Face-Off: A New Identification Procedure for Child Eyewitnesses

Heather L. Price[1] and Ryan J. Fitzgerald[2]

[1]University of Regina

[2]University of Portsmouth

This article is published in *Journal of Experimental Psychology: Applied*.

This research was supported by a Natural Sciences and Engineering Research Council (NSERC) Discovery Grant to the first author, a NSERC CGS-D grant to the second author, and a University of Regina Dean's Research Award. The authors thank the Educating Youth in Engineering and Science Camp at the University of Regina, Ben Freitag, Jordan Nixdorf, Brittany Whiting, Kaila Bruer, Natalie, Therrien, Briana Vaags, Matthew Pechey, Caitlin Hunter, Jordan Dean, the research assistants, and camp counsellors who assisted with data collection, and the parents and children for their participation in the research.

Ryan J. Fitzgerald, Department of Psychology, University of Portsmouth; Heather L. Price, Department of Psychology, Thompson Rivers University; Tim Valentine, Department of Psychology, Goldsmiths, University of London

Correspondence concerning this article should be addressed to Heather L. Price, who is now at Department of Psychology, Thompson Rivers University, 900 McGill Road, Kamloops, BC, Canada V2C 0C8. E-mail: hprice@tru.ca

Abstract

In 2 experiments, we introduce a new "face-off" procedure for child eyewitness identifications. The new procedure, which is premised on reducing the stimulus set size, was compared with the showup and simultaneous procedures in Experiment 1 and with modified versions of the simultaneous and elimination procedures in Experiment 2. Several benefits of the face-off procedure were observed: it was significantly more diagnostic than the showup procedure; it led to significantly more correct rejections of target-absent lineups than the simultaneous procedures in both experiments, and it led to greater information gain than the modified elimination and simultaneous procedures. The face-off procedure led to consistently more conservative responding than the simultaneous procedures in both experiments. Given the commonly cited concern that children are too lenient in their decision criteria for identification tasks, the face-off procedure may offer a concrete technique to reduce children's high choosing rates.

Keywords: *face-off, simultaneous, showup, elimination, children*

Supplemental materials: http://dx.doi.org/10.1037/xap0000091.supp

Face-Off: A New Identification Procedure for Child Eyewitnesses

Although research on children's eyewitness identifications has been thriving for decades, there have been relatively few procedural advances aimed specifically at improving children's identification accuracy. For many years, researchers examined children's identification behavior on lineup tasks that were intended for adults; however, it is clear that children require procedural accommodations. A consistent concern expressed about child eyewitnesses is their tendency to choose from target-absent lineups (see Fitzgerald & Price, 2015). In the present work, we introduce a new identification procedure developed to manage this tendency to choose by relying on past eyewitness identification literature and knowledge of children's developing cognitive skills.

Existing Lineup Procedures for Children

Although children's propensity to choose could even have a positive impact on lineup tasks in which the target is present (target-present lineups), it has a detrimental effect on lineups that do not include the target (target-absent lineups). Such false eyewitness identifications have been clearly linked to wrongful convictions (Innocence Project, 2015). Researchers have employed a variety of procedures in the hopes of reducing children's false identification rate, but these attempts have been met with mixed success.

Children have not performed well on identification tasks designed for adults. Children's liberal response bias was first revealed using a simultaneous lineup in which all photos are presented at one time (Parker & Carranza, 1989). Given that adults are known to choose less from sequential lineups (in which each lineup member is presented one at a time) than simultaneous lineups (Lindsay & Wells, 1985; Meissner, Tredoux, Parker, & MacLin, 2005; Palmer & Brewer, 2012), early interventions with children naturally focused on the potential utility of sequential presentation. However, sequential presentation has not resulted in reduced choosing in children. Rather, young children (4–6 years) have been reported to commonly choose the first sequential lineup member presented and older children (8–15 years) have been reported to commonly choose multiple members from sequential lineups (Humphries, Holliday, & Flowe, 2012; Lindsay, Pozzulo, Craig, Lee, & Corber, 1997; Parker & Ryan, 1993).

The showup is another procedure commonly used with adults that has been deemed inappropriate for children (Lindsay et al., 1997). At a showup, witnesses view only the person under investigation (i.e., the suspect) and decide whether that person is culprit or not. Although children are more likely to correctly reject a showup than a lineup (Beal, Schmitt, & Dekle, 1995; Dekle, Beal, Elliot, & Huneycutt, 1996; Lindsay et al., 1997), showups are problematic because there are no known-innocent fillers to “siphon” misidentifications away from suspects who are innocent (Wells, Smalarz, & Smith, 2015). Consequently, even when showups reduce the overall false positive rate, innocent suspects are chosen more from showups than from lineups (Lindsay et al., 1997; Yarmey, Yarmey, & Yarmey, 1996).

The inadequacy of procedures intended for adults has led to the development of child-specific identification procedures. One procedure, the elimination lineup, involves dividing a simultaneous lineup into two phases (Pozzulo & Lindsay, 1999). In Phase 1, children choose the lineup member who most resembles the target (fast elimination) or eliminate the lineup member who least resembles the target until only one remains (slow elimination). In Phase 2, children decide whether the remaining lineup member is the target or not. Breaking the decision into two stages was intended to encourage children to first use relative judgments and then make a final absolute judgment. An alternative procedure, the wildcard, focuses on increasing the salience of the reject option by embedding a blank silhouette in the lineup that children can select to report that the target is absent (Zajac & Karageorge, 2009). It is proposed that the active process of choosing a “no choice” option is more appropriate for children and also reminds them visually about the option of not selecting a photo. Elimination and wildcard procedures have been successful in reducing children's false identifications (e.g., Beal et al., 1995; Davies, Tarrant, & Flin, 1989; Dunlevy & Cherryman, 2013; Karageorge & Zajac, 2011; Pozzulo & Balfour, 2006; Pozzulo, Dempsey, & Crescini, 2009; Pozzulo & Lindsay, 1999), indicating that many children require, and can make use of, assistance with resisting the urge to identify a lineup member.

Additional Considerations for Child Witnesses

One area that has not yet been fully explored in the child eyewitness identification literature is reducing the stimulus set size. The elimination and wildcard procedures both begin by presenting all the lineup members (e.g., six to 12) to the witness. This initial presentation of a large stimulus set may perceptually overwhelm children and discourage a thorough evaluation of each alternative.

Evidence from the information and visual search literatures (e.g., Bereby-Meyer, Assor, & Katz, 2004; Hommel, Li, & Li, 2004) suggests that there are robust developmental changes in children's strategies/abilities when searching through alternatives. One factor that contributes to developmental differences in search accuracy across multiple domains is set size. Children, and younger children in particular, have difficulty in accurately choosing among more, relative to fewer, options. For instance, Bereby-Meyer, Meyer, Assor, and Katz (2004) concluded that the cognitive demands of choosing from four, rather than two options led to an increase in use of ineffective decision strategies in 8- to 9- and 12- to 13-year-olds, with older children better able to cope with more options. A common explanation for this difficulty is the increased cognitive demands of searching through a larger stimulus set.

Although children may spend more absolute time searching through a larger than smaller stimulus set, they may not spend as much time on each alternative when the set size is larger (see Davidson, 1991). This is often discussed as a cost/benefit judgment in which participants weigh the benefits of an accurate decision (choosing the right alternative) against the cognitive effort of searching through many options. With a large number of alternatives, the cost of a thorough search becomes too high and each alternative is not evaluated with the same cognitive effort as if the set size was smaller. Such problems with set size are exacerbated in children, or at least become a problem at a smaller minimum set size, because of their already impoverished cognitive abilities. That is, we would expect an increase in set size to have a disproportionately large effect on children relative to adults, and on younger children relative to older children, because of the younger group's already lower attentional and perceptual capacities (see Lavie, Hirst, de Fockert, & Viding, 2004).

There is evidence that children's lineup identification tasks may present too many options and that children may benefit from smaller decision pieces. In their use of a showup procedure with children aged 5- to 6-years-old, Beal, Schmitt, and Dekle (1995; Exp. 2, see also Dekle et al., 1996) found that children were more likely to correctly reject a target-absent showup than a target-absent simultaneous lineup. The authors hypothesized that making a single judgment about a single photograph was easier for children than making multiple judgments about multiple photographs. Other researchers have made similar speculations about the difficulty of having many lineup options, versus few (Beresford & Blades, 2006). Thus, if a lineup task can be divided into more manageable decision pieces, accuracy may increase.

Face-Off: A New Procedure for Child Witnesses

In the present work, we compared a new "face-off" procedure with existing lineup procedures. The face-off procedure is premised on reducing the array of choices. Building on the elimination procedure (Pozzulo & Lindsay, 1999), the face-off procedure adopts a new method for determining the lineup member who best matches the witness' memory of the target. This new method involves breaking the task into a series of binary decisions and never presenting the entire set of lineup members at once.

The face-off procedure comprises several rounds of decision-making. In the first round, children are presented with two photographs and asked to choose the one that looks most like the target. This procedure is repeated for three additional pairs. The chosen photographs from the first round proceed to a subsequent round of face-offs and the nonchosen photographs are eliminated (somewhat analogous to the slow elimination procedure). The face-offs are repeated until only one photograph remains, at which point a blank silhouette (wildcard) is placed beside the surviving photograph. Prior to making a final decision, the administrator reminds the child that none of the pictures may have been of the target, but that if s/he believed the target is the remaining picture the child can point to the picture. Children who do not believe the picture is of the target can point to the silhouette.

Two experiments are reported. In Experiment 1, we compare the face-off procedure with the simultaneous and showup procedures— the two procedures that are reportedly the most frequently used in the field (Police Executive Research Forum, 2013). In Experiment 2, we explore mechanisms of the face-off procedure by comparing it to modified versions of the simultaneous and elimination procedures.

Experiment 1

Method

Children aged 6- to 11-years-old (M_{age} = 8.51, *SD* = 1.22), *N* = 243, were recruited from a summer science camp. Participants witnessed a live art show containing two target people. One day later, participants were randomly assigned to participate in two identification tasks, one for each target.[1] This study was a 3 (identification procedure: simultaneous, showup, face-off) × 2 (target: present, absent) design. Age distributions for each condition were highly similar: simultaneous (*M* = 8.38, *SD* = 1.06, range = 6–11), showup (*M* = 8.58,

[1] Two children withdrew from the experiment after the first lineup task.

SD = 1.26, range = 6–11), and face-off (*M* = 8.59, *SD* = 1.33, range = 6–11). Identification rates for each age are reported in the supplementary materials (Table SM1).

Lineups and showups. Photographs of 18 individuals were required to construct the lineups and showups. The man and woman who acted as targets were photographed locally. For each target person, one innocent suspect and seven fillers were selected from the Glasgow Unfamiliar Face Database (Burton, White, & McNeill, 2010). Although lineups with fewer members are common in some jurisdictions (Police Executive Research Forum, 2013), eight-member lineups have been used in previous research with children (e.g., Keast, Brewer, & Wells, 2007) and in some jurisdictions larger lineups are required (in England the minimum is nine; Police & Criminal Evidence Act, 1984 and accompanying Codes of Practice, 2011).

Lineup members were selected by first placing all individuals from the database who matched the target's general description in an electronic folder. Those photographs were then directly compared with the target person and subjectively ranked according to similarity. The selection of an innocent suspect is an important methodological decision. Wells and Penrod (2011) describe specific examples of applied situations that would lead to an innocent suspect who strongly resembles the culprit, but also note that an innocent suspect may simply match a verbal description of the culprit and that in most instances an innocent person does not become the suspect because he or she is particularly similar to the culprit. We opted to select a midranked lineup member with the aim of producing innocent suspects who were plausible, but not the lineup members who most resembled the culprit.

Photographs of the target, the innocent suspect, and the seven fillers for the female and male lineups are provided in the online supplementary materials (Figures SM1 and SM2, respectively). The composition of the lineups in the face-off and simultaneous conditions was identical. Only the suspect images (i.e., either a target or an innocent suspect) were used in the showup condition. The monochrome images were 2.95” in height and 2.20” in width and printed on cardstock. The order/location of the lineup images was randomly determined across participants.

General procedure. Two research assistants (one male, one female) visited the children’s summer science camp to perform an art show for groups of 15–20 children. The show began with the two research assistants introducing themselves and explaining their roles. The female research assistant performed the role of the artist and indicated that she would execute two art tricks, and then show a video of a messy art trick. While she set up for each trick, her assistant (the male) performed physical activities with the children (e.g., stretching, lunges). Each research assistant was the focus of the children’s attention for approximately half of the show. Prior to the second art trick, the female research assistant spilled water on a laptop computer, which resulted in an inability to show the planned video. Both research assistants expressed worry about the accident and the art show ended early as a result. The entire event lasted approximately 10 min.

The identification tasks were completed individually the following day in one-on-one interviews using paper stimulus materials. Research assistants who administered the tasks were not informed of target identities, however, they may have been able to surmise the identity after administering the task several times. In eyewitness identification research with adults, computer-administered identification tasks are typically used to prevent the possibility of administrator influences on eyewitness behavior. However, for the present investigation, paper materials were deemed more desirable because they allowed the children to observe how their decisions influenced the progression of the face-off procedure. To minimize the risk of influence during the interviews, a relatively large group of research assistants was recruited (19). This reduced the number of identifications per interviewer, which in turn reduced the likelihood that the interviewer would infer the target identities. Interviewers were also trained in best practices in administration of identification tasks (e.g., avoiding bias, not giving feedback). Despite these efforts, it remains possible that experience with prior witnesses may have led to administrator influence (Douglass, Smith, & Fraser-Thill, 2005).

After obtaining verbal assent from children with parental consent, research assistants reminded the children about the visitors who performed the art show the previous day. Children were each administered two identification tasks (one for the male and one for the female research assistant), the order of which was counterbalanced. The same procedure was administered for the two identification tasks (e.g., both simultaneous lineups). The target was always present for one task and absent for the other task, the order of which was counterbalanced across participants.[2] Children in all conditions were warned that the target's picture may or may not be present. Children received a small prize in thanks for their participation.

Simultaneous procedure. The eight photographs were shuffled and placed in a 3 × 3 array, with the center position empty. After presenting the array, children received the following instructions: "I want you to tell me if [target]'s picture is there or not. Remember, [target]'s picture might be here or [target]'s picture might not be here. If you see [target]'s picture, you can point to it. If you do not see [target]'s picture, you can tell me [s]he's not here." Administrators recorded the child's response on a response sheet.

Showup procedure. Children received the following instructions: "When I show you the picture, I want you to tell me if the person is [target] or someone else. Remember, it might be [target] or it might be someone else. Is this [target]?" Children's decisions were then recorded.

Face-off procedure. In anticipation of eventually comparing the face-off procedure with the most conceptually similar existing procedure, the elimination lineup, we examined

[2] For two participants, experimenter error resulted in the administration of two target-absent lineups.

the instructions from the seminal elimination procedure study (Pozzulo & Lindsay, 1999) and adapted them to suit the structure of the face-off procedure. The administrator began with the following instructions: “I’m going to show you two pictures at a time. Each time I show you two pictures, I want you to pick the person who looks most like [target]. It’s important to remember that for now, I don’t need you to pick [target], just the picture that looks most like [him/her].” The eight cards were shuffled and at the presentation of each pair of photos, the administrator reinforced the task instruction: “Which of these looks MOST like [target]?” Four pairs of photographs were shown, with only one pair in view at one time. The selected picture from each pairing was set aside, and then another round of two pairs of the previously selected photographs was shown. The final pairing comprised the last surviving photographs from the second round. In the final round, the remaining photo was then laid in front of the child with the following instructions: “You told me that this picture looked more like [target] than some of the other pictures, but that doesn’t necessarily mean it’s [target]’s picture. Remember, [target]’s picture might not have even been in the pile at all, so this might be a picture of [target] or it might be a picture of someone else. Think back to what [target] looks like. I want you to compare your memory of [target] to this picture. Now I want you to tell me if you think this is [target]’s picture or it is a picture of someone else.” At this final stage, the wildcard picture was introduced with this instruction, “If you do think it’s [target]’s picture, point to this picture. If you don’t think it’s [target]’s picture, point to this question mark [wildcard].”

Results

Table 1 displays identification response rates for the face-off, showup, and simultaneous procedures. Identification responses were categorized as suspect identifications, filler selections, or lineup rejections. Suspect identifications refer to correct identifications of the target in the target-present condition and false identifications of the innocent suspect in the target-absent condition. A lineup rejection occurred if none of the lineup members were chosen, which was the correct response for target-absent lineups and an incorrect response for target-present lineups. Filler selections were always errors. We conducted hierarchical loglinear (HILOG) analyses to explore associations between the procedures and the identification responses. All HILOG analyses were conducted separately for target-present and target-absent conditions. Odds ratios (*OR*), accompanied by 95% confidence intervals in brackets, were computed as an effect size for differences between procedures. Confidence intervals that do not overlap with 1.00 indicate a significant difference ($\alpha = .05$).

The first objective was to determine whether the identification procedure influenced correct identifications (i.e., suspect identification in the target-present condition). A 3 (procedure: face-off vs. showup vs. simultaneous) × 2 (actor: male target vs. female target) × 2 (response: suspect identified vs. not identified)[3] HILOG analysis indicated the three-way interaction was not significant, $\chi^2(2) = 3.07$, $p = .22$. The highest order effect to reach significance was a two-way interaction. Partial associations revealed a significant association between procedure and suspect identifications, $\chi^2(2) = 26.47$, $p = .001$, and a nonsignificant association between actor and suspect identifications, $\chi^2(2) = 1.70$, $p = .19$. Children in the showup condition (84%) were more likely to correctly identify the target than were children in the simultaneous and face-off conditions (both 51%). The odds that the target would be chosen from a showup were nearly five times the odds for the simultaneous and face-off procedures, *OR* = 4.93 [2.36, 10.32].

The next objective was to determine whether the identification procedure influenced correct rejections when the target was absent. A 3 (procedure) × 2 (actor) × 2 (response: lineup rejected vs. not rejected) HILOG analysis indicated the three-way interaction did not exceed the criterion for significance, $\chi^2(2) = 5.60$, $p = .06$, and the highest order effect to

[3] HILOG analyses can typically be used to assess the effect of a variable on the entire pattern of identification responses (i.e., suspect vs. filler vs. no identification); however, this approach would be not appropriate for the current design because one of the responses (filler identification) is not possible for showups.

exceed that criterion was a two-way interaction. Partial associations revealed a significant association between procedure and rejections, $\chi^2(2) = 16.64$, $p = .001$, and a nonsignificant association between actor and rejections, $\chi^2(2) = 0.57$, $p = .45$. The correct rejection rates were higher in the showup (74%) and face-off (65%) conditions than in the simultaneous condition (44%). Relative to the odds for the simultaneous procedure, the odds of a correct rejection were 3.67 [1.89, 7.14] times as great for the showup procedure and 2.39 [1.27, 4.50] times as great for the face-off procedure. The odds ratio for the comparison between the showup and face-off procedure had confidence intervals overlapping with 1.00, indicating a nonsignificant difference, *OR* = 1.53 [0.78, 3.00]. The identity of the actor influenced whether the face-off or showup procedure had the highest correct rejection rate (male target: showup = 81%, faceoff = 55%; female target: showup = 67%, face-off = 74%). The correct rejection rate for the simultaneous procedure was always the lowest (male target = 40%; female target = 48%).

In the preceding analyses, the showup procedure led to the highest accuracy rates. This suggests that the presence of fillers makes it more difficult for children to make the correct decision. However, focusing strictly on accuracy would not be sufficient because some identification errors have more grave consequences than others. The showup procedure led to the highest correct identification rate, but it also led to a much higher misidentification rate of the designated innocent suspect (26%) than did the simultaneous (26% vs. 3%, *OR* = 13.63 [3.08, 60.42]) and face-off (26% vs. 2%, *OR* = 14.21 [3.20, 63.13]) procedures. Contrary to filler misidentifications, which are known errors, innocent suspect misidentifications increase the risk of wrongful conviction.

Another way to assess the risk to an innocent suspect is to compute a “worst case scenario” analysis (Pryke, Lindsay, Dysart, & Dupuis, 2004; Valentine, Darling, & Memon, 2007) in which the most commonly identified filler is designated post hoc as the innocent suspect. This analytical approach gives insight into how the consequences might change if an innocent suspect happens to be the person who most closely resembles the criminal. In the worst case scenario analysis, the post hoc designated innocent suspects were misidentified more often from the simultaneous procedure (male = 42%, female = 14%) than from the face-off procedure (male = 30%, female = 12%), suggesting greater dispersion of errors across the lineup for the face-off procedure, but the differences between procedures were not significant, male lineup: *OR* = 1.67 [0.67, 4.32]; female lineup: *OR* = 1.27 [0.36, 4.52]. There were no fillers for the showup procedure, so the worst case analysis was only applicable to the two lineup procedures.

To assess for global choosing/rejecting patterns, incorrect rejection rates in the target-present condition need to be taken into account. When the target was present, the incorrect rejection rate was lower for showups (16%) than for simultaneous lineups (23%), *OR* = 1.50 [0.68, 3.32]. This trend is not consistent with a response bias explanation for the showup advantage in rejection rates because children in the showup condition only rejected more often than in the simultaneous condition when the target was absent. However, in the comparison between the face-off and simultaneous procedures, the rejection rate for the face-off procedure was higher not only when the target was absent but also when the target was present (face-off = 35%; simultaneous = 23%), *OR* = 1.85 [0.93, 3.72]. This pattern is consistent with a response bias explanation for the face-off advantage in correct rejections.

Diagnosticity. An ideal identification procedure would produce high suspect identification rates when the target is present and low suspect identification rates when the target is absent. Such a procedure would produce suspect identifications that are diagnostic of the suspect’s guilt. In eyewitness identification experiments, a procedure’s diagnosticity is typically judged by the ratio of guilty-to-innocent suspect identifications. However, filler selections and rejections can also be diagnostic of the suspect’s innocence (Wells & Lindsay, 1980; Wells & Olson, 2002) and the almost exclusive focus on the incriminating function of suspect identifications was recently criticized in a major review of the literature (National Research Council, 2014).

To explore the incriminating value of suspect identifications and the exonerating value of filler selections and rejections, diagnosticity ratios for all identification responses were computed (see Table 2). The ratio of relative risk (*RRR*), a statistical test for comparing diagnosticity ratios (Altman & Bland, 2003), indicated that the diagnosticity of suspect identifications from the face-off and simultaneous procedures was greater than the diagnosticity of suspect identifications obtained via the showup procedure, *RRR* = 6.59 [1.56, 27.84] and *RRR* = 6.33 [1.50, 26.44], respectively (confidence intervals that do not overlap with 1.00 indicate statistical significance). Conversely, rejections of showups were more diagnostic of innocence than were rejections of lineups presented with either the face-off or simultaneous procedure, *RRR* = 2.44 [1.32, 4.52] and *RRR* = 2.34 [1.16, 4.70], respectively. For all response types, the difference in diagnosticity between the face-off and simultaneous procedures was negligible. Both the face-off and simultaneous procedures had diagnosticity ratios around 2.0 for lineup rejections. Thus, although the correct rejection rate was higher for the face-off procedure than for the simultaneous procedure, both procedures had approximately twice as many correct rejections as incorrect rejections.

Information gain. In addition to diagnosticity ratios, the base rate of suspect guilt is required to estimate the likelihood that an identification response is indicative of guilt or innocence. The diagnosticity ratios in the previous section assume that the suspect is guilty half of the time, but the actual base rate of guilt is unknown and subject to variation across jurisdictions. To incorporate the diagnosticity ratio and its interaction with the range of possible base rates, an information gain analysis is required.

The information gained from a suspect identification response can be represented as the absolute difference between the base rate of suspect guilt and the posterior probability of suspect guilt given that the suspect was identified (Wells & Lindsay, 1980). Information gain can also be computed for filler and rejection responses, which in some circumstances are more informative than suspect identifications (Wells & Olson, 2002). The information gained can then be plotted as curves, with the range of possible base rates on the x-axis. The information gain curves for the face-off, simultaneous, and showup procedures are displayed in Figure 1. The curves for the simultaneous and face-off procedures are barely distinguishable. Suspect identifications for both lineup procedures are more informative than for the showup procedure, particularly when the base rate of guilt is low. However, rejections from showups are more informative than rejections from the simultaneous and face-off procedures, particularly when the base rate of guilt is high. For those interested in the change in base rate needed to improve the diagnosticity from showups to the more diagnostic procedures (simultaneous and face-off), base-rate effect-equivalency (BREE) curves are reported in the Supplementary Materials (Wells, Yang, & Smalarz, 2015).

***d'*.** As an alternative to the diagnosticity ratio, Mickes, Moreland, Clark, and Wixted (2014) recommend computing a measure derived from signal detection theory: $d' = z\text{Hits} - z\text{False Alarms}$. The proportion of guilty suspect identifications in the target-present condition is treated as the hit rate and the proportion of innocent suspect misidentifications in the target-absent condition is treated as the false alarm rate (Clark, 2012). Because only a misidentification of the innocent suspect is considered a false alarm, erroneous filler selections in the target-absent condition are effectively treated as a correct response and therefore this measure should not be interpreted as a measure of the participant's underlying discriminability (Wells et al., 2015). Instead, *d'* should be interpreted as an indication of how well a procedure sorts between guilty and innocent suspects (Wixted & Mickes, 2015).

Calculation of *d'* produced results that corresponded with those obtained from the diagnosticity ratio. The face-off and simultaneous conditions yielded the highest values (*d'* = 2.01 and *d'* = 1.99, respectively), followed by the showup condition (*d'* = 1.63). A statistical test for comparing two *d'* scores (Gourevitch & Galanter, 1967) indicated none of the differences were significant, *G*s = 0.95, *p*s = .34. We also computed a measure of suspect response bias (Fitzgerald & Price, 2015), the analyses for which can be found in the Supplementary Materials.

Age. Identification rates for each age are reported in the Supplementary Materials. There were not enough children of each age to perform inferential statistics that would be meaningful. A solution to this problem is to group children of several ages together. Although applying an arbitrary cut-off for younger and older children has its drawbacks, looking for differences between the younger and older children may contribute to an understanding of mechanisms of effectiveness. We applied the cut-off used in the most recent meta-analysis of age effects in eyewitness identification (Fitzgerald & Price, 2015) to divide the children into younger (6–8 years) and older (9–12 years) groups. No significant associations involving age were found. Descriptive and inferential statistics for these analyses can be found in the Supplementary Materials.

Survival rates. To explore the path of a suspect through the face-off procedure, Table 3 presents the suspect's survival rate for each round in the target-present and target-absent conditions (see Pozzulo & Lindsay, 1999). Table 3 also presents conditional survival rates, which are survival rates calculated for only participants who selected the suspect in the preceding round. In the target-present condition, the survival rate predictably decreased as participants progressed through the face-off procedure (increases are not possible). Although the difference between survival rates tended to get smaller with each progressive round, the conditional survival rates at each round were relatively stable (84%–87%) and consistent with the rate at which the target survived to the second round (84%). Note that when participants in the showup condition were similarly presented with the target and asked to make a binary decision, the selection rate was also 84% (see Table 1). In the target-absent condition, the innocent suspects survived the first round 51% of the time. The chance expectancy of survival past Round 1 for any lineup member is 50%. Given that the designated innocent suspects were selected as midranked lineup members (in terms of similarity to the target), the correspondence between the survival rate and the chance expectancy is not surprising. The innocent suspect only survived to the wildcard round for nine of the participants (i.e., 11% survival rate). The conditional survival rate for the final choice was 22%, which is similar to the 26% rate of innocent suspect misidentifications in the showup condition (see Table 1).

Discussion

When compared with showups, the face-off procedure resulted in lower rates of correct identifications and lower rates of innocent suspect identifications, but suspect identifications in the face-off procedure were far more diagnostic of suspect guilt and the information gain curves show a clear disadvantage for the showup procedure. The higher correct identification rate and numerically higher correct rejection rate in the showup procedure suggest that making one binary judgment is easier than making several binary judgments. Indeed, the conditional survival rates show that whenever the target was presented for a face-off, children correctly selected that target at rates (84%–87%) that were either the same or very similar to the showup correct identification rate (84%). However, as the number of decisions in the face-off procedure increases, the cumulative probability of a suspect identification decreases. This produced a cost when the suspect was guilty and a benefit when the suspect was innocent (Clark, 2012). The benefit was a reduction in innocent suspect identifications from 26% for the showup procedure to 2% for the face-off procedure. Such a high rate of innocent suspect misidentification for showups underscores the danger that this procedure poses for police suspects who did not actually commit the crime and reinforces previously raised concerns about the procedure (Lindsay et al., 1997).

Compared with the simultaneous procedure, the face-off procedure resulted in better decision-making but did not improve diagnosticity. Given that the face-off procedure led to more correct rejections of target-absent lineups and just as many correct identifications from target-present lineups, decision-making with the face-off procedure was more accurate overall than with the simultaneous procedure. In addition to increasing the correct rejection rate, the face-off procedure increased the incorrect rejection rate, which suggests the face-off procedure induced a more conservative response bias. Diagnosticity was not affected because the more conservative response bias was associated with neither a decrease in correct identifications nor a decrease in misidentification of the innocent suspect. The only observed effect of the conservative response bias was a reduction in filler identifications. From a practical perspective, filler identifications are less worrisome than innocent suspect misidentifications but may nevertheless have undesirable effects on case outcomes. For instance, if an investigation leads to a new suspect, the investigators may require a witness who has already attempted one identification procedure to make another identification attempt (Behrman & Davey, 2001). If the witness identified a filler from the first lineup, the credibility of any subsequent identification could be called in question. Conversely, if the witness rejected the first lineup, the witness could then make a valid identification from the second lineup (Tunnicliff & Clark, 2000).

Experiment 2

The data from Experiment 1 support further exploration of the face-off procedure. Relative to the most commonly used procedure, the simultaneous lineup, the face-off procedure increased correct rejections without negatively affecting correct identifications. The finding that children made more correct decisions in the face-off condition relative to the simultaneous condition was unequivocal. However, consideration of the forensic implications of the children's choices shifts the focus from decision accuracy to suspect identification rates. The face-off and simultaneous procedures elicited highly similar suspect identification rates, both for innocent and guilty suspects. Therefore, in terms of prosecuting the guilty and protecting the innocent, the results of Experiment 1 suggest that these two lineup procedures would have similar consequences.

In Experiment 1 the designated innocent suspects were not intended to be the lineup members who most closely resembled the target. Given that the target-absent misidentification (filler identification + innocent suspect identification) rate was 20% higher in the simultaneous condition than in the face-off condition, a difference in innocent suspect misidentifications between these two procedures might be predicted for lineups containing an innocent suspect who resembles the target more than any of the fillers (Fitzgerald, Price, Oriet, & Charman, 2013). To test this possibility, we designated the lineup member who most resembled the target as the innocent suspect in Experiment 2. We then also manipulated the resemblance of the fillers to the target to vary the degree to which the innocent suspect would stand out from the fillers.

Another feature of Experiment 1 was that only the face-off procedure included a salient rejection option. For applied reasons, comparing the face-off procedure with the two identification procedures that are currently in practice was important, but it nevertheless raises questions about what component of the face-off procedure facilitated the change in children's identification responses. In Experiment 2, we included the wildcard in two comparison procedures: simultaneous and elimination. If the wildcard was responsible for the face-off advantage in correct rejections over the simultaneous procedure in Experiment 1, then the face-off and simultaneous procedures should yield similar correct rejection rates when both procedures include the wildcard. The inclusion of the elimination procedure for comparison was also intended to tease apart the mechanisms underlying the face-off procedure's effect. Both the face-off and the elimination procedure involve a relative judgment phase followed by an absolute judgment phase, whereas the simultaneous procedure is not broken down into distinct phases. Therefore, if both the elimination and face-off procedures outperform the simultaneous procedure, it could be an indication that the separation of relative and absolute decision phases contributed to the face-off advantage. Moreover, if the face-off procedure outperforms the elimination procedure, it would suggest features that are unique to the face-off procedure facilitate children's identification accuracy.

Method

Children aged 6- to 15-years, *N* = 503, were recruited from a university summer science camp. Participants witnessed a magic show interspersed with five video clips containing a target person. One day later, participants were randomly assigned to participate in a lineup task. This study was a 3 (identification procedure: simultaneous + wildcard, elimination + wildcard, face-off) × 2 (target: present, absent) × 2 (target-filler similarity: higher, lower) design. There were no differences in the age distribution across procedures (simultaneous: *M* = 9.92, *SD* = 2.13, range = 6–15; elimination: *M* = 9.90, *SD* = 2.09, range = 6–15; face-off: *M* = 9.89, *SD* = 2.10, range = 6–14). Identification rates for each age are reported in the supplementary materials (Table SM2).

Lineups. The filler selection procedure involved collecting pairwise similarity ratings between photographs of the target person and 200 individuals of the same race and sex. Participants (*N* = 35), who were otherwise independent from the study, were given the following instructions: "In terms of physical appearance, how similar are these two individuals?" The similarity judgments were made on a scale from 1 (highly dissimilar) to 10 (highly similar). Similarity ratings for the set ranged from 1.49 to 6.06 (*M* = 3.51, *SD* = 0.78). Each lineup comprised one suspect, seven fillers, and the wildcard. In target-present lineups, the suspect was the target person. In target-absent lineups, the target person was replaced by an innocent suspect. In this experiment, the person rated to be most similar to the target person (*M* = 6.06) was chosen to be the innocent suspect. Lineups were constructed to have fillers of either "lower" or "higher" similarity to the target person. Both sets of fillers matched a general description of the target person, but target-filler similarity ratings were lower in the low similarity condition (*M* = 2.73, *SD* = 0.34) than in the high similarity condition (*M* = 4.34, *SD* = 0.37).

General procedure. A female magician visited the children's summer science camp to perform four magic tricks for small groups of children (15–20). Prior to the beginning of the show, an introductory video clip (28 s) was shown in which a 29 year-old male (the target) introduced the magician and informed viewers that he would read four lists of words (14 per list) aloud, one after each magic trick, that would later need to be recalled. The magician then performed her first trick, which was followed by a word list (each 23 s). This sequence repeated until four word lists were read and four magic tricks were performed. In each video clip, the target looked directly at the camera and his head and shoulders were in view. In the introductory clip, the target's hair was visible. In each subsequent video, the target wore a different type of hat (jester hat, Viking hat, hockey helmet, baseball cap) to allow for cuing of each word list for an unrelated study. All videos were projected onto a large screen in a university classroom, ensuring good viewing conditions.

The identification task was completed the following day. After obtaining verbal assent from children with parental consent, administrators reminded children about the man who read the word lists the previous day. Children were told that they would view some pictures and would be asked if the man from the videos was in any of the pictures. Children in all conditions were warned that the man's picture may or may not be present. Children received a prize in thanks for their participation.

Procedure conditions. Instructions across procedures were kept as similar as possible. For both the simultaneous and elimination procedures, the eight photographs were shuffled and randomly placed in a 3 × 3 array. For the simultaneous lineup, the wildcard was placed in the center of the array. For the elimination procedure, the center of the array was empty. The administrator then recorded the position of each photograph. For the face-off procedure, cards were shuffled and presented randomly in pairs.

Simultaneous + wildcard procedure. After presenting the 3 × 3 array, children received the following instructions:

> *Think back to what Jordan looks like. I want you to compare your memory of Jordan to each of these pictures. If you see Jordan's picture, you can point to it. If you do not see Jordan's picture, you can point to the question mark in the middle [wildcard].*

Elimination + wildcard procedure. After presenting the photo array, children were told that they were to select the person who looked most like the target. The chosen photograph was then recorded and all other photographs were removed. The administrator then instructed the child:

> *You told me that this picture looked most like Jordan, but that doesn't necessarily mean it's Jordan's picture. Remember, Jordan's picture might not have even been in the pile at all, so this might be a picture of Jordan or it might be a picture of someone else. Think back to what Jordan looks like. I want you to compare your memory of Jordan to this picture. Now I want you to tell me if you think this is Jordan's picture or it is a picture of someone else.*

The administrator then introduced the wildcard and said: "If you think it's Jordan's picture, point to it. If you do not think it's Jordan's picture, point to this question mark [wildcard]."

Face-off procedure. The face-off procedure was the same as in Experiment 1.

Results

Table 4 displays identification response rates in Experiment 2. For the target-present condition, a 3 (face-off vs. elimination vs. simultaneous) × 2 (lower similarity vs. higher similarity) × 3 (suspect vs. filler vs. rejection) HILOG analysis indicated the model that included all three variables did not provide an adequate fit for the data, $\chi^2(4) = 0.92$, $p = .92$. Further, no significant two-way associations were detected, $\chi^2(12) = 7.97$, $p = .79$.

For the target-absent condition, a 3 (face-off vs. elimination vs. simultaneous) × 2 (lower similarity vs. higher similarity) × 3 (suspect vs. filler vs. rejection) HILOG again produced a model with all three variables that did not adequately fit the data, $\chi^2(4) = 1.46$, $p = .83$. The highest order significant effect was a two-way interaction, $\chi^2(12) = 21.54$, $p = .04$. The relation between similarity and lineup response in the partial association analysis did not exceed the significance threshold, $\chi^2(2) = 5.09$, $p = .08$. The largest effect of the similarity manipulation was an increase in filler selection rates in the higher similarity condition (25%) relative to the lower similarity condition (15%), *OR* = 1.97 [1.04, 3.74]. More importantly for the present purposes, partial associations revealed a significant association between lineup procedure and lineup response, $\chi^2(2) = 15.07$, $p = .005$. For correct rejections, the face-off (79%) and elimination (76%) procedures both led to higher rates than the simultaneous procedure (56%), *OR* = 3.01 [1.52, 5.98] and *OR* = 2.43 [1.25, 4.72], respectively. The innocent suspect was also more frequently selected from the simultaneous procedure (16%) than from the face-off (4%) and elimination (10%) procedures, *OR* = 4.78 [1.32, 17.24] and OR = 1.69 [0.66, 4.31], respectively. The designated innocent suspect was the most commonly identified target-absent lineup member in both the lower and higher similarity conditions, so the identity of the innocent suspect would not have changed if we had conducted the type of “worst case scenario” analysis that was reported in Experiment 1.

Diagnosticity and information gain. Diagnosticity ratios were computed to assess the incriminating value of suspect identifications and the exonerating value of filler selections and rejections. Suspect identifications were most diagnostic of guilt when obtained from the face-off procedure, followed by the elimination procedure and, last, the simultaneous procedure (see Table 5). This trend was consistent across the lower and higher similarity conditions. The difference in suspect identification diagnosticity between the face-off and simultaneous procedures (collapsed across similarity) was significant, *RRR* = 3.62 [1.05, 12.47]. No other differences between diagnosticity ratios were significant, *p*s = .07.

Information gain curves for the low and high similarity conditions are displayed in Figures 2 and 3, respectively. For suspect identifications, the curves show that when a lineup has lower similarity fillers and the base rate of guilt is also on the lower end of the spectrum, more information is gained from the face-off procedure than from the elimination and simultaneous procedures. When the base rate is high or when the fillers are high in similarity, the information gained from face-off suspect identifications is more comparable to the information gained from elimination and simultaneous suspect identifications. When a filler was identified from a lower similarity lineup, the face-off procedure led to higher information gain than the elimination and simultaneous procedures. When a filler was identified from a higher similarity lineup, the information gained from the face-off and simultaneous lineups was comparable, and again the elimination procedure led to the lowest information gain. The information gained from face-off filler selections tends to increase along with increases in the base rate. All of the procedures had similar information gain curves for rejections. BREE curves showing the difference in base rate needed to produce suspect identifications with equivalent diagnosticity ratios across procedures are reported in the Supplementary Materials.

***d'*.** As with Experiment 1, calculation of *d'* produced results that corresponded with those obtained from the diagnosticity ratio, and again statistical comparisons between procedures indicated none of the differences in *d'* were significant, *G* = 1.71, *p* = .08. Overall, the face-off procedure yielded the largest *d'* score (*d'* = 1.90), followed by the elimination procedure (*d'* = 1.49), and, finally, the simultaneous procedure (*d'* = 1.35). The direction of this trend was not affected by the similarity manipulation (low similarity condition, face-off: *d'* = 2.04; elimination: *d'* = 1.28; simultaneous: *d'* = 1.11; high similarity condition, face-off: *d'* = 1.82; elimination: *d'* = 1.73; simultaneous: *d'* = 1.64). Suspect response bias analyses are reported in the Supplementary Materials.

Age. We applied the cut-offs used in the most recent meta-analysis of age effects in eyewitness identification (Fitzgerald & Price, 2015) to divide the participants into three groups: 6–8, 9–13, and 14–15 years. Descriptive statistics for each group can be found in the Supplementary Materials. Given the small number of participants in the oldest group (n 19), inferential statistics involving age as a factor were limited to comparisons between 6- to 8- and 9- to 13-year-olds.

Two 3 (procedure: face-off vs. showup vs. simultaneous) × 2 (age: 6–8 vs. 9–13 years) × 3 (response: suspect vs. filler vs. reject) HILOG analyses were performed, one for target-present lineups and one for target-absent lineups. In both cases, the model that retained all three factors was significant: target-present: $\chi^2(2) = 14.61$, *p* = .006; target-absent: $\chi^2(2) = 12.73$, *p* = .01. When the target was present, the age-related increase in correct identifications in the elimination condition (younger 35% vs. older 67%), *OR* = 3.75 [1.35, 10.40], was larger than the age-related increases in the face-off (younger = 41% vs. older = 56%) and simultaneous (younger = 60% vs. older = 63%) conditions, *OR* = 1.81 [0.68, 4.89] and *OR* = 1.12 [0.40, 3.17], respectively. When the target was absent, the advantage in correct rejections for the face-off procedure relative to the simultaneous procedure was larger for the comparison involving older children (face-off = 77% vs. simultaneous = 50%), *OR* = 3.31 [1.47, 7.45], than for the comparison involving younger children (faceoff = 84% vs.

simultaneous = 68%), *OR* = 2.47 [0.63, 9.62]. In the comparison between the elimination and simultaneous procedures, the elimination procedure yielded a higher correct reject rate in the comparison involving older children (elimination = 79% vs. simultaneous = 50%), *OR* = 3.66 [1.60, 8.37], but the two procedures yielded similar correct reject rates in the comparison involving young children (simultaneous = 68% vs. elimination = 65%), *OR* = 1.13 [0.34, 3.77].

Survival rates. The survival rate for each round represents the proportion of witnesses who choose the suspect from the lineup (see Table 6). For analyses of survival rates, only the elimination and face-off lineups can be considered because only a single decision is made in a simultaneous lineup. For the face-off procedure, the wildcard round is particularly informative. At this phase, a selected photograph has survived all pairings. That is, a photograph surviving through to the wildcard round indicates that this photograph was chosen as the most similar to the suspect of all the presented photographs. Surviving to the wildcard round for both procedures, then, involves selection of the "most similar" photograph. Thus, we compared the wildcard round survival rates for the elimination and face-off procedures. For target-present lineups, the differences in survival rates were small and nonsignificant (lower similarity: face-off = 79% vs. elimination = 70%, *OR* = 1.61 [0.62, 4.20]; higher similarity: elimination = 70% vs. face-off = 61%, *OR* = 1.49 [0.59, 3.75]). The differences in survival rates for the innocent suspect in the target-absent lineup with higher similarity fillers was also small and nonsignificant, elimination = 30% versus face-off = 26%, *OR* = 1.22 [0.47, 3.15]. However, in the lower similarity condition, the odds that the innocent suspect would survive to the wildcard round for the elimination procedure were almost three times as great as the odds for the face-off procedure, elimination = 56% versus face-off = 30%, *OR* = 2.97 [1.18, 7.50].

Discussion

Experiment 2 helped to elucidate the mechanisms underlying the face-off procedure's effects. The face-off advantage in correct rejections over the simultaneous procedure was replicated, and this time both procedures included a wildcard. Thus, the observed improvement in children's correct rejection rates cannot be explained by the presence of the wildcard in the face-off procedure. Correct rejection rates were comparable for the face-off and elimination procedures, which both began with a relative judgment phase and progressed to a final absolute judgment of a single photo against a wildcard. Although further research is required to fully tease apart the face-off procedure and link its components to its effects, the comparison between the face-off and elimination procedures suggest that the separation of relative and absolute judgments plays a role. This finding may be an indication that the inclusion of an absolute judgment phase encouraged children to adopt a more stringent criterion for making a positive identification. The suggestion that absolute judgments encourage conservative responding is consistent with findings in the adult eyewitness identification literature. For example, the sequential procedure is theorized to promote an absolute judgment strategy (Lindsay & Wells, 1985) and has been shown to reduce choosing in comparison with the simultaneous procedure (Palmer & Brewer, 2012). Clark (2005, 2012) has argued that a criterion shift will tend to affect correct and false identifications, which is consistent with the data in Table 4. Specifically, in addition to reducing false positives, the face-off and elimination procedures had numerically lower correct identification rates than the simultaneous procedure. However, the differences in the target-present condition were small and nonsignificant, whereas the differences in the target-absent condition were larger and significant.

Given the comparable rates of correct identifications and correct rejections for the face-off and elimination procedures, one might question whether there is any benefit of dividing the task into face-offs. To address this point, we draw attention to the higher diagnosticity ratios and the greater information gain observed for suspect identifications elicited from the face-off procedure when both filler similarity and the base rate of suspect guilt were low. These increases in diagnosticity and information gain were primarily a consequence of the lower innocent suspect identification rate that was observed when low similarity fillers were used with the face-off procedure (3%) relative to when the same fillers were used with the elimination procedure (13%).

The reduced innocent suspect misidentification rate for the face-off procedure may be an indication that parsing the lineup into binary tasks can help to mitigate the biasing effect of low similarity fillers. When presented with a simultaneous lineup containing an innocent suspect who resembles the target and fillers who do not resemble the target, witnesses can experience a “pop out” effect and misidentify the innocent suspect (Ross, Benton, McDonnell, Metzger, & Silver, 2007). Children in both the simultaneous and elimination conditions were presented with all lineup members at once, creating an opportunity for the innocent suspect to stand out from the fillers. By contrast, children in the face-off condition were never presented with all lineup members at once. Thus, dividing the lineup into smaller tasks may have reduced innocent suspect misidentifications by bypassing the biasing effect of simultaneously presenting a strong match with a group of weaker matches. Examination of the survival data provides further support for this point: in the low filler similarity condition, significantly fewer innocent suspects survived to the absolute judgment phase in the face-off procedure than in the elimination procedure. This mitigation of the “pop out” effect mirrors a benefit of the sequential lineup, without the challenges of the sequential lineup previously observed in children (e.g., Lindsay et al., 1997).

In the analysis of filler identifications from low similarity lineups, the face-off procedure also led to higher diagnosticity ratios and information gain than did the other two procedures. The diagnosticity ratios for filler identifications were calculated such they indicated the likelihood that the suspect is innocent. The increase in diagnosticity (and information gain) for filler identifications via the face-off procedure was primarily a consequence of the low filler identification rate when the culprit was present (2% compared with 13% when the culprit was absent). Thus, a filler was much more likely to be chosen from a face-off if the target was absent than if he was present. Statistical considerations are also relevant here. The diagnosticity ratio is a relative measure that is sensitive to the size of the proportion in the denominator (in this case, the target-present filler selection rate). In absolute terms, the difference between the target-absent and target-present filler rates for the simultaneous procedure was slightly larger than the difference for the face-off procedure. To some extent, statistical relativity is also relevant to the face-off advantage in suspect identification diagnosticity for low similarity lineups; however, the absolute difference between the target-present and target-absent suspect identification rates was also larger for the face-off condition than for the other two conditions.

General Discussion

The present data provide early evidence that the face-off procedure controls children's propensity to choose. Children were less likely to mistakenly choose a lineup member from a face-off than from a simultaneous lineup and, importantly, this reduction in choosing did not substantially impact correct identification rates. The result was identification decisions that were more diagnostic of guilt in the face-off procedure than the comparison procedures. The reduction in set size is a key feature of the face-off procedure. Other identification techniques for children, such as the elimination and wildcard procedures, begin by simultaneously presenting all of the lineup members. However, children may have difficulty processing the entire stimulus set at once and using that information to make a decision. Face-off decisions never require choosing from more than two options. Although set size reductions in other domains have led children to use more effective decision-making strategies (Bereby-Meyer, Assor, & Katz, 2004), further work will be required to fully understand how the face-off structure affects lineup responses.

One consideration is that the face-off procedure provides a clear structure for making relative judgments. With simultaneous presentation, children are left to their own devices and we do not know how systematically they compare all lineup members. This may lead children to make an identification because one lineup member is a much better match to their memory than one of the other lineup members. Conversely, if the lineup members are randomly sorted into face-off pairings, as was the case in the present research, the targets and innocent suspects who survive through the rounds should, on average, face off against increasingly stronger competition as the procedure advances. Thus, the more conservative response pattern observed with the face-off procedure may have been the result of having to decide between two strong matches to memory for the final relative decision. This interpretation could be explored in future research by constructing the face-off pairs in a more deliberate fashion. If the pairs were arranged such that the target would initially compete with weak matches and then face off against a strong match for the final relative decision, we would expect the conservative response pattern to be replicated. Conversely, if the pairs were arranged such that the target would initially compete with strong matches and then face off against a weak match for the final relative decision, a more liberal response pattern would be expected to emerge.

A Change in Decision Strategies?

We proposed that one of the primary mechanisms through which the face-off procedure would produce better decision-making in children was by carving the task into smaller decision pieces. The anticipated reduction in cognitive demand as a result of reducing the stimulus set size is not an entirely new idea in children's eyewitness identification research. Humphries, Holliday, and Flowe (2012) speculated that the current video identification system in the United Kingdom may reduce cognitive demand for child witnesses through similar processes. The video lineup procedure requires viewing of the full sequential set of faces at least twice before a selection is made (Police & Criminal Evidence Act, 1984 and accompanying Codes of Practice, 2011). Humphries et al. (2012) suggested that multiple viewings of lineup members facilitated children's ability to compare their memory with the appearance of each lineup member. In the face-off procedure, the target's photo is viewed three times (using relative judgment) before children are encouraged to compare their memory of the target with the presented photo (using absolute judgment) in the final decision. These multiple viewing opportunities may have facilitated a better balance between the absolute and relative judgments offered by the sequential and simultaneous lineups, respectively, with fewer of the noted risks of those procedures for children (Humphries et al., 2012; Lindsay et al., 1997; Parker & Ryan, 1993).

In addition to multiple viewing opportunities, children also have multiple choosing opportunities with the face-off procedure. These multiple choosing rounds allow for exploration of the suspect's path through the procedure. In Tables 3 and 6, we presented the Experiment 1 and 2 survival rates, respectively, for the suspect across the four rounds of the face-off procedure. In the target-present lineup, it is clear that the target remained a strong contender for most of the rounds, ultimately leading to correct identification rates that were similar to the simultaneous procedure. In the target-absent lineup, the initial face-off round weeded out approximately half of the innocent suspects in Experiment 1, and a quarter of the innocent suspects in Experiment 2, with each subsequent round resulting in an approximate 50% reduction in the likelihood that the innocent suspect would be identified for Experiment 1 and a reduction of about a third for each round in Experiment 2. These latter comparisons allay concerns that the multiple opportunities for choosing in the face-off procedure may lead to a commitment effect to a particular photograph.

Procedural Superiority?

The face-off procedure outperformed the comparison procedures in several respects. In Experiment 1, the face-off procedure led to a higher correct rejection rate than the simultaneous procedure and a higher diagnosticity ratio than the showup procedure. In Experiment 2, it again produced a higher correct rejection rate than the simultaneous procedure, and the information gained from the face-off procedure was greater than for the elimination procedure when the lineup contained low similarity fillers. These findings are certainly encouraging.

However, it is premature to suggest that the face-off procedure is superior to the alternatives. Identification procedures need to be tested under a variety of conditions and circumstances before such judgments can be made. Thus, the current data do not solve the problem of which lineup procedure to recommend for children. However, we firmly believe that continuing to innovate in the lineup literature is critical to moving the field toward a better understanding of children's eyewitness identifications, as well as to the development of procedural advances. Even without a consistent advantage in all response options over existing procedures, there was evidence that this new approach may help children to make better identification decisions. As Brewer and Wells (2011) convincingly argued, reevaluating current normative practice is a critical way to move forward in this field (e.g., Sauer, Brewer, & Weber, 2008).

Future Directions

Given the benefits observed with the face-off procedure in the present study, we believe it is worth extending the investigation of the procedure in several ways. Further comparisons between the face-off and elimination procedures may be particularly informative. We only compared the face-off procedure to the fast elimination procedure and it is unclear how it might compare with the slow elimination procedure. Although the slow elimination procedure has been effectively dropped from the literature since it was first introduced, empirical data on its effectiveness is sparse and it is possible that its retirement was premature.

As with any new identification procedure, the face-off procedure should be attempted with witnesses across the life span. Although our focus in the present work was to address child witnesses specifically, the same procedural benefits may indeed apply to witnesses of all ages. Life span explorations of the face-off procedure as it compares to simultaneous lineups of varying sizes may be particularly informative. Further, the patterns observed across the life span can both inform the generalizability of the procedure to other age groups as well as contribute to a more advanced understanding of memory differences in face recognition (see Fitzgerald & Price, 2015). An important benefit of exploring new procedures with all ages lies in the application to the legal system. Procedural recommendations are difficult to make if they differ depending on witness age. At what point does a child become an adult and warrant recommendation of an entirely new procedure? Any new procedure must undergo these practice-driven investigations.

If adapting this procedure for use with adults, researchers could consider equalizing the instructional emphasis on the likelihood that the target was not in the lineup. In the face-off procedure, we reminded children at three critical points spaced over the procedure about the possibility of target absence, whereas in the simultaneous lineup, this warning was issued at two critical points. For child witnesses, we felt strongly that the instruction needed to be repeated during each critical phase to maintain attention. This methodological decision precludes us from ruling out instruction as a potential mechanism of effectiveness of the face-off procedure. However, adults are likely to be better able to hold the instructions in memory and may thus be responsive to a more simplified form of the instructions.

Independent replication of the face-off procedure with new stimuli is essential. Although the present article describes the results of two studies involving three targets and two target events (one live, one video), we encourage interested researchers to explore the procedure themselves. The present exposure times, for example, are longer than those typically used in eyewitness identification research, and larger group differences may be observed when exposure is more fleeting. Further, as with all lineup innovations, to increase the audience for which the procedure is relevant, replication with video lineups is needed. Other lineup procedure innovations may be more challenging to implement with videos (see Beresford & Blades, 2006; Havard, Memon, Clifford, & Gabbert, 2010), but the binary comparisons of the face-off procedure could likely be easily implemented in a side-by-side video format. Extending the face-off procedure to different presentation modalities may assist with more than mere generalizability, but rather may increase our understanding of how children's lineup decisions are made.

Conclusion

The face-off procedure was developed in an attempt to effect change in children's lineup decision processes. Given the commonly cited concern that children are too lenient in their decision criteria for identification tasks, the face-off procedure may offer a concrete technique to reduce children's high choosing rates. Although more work will be needed to determine which child identification procedure performs best across the range of eyewitness circumstances, our findings indicate the face-off procedure is a viable candidate.

References

Beal, C. R., Schmitt, K. L., & Dekle, D. J. (1995). Eyewitness identification of children: Effects of absolute judgments, nonverbal response options, and event encoding. *Law and Human Behavior, 19*, 197–216. doi: 10.1007/BF01499325

Bereby-Meyer, Y., Assor, A., & Katz, I. (2004). Children's choice strategies: The effects of age and task demands. *Cognitive Development, 19*, 127-146. doi: 10.1016/j.cogdev.2003.11.003

Beresford, J., & Blades, M. (2006). Children's identification of faces from lineups: The effects of lineup presentation and instructions on accuracy. *Journal of Applied Psychology, 91*, 1102–1113. doi: 10.1037/0021-9010.91.5.1102

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

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

Brewer. N. & Wells, G. L. (2011). Eyewitness identification. *Current Directions in Psychological Science, 20*, 24-27. doi: 10.1177/0963721410389169

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. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior*, *29*, 395–424. doi:10.1007/s10979-005-5690-7

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

Davidson, D. (1991). Children's decision-making examined with an information-board procedure. *Cognitive Development, 6*, 77-90.

Davies, G., Tarrant, A., & Flin, R. (1989). Close encounters of the witness kind: Children's memory for a simulated health inspection. *British Journal of Psychology, 80,* 415-429.

Dekle, D. J., Beal, C. R., Elliott, R., & Huneycutt, D. (1996). Children as witnesses: A comparison of lineup versus showup identification methods. *Applied Cognitive Psychology, 10,* 1–12. doi: 10.1002/(SICI)1099-0720 (199602)10:1<1::AID-ACP354>3.0.CO;2-Y

Douglass, A. B., Smith, C., & Fraser-Thill, R. (2005). A problem with double-blind photospread procedures: Photospread administrators use one eyewitness's confidence to influence the identification of another eyewitness. *Law and Human Behavior, 29,* 543-562. doi: 2048/10.1007/s10979-005-6830-9

Dunlevy, J. R., & Cherryman, J. (2013). Target-absent identification lineups: Why do children like to choose. *Psychiatry, Psychology, and Law, 20,* 284-293. doi: 10.1080/13218719.2012.671584

Fitzgerald, R. J., & Price, H. L. (2015). Eyewitness identification across the lifespan: A meta-analysis of age differences. *Psychological Bulletin, 14,* 1228-1265. doi: 10.1037/bul0000013

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

Gourevitch, V., & Galanter, E. (1967). A significance test for one parameter isosensitivity functions. *Psychometrika, 32,* 25-33. doi: 10.1007/BF02289402

Havard, K., Memon, A., Clifford, B. R., & Gabbert, F. (2010). A comparison of video and static photo lineups with child and adolescent witnesses. *Applied Cognitive Psychology, 24,* 1209-1221. doi: 10.1002/acp.1645

Hommel, B., Li, K. Z. H., & Li, S-C. (2004). Visual search across the lifespan. *Developmental Psychology, 40*, 545-558. doi: 10.1037/0012-1649.40.4.545

Humphries, J. E., Holliday, R. E., & Flowe, H. D. (2012). Faces in motion: Age-related changes in eyewitness identification performance in simultaneous, sequential, and elimination video lineups. *Applied Cognitive Psychology, 26*, 149-158. doi: 10.1002/acp.1808

Innocence Project (2015). Innocence project. Retrieved from http://www.innocenceproject.org

Karageorge, A., & Zajac, R. (2011). Exploring the effects of age and delay on children's person identifications: Verbal descriptions, lineup performance, and the influence of 'wildcards'. *British Journal of Psychology, 102,* 161-183. doi:10.1348/000712610X507902

Keast, A., Brewer, N., & Wells, G. L. (2007). Children's metacognitive judgements in an eyewitness identification task. *Journal of Experimental Child Psychology, 97,* 286-314. doi:10.1016/j.jecp.2007.01.007

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General, 133*, 339-354. doi: 10.1037/0096-3445.133.3.339

Lindsay, R. C. L., Pozzulo, J. D., Craig, W., Lee, K., & Corber, S. (1997). Simultaneous lineups, sequential lineups, and showups: Eyewitness identification decisions of adults and children. *Law and Human Behavior, 21,* 391-404. doi: 10.1016/j.jecp.2007.01.007

Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70,* 556–564.

Meissner, C. A., Tredoux, C. G., Parker, J. F., & MacLin, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory analysis. *Memory & Cognition, 33,* 783–792. doi: 10.3758/BF03193074

Mickes, L., Moreland, M. B., Clark, S. E., & Wixted, J. T. (2014). Missing the information need to perform ROC analysis? Then compute *d'*, not the diagnosticity ratio. *Journal of Applied Research in Memory and Cognition, 3,* 58-62. doi: 10.1016/j.jarmac.2014.04.007

National Research Council. (2014). *Identifying the culprit: Assessing eyewitness identification*. Washington, DC: The National Academies Press.

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. doi:10.1037/h0093923

Parker, J. F., & Carranza, L. E. (1989). Eyewitness testimony of children in target-present and target-absent lineups. *Law and Human Behavior, 13,* 133–149. doi: 10.1007/BF01055920

Parker, J. F., & Ryan, V. (1993). An attempt to reduce guessing behavior in children's and adults' eyewitness identifications. *Law and Human Behavior, 17,* 11–26. doi: 10.1007/BF01044534

Police and Criminal Evidence Act 1984 and accompanying Codes of Practice (2011). Retrieved from http://www.homeoffice.gov.uk/publications/police/operational-policing/pace-codes/

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

Pozzulo, J. D., & Balfour, J. (2006). The impact of change in appearance on children's eyewitness identification accuracy: Comparing simultaneous and elimination lineup procedures. *Legal and Criminological Psychology, 11*, 25–34. doi: 10.1348/135532505X52626

Pozzulo, J. D., Dempsey, J., & Crescini, C. (2009). Preschoolers' person description and identification accuracy: A comparison of the simultaneous and elimination lineup procedures. *Journal of Applied Developmental Psychology*, *30,* 667-676. doi: 10.1016/j.appdev.2009.01.004

Pozzulo, J. D., & Lindsay, R. C. L. (1999). Elimination lineups: An improved identification procedure for child eyewitnesses. *Journal of Applied Psychology, 84*, 167–176. doi: 10.1037/0021-9010.84.2.167

Pryke, S., Lindsay, R. C. L., Dysart, J. E., & Dupuis, P. (2004). Multiple independent identification decisions: A method of calibrating eyewitness identifications. *Journal of Applied Psychology, 89,* 73-84. doi: 10.1037/0021-9010.89.1.73

Ross, D. F., Benton, T. R., McDonnell, S., Metzger, R., & Silver, C. (2007). When accurate and inaccurate eyewitnesses look the same: A limitation of the 'pop-out' effect and the 10- to 12-second rule. *Applied Cognitive Psychology*, *21,* 677–690. doi: 10.1002/acp.1308

Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General*, *137,* 528-547. doi: 10.1037/a0012712

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

Valentine, T., Darling, S., & Memon, A. (2007). Do strict rules and moving images increase the reliability of sequential identification procedures? *Applied Cognitive Psychology, 21,* 933-949. doi: 10.1002/acp.1306

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

Wells, G. L., & Olson, E. A. (2002). Eyewitness identification: Information gain from incriminating and exonerating behaviors. *Journal of Experimental Psychology: Applied, 8,* 155–167. doi: 10.1037/1076-898X.8.3.155

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*. John Wiley and Sons, Hoboken, NJ.

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

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

Yarmey, A. D., Yarmey, M. J., & Yarmey, A. L. (1996). Accuracy of eyewitness identifications in showups and lineups. *Law and Human Behaviour, 20,* 459–477. doi: 10.1007/BF01498981

Zajac, R., & Karageorge, A. (2009). The wildcard: A simple technique for improving children's target-absent lineup performance. *Applied Cognitive Psychology*, *23*, 358-368. doi: 10.1002/acp.1511

Table 1

Identification response rates in Experiment 1

		Identification Response			
Target	**Procedure**	**Suspect**	**Filler**	**Reject**	***n***
Present	Face-Off	.51	.14	.35	80
	Showup	.84	-	.16	80
	Simultaneous	.51	.26	.23	80
Absent	Face-Off	.02	.33	.65	83
	Showup	.26	-	.74	81
	Simultaneous	.03	.54	.44	80

Table 2

Diagnosticity Ratios in Experiment 1

	Diagnostic of Guilt			Diagnostic of Innocence					
	Suspect Identification			Filler Selection			Rejection		
Procedure	***DR***	***LL***	***UL***	***DR***	***LL***	***UL***	***DR***	***LL***	***UL***
Face-Off	21.33	5.32	85.52	2.36	1.26	4.42	1.86	1.33	2.61
Showup	3.24	2.21	4.74	-	-	-	4.57	2.73	7.66
Simultaneous	20.48	5.13	81.82	2.05	1.35	3.13	1.95	1.21	3.14

Note. DR = diagnosticity ratio. LL = lower limit 95% confidence interval. UL = upper limit of 95% confidence interval. For suspect identifications, diagnosticity was computed as the ratio of target-present/target-absent responses. For filler selections and rejections, diagnosticity was computed as the ratio of target-absent/target-present responses. Diagnosticity ratios were computed as relative risks because the sampling distributions are known and can be used to compute confidence intervals.

Table 3

Suspect survival rates in Experiment 1 (conditional survival rates in parentheses)

	n	**Survive to Second Pairings**	**Survive to Third Pairings**	**Survive to Wildcard Round**	**Final Choice**
Target-Present	80	.84	.70 (.84)	.59 (.84)	.51 (.87)
Target-Absent	83	.51	.28 (.55)	.11 (.39)	.02 (.22)

Table 4

Procedure and similarity effects on identification response rates in Experiment 2

			Identification Response			
Target	**Procedure**	**Similarity**	**Suspect**	**Filler**	**Reject**	***n***
Present	Face-Off	Lower	.53	.02	.45	47
		Higher	.56	.05	.39	41
		Total	**.55**	**.03**	**.42**	**88**
	Elimination	Lower	.56	.05	.40	43
		Higher	.60	.10	.30	40
		Total	**.58**	**.07**	**.35**	**83**
	Simultaneous	Lower	.59	.09	.33	46
		Higher	.68	.08	.24	38
		Total	**.63**	**.08**	**.29**	**84**
Absent	Face-Off	Lower	.03	.13	.85	40
		Higher	.05	.21	.74	42
		Total	**.04**	**.17**	**.79**	**82**
	Elimination	Lower	.13	.08	.80	39
		Higher	.07	.21	.72	43
		Total	**.10**	**.15**	**.76**	**82**
	Simultaneous	Lower	.19	.23	.58	43
		Higher	.12	.34	.54	41
		Total	**.16**	**.29**	**.56**	**84**

Table 5

Diagnosticity Ratios in Experiment 2

		Diagnostic of Guilt			Diagnostic of Innocence					
		Suspect Identification			Filler Selection			Rejection		
Procedure	**Similarity**	***DR***	***LL***	***UL***	***DR***	***LL***	***UL***	***DR***	***LL***	***UL***
Face-Off	Lower	21.28	3.02	150.14	5.95	0.72	49.45	1.90	1.35	2.68
	Higher	11.69	2.96	46.17	4.37	1.01	18.95	1.89	1.24	2.89
	Total	**14.73**	**4.80**	**45.17**	**5.03**	**1.50**	**16.89**	**1.89**	**1.44**	**2.47**
Elimination	Lower	4.36	1.84	10.32	1.64	0.29	9.24	2.01	1.35	3.01
	Higher	8.57	2.80	26.23	2.09	0.70	6.26	2.40	1.45	4.00
	Total	**5.90**	**2.98**	**11.66**	**2.03**	**0.80**	**5.16**	**2.17**	**1.58**	**2.98**
Simultaneous	Lower	3.16	1.61	6.17	2.68	0.91	7.90	1.78	1.10	2.90
	Higher	5.61	2.40	13.11	4.32	1.35	13.85	2.27	1.20	4.29
	Total	**4.07**	**2.41**	**6.89**	**3.45**	**1.56**	**7.57**	**1.96**	**1.33**	**2.89**

Note. DR = diagnosticity ratio. LL = lower limit 95% confidence interval. UL = upper limit of 95% confidence interval. For suspect identifications, diagnosticity was computed as the ratio of target-present/target-absent responses. For filler selections and rejections, diagnosticity was computed as the ratio of target-absent/target-present responses. Diagnosticity ratios were computed as relative risks because the sampling distributions are known and can be used to compute confidence intervals.

Table 6
Suspect survival rates in Experiment 2 (conditional survival rates in parentheses)

	Similarity	*n*	Survive to Second Pairings	Survive to Third Pairings	Survive to Wildcard Round	Final Choice
Target-Present						
Elimination	Lower	43	-	-	.70	.56 (.80)
	Higher	40	-	-	.70	.60 (.86)
Face-Off	Lower	47	.89	.83 (.93)	.79 (.95)	.53 (.67)
	Higher	41	.88	.76 (.86)	.61 (.80)	.56 (.92)
Target-Absent						
Elimination	Lower	39	-	-	.56	.13 (.23)
	Higher	43	-	-	.30	.07 (.23)
Face-Off	Lower	40	.80	.48 (.60)	.30 (.63)	.03 (.10)
	Higher	42	.74	.36 (.49)	.26 (.72)	.05 (.19)

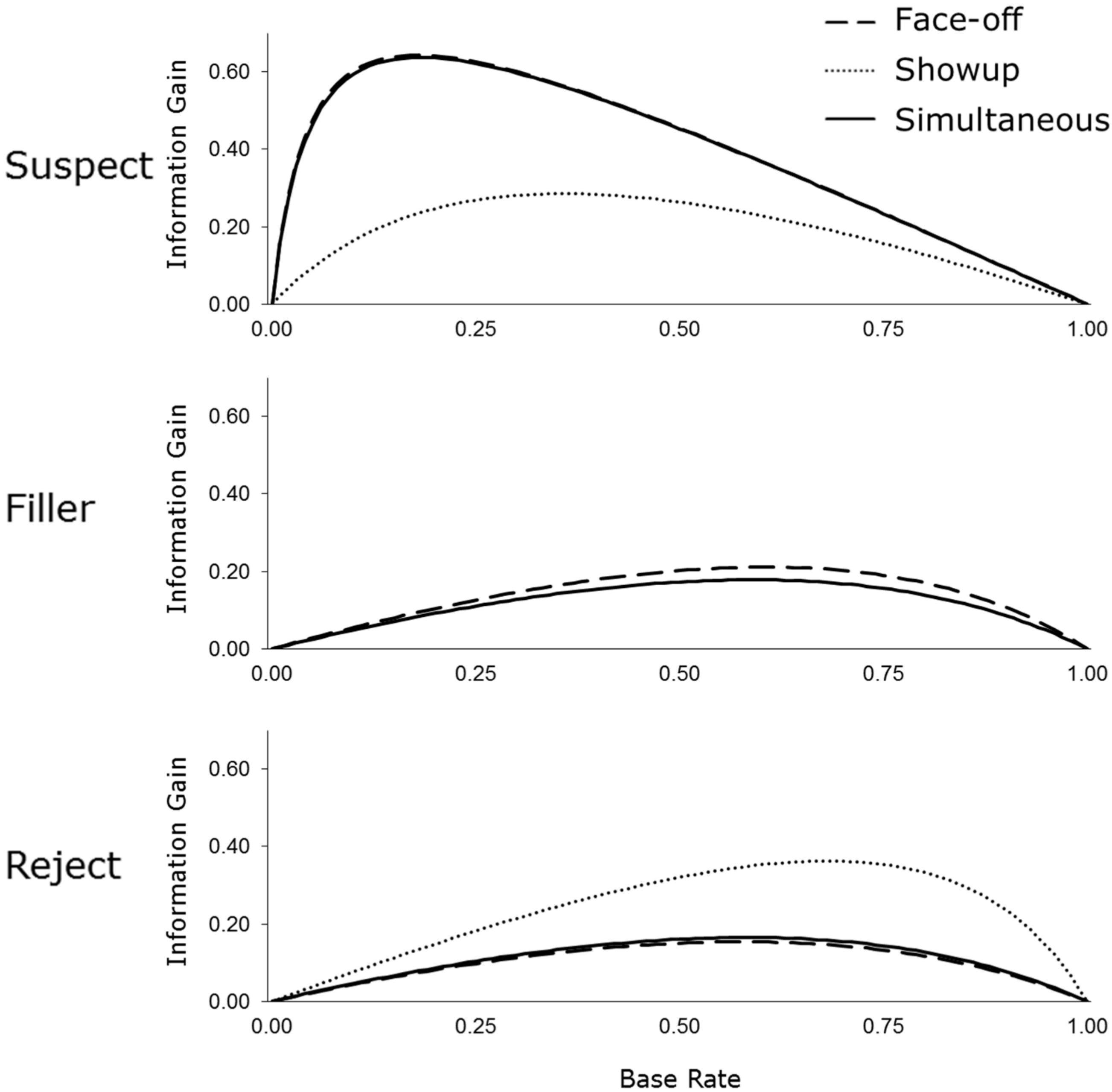

Figure 1. Information gain in Experiment 1.

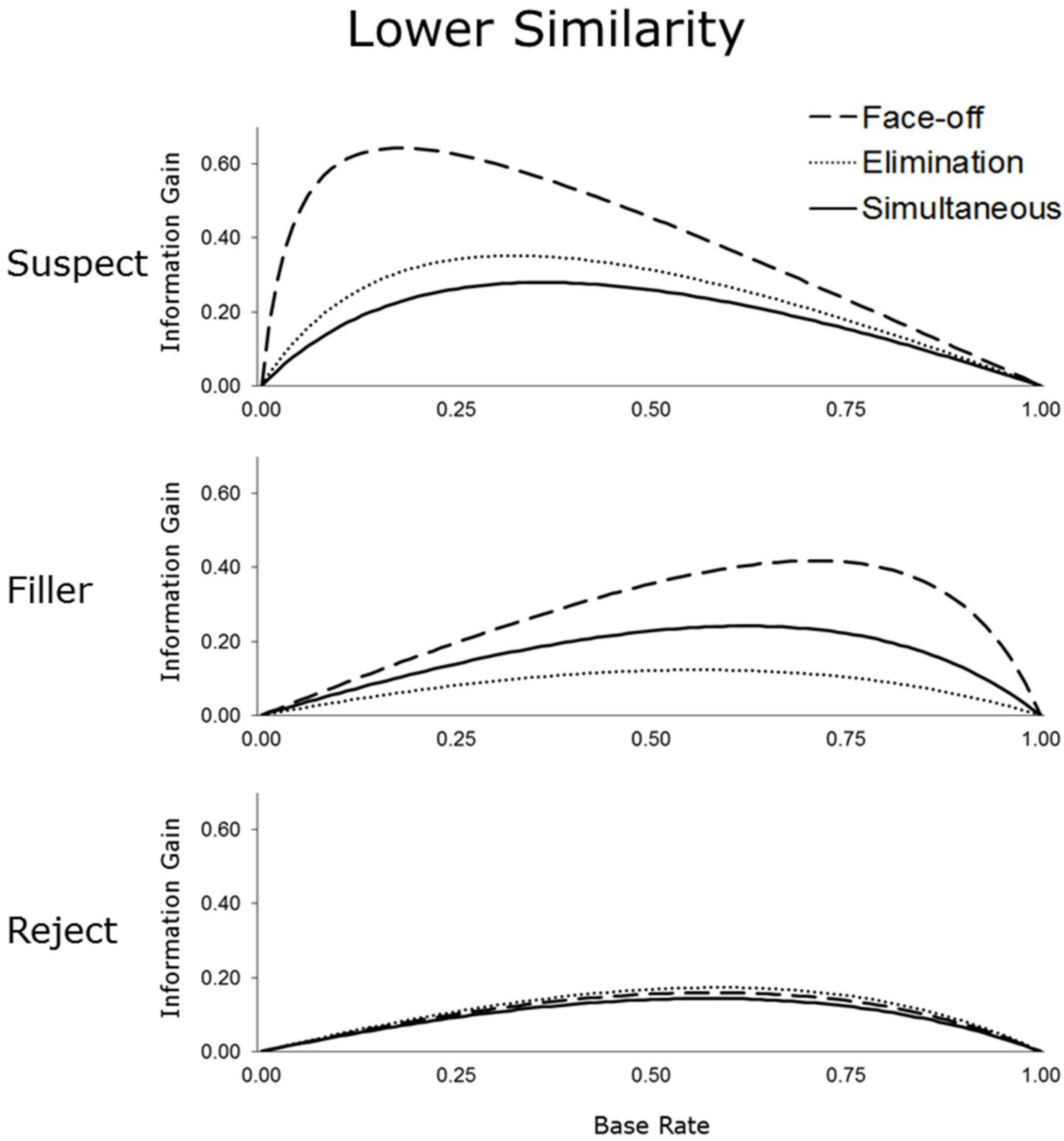

Figure 2. *Information gain for lower similarity lineups in Experiment 2.*

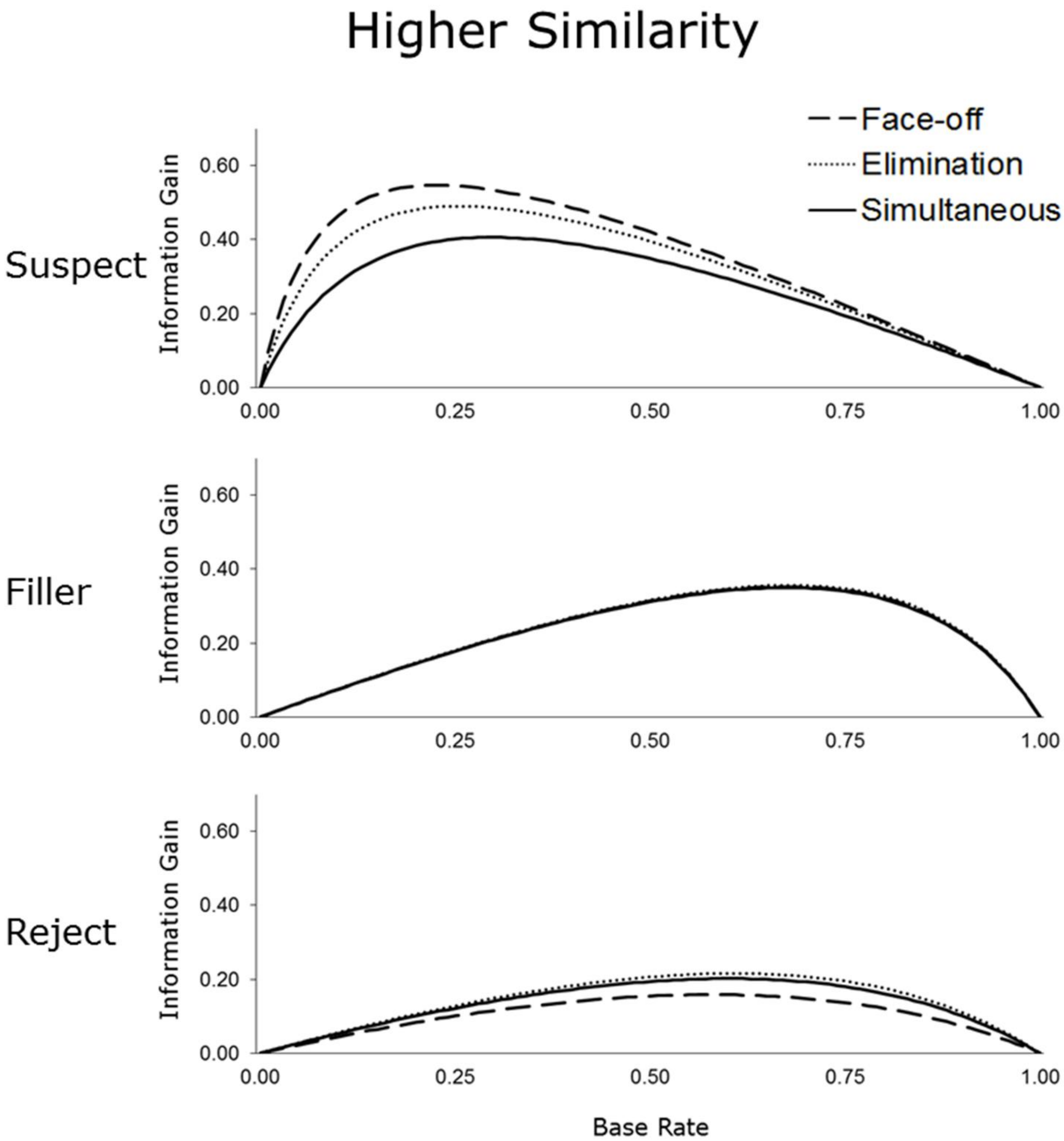

Figure 3. *Information gain for higher similarity lineups in Experiment 2.*