The single lineup paradigm:

A new way to manipulate target presence in eyewitness identification experiments

by

Chris Oriet

University of Regina

and

Ryan J. Fitzgerald

University of Portsmouth

This article has been accepted for publication in *Law and Human Behavior*.

Chris Oriet, Department of Psychology, University of Regina
Ryan J. Fitzgerald, Department of Psychology, University of Portsmouth.

This research was supported by a Natural Sciences and Engineering Council of Canada Discovery Grant to C.O.

Correspondence concerning this article should be addressed to Chris Oriet, Department of Psychology, University of Regina, Regina, Saskatchewan, Canada, S4S 0A2. E-mail: chris.oriet@uregina.ca.

Abstract

The suspect in eyewitness lineups may be guilty or innocent. These possibilities are traditionally simulated in eyewitness identification studies using a dual-lineup paradigm: all witnesses observe the same perpetrator and then receive one of two lineups. In this paradigm, the suspect's guilt is manipulated by including the perpetrator in one lineup and an innocent suspect in the other. The lineup is then filled with people matched to either the suspect (resulting in different fillers in perpetrator-present and perpetrator-absent lineups) or to the perpetrator (resulting in the same fillers in each lineup). An inescapable feature of the dual-lineup paradigm is that the perpetrator-present and perpetrator-absent lineups differ not only in the suspect's guilt, but also in their composition. Here, we describe a single-lineup paradigm: subjects observe one of two perpetrators and then receive the same lineup containing one of the perpetrators. This alternative paradigm allows manipulation of the suspect's guilt without changing the lineup's composition. In three experiments, we applied the single-lineup paradigm to explore suspect-filler similarity and consistently found that increasing similarity reduced perpetrator identifications but did little to prevent innocent suspect misidentifications. Conversely, when fillers were matched to the perpetrator using a dual-lineup paradigm, increasing similarity reduced identification of perpetrators and innocent suspects. This finding suggests that the effect of filler similarity may depend on the person to whom the fillers are matched. We suggest that the single-lineup paradigm is a more ecologically-valid and better-controlled approach to creating suspect-matched lineups in laboratory investigations of eyewitness memory than existing procedures.

Keywords: eyewitness memory, recognition memory, lineups, similarity, single lineup paradigm

The single lineup paradigm:

A new way to manipulate target presence in eyewitness identification experiments

An eyewitness identification - or confident rejection of a lineup – can provide compelling corroborating (or exonerating) evidence. However, factors other than the innocence or guilt of the suspect can bias identification choices (e.g., Buckhout & Friere, 1975; Lindsay, Wallbridge, & Drennan, 1987). Of particular concern are factors that increase identifications of innocent suspects. Because the guilt or innocence of the suspect included in the lineup is unknown when the lineup is administered, care must be taken not to unfairly bias witness choices and risk incriminating an innocent person.

To investigate factors affecting lineup choices, researchers ask witnesses to a staged crime to identify the perpetrator from a perpetrator-present or perpetrator-absent lineup. Although rejection of the perpetrator-absent lineup is the only correct response, a surprising number of witnesses identify an innocent lineup member as the perpetrator. A meta-analysis of staged-crime experiments suggests that approximately 50% of witnesses presented with a perpetrator-absent lineup select an innocent lineup member (Clark, Howell, & Davey, 2008), although this rate is almost certainly influenced by the tendency of experimenters to stage events in a way that yields high variability in identification accuracy. Of greater concern, field studies with witnesses to real crimes indicate that approximately 33% of those who made an identification select a known-innocent filler (e.g., Valentine, Pickering, & Darling, 2003; Wells, Steblay, & Dysart, 2015). Filler identifications can be rejected as errors by researchers and police alike, but experiments are needed to establish the ground truth of suspect identifications. In this paper, we revisit the paradigm traditionally used in laboratory experiments to simulate the innocence vs. guilt of a suspect in a crime, and apply an alternative paradigm to explore how a manipulation of filler similarity might lead to different conclusions depending on the paradigm used to study its effects.

Experimental methodology: Dual- and single-lineup paradigms

Researchers typically manipulate the presence of the perpetrator using a dual-lineup paradigm. In the traditional dual-lineup paradigm, witnesses are randomly assigned to a perpetrator-present condition in which the perpetrator appears among several fillers, or a perpetrator-absent condition in which the perpetrator is replaced with an innocent suspect in an otherwise identical lineup. By using two lineups, researchers can simulate a situation in which the suspect identified by police is actually the perpetrator and one in which the suspect is innocent. The practice of replacing the perpetrator with an innocent suspect conforms to the principle that an experimental manipulation should change only one variable and hold all others constant, but has been criticized for its inconsistency with lineup construction practices in the field (Clark & Tunnicliff, 2001).

Although there have been calls for matching fillers to the witness description of the perpetrator (Luus & Wells, 1991; Wells, Rydell, & Seelau, 1993), the most commonly reported strategy by investigators in the field is to match fillers to the suspect's appearance (Police Executive Research Forum, 2013). To simulate this in a dual-lineup paradigm, researchers would need to select fillers who resemble the perpetrator in perpetrator-present lineups and fillers who resemble the innocent suspect in perpetrator-absent lineups; i.e., different fillers across conditions. However, Clark and Tunnicliff (2001) reviewed 25 experimental studies and found near-universal use of the same fillers across perpetrator-present and perpetrator-absent conditions.

Clark and Tunnicliff (2001) identified two problems with the traditional dual-lineup paradigm. First, unlike researchers, police do not have the luxury of choosing whether to match fillers to the appearance of the perpetrator or to an innocent suspect, as the appearance of the perpetrator is unknown if the suspect turns out to be innocent. Second, researchers who match fillers to the perpetrator's appearance using the traditional dual-lineup paradigm are implicitly accepting a confound: because fillers are matched to the perpetrator in both the perpetrator-present and

perpetrator-absent lineups, there is always greater similarity between the perpetrator and the fillers than there is between the innocent suspect and the same fillers.

Clark and Tunnicliff (2001) revised the dual-lineup paradigm to address this confound and equate lineup construction practices in research and in the field. In the revised dual-lineup paradigm, fillers are matched to the perpetrator in perpetrator-present lineups, but matched to the innocent suspect in perpetrator-absent lineups. Further, pilot studies carefully measure the perceived similarity among the lineup members and suspects to create perpetrator-present and perpetrator-absent lineups equated for suspect-filler similarity. Given that police do not always have knowledge of the perpetrator's appearance, matching fillers to the suspect in each lineup (rather than to the perpetrator) is a more ecologically-valid filler selection strategy. However, because the two lineups contain completely different lineup members, this gain in ecological validity comes at the cost of experimental control.

Although measures can be taken to ensure the average similarity of the fillers to the suspect is comparable across the lineups, the lineups may differ in other ways, such as in perceived friendliness or trustworthiness. As such, a suspect who is comparable to other lineup members in overall appearance might nevertheless draw more choices for reasons other than culpability. The problem becomes intractable as any attempt to equate perpetrator present and absent lineups on one variable will likely cause them to differ on another. Moreover, even if the lineups could somehow be equated on a number of different dimensions simultaneously, they could still differ on some dimension not considered by the researchers. It seems that modifications to fix one problem with the dual-lineup paradigm could potentially introduce new problems.

Here, we propose a single-lineup paradigm that allows researchers perfect experimental control over lineup composition across conditions without compromising ecological validity. Rather than using one perpetrator and two suspects, the single-lineup paradigm uses two perpetrators and one suspect (Figure 1). In this new paradigm all lineups contain the same suspect, but the suspect's

guilt depends on which of the two perpetrators was previously observed. Like the revised dual-lineup paradigm (Clark & Tunnicliff, 2001), the single-lineup paradigm uses a suspect-matched lineup. Unlike this paradigm (Clark & Tunnicliff, 2001), however, the suspect is always the same person, and witnesses are presented with the exact same lineup irrespective of whether they are assigned to the perpetrator present or absent condition. Thus, unlike in the traditional dual-lineup paradigm, in the single-lineup paradigm fillers are chosen in the same way they are chosen in police investigations: they are matched to the person suspected of the crime, whose actual guilt or innocence in a real investigation is unknown.

Because the same lineup is presented in the perpetrator-present and perpetrator-absent conditions, this paradigm could be especially useful for testing manipulations of retrieval conditions. Research on lineup composition, in particular, would provide a good application of this approach. In the traditional dual-lineup paradigm, fillers are matched to the identity of the perpetrator in both the perpetrator-present and perpetrator-absent conditions; in the single-lineup paradigm, fillers are matched to the suspect in both the perpetrator-present and perpetrator-absent conditions and the guilt of the suspect is determined by which video the witness saw. In the empirical research reported here, we examine filler similarity and the importance of whether fillers are matched to the suspect or to the perpetrator.

Filler similarity in suspect-matched and perpetrator-matched lineups

Meta-analytic reviews suggest the typical effect of increasing similarity is to reduce identifications of the suspect, regardless of whether that suspect is guilty or innocent (Clark, 2012; Fitzgerald, Price, Oriet, & Charman, 2013). Researchers generally manipulate the similarity between the fillers and the perpetrator, and this is normally achieved by varying the degree to which fillers match either the perpetrator's appearance or an accurate description of the perpetrator (e.g., Lindsay & Wells, 1980; Tredoux, Parker, & Nunez, 2007). With the traditional dual-lineup paradigm, increasing perpetrator-filler similarity should draw more choices away from the suspect (regardless

of guilt) and toward the fillers, a process known as filler siphoning (Smith, Wells, Lindsay, & Penrod, 2017; Wells, Smalarz, & Smith, 2015). This practice of matching fillers to the perpetrator in perpetrator-present and perpetrator-absent lineups may contribute to the symmetrical effect of similarity on guilty and innocent suspects.

A different effect of similarity may occur if fillers are matched to the suspect. Using a feature-matching model, Clark and Tunnicliff (2001) proposed that fillers matched to an innocent suspect would resemble the perpetrator less than would fillers matched to the perpetrator. The model specified that a perpetrator and an innocent suspect would share some but not all features; that fillers would be matched on some but not all of the suspect's features; and that, of the features the perpetrator-absent lineup fillers are matched on, some would be shared by the perpetrator and innocent suspect but others would be possessed by only the innocent suspect. To illustrate their case, Clark and Tunnicliff created an example with a perpetrator and an innocent suspect who shared four out of eight features with each other. In this example, the fillers were always matched to four features possessed by the suspect in their respective lineup. Thus, a filler in the perpetrator-present lineup shared four features with the perpetrator and a filler in the perpetrator-absent lineup shared four features with the innocent suspect. Of the four features on which the perpetrator-absent filler was matched, Clark and Tunnicliff specified that two were shared by the perpetrator and innocent suspect and two were specific to the innocent suspect. As a consequence, the perpetrator-present filler shared two more features with the perpetrator than did the perpetrator-absent filler. Although we could imagine how this might play out differently in exceptional circumstances, it is hardly controversial to suggest that fillers matched to the perpetrator would normally resemble the perpetrator more than would fillers matched to an innocent suspect.

The single-lineup paradigm is a prime candidate for studying suspect-filler similarity in suspect-matched lineups. The only alternative for suspect-matching, the revised dual-lineup paradigm (Clark & Tunnicliff, 2001), would require a minimum of four lineups with different fillers

in each lineup: lower similarity + perpetrator, lower similarity + innocent suspect, higher similarity + perpetrator, higher similarity + innocent suspect. More often than not, we suspect this approach would serve its intended purpose and inform about suspect-filler similarity effects. But the revised dual-lineup paradigm requires an implicit assumption that the fillers in the perpetrator-present lineup are functionally equivalent to the fillers in the perpetrator-absent lineup. By contrast, the single-lineup paradigm requires a minimum of only two lineups and functional equivalence across perpetrator-present and perpetrator-absent lineups need not be assumed because the lineups are identical. This permits a more direct test of the diagnostic value of a specific set of fillers. That is, with the single lineup paradigm, researchers can assess how one specific set of fillers affect identifications of one specific suspect, both when innocent and when guilty.

Present Research

We manipulated suspect-filler similarity in three experiments using the single-lineup paradigm. In Experiment 1, we compared three levels of suspect-filler similarity in perpetrator-present and perpetrator-absent lineups. In Experiment 2, we directly contrasted the effects of a similarity manipulation in the suspect-matched lineups (using our single-lineup paradigm) and perpetrator-matched lineups (using the traditional dual-lineup paradigm). The encouraging results of Experiments 1 and 2 led us to carry out a much more rigorous test in Experiment 3, which used stimulus sampling to test the generalizability of the preceding findings. We regard the first two experiments as a demonstration of the method and treat the conclusions from them as tentative, drawing firmer conclusions from Experiment 3.

---------- Insert Figure 1 about here ----------

Experiment 1

The purpose of Experiment 1 was to explore the utility of the single-lineup paradigm in investigating the effect of suspect-filler similarity, a variable whose effects in the dual-lineup paradigm have been well-documented (Lindsay & Wells, 1980). Similarity has been characterized as

a kind of double-edged sword: increasing suspect-filler similarity protects innocent lineup members, but at the expense of valuable guilty suspect identifications (Clark, 2012). Thus, we expected to observe the usual effect of increasing suspect-filler similarity, namely a decrease in correct perpetrator identifications and a corresponding decrease in mistaken innocent suspect identifications.

Method

All procedures were approved by the University of Regina Research Ethics Board and the University of Portsmouth University Ethics Committee. Informed consent was obtained from all participants prior to data collection. Based on a similar study that used suspect-matched lineups (Clark & Tunnicliff, 2001), we aimed to test 60 subjects per cell (3 levels of similarity × 2 video events = 6 cells). Because we were recruiting from a panel of online respondents (*SurveyMonkey Audience*) we anticipated that data from some subjects would not be usable (i.e., that they would fail to report a detail from the video correctly). To ensure adequate sample size, we allowed enrollment in the study to proceed for the remainder of the day on which we reached 360 subjects; this resulted in a total of 415 subjects. To be included in the study, subjects were required to be between the ages of 18 and 37[1], proficient in English, and currently residing in the U.S. An exclusion criterion was that subjects needed to report at least one accurate descriptor of the target actor.

For the description task, we instructed participants to provide as much detail as possible about the actor and specifically requested information about his sex, age, ethnicity, hair color, attractiveness, and facial features. The response option was open-ended, allowing respondents to give as much or as little information as they wished. Of the 415 respondents, the majority (92%, $n = 382$) responded to most or all of the specific probes and reported at least two descriptors that were unambiguously correct (e.g., White, man). The remaining respondents either did not report a single descriptor that could possibly counted as correct ($n = 22$) or reported only one or two correct

[1] Because we are testing a new paradigm, we opted to restrict age to the range of ages we observed in our other studies using the traditional paradigm, which happened to be 18-37.

responses and ignored four or five of the specific probes (n = 11). We excluded only the former, who either were clearly not following the instructions (e.g., responding "fvfvfv") or had such a poor memory of the actor that they could not recall something as basic as his sex. After removal of those 22 subjects and another three who reported that they were outside the specified age range, the final sample comprised 390 subjects (200 women; 18-37 years, M = 27.78, SD = 6.03).

To disguise the purpose of the study, subjects were advised that they would be viewing a series of videos and judging the suitability of the video actors for a variety of roles in an upcoming film. All subjects viewed a video with a male actor (Person 1 or Person 2) followed by two videos, each with a female actor. Both Person 1 and Person 2 matched the description of a 25-year old clean-shaven Caucasian male with brown hair and no visible piercings or tattoos. After watching each video, subjects described the actor and judged the appropriateness of the actor for various film roles. After providing judgments for the third video, subjects learned the true purpose of the experiment and that they would be asked to try to identify the male actor from a lineup. Instructions specified that he may or may not be in the upcoming lineup. The lineup task allowed participants to select one or none of the lineup members. Finally, participants provided a post-identification confidence rating, ranging from 1 (not at all confident) to 6 (highly confident).

All lineups contained one suspect and five fillers. The suspect was always Person 1. Whether the suspect was guilty or innocent depended on which male actor was viewed previously. Subjects in the perpetrator-present condition viewed a video featuring Person 1, the male actor who would become the guilty suspect for these subjects by virtue of his presence in the subsequent lineup. Subjects in the perpetrator-absent condition viewed a video featuring Person 2, another male actor who would not become the suspect and did not appear in any lineup. These subjects were then presented exactly the same lineup as subjects in the perpetrator-present condition, with Person 1 and fillers matched to him. The perpetrator-absent condition is analogous to an investigation in which an innocent person (Person 1) is suspected of committing the crime and the true perpetrator (Person 2)

has evaded investigation. The videos of Person 1 and Person 2 were both 3 s in duration and displayed a clear view of the person. The videos of each actor depicted only their appearance (no actions).

The similarity between fillers and the suspect was manipulated across lineups using ratings obtained in a pilot study. Judges (n = 20) made pairwise comparisons between Person 1 and 277 faces on a scale from 0 (not at all similar) to 10 (highly similar). Mean similarity ratings for the set ranged from 0.45 to 5.70 (M = 2.27, SD = 0.97). Rater consistency was assessed in a two-way random-effects model, ICC (2, 20) = .87 [.84, .89]. The similarity ratings were used to guide construction of three lineups: Very High similarity, using faces with ratings at the 99th percentile (M rating = 5.34, SD = 0.42, range = 4.75 – 5.70); High similarity, using faces with ratings at the 65th – 83rd percentile (M = 2.87, SD = 0.25, range = 2.65 – 3.20); and, Moderate similarity, using faces at the 46th – 53rd percentile (M = 2.11, SD = 0.08, range = 2.05 – 2.25). The similarity rating for Person 2 was at the 95th percentile (M = 3.90).

Analyses

We first performed hierarchical loglinear (HILOG) analyses to test for any differences among the three similarity conditions, and then proceeded with tests of pairwise differences if necessary. Odds ratios and 95% confidence intervals [LL, UL] are reported as a measure of effect size for rate differences. We converted all odds ratios below 1.00 to their inverse to facilitate comparisons across analyses.

Results

The effect of similarity on suspect identifications depended on whether the suspect was guilty or innocent (Table 1). When innocent (i.e., a perpetrator-absent lineup), the HILOG revealed no significant association between suspect-filler similarity and suspect identification, $\chi^2(1) = 0.07$, $p >$ .250. Regardless of suspect-filler similarity, suspect identifications in the perpetrator-absent condition were consistently low. When guilty (i.e., a perpetrator-present lineup), the HILOG analysis

revealed a significant association between suspect-filler similarity and suspect identification, $\chi^2(1) = 18.68$, $p < .001$. Follow-up pairwise analyses indicated the guilty suspect was correctly identified more frequently in both the Moderate and High similarity conditions than in the Very High similarity condition, $z = 4.11$, $p < .001$, $OR = 4.30$ [2.00, 9.92] and $z = 3.92$, $p < .001$, $OR = 4.09$ [1.90, 8.80], respectively. The low correct identification rate in the Very High similarity condition coincided with a high rate of filler selections.

Similarity also affected rejections differently in perpetrator-present and perpetrator-absent conditions. In the perpetrator-absent condition, the HILOG revealed a significant association between suspect-filler similarity and lineup rejections (identification vs. no identification), $\chi^2(2) = 7.17$, $p = .028$. The Moderate and High similarity lineups were correctly rejected more than was the Very High similarity lineup. The difference between the Moderate and Very High similarity conditions was significant, $z = 2.62$, $p = .009$, $OR = 2.61$ [1.25, 5.41], but the difference between the High and Very High similarity conditions was not, $z = 1.77$, $p = .077$, $OR = 1.96$ [0.92, 4.20]. In the perpetrator-present condition, the HILOG revealed no significant association between suspect-filler similarity and lineup rejections, $\chi^2(2) = 1.99$, $p > .250$.

---------- Insert Table 1 about here ----------

Discussion

In Experiment 1, all subjects were assigned to a lineup with Person 1 as the designated suspect. When subjects were looking for Person 1, he was hard to identify when surrounded by highly similar fillers. But when looking for Person 2, identifications of Person 1 were unaffected by how much the fillers resembled him. In other words, matching fillers closely to the suspect's appearance only reduced identification of the suspect if he was guilty.

This pattern of results is not without precedent. Compared with fillers who matched a witness description but were otherwise dissimilar to the suspect, Wells et al. (1993) found that fillers closely matched to the suspect had no effect on innocent suspect identifications and only reduced guilty

suspect identifications. Similarity manipulations do not always have such asymmetrical effects in perpetrator-present and perpetrator-absent lineups. Indeed, meta-analyses of the wider literature show that increasing similarity tends to reduce identifications of both guilty and innocent suspects (Clark, 2012; Fitzgerald et al., 2013).

We point to two factors that may have contributed to the absence of an effect of similarity on innocent suspect identifications. First, although Persons 1 and 2 were rated high in similarity, the vast majority of subjects looking for Person 2 did not misidentify Person 1. Innocent suspect identifications in Experiment 1 were consistently low and with only 7% selecting the innocent suspect in the low similarity condition, there was little room for improvement as similarity increased. Second, we matched fillers in the perpetrator-absent lineup to the innocent suspect (see also Wells et al., 1993). The more common approach is to use the same fillers in perpetrator-present and perpetrator-absent lineups. Therefore, perpetrator-matched designs may be more likely to yield symmetrical effects of similarity (reductions in perpetrator and innocent suspect identifications) and suspect-matched designs may be more likely to yield asymmetrical effects (reductions in perpetrator identifications only).

Experiment 2

In Experiment 2, we directly compared performance in suspect-matched and perpetrator-matched designs. Suspect-matching was implemented with the single-lineup paradigm: Subjects were randomly assigned to view a video with Person 1 or Person 2, and then chose from a lineup with Person 1. Perpetrator-matching was implemented with the traditional dual-lineup paradigm: Subjects watched a video of Person 1, and then were randomly assigned to choose from a lineup containing Person 1 or Person 2. In both paradigms, fillers were matched to Person 1.

Method

A total of 401 subjects were recruited from *SurveyMonkey Audience* with the aim of testing 60 subjects in each of the 6 cells of the design. The inclusion and exclusion criteria were the same as

in Experiment 1, with the added constraint that subjects could not have participated in Experiment 1. Fifteen subjects failed to report an accurate descriptor and four reported their age as outside the specified range, resulting in the final sample of 382 subjects (189 women; 18-37 years, M = 28.15, SD = 5.49).

The procedure and stimuli were the same as in Experiment 1 with two exceptions:

(1) Decreasing suspect-filler similarity improved identification outcomes in Experiment 1, so we tested the limits of this effect by comparing the High similarity lineup from Experiment 1 (65^{th}-83^{rd} percentile, M = 2.87, SD = 0.25, range = 2.65 – 3.20) with a Low similarity lineup comprised of fillers who received some of the lowest ratings in the set (5^{th} percentile or lower, M = 0.86, SD = 0.23, range = 0.45 – 1.00).

(2) Subjects were randomly assigned to view a perpetrator-present lineup, a suspect-matched perpetrator-absent lineup using the single-lineup paradigm, or a perpetrator-matched perpetrator-absent lineup using the dual-lineup paradigm. The first two conditions used the same methodology as in Experiment 1. In the third condition, subjects viewed Person 1 at encoding and were subsequently administered a perpetrator-absent lineup in which Person 2 was the suspect and the fillers were matched to Person 1.

Results

In the perpetrator-present condition, the correct identification rate was significantly greater in the Low similarity condition (77%) than in the High similarity condition (60%), z = 2.21, p = .030, OR = 2.30 [1.08, 4.88]. For perpetrator-absent lineups, a 2 (filler matching: suspect-matched vs. perpetrator-matched lineup) × 2 (similarity: Low vs. High) × 2 (innocent suspect: identified vs not identified) HILOG indicated the highest order effect was a 2-way interaction, $\chi^2(4) = 32.38$, p = .001. Partial associations showed that suspect identifications were influenced both by filler matching, $\chi^2(1) = 26.54$, p = .001, and by similarity, $\chi^2(1) = 5.36$, p = .021. The effect of filler matching reflected the higher rate of innocent suspect identifications in the perpetrator-matched lineup (27%) than in the

suspect-matched lineup (5%), $z = 5.04$, $p < .001$, $OR = 7.63$ [3.06, 18.98]. The effect of similarity reflected an increase in innocent suspect identifications in the Low similarity condition (21%) compared with the High similarity condition (11%), $z = 2.14$, $p = .033$, $OR = 2.16$ [1.06, 4.39].

Similarity affected innocent suspect identifications in perpetrator-matched lineups, but not in suspect-matched lineups. When the innocent suspect (Person 2) was placed in a lineup with fillers who were matched to the perpetrator (Person 1), the innocent suspect was significantly more likely to be mistakenly identified in the Low similarity condition (38%) than in the High similarity condition (18%), $z = 2.54$, $p = .010$, $OR = 2.89$ [1.25, 6.69]. The reduction in innocent suspect misidentifications for the High similarity condition coincided with a high rate of filler selections. Consistent with Experiment 1, the innocent suspect was rarely selected from a suspect-matched lineup, regardless of whether it contained High (4.5%) or Low (4.8%) similarity fillers, $z = 0.08$, $p > .250$, $OR = 1.07$ [0.21, 5.50]. But, as we elaborate in the Discussion, the comparison between filler-matching procedures is muddied by the generally higher rates of innocent suspect identification rates in the perpetrator-matched conditions.

The higher rates of innocent suspect identifications for perpetrator-matched lineups corresponded with lower correct rejection rates. For perpetrator-absent lineups, a 2 (filler matching: suspect-matched vs. perpetrator-matched lineup) × 2 (similarity: Low vs. High) × 2 (lineup rejection: identification vs no identification) HILOG indicated the highest order effect was a 2-way interaction, $\chi^2(4) = 22.35$, $p < .001$. Partial associations showed the correct rejection rate (81% [74%, 88%]) was significantly higher for suspect-matched lineups than for perpetrator-matched lineups (55% [45%, 64%]), $\chi^2(1) = 20.87$, $p < .001$, $OR = 3.61$ [2.04, 6.39], and that similarity was not significantly associated with correct rejections, $\chi^2(1) = 0.32$, $p > .25$, $OR = 1.17$ [0.68, 1.99]. When the perpetrator was present, similarity was also not significantly associated with incorrect lineup rejections, $z = 0.94$, $p > .250$, $OR = 1.45$ [0.66, 3.14].

We examined only those who made an identification from a perpetrator-absent lineup (choosers) to test whether fillers comprised a greater proportion of the identifications in the High similarity lineups than in Low similarity lineups. In the perpetrator-matched lineup, the proportion of choosers who selected a filler in the High similarity condition (61% [42%, 76%], $n = 28$) was over three times as great as the proportion in the Low similarity condition (19% [8%, 37%], $n = 27$), $z = 3.49$, $p < .001$, $OR = 6.80$ [1.98, 23.31]. Conversely, in the suspect-matched lineup, similarity's effect on the proportion of choosers who selected a filler was attenuated (High: 80% [55%, 93%], $n = 15$; Low: 67% [35%, 88%], $n = 9$), $z = 0.68$, $p > .250$, $OR = 2.00$ [0.31, 13.06]. Although increasing filler similarity led to a smaller increase in the proportion of choosers who selected a filler in suspect-matched lineups than in perpetrator-matched lineups, a 2 x 2 x 2 HILOG showed the three-way interaction was nonsignificant, $\chi^2(1) = 1.14$, $p > .25$.

---------- Insert Table 2 about here ----------

Discussion

The effects of similarity we observed in Experiment 1 replicated in Experiment 2 when fillers were matched to the suspect (single lineup paradigm), but we found a different result in the perpetrator-absent condition when fillers were matched to the perpetrator (traditional dual lineup paradigm). These results are consistent with the suggestion that increasing filler similarity to the perpetrator in the perpetrator-absent lineup reduces innocent suspect identifications, but increasing filler similarity to the suspect does not. However, subjects looking for Person 2 were not inclined to choose Person 1 from the suspect-matched lineup. Therefore, as with Experiment 1, the generally low rates of innocent suspect identifications could explain why similarity had no effect on suspect identifications in suspect-matched, perpetrator-absent lineups. Interestingly, in the perpetrator-matched condition, a relatively large proportion of the subjects looking for Person 1 ended up choosing Person 2. So Person 2 was easily mistaken for Person 1, but for some reason Person 1 was not a plausible replacement for Person 2. Because the actors were not counterbalanced across

conditions, we cannot discount the possibility this facilitated the effect of similarity in the perpetrator-matched lineup condition and precluded the same effect in the suspect-matched condition.

Experiment 3

In Experiment 3, we attempted to replicate the findings from Experiment 2 with a broader range of counterbalanced stimuli. In this experiment, we used three pairs of perpetrators and ensured that each person appearing as a perpetrator took the roles of Person 1 and Person 2 (see Online Supplementary Materials for all lineup stimuli and a detailed explanation of the counterbalancing procedure for this experiment). As in Experiment 2, we compared identification responses in the single-lineup paradigm with suspect-matched fillers to identification responses in the dual-lineup paradigm with perpetrator-matched fillers

Method

A total of 363 subjects were recruited from *SurveyMonkey Audience*. The inclusion and exclusion criteria were the same as in Experiments 1 and 2, except that, because Experiment 3 was designed to test the generality of the effect, the upper bound of the age criterion was removed. Seven subjects failed to report an accurate descriptor, resulting in the final sample of 356 subjects (232 women; 18-80 years, $M = 43.99$, $SD = 16.87$).

In this experiment, the design from Experiment 2 was repeated with a wider range of stimuli. All subjects viewed only one perpetrator and one lineup, but the identity of perpetrators was sampled across subjects to reduce the possibility that our findings would be influenced by idiosyncratic characteristics of a particular stimulus set (Wells & Windschiltl, 1999). In all, we used six video actors to create three perpetrator-innocent suspect pairs. Each video actor served as both Person 1 and Person 2, and thus was a perpetrator for some participants, and an innocent suspect for others. We also made minor changes to the encoding and filler tasks, such that the duration of videos was increased to 10 s and the number of female actors presented after the target male actor was increased

to three. In each video, the actor could be seen walking towards the camera, staring into the camera with a neutral expression for 2 s, then turning around and walking away from the camera.

We selected a pool of potential suspects and fillers from the Glasgow Unfamiliar Face Database (Burton, White, & McNeill, 2010). The database consists of video and photo images of approximately 300 individuals. To create a pool of fillers who would match a coarse description of the suspect, we removed all individuals from the database who did not have the following characteristics: White British, male, 18- to 28-years-old, short hair, dark hair, and minimal or no facial hair. The database comes with ground-truth information about age, sex, and ethnicity. Hair length and shade were categorized by two coders. Removal of the nonmatching individuals resulted in a pool of 69 potential suspects and fillers.

The selection of suspects and fillers was informed by similarity judgments. The database includes a matrix of pairwise similarity scores that were obtained from judges (N = 30) who sorted the images into piles (see Bruce et al., 1999). Judges were instructed to place photographs of similar-looking individuals into piles, using as many or as few piles as they wished to do so. The similarity scores, which ranged from 0%-53%, represent the proportion of judges who put two individuals into the same pile. A limitation of the similarity ratings used to select the innocent suspect in Experiments 1 and 2 is that even if our innocent suspect received one of the highest ratings of similarity to our perpetrator, it is not necessarily the case that people would spontaneously indicate that they look alike. Using the metric provided by Bruce et al. ensures that the actors chosen to serve as Person 1 and 2 were actually judged to be similar in appearance by a given proportion of judges.

We selected target-pairs with scores at the higher end of the range (40%-53%) to increase the likelihood that the innocent suspect in the perpetrator-absent condition would resemble the perpetrator. In real investigations, innocent suspects could be expected to resemble the perpetrator if they are under investigation because their appearance corresponds with an eyewitness description, a composite drawing, or CCTV video of the perpetrator (Wells & Penrod, 2011). Conversely, innocent

people who become suspects for other reasons (e.g., a history of committing similar crimes) may look nothing like the perpetrator. We focused on the former scenario because there are no theoretical grounds to predict an effect of filler similarity on identification of an innocent suspect who is dissimilar to the perpetrator and unlikely to be chosen in the presence of either high or low similarity fillers. Further, we wanted to test whether the data pattern observed in Experiment 2 would replicate for innocent suspects at the greatest risk of false identification. The Low and High similarity conditions contained the individuals with the lowest (range: 0-4%) and highest (range: 30-37%) scores, respectively, with the added condition that the same filler could not be in more than one lineup per actor-pair.

As in Experiment 2, subjects were randomly assigned to choose from a perpetrator-present lineup or one of two types of perpetrator-absent lineups. In the perpetrator-absent lineups, fillers were either matched to the innocent suspect using the single-lineup paradigm or matched to the perpetrator using the dual-lineup paradigm. In the single-lineup condition, subjects viewed Person 1 or Person 2 and were subsequently administered a lineup in which Person 1 was the suspect and the fillers were matched to Person 1. In the dual-lineup condition, subjects viewed Person 1 and were subsequently administered a lineup in which Person 1 or Person 2 was the suspect and the fillers were matched to Person 1. The actors who took the roles of Person 1 and Person 2 were counterbalanced within actor-pairs across subjects.[2]

Results

The identification response pattern was similar to that observed in Experiment 2 (Table 2). Once again, increasing similarity led to a significant decrease in correct identifications of guilty suspects from perpetrator-present lineups (Low = 77%, High = 60%), $z = 2.54$, $p = .011$, $OR = 2.81$

[2] Note: A fully-counterbalanced design would require a condition in which the lineup administered in the single-lineup paradigm contains the perpetrator accompanied by fillers matched to the known-innocent video actor rather than to the suspect appearing in the lineup. A real police lineup with a guilty suspect would never contain fillers matched to a known-innocent suspect, and thus this condition was excluded from the present design.

[1.22, 6.47], and the effect of similarity on mistaken identifications of innocent suspects depended on whether fillers were matched to the suspect or perpetrator. For the perpetrator-absent condition, a 2 (filler matching: suspect-matched vs. perpetrator-matched) × 2 (similarity: low vs high) × 2 (suspect: identified vs not identified) HILOG analysis revealed a significant 3-way interaction, $\chi^2(1) = 4.94$, $p = .026$. In perpetrator-absent lineups with perpetrator-matched fillers, increasing similarity led to a significant decrease in mistaken identifications of innocent suspects (Low = 35%, High = 8%), $z = 3.79$, $p < .001$, $OR = 6.15$ [1.97, 19.24]. However, in perpetrator-absent lineups with suspect-matched fillers, similarity had no significant effect on mistaken identifications of innocent suspects (Low = 26%, High = 21%), $z = 0.69$, $p > .250$, $OR = 1.33$ [0.59, 3.04]. Importantly, the null effect of similarity on innocent suspect misidentifications in the single-lineup condition using suspect-matched lineups was replicated with a much higher base rate of innocent suspect misidentifications than in Experiment 2, likely owing to the similarity between the actors who took the roles of Person 1 and Person 2.

As a follow-up analysis, we treated each perpetrator-innocent suspect pair as a unit of analysis and assessed the summary effect in a random-effects meta-analysis. This allowed us to include target pair as a factor in the design. This approach led to findings similar to those reported above. In the perpetrator-present condition, the correct identification rate was significantly higher in the Low similarity condition than in the High similarity condition, $z = 2.23$, $p = .026$, $OR = 2.65$ [1.13, 6.24]. No significant heterogeneity in effect sizes was detected, $Q(2) = 0.32$, $p > .25$, $I^2 = .00$ (a forest plot is reported in the Online Supplementary Materials). In the perpetrator-absent condition, increasing similarity produced a large and significant decrease in innocent suspect misidentifications with perpetrator-matched lineups, $z = 2.24$, $p = .025$, $OR = 3.92$ [1.19, 12.94], compared with a much smaller and nonsignificant decrease in innocent suspect identifications with suspect-matched lineups, $z = 0.86$, $p > .25$, $OR = 1.45$ [0.62, 3.39]; a moderator test comparing the two summary effects was nonsignificant, $Q = 2.89$, $p = .17$. No significant heterogeneity in effect sizes was detected for either

the suspect- or perpetrator-matched lineups, both $Q(2) < 3.11$, $p > .20$, $I^2 < .36$ (forest plots are reported in the Online Supplementary Materials).

Next, we looked at rejection rates. For perpetrator-absent lineups, a 2 (filler matching: suspect-matched vs. perpetrator-matched lineup) × 2 (similarity: Low vs. High) × 2 (lineup rejection: identification vs no identification) HILOG indicated the highest order effect was a 3-way interaction, $\chi^2(1) = 6.56$, $p = .010$ (see Table 2). For suspect-matched lineups the correct rejection rate was significantly higher for Low similarity lineups than for High similarity lineups, $z = 2.23$, $p = .026$, $OR = 2.22$ [1.09, 4.55], whereas for perpetrator-matched lineups there was a nonsignificant effect in the other direction, $z = 1.45$, $p = .145$, $OR = 1.78$ [0.81, 3.91]. When the perpetrator was present, similarity was not significantly associated with incorrect lineup rejections, $z = 0.33$, $p > .250$, $OR = 1.16$ [0.47, 2.81].

Focusing on choosers, we found a pattern similar to that observed in the perpetrator-absent condition in Experiment 2. In the dual-lineup condition with perpetrator-matched lineups, the proportion of choosers who selected a filler in the High similarity condition (71% [45%, 88%], $n = 14$) was almost five times as great as the proportion in the Low similarity condition (15% [6%, 32%], $n = 27$), $z = 3.94$, $p < .001$, $OR = 14.38$ [2.98, 69.25]. As with Experiment 2, the difference was smaller in the single-lineup condition with suspect-matched fillers (High: 61% [43%, 75%], $n = 33$; Low: 23% [10%, 43%], $n = 22$), $z = 3.01$, $p = .002$, $OR = 5.23$ [1.55, 17.67]. Although increasing filler similarity once again led to a smaller increase in the proportion of choosers who selected a filler in suspect-matched lineups than in perpetrator-matched lineups, a 2 x 2 x 2 HILOG showed the three-way interaction was nonsignificant, $\chi^2(1) = 1.01$, $p > .25$.

An advantage of Bruce et al.'s (1999) similarity scoring procedure is that the resulting matrix can be used to assess perpetrator-filler similarity in the perpetrator-absent lineups for each of the conditions. The increase of suspect-filler similarity may have led to an indirect increase in perpetrator-filler similarity in the single-lineup, perpetrator-absent condition, because we chose

perpetrator-innocent suspect pairs who were judged to be similar to one another. Consistent with this suggestion, the manipulation of suspect-filler similarity in this condition led to higher perpetrator-filler similarity scores (with similarity measured as the proportion of raters who grouped each filler's photo in the same pile as the perpetrator's, averaged over all fillers) in the High similarity condition (M = .26, SD = .04) than in the Low similarity condition (M = .07, SD = .02).

Given that perpetrator-filler similarity was directly manipulated in the dual-lineup condition, it is not surprising that this led to a larger difference in perpetrator-filler similarity between the High (M = .32, SD = .03) and Low (M = .02, SD = .01) similarity conditions. These differences in perpetrator-filler similarity scores would seem to account for the different effects of similarity on the proportion of choosers who selected a filler from suspect-matched and perpetrator-matched lineups. Note also that, as reported above, increasing similarity in the single-lineup condition led to a significant decrease in the correct rejection rate. Thus, although increasing similarity in the single-lineup condition with suspect-matched fillers led to an increase in the proportion of choosers who selected a filler in the perpetrator-absent condition, it also led to a greater number of choosers overall in the perpetrator-absent condition.

Discussion

The general pattern observed in Experiment 2 replicated in Experiment 3. High similarity fillers reduced correct identifications in the perpetrator-present lineup and also innocent suspect identifications when fillers were matched to the perpetrator. But when fillers were matched to the innocent suspect no significant reduction in innocent suspect identifications was observed with increased similarity. Rather, increasing the similarity between the fillers and the innocent suspect in the suspect-matched condition led to more filler selections and fewer correct rejections, the same pattern that occurred in the first two experiments. Importantly, in contrast to the first two experiments, the rate of innocent suspect identifications from the Low similarity suspect-matched lineups was high enough that it could have been reduced by an increase in suspect-filler similarity.

Indeed, using more similar suspect-matched fillers did lead to a small reduction of innocent suspect identifications, but the effect of similarity was notably stronger when fillers were matched to the perpetrator.

General Discussion

The key conclusions of our investigation can be summarized as follows:

1. The single-lineup paradigm can be used to manipulate perpetrator-presence by changing only the identity of the perpetrator, rather than changing the identity of the suspect.
2. The single-lineup paradigm could be especially useful for studying the effects of manipulations of lineup composition.
3. One such manipulation, filler similarity, yielded quite different findings in the single and dual-lineup paradigms. In the single-lineup paradigm, with fillers matched to the suspect, increasing similarity consistently reduced correct identifications without significantly reducing innocent suspect identifications. In the dual-lineup paradigm, with fillers matched to the perpetrator, increasing similarity consistently reduced both perpetrator and innocent suspect identifications.

Why did the person to whom fillers were matched have such a dramatic effect on response patterns? This may be an instance of differential filler siphoning. Smith et al. (2017) used this phrase to describe how good fillers draw more choices from innocent suspects than from guilty suspects. They argued that differential filler siphoning occurred because fillers could compete better with an innocent suspect than with the perpetrator. Our findings could be described as another type of differential filler siphoning. Of the subjects who made an identification in Experiments 2 and 3, increasing the similarity of perpetrator-matched fillers resulted in larger increases in filler identifications (i.e., more siphoning) than did increasing the similarity of suspect-matched fillers. Fillers matched to the perpetrator would normally resemble the perpetrator more than would fillers matched to the innocent suspect (Clark & Tunnicliff, 2001). Therefore, differential filler siphoning would occur because perpetrator-matched fillers would be able to compete with the innocent suspect

better than would suspect-matched fillers. Much as they did with the perpetrator, perpetrator-matched fillers drew choices away from the innocent suspect.

Our direct comparison of similarity effects in suspect-matched and perpetrator-matched lineups points to an explanation for the discrepant findings reported in the filler similarity literature. With the traditional dual-lineup paradigm in Experiments 2 and 3 using perpetrator-matched lineups, we found that increasing filler similarity reduced mistaken identifications of the innocent suspect. This replicates previous findings obtained when similarity has been manipulated using the traditional dual-line-up paradigm (Fitzgerald, Whiting, Therrien, & Price, 2014; Lindsay & Wells, 1980). We know of only two studies that have manipulated the similarity between the innocent suspect and the fillers in a perpetrator-absent lineup. In the first study, Wells et al. (1993) tested lineups containing perpetrator-absent fillers who either matched the innocent suspect's appearance or who matched a general description of the perpetrator but were otherwise dissimilar from the innocent suspect. In the second study, Fitzgerald, Oriet, and Price (2015) morphed fillers with the innocent suspect to a higher or lower degree. The effect of the similarity manipulation was consistent across these two studies: increasing similarity led to more perpetrator-absent filler selections and fewer correct lineup rejections, but had minimal (if any) effect on innocent suspect misidentifications. This is precisely the pattern we obtained in all three experiments when fillers were matched to the innocent suspect via the single lineup paradigm. Our direct comparison of approaches to manipulating filler similarity (approaches only tested previously in isolation) suggest the discrepant findings in the literature may have occurred because perpetrator-absent fillers were matched to the innocent suspect in some experiments but matched to the perpetrator in others.

Applications and Benefits of the Single-Lineup Paradigm

The single-lineup paradigm should be considered for future research in the field of eyewitness identification. The single-lineup paradigm may be less desirable for manipulations involving witnessing conditions (e.g., Memon, Hope, & Bull, 1993) as the paradigm requires

different actors in each event, reintroducing the confound it was intended to eliminate. For instance, it would seem to be unsuitable for testing the weapon-focus effect, where additional videos manipulating the presence or absence of a weapon are required. However, the paradigm would seem to be ideal in experiments testing effects at retrieval, particularly if the primary manipulation concerns the lineup itself, such as testing the effect of a distinguishing mark and whether it is concealed, or present and replicated, in the fillers. Previously, attempts to equate similarity in perpetrator-present and perpetrator-absent lineups (the revised dual-lineup paradigm; Clark & Tunnicliff, 2001) involved creating both perpetrator-present and perpetrator-absent lineups with fillers matched to the suspect in each lineup. An experiment with three levels of suspect-filler similarity, then, would require six lineups of entirely different individuals. Given the problems inherent in comparing choices made from lineups that contain different individuals outlined above, it would seem beneficial to minimize the number of lineups required. Moreover, the single-lineup procedure reduces the difficulty of trying to balance perpetrator-filler, innocent suspect-filler, and perpetrator-innocent suspect similarity simultaneously to create equivalent perpetrator-present and perpetrator-absent lineups. Finally, because fillers are chosen in the same way as they are in police investigations – that is, by matching them to the one and only suspect in the investigation to create a single lineup – the single-lineup paradigm retains the improved ecological validity of the revised (i.e., suspect-matched) dual-lineup paradigm.

Beyond these, the single-lineup paradigm, in which fillers are matched to the same suspect in the perpetrator present and absent conditions, offers another important advantage. In the traditional dual-lineup paradigm, subjects in the perpetrator absent condition see the person we are calling Person 2 for the first time presented among fillers matched to Person 1. If Person 2 is the most similar alternative to the perpetrator available, Person 2 will be more similar to Person 1 than other lineup members and thus more likely than other members to be chosen. This is also true of the single-lineup paradigm. However, in the single lineup paradigm, Person 2 does not appear in any

lineup. In the traditional dual-lineup paradigm, in contrast, Person 2 appears in a lineup surrounded by fillers matched to Person 1. Hence, the fillers will on average be more similar to the suspect in a perpetrator-present lineup containing Person 1 than in a perpetrator-absent lineup containing Person 2, presenting an intractable confound between similarity and perpetrator presence that is eliminated in the single-lineup paradigm. Although a paradigm that offers better experimental control should be preferred on logical grounds alone, our data do not allow us to determine whether the single-lineup paradigm better simulates the outcome of a real police investigation than other approaches. Nevertheless, because it allows for creating suspect-matched lineups without confounding perpetrator presence with lineup composition, it should be considered for use in laboratory experiments in which the manipulation concerns the lineup itself.

Implications for Lineup Construction

We found detrimental effects of using fillers who were similar to the suspect, but how likely are police to select fillers like these in practice? Perhaps the best answer to this question comes from a nationally-representative survey of U.S. police agencies, which indicated that 31% select "fillers who look as much like the suspect as possible" (p. 59, Police Executive Research Forum, 2013). Our findings could be interpreted as evidence that this approach would make it difficult for a witness to identify guilty suspects and provide little, if any, benefit to innocent suspects (see also Tredoux, 2002; Wells et al., 1993). Nevertheless, for reasons outlined below, our findings need to be interpreted carefully.

We found the best results with "low" similarity fillers, but this label was a necessary convenience to denote relative differences in similarity. Although reducing similarity did not increase rates of innocent suspect identification in our experiments, there is almost certainly a point at which fillers become so dissimilar that the lineup is no longer fair. It is important to remember that the fillers in all the lineups we tested matched a basic description of the suspect and that biased lineups – those containing lineup members who do not match the description of the perpetrator –

must be avoided (Luus & Wells, 1991). Furthermore, in some specific situations fillers should probably be matched to much more than a basic description of the suspect.

When selecting fillers, one must consider how the person under investigation became a suspect (Wells & Penrod, 2011; Wixted & Wells, 2017). Our research would seem to be especially applicable when there is no reason other than coincidence that the suspect, if innocent, would closely resemble the culprit. For example, if an innocent man named Steve became a suspect in a burglary investigation only because Steve committed similar crimes in the past, he should not normally bear a strong resemblance to the real culprit. But it would be a different story if he became a suspect because someone viewed a CCTV recording of the crime and thought, "That kinda looks like my mechanic, Steve." Under these circumstances, there would be a nontrivial risk that Steve would resemble the real culprit. Otherwise, the CCTV footage would not have led the viewer to think of Steve. The same logic applies to suspects under investigation because they match a composite image or a good eyewitness description of the culprit. In these situations, matching fillers to a basic description of the suspect may not provide enough protection for an innocent suspect.

A final point to consider is that low similarity lineups may yield the right outcomes, but for the wrong reasons. If the perpetrator appears in a lineup with a set of duds, an eyewitness might correctly identify that perpetrator even in the absence of recognition. In other words, low similarity may just increase the number of lucky guesses that land on the perpetrator (Wells, Steblay, & Dysart, 2012).

Limitations and Future Directions

Although we conducted three experiments with fairly consistent results, we encourage caution interpreting these findings until future research can be completed to further explore all the relevant parameters of lineup similarity. We are limited in what we can conclude about the effect of suspect-filler similarity on innocent suspect identifications in Experiments 1 and 2 due to the low innocent suspect identification rates in those experiments. Even Experiment 3, which had relatively

high innocent suspect identification rates, may not be conclusive. Increasing suspect-filler similarity did not seem to reduce the innocent suspect identification rate much, but maybe we simply did not increase similarity enough.

Furthermore, suspect-filler similarity is only one dimension of similarity that can be examined in laboratory experiments of eyewitness memory (e.g., perpetrator-innocent suspect similarity; perpetrator-at-event vs. perpetrator-in-lineup similarity). Although all actors matched the same basic description, a limitation of the present research is that we did not manipulate the similarity between the pairs of video actors. The experiments reported here are, to our knowledge, the first manipulations of both filler matching (to perpetrator vs. to suspect) and filler similarity, but computational modelling has been used to simulate the effect of these variables on eyewitness identification decisions (Clark, Erickson, & Breneman, 2011). Clark et al. (2011) modelled outcomes from suspect-matched and perpetrator-matched lineups in WITNESS (Clark, 2003), including parameters such as memory strength, filler similarity, and perpetrator-innocent suspect similarity. Their main focus was on the eyewitness decision strategies, but the simulations nevertheless suggest a complex interaction among the variables examined. Additional factors such as inter-filler similarity, which has received minimal empirical attention, could also come into play.

We chose Person 1-Person 2 pairs that pilot subjects rated as high in similarity – meaning, Person 2 possesses a number of features in common with Person 1, and matches a basic description of him. It does not necessarily follow, however, that the features that make Person 1 memorable are the same features that would be remembered about Person 2, as suggested by the low rate of innocent suspect misidentification in the single-lineup condition of Experiments 1 and 2.

This possibility highlights an important difference between the role of the innocent suspect in the single-lineup and dual-lineup paradigms. In the single-lineup paradigm, subjects viewing a video with Person 2 likely encode features that allow him to be distinguished from similar others (say, his piercing eyes). If Person 1 doesn't have piercing eyes, subjects can quickly and confidently reject a

lineup containing Person 1. Thus, the *absence* of Person 2's memorable feature can become the basis for a lineup rejection. Now consider the role of Person 2 in the traditional dual-lineup paradigm. All subjects view a video with Person 1 and (presumably) encode his most memorable feature (say, his short spiky hair). Person 2 likely shares this feature; he was chosen to serve as the innocent suspect because he was rated as highly similar to Person 1. When Person 2 appears in the perpetrator-absent lineup, subjects are looking for – and find – someone with short, spiky hair. They may notice his piercing eyes but assume that they failed to notice this feature in the video, or that they've forgotten it. Thus, the *presence* of Person 2's memorable feature is not likely to become the basis for a lineup rejection. In the traditional dual-lineup paradigm, the same perpetrator (and therefore, the same features) are encoded by all subjects, irrespective of whether they are assigned to choose from a perpetrator-present or perpetrator-absent lineup. Thus, the dual-lineup procedure in its traditional, single-event form does not allow the possibility of testing for this intransitivity in similarity between the perpetrator and innocent suspect. The effect of perpetrator-innocent suspect transitivity on lineup identification choices is an interesting direction for future exploration of the single-lineup paradigm.

A further issue that is not directly addressed in the present manuscript concerns lineup fairness. The potential for lineups to differ in fairness exists whenever different suspects are involved, and the extent to which a given lineup is fair varies with what the witness recalls about the identity of the perpetrator (Malpass, 1981; Tredoux, 1998). Rather than manipulating perpetrator presence or absence in an otherwise identical lineup as in the single and dual-lineup paradigms, we believe a proper test of lineup fairness can only be carried out by tailoring the lineup to the description provided by the witness (e.g., Wells, Rydell, & Seelau, 1993), which in turn may require as many lineups as there are witnesses. Thus, neither the single-lineup paradigm nor the dual-lineup paradigm would seem to be well-suited to exploring the issue of lineup fairness.

Concluding remarks

We have introduced a new single-lineup paradigm for the scientific investigation of eyewitness identification. The new paradigm improves upon the suspect-matching approach introduced by Clark and Tunnicliff (2001), while retaining its benefits. Applied to filler selection strategies, our findings remind us of the importance of maintaining experimental control while striving for more ecologically-valid research designs. Going forward, we believe researchers who wish to study manipulations that affect retrieval in laboratory experiments should seriously consider using the single-lineup paradigm; whether these experiments ultimately serve as better simulations of real police investigations remains to be seen.

References

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied, 5,* 339-360. doi: 10.1037/1076-898X.5.4.339

Buckhout, R., & Friere, V. (1975). *Suggesting in lineups and photospreads: A casebook for lawyers.* New York, Center for Responsive Psychology. Brooklyn College, Brooklyn, CUNY.

Burton, A. M., White, D., & McNeill, A. (2010). 'The Glasgow Face Matching Test'. *Behavior Research Methods, 42,* 286 – 291. doi: 10.3758/BRM.42.1.286

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,* 364–380. doi: 10.1007/s10979-010-9245-1

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

Clark, S. E., & Tunnicliff, J. L. (2001). Selecting lineup foils in eyewitness identification experiments: experimental control and real-world simulation. *Law and Human Behavior, 25,* 199 – 216. doi: 10.1023/A:1010753809988

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

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

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

Lindsay, R. C. L., Wallbridge, H., & Drennan, D. (1987). Do clothes make the man? An exploration of the effect of lineup attire on eyewitness identification accuracy. *Canadian Journal of Behavioural Science, 19,* 463 – 478. doi: 10.1037/h0079998

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. 10.1007/BF01040622

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

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

Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology, 94,* 339 – 354. doi: 10.1348/000712603767876262

Police Executive Research Forum. (2013). *A National Survey of Eyewitness Identification Processes in Law Enforcement Agencies*. Washington, DC: Author. Retrieved from http://policeforum.org/library/eyewitness-identification/NIJEyewitnessReport.pdf

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.

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

Tredoux, C. (2002). A direct measure of facial similarity and its relation to human similarity perceptions. *Journal of Experimental Psychology: Applied, 8,* 180-193. doi: 10.1037//1076-898X.8.3.180

Tredoux, C. G., Parker, J. F., & Nunez, D. T. (2007). Predicting eyewitness identification accuracy with mock witness measures of lineup fairness: Quality of encoding interacts with lineup format. *South African Journal of Psychology, 37,* 207–222. doi: 10.1177/008124630703700201

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology, 17,* 969–993. doi:10.1002/acp.939

Wells, G. L., & Penrod, S. D. (2011). Eyewitness identification research: Strengths and weaknesses of alternative methods. In B. Rosenfeld & S. D. Penrod (Eds.), *Research methods in forensic psychology* (pp. 237-256). Hoboken, NJ: Wiley and Sons.

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

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

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2012). Eyewitness identification reforms: Are suggestiveness-induced hits and guesses true hits? *Perspectives on Psychological Science, 7,* 264-271. doi: 10.1177/1745691612443368

Wells, G. L., & Turtle, J. W. (1986). Eyewitness identification: The importance of lineup models. *Psychological Bulletin, 99,* 320–329. doi: 10.1037/0033-2909.99.3.320

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25,* 1115-1125. doi: 10.1177/01461672992512005

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: 10.1177/1529100616686966

Figure 1. Top: Traditional dual-lineup paradigm. Bottom: New single-lineup paradigm.

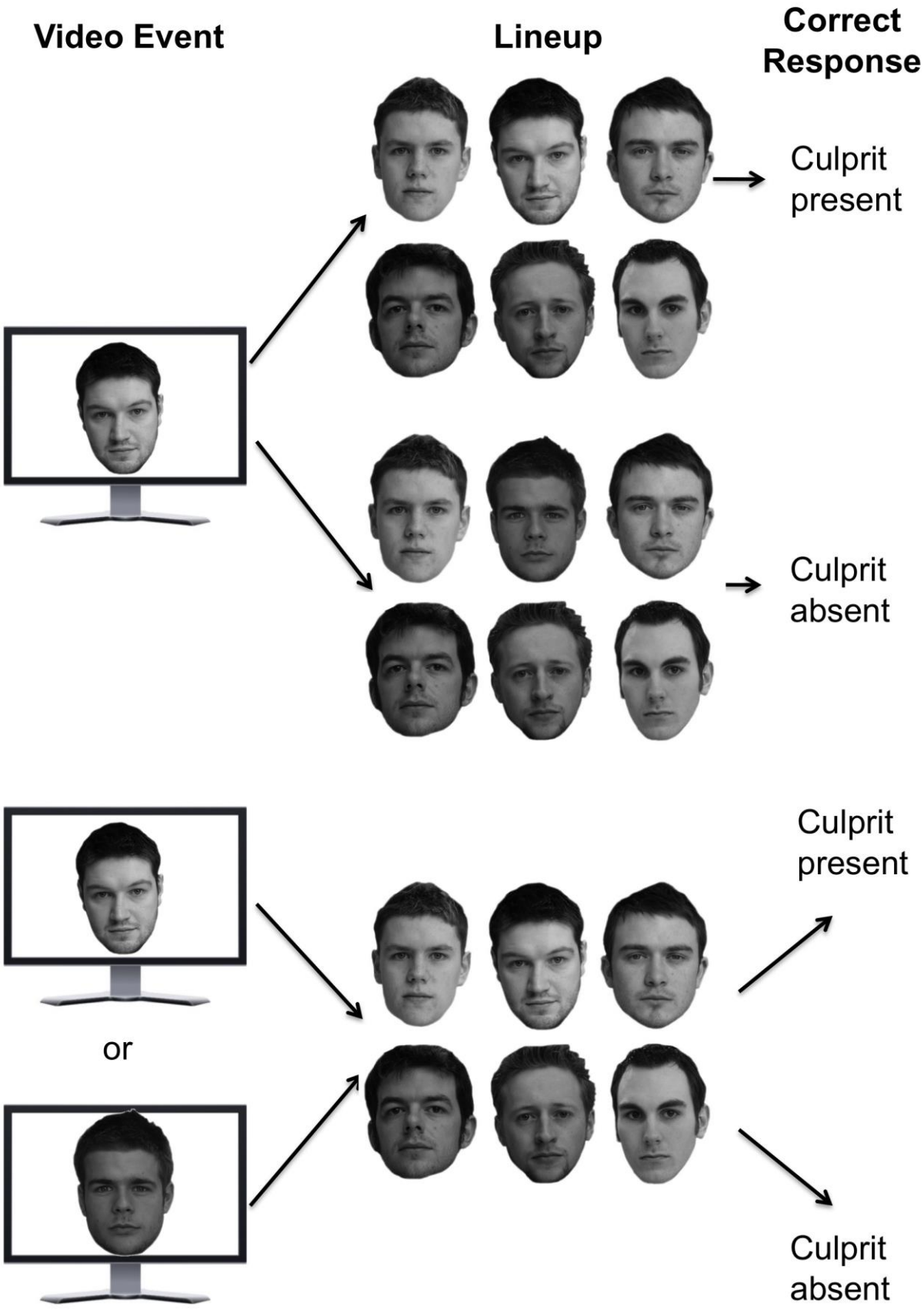

Table 1. *Identification Decisions (%) in Experiment 1*

Perpetrator	Suspect-Filler Similarity	Suspect	Filler	Reject	*n*
Present	Moderate	57.7	14.1	28.2	71
		[46.1, 68.5]	[7.8, 24.0]	[19.0, 39.5]	
	High	56.5	21.7	21.7	69
		[44.8, 67.6]	[13.6, 32.8]	[13.6, 32.8]	
	Very High	24.1	43.1	32.8	58
		[15.0, 36.5]	[31.2, 55.9]	[22.1, 45.6]	
Absent	Moderate	6.9	16.7	76.4	72
		[3.0, 15.2]	[9.8, 26.9]	[65.4, 84.7]	
	High	7.3	21.8	70.9	55
		[2.9, 17.3]	[12.9, 34.4]	[57.9, 81.2]	
	Very High	6.2	38.5	55.4	65
		[2.4, 14.8]	[27.6, 50.6]	[43.3, 66.8]	

Note. Brackets contain 95% confidence intervals. Fillers in the perpetrator-absent lineups were matched to the innocent suspect.

Table 2. *Identification Decisions (%) in Experiments 2 and 3*

Perpetrator	Paradigm	Similarity	Experiment 2				Experiment 3			
			Suspect	Filler	Reject	*n*	Suspect	Filler	Reject	*n*
Present	-	Low	77.3	0.0	22.7	66	78.4	0.0	21.6	51
			[65.8, 85.7]	[0.0, 5.5]	[14.3, 34.2]		[65.4, 87.5]	[0.0, 7.0]	[12.5, 34.6]	
		High	59.7	10.4	29.9	67	56.5	19.4	24.2	62
			[47.7, 70.6]	[5.2, 20.0]	[20.2, 41.7]		[44.1, 68.1]	[11.4, 30.9]	[15.2, 36.2]	
Absent	Suspect-Matched	Low	4.8	9.7	85.5	62	26.2	7.7	66.2	65
			[1.7, 13.3]	[4.5, 19.5]	[74.7, 92.2]		[17.0, 38.0]	[3.3, 16.8]	[54.0, 76.5]	
		High	4.5	18.2	77.3	66	21.0	32.3	46.8	62
			[1.6, 12.5]	[10.7, 29.1]	[65.8, 85.7]		[12.7, 32.6]	[22.0, 44.6]	[34.9, 59.0]	
	Perpetrator-Matched	Low	37.9	8.6	53.4	58	34.8	6.1	59.1	66
			[26.6, 50.8]	[3.7, 18.6]	[40.8, 65.7]		[24.5, 46.9]	[2.4, 14.6]	[47.0, 70.1]	
		High	17.5	27.0	55.6	63	8.0	20.0	72.0	50
			[10.0, 28.6]	[17.6, 39.0]	[43.3, 67.2]		[3.2, 18.8]	[11.2, 33.0]	[58.3, 82.5]	

Note. Brackets contain 95% confidence intervals.