dx.doi.org/10.1037/lhb0000096

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

Received 9 January 2019;
accepted 6 March 2019
Available online xxx

Please cite this article in press as: Smalarz, L., et al. Identification Performance from Multiple Lineups: Should Eyewitnesses Who Pick Fillers Be Burned? *Journal of Applied Research in Memory and Cognition* (2019), https://doi.org/10.1016/j.jarmac.2019.03.001