American Psychological Association
American Psychology-Law Society
Law and Human Behavior

ISSN: 0147-7307

2020, Vol. 44, No. 3, 223–237
http://dx.doi.org/10.1037/lhb0000364

Using Machine Learning Analyses to Explore Relations Between Eyewitness Lineup Looking Behaviors and Suspect Guilt

Heather L. Price
Thompson Rivers University

Kaila C. Bruer
University of Regina

Mark C. Adkins
York University

Objective: We conducted 2 experiments using machine learning to better understand which lineup looking behaviors postdict suspect guilt., ***Hypotheses***: We hypothesized that (a) lineups with guilty suspects would be subject to shorter viewing duration of all images and fewer image looks overall than lineups with innocent suspects, and (b) confidence and accuracy would be positively correlated. The question of which factors would combine to best postdict suspect guilt was exploratory. ***Method:*** Experiment 1 included 405 children (6–14 years; 43% female) who each made 2 eyewitness identifications after viewing 2 live targets. Experiment 2 included 342 adult participants (M_{age} = 21.00; females = 75%) who each made 2 identifications after viewing a video including 2 targets. Participants made identifications using an interactive touchscreen simultaneous lineup in which they were restricted to viewing one image at a time and their interaction with the lineup was recorded. ***Results:*** In Experiment 1, five variables (filler look time, suspect look time, number of suspect looks, number of filler looks, and winner look time) together postdicted (with a 67% accuracy score) target presence. In Experiment 2, four variables (number of suspect looks, number of filler looks, number of loser looks, and winner looks) together postdicted (with a 73% accuracy score) target presence. ***Conclusions:*** Further exploration of witness search behaviors can provide context to identification decisions. Understanding which behaviors postdict suspect guilt may assist with interpretation of identification decisions in the same way that decision confidence is currently used.

Public Significance Statement
These experiments suggest that the way in which eyewitnesses visually explore a lineup may help investigators evaluate the likelihood that the guilty perpetrator of a crime is, or is not, in the lineup. A witness' visual exploration of a lineup may help researchers better understand witness decision processes.

Keywords: eyewitness identification, looking behavior, machine learning

Supplemental materials: http://dx.doi.org/10.1037/lhb0000364.supp

When a witness selects from, or rejects, a traditional eyewitness identification lineup, researchers record the identification decision, and this decision is often accompanied by a confidence judgment. At times, decision speed is also measured (e.g., Sporer, 1992). Yet there is typically little other observable behavior recorded that can provide context to the identification decision. In the present research, we adapted a traditional simultaneous lineup to allow for tracking of eyewitness lineup navigation behavior. We tracked which faces children (Experiment 1) and adults (Experiment 2) looked at, how often they looked at each face, and for how long each face was viewed. Using a novel application of machine learning analyses, we then examined which of these looking be-

This article was published Online First February 27, 2020.
Heather L. Price, Department of Psychology, Thompson Rivers University; Kaila C. Bruer, Department of Psychology, University of Regina; Mark C. Adkins, Department of Psychology, York University.
This research was supported by the Canada Research Chairs program and a Natural Sciences and Engineering Research Council of Canada Discovery Grant to Heather L. Price and Social Sciences and Humanities Research Council of Canada Postdoctoral Fellowship to Kaila C. Bruer. The authors thank the many research assistants, children, parents, and teachers for their participation in the research.
Portions of these data were presented at the 2017 Canadian Society for Brain, Behaviour, and Cognitive Science annual conference.
Correspondence concerning this article should be addressed to Heather L. Price, Department of Psychology, Thompson Rivers University, 805 TRU Way, Kamloops, BC V2C 0C8, Canada. E-mail: hprice@tru.ca

haviors optimally discriminated between lineups that did and did not include the guilty suspect (i.e. target presence).

Lineup Behavior

There is growing interest in collecting more nuanced information about witnesses' decision processes by observing behavior during an identification task. For instance, developmental researchers have adapted lineup procedures to track children's responses to individual faces during a lineup task to identify mechanisms related to accuracy (e.g., Bruer, Fitzgerald, Price, & Sauer, 2017). Bruer and Price (2017) showed children lineup members several times in pairs and tracked how often children selected each face as the "most similar" face of the pair to the perpetrator. Children's pattern of suspect selection was predictive of older (9–11 years), but not younger (6–8 years), children's identification accuracy.

Other efforts, primarily with adult witnesses, have focused on obtaining data from eyewitness behavior during a traditional lineup. This work has shown, for example, that faster response times related to greater accuracy among witnesses who chose from the lineup (Brewer & Weber, 2008; Sporer, 1992; Weber, Brewer, Wells, Semmler, & Keast, 2004). Responses that are made more quickly are associated with automaticity of recognition processes and thus with strong memory for the target (e.g., Brewer & Weber, 2008). However, most of this research has explored latency to final decision (i.e. identify a face or reject the lineup) and was not able to individuate looking time for each lineup image, thus limiting our understanding of how viewing behavior may differ across faces (e.g., suspects vs. fillers). Inferences can be made, however, from studies that have explored relations between decision latency and accuracy in showup lineups (lineups in which only one photograph is shown) and found similar patterns (e.g., Sauerland, Sagana, Sporer, & Wixted, 2018). Showups obviously exclude potentially distracting fillers and are an imperfect test of how a witness may interact with a lineup of several images but nonetheless provide complementary data about how responses to an individual image might relate to witness accuracy.

In addition to latency, other information can be obtained about decision processes and memory strength from a systematic observation of lineup behavior. Dunning and Stern (1994) explored the self-reported automaticity of lineup decisions and found that identification decisions that were reportedly made more automatically were more accurate than those made with slower decision processes (see also Dunning & Perretta, 2002). A recent extension of this line of thinking by Charman and Cahill (2012) demonstrated that the strength of memory for fillers can be used to infer the automaticity of the lineup decision process. Stronger filler memory implies that the witness spent more time looking at fillers, which itself implies that the suspect identification involved relatively less automaticity than a witness with weaker filler memory. A decision made with relatively less automaticity implies a more difficult decision, which itself is likely related to lower accuracy. Charman and Cahill suggested that memory for fillers provided a measure of deliberative processing that, in combination with other factors (e.g., decision time), should contribute to a better understanding of automaticity or deliberative processing.

Tracking a witness' lineup navigation and search behavior can be done using eye tracking technology. Three studies have explored visual interaction with lineup images using eye trackers and the findings across studies are relatively consistent: Reduced visual exploration of lineups is related to greater decision accuracy, with accurate witnesses making fewer comparisons across lineup members than inaccurate witnesses (Flowe, 2011; Flowe & Cottrell, 2011; Mansour, Lindsay, Brewer, & Munhall, 2009). These findings support the concept of a "pop-out" effect in which one lineup member immediately stands out from the others (Dunning & Stern, 1994). This initial foray into eye tracking as a method to better understand lineup decisions has not been picked up after these early studies, perhaps because the lack of mobility and the cost and expertise required to use eye tracking technology present a challenge. To address these concerns, we developed a method to track eyewitness navigation behavior using readily accessible technology (e.g., a tablet). By adapting a traditional, simultaneous procedure—hereafter referred to as the Interactive Simultaneous Procedure—we presented witnesses with a lineup identification procedure that gave them control over the lineup (e.g., who they look at, how often they look) using an interactive touch-screen program (as described below).

Machine Learning

Collecting information about how witnesses navigate a lineup allows for the observation of several types of behaviors including, for example, if a witness interacted with the suspect image differently than filler images. Traditionally, such multiple measures of behavior would be entered into a predictive regression model (e.g., stepwise logistic regression) to identify which behaviors predict accurate decisions. In the present research, we instead applied a machine learning approach called support vector machines to achieve a sophisticated classification model based on repeated testing and learning from the association between lineup behavior and target presence (see detailed description below). Both regression and machine learning analyses allow for robust classification of data; however, there are also key differences. Machine learning, specifically support vector analyses, allows for multidimensional analyses of complex data. It is less sensitive to outliers, handles correlated data better than logistic regression, and can handle smaller sample sizes (Salazar, Vélez, & Salazar, 2012). In prior work, support vector machine (SVM) has demonstrated more accurate predictions of human behavior than logistic regression (Amini, Ahmadinia, Poorolajal, & Moqaddasi Amiri, 2016). But perhaps the most essential difference is that logistic regression does not make a "true" prediction (produces probabilities) while SVM does (produces classifications of 1 or 0). Should a researcher lack confidence in the data (e.g., small sample size, confounds) or only be interested in estimates, logistic regression is an appropriate choice. If, however, researchers are interested in robust predictions, SVM may provide a more reliable prediction. In the present study, we introduce SVM as a viable means to predict suspect guilt. Importantly, using machine learning allows us to not only postdict suspect guilt in our sample but also make predictions about likely suspect guilt in future lineups. Throughout the article, we use both of these terms when discussing our particular findings (postdict) and broader implications (predict).

Developmental Considerations

In addition to our interest in exploring a wider range of lineup navigation behaviors, we were also interested in how children and adults might similarly, and differently, approach a lineup task. Child eyewitnesses pose a particular challenge because they are consistently less accurate than adults on lineup tasks (Fitzgerald & Price, 2015). Several efforts have been made over the past decade to improve children's identification performance with some, albeit limited, success (e.g., Brewer, Keast, & Sauer, 2010; Pozzulo & Lindsay, 1999; Zajac & Karageorge, 2009). Children's difficulty with lineups may arise, in part, because lineup tasks have traditionally been developed for adult witnesses, without children's particular needs taken into consideration. There is emerging evidence that this is perhaps a more substantive problem than mere consideration of task difficulty. In the past several years, it has become clear that there are aspects of lineup composition that may uniquely impact children's accuracy.

Stimulus Set Size

Some of the attempts to adapt lineups for children were premised on reducing decision load through reducing stimulus set size (Pozzulo & Lindsay, 1999; Price & Fitzgerald, 2016). Pozzulo and Lindsay (1999) first proposed an elimination lineup that reduced the stimulus set size to a single face (after an initial relative judgment from a standard simultaneous lineup) when children make a final identification decision. Multiple studies by Pozzulo and her colleagues (e.g., Pozzulo & Balfour, 2006) have reported that reducing the number of stimuli to one during the final decision enhances children's decision accuracy. In addition, Price and Fitzgerald (2016) introduced the face-off procedure in which children viewed random pairings of eight photographs and chose which photograph was most similar to a target, after which the nonchosen photographs were discarded. This process continued until only one photograph remained, and children made a final identification decision. Price and Fitzgerald found that this procedure was particularly effective for younger children (6–8 years), a finding they speculated resulted from reducing the stimulus set size. This reduction to more manageable stimulus sets may assist children in focusing more intently on the one or two images presented at a time. However, given the simplistic data gathered about children's decisions when using these procedures, we cannot directly assess if increased attention (e.g., longer views) on each image in a smaller set is the mechanism driving increases in accuracy.

Filler Similarity

Children may also differ from adults in the optimal level of similarity among lineup members (Fitzgerald, Whiting, Therrien, & Price, 2014). Unlike adults, for whom highly similar lineup fillers increase the diagnosticity of suspect identifications (Fitzgerald, Price, Oriet, & Charman, 2013), a lineup of highly similar fillers appears to make an accurate decision out of reach for most children. Fitzgerald et al. (2014) hypothesized that children who were given lineups with very similar fillers may have essentially "given up" due to task difficulty, resulting in much lower accuracy than adults. However, with no other information available in these studies to assess children's decisions (e.g., a measure of children's looking time or decision speed), this explanation was speculative.

It is worth considering the impact of both filler similarity and set size on the automaticity of decisions across all ages. With increases in both set size and filler similarity, the lineup may become a more challenging task, thus increasing the likelihood of deliberative decision processes. However, perhaps a strong target memory can override the difficulty posed by a challenging lineup. In either case, observable behavior, such as decision latency, may reflect memory strength. Thus, in the present study, we manipulated both filler similarity and set size to assess the impact of what we anticipated would be varying levels of lineup difficulty as it relates to observable lineup behavior.

The Interactive Simultaneous Procedure

The Interactive Simultaneous Procedure involved two steps. First, in the *fuzzy* lineup, witnesses saw a simultaneous lineup on a touch-screen tablet with all faces blurred to the point that only the general shape of the face and hair of each image was discernable. To see a face clearly, witnesses were instructed to touch the face. Before looking at the next face, witnesses were required to touch the face again to "reblur" the image. That is, the Interactive Simultaneous Procedure allowed witnesses to view only one clear image at a time, while all other images remained visually available (though heavily blurred).

Witnesses were aware of how many faces there were to choose from, and there were enough visible cues that witnesses could recall which of the previously viewed faces they were interested in considering further. This restricted viewing served three purposes. First, it allowed us to track which face each witness looked at, how often each face was looked at, and for what length of time without the use of eye tracking technology. Second, it allowed us to ensure that each face in the lineup was viewed at least once. Third, though a procedure exists through which the opportunity to make relative judgment comparisons between lineup members is restricted (i.e. the sequential lineup), prior research on the use of the sequential lineup with children has indicated that they struggle with the task, for example, by frequently selecting the first photo presented (e.g., Lindsay, Pozzulo, Craig, Lee, & Corber, 1997). By blurring the images, the task limits the ability to make direct relative comparisons while also minimizing challenges associated with presentation order. Thus, the stimulus set size was functionally reduced by requiring focus on a single image, but access to a complete set of images remained available for navigation.

After viewing each face at least once, witnesses could either revisit images or make their lineup decision (identify the suspect, the filler, or reject the lineup). After making their initial *fuzzy* lineup decision, the second step of the procedure was presented, the *clear* lineup. For this step, witnesses were shown the same lineup again (with faces in a new random order)—but all faces were clear. Witnesses could then affirm or change their original decision. Including the *clear* or second decision allowed us to ascertain whether the procedure associated with the first decision phase impacted children's lineup decision.

The Present Research

We tested the Interactive Simultaneous Procedure with children (Experiment 1) and adults (Experiment 2). Unlike other adapted lineup procedures, the procedure was not designed to impact lineup identification performance but rather to observe lineup navigation behavior. Using the novel machine learning analyses, our primary research goals were to: (a) examine which lineup navigation behaviors could be used to postdict suspect guilt in both children and adults, if any; and (b) assess whether this model was impacted by two important variables—stimulus set size and similarity between lineup members. When an investigator presents a lineup to a witness, as long as the lineup comprises one suspect and known-innocent fillers (consistent with best practice), the witness' decision will provide information about decision accuracy when a filler is identified. Thus, the remaining question about the accuracy of a witness' decision is whether the guilty suspect is present in, or absent from, a lineup. Given that investigators will most often lack conclusive information about whether a suspect is guilty or innocent, we sought to explore the possibility that witness behaviors could postdict suspect guilt. Given the prior literature's large focus on decision accuracy, we draw similar links to our analysis of postdicting target presence (i.e. suspect guilt).

In several studies, when adults and children have been compared, different decision patterns have been evident (e.g., Fitzgerald et al., 2014; Lindsay et al., 1997), thus calling into question how patterns observed in the adult witness literature can be applied to the child witness literature. Given the paucity of research exploring observable witness behavior in children, we did not develop specific hypotheses and considered our child sample to be exploratory. However, for adult witnesses, we can rely on some of the previously established relations with accuracy to speculate which witness behaviors may help us predict whether the suspect is guilty. With the previously established relationship between decision latency and identification accuracy (e.g., Weber et al., 2004), we hypothesized that shorter total lineup looking durations and fewer face views overall would be indicative of suspect guilt. Relatedly, we anticipated that participants viewing guilty suspects would have lower numbers of suspect looks and less time looking at the suspect overall than those viewing innocent suspects. Further, given Charman and Cahill's (2012) finding that strong filler memory is related to lower identification accuracy, we anticipated that participants viewing guilty suspects would evince decreased length and frequency of filler looks than those viewing innocent suspects. The potential differences between number of looks and time spent overall is not something that, to our knowledge, has been explored in prior work and thus was left exploratory.

Given prior relations between witness accuracy and confidence in adults, at least under "pristine" testing conditions (Wixted & Wells, 2017), we also included witnesses' self-reported confidence in our models. With our controlled experimental environment, we anticipated that adults viewing guilty suspects would show increased confidence relative to those viewing innocent suspects. Finally, we anticipated that the combination of several factors relating to the automaticity of decisions would collectively increase the ability to postdict suspect guilt (Charman & Cahill, 2012; Sauerland et al., 2018). However, given that little research has previously examined the relative importance of various looking behaviors as they relate to accuracy, we did not make specific predictions about which behaviors might optimally combine to postdict suspect guilt.

Experiment 1

Method

Participants and design. Experiment 1 included 405 children, 149 younger children (aged 6–8; $M_{age} = 7.39$; 42% female) and 256 older children (aged 9–14; $M_{age} = 9.97$; 44% female) who were quasi-randomly assigned to conditions in a 3 (stimulus set size: 4, 6, 8) × 2 (filler similarity: low, high) × 2 (target: present, absent) design. Each of the 405 children made two identifications (one male target and one female target), for a total of 789 identifications (21 identifications were not completed due to experimenter or technical error). There were an additional 86 participants (aged 6–12; $M_{age} = 8.90$; 30% female), each with two identifications for a total of 161 identifications (11 identifications were not completed due to experimenter or technical error), who served as a comparison sample and completed only a traditional, 8-person, low-similarity, simultaneous lineup procedure (both target-present and target-absent). This comparison sample was included to examine the impact, if any, of the Interactive Simultaneous Procedure on identification accuracy. To obtain stable prediction results with our large number of predictors, we sought a minimum of 300 identification decisions (Gündüz & Fokoué, 2015). This research was approved by the Institutional Research Ethics Board.

Target event and lineup. Children viewed two targets (one male, one female) during a live 10-min event. The following day, children completed the Interactive Simultaneous Procedure two times, once for each target. Participants were told that the target may or may not be present within the lineup. Lineups included one suspect and the remainder fillers. Participants were randomly assigned to view either a 4-, 6-, or 8-person lineup for both identification tasks. The location of images, whether the target was present within the array, and filler similarity (low, high) were all randomized for each lineup task. Fillers and innocent suspects (16 photographs) were selected from the Glasgow Unfamiliar Face Database (Burton, White, & McNeill, 2010). To select the pictures, 100 photographs which matched the two targets' sex and race were preselected (50 females and 50 males). Thirteen independent adult judges were asked to provide pairwise similarity ratings between photographs of each target with its associated 50 potential fillers on a 10-point Likert-type scale (1 = *not at all similar*, 10 = *highly similar*). For the male target, similarity ratings ranged from 2.07 through 6.83. Low similarity lineup members had an average rating of 2.76, while high similarity lineup members had an average rating of 5.86. For the female target, similarity ratings ranged from 1.75 through 5.41. Low similarity lineup members had an average rating of 2.69, while high similarity lineup members had an average rating of 5.03. For both male and female targets, low and high similarity average ratings differed significantly [Male: $t(12) = 12.37$, $p < .001$; Female: $t(12) = 21.61$, $p < .001$]. Similarity ratings for all images are available in the online supplementary materials. For both male and female targets, the target replacement (innocent suspect) was chosen as the most-similarly rated filler (female target replacement = 5.41; male = 6.83). All lineup images (180H × 288W pixels) were displayed on an 11-in. touch screen tablet and shown using *Eprime 2.0* software that recorded participants' responses. Given the prior success observed with salient rejection options, all lineups also included a salient rejection option (i.e., the wildcard; Zajac & Karageorge, 2009).

Procedure. Children attending a week-long science camp experienced the live event in small groups (n = 12–15 per group) during which two targets posed as camp workers who were traveling to each camp group holding a contest to find the "smartest" group. For each group, the targets presented riddles to solve. Targets passed out scorecards and interacted (e.g., gave hints) with each child to ensure children attended to their faces. After tallying the scores, the visitors left. The following day, children with parental consent, and who themselves assented, were invited to work with a researcher individually who administered the procedure. After the experiment was completed, the visitors returned to announce that all groups were winners and provided prizes.

The Interactive Simultaneous Procedure involved two steps. For the first step, the *fuzzy* lineup, children viewed a 4-, 6-, or 8-person simultaneous lineup (with either high or low similarity fillers) on a touch-screen tablet (see Figure 1 for a sample lineup). All faces were blurred, but face/hair outline was discernable, as was the number of faces. To see a face clearly, children touched the face. Before looking at the next face, children were required to touch the face again to "reblur" the image. The program did not allow children to make a decision if they did not look at all faces. Prior to undertaking the *fuzzy* lineup, children practiced with a lineup of puppies to ensure they understood how to navigate the lineup. Then, they received the following instructions:

> Think back to what the two visitors who came to your camp looked like. The computer is going to show you some pictures and a shadow. Out of all of the pictures, there might be a picture of the man/lady or there might not be a picture of the man/lady. Think back to what man/lady looked like. You need to decide if you think the man/lady's picture is there or not.
>
> Just like the puppy pictures, the pictures will be hard to see at first. To see a picture clearly, just touch the picture. You can look at the picture for as long as you want. Once you are done looking, click on the picture again—and it will be hard to see again. Make sure to do this for each picture so that you can look at each person easily. You can go back and look at any picture you want, as many times as you want.
>
> Whenever you are all done looking and are ready to make a decision, press the "I'm Ready" button and tell the computer what you decided by picking the number that matches the picture. If do not see the man/lady's picture, press the "Not Here" button.

Figure 1. Sample *Fuzzy* Lineup with all pictures blurred out except for the guilty suspect (target) face (bottom center). The *wildcard* option was displayed in the center of every lineup. Images are from the Glasgow Unfamiliar Face Database (Burton, White, & McNeill, 2010). Reprinted with permission.

Figure 2. Sample *Clear* target-absent lineup (innocent suspect is bottom center). Images are from the Glasgow Unfamiliar Face Database (Burton, White, & McNeill, 2010). Reprinted with permission.

After children made a decision (i.e. suspect, filler, rejection) they were asked to rate their confidence (from 0 to 10) in their decision.

For the second step, the *clear* lineup, children were shown the same lineup again (in a new random order), but all faces were clearly visible (see Figure 2 for a sample lineup). Children could then affirm or change their original decision (without a reminder of the original decision) and rate their confidence. Children received the following instructions:

> Now, you will see the SAME pictures again. This time, all of the pictures will be easy to see. Look at each picture. Remember, there might be a picture of the man/lady shown or there might not be. One more time, you need to decide if you think the man's/lady's picture is there or not. You can change your answer from the last decision OR you can make the same decision . . . If you don't see the man/lady's picture, press the "Not Here" button.

Results

We divide the results section into two parts. First, we explored children's overall identification accuracy for the Interactive Simultaneous Procedure. We compared the identification accuracy of children's decisions during each phase of the procedure with the decision accuracy of an independent group of children who completed a standard lineup task (traditional simultaneous procedure). Accuracy was defined as the correct lineup decision—either a correct identification in target-present lineups or a correct rejection in target-absent lineups. Second, we explored children's lineup navigation behavior in the *fuzzy* decision phase. This second part of the results is where we systematically examined which looking behaviors postdicted identification target presence (i.e. suspect guilt) in our sample and how this relationship varied as a function of age, stimulus set size, and filler similarity. Our primary focus was not to explore identification accuracy as a function of the manipulated variables as is traditionally done in lineup identification research. Thus, we excluded those traditional analyses from the article, but they are available in the online supplementary materials for interested readers. Identification decisions as a function of similarity and stimulus set size are available in Table 1. All statistical tests were conducted using a significance level of $\alpha = .05$.

How does the interactive simultaneous procedure impact children's lineup accuracy? Children's overall identification accuracy during both decision phases of the procedure were not significantly different (*fuzzy* lineup accuracy = .50; *clear* lineup accuracy = .47, $z = 1.11$, $p = .27$, Cohen's $h = .06$). Recall that after children made a decision in the *fuzzy* lineup, they could confirm or change this decision once they saw the *clear* lineup (same photos presented in a different, random order). Between the *fuzzy* and *clear* lineup phases, relatively few children changed their response (target-present: 16%; target-absent: 20%). Of those changes, fewer than half were to a correct response (target-present: 49%; target-absent: 37%).

We compared children's overall decision accuracy during both the *fuzzy* and *clear* lineup phases with responses from children who were in the independent simultaneous group. In the *clear* phase, for target-present lineups, there was no significant difference in accuracy between the Interactive Simultaneous (.57) and simultaneous (.48) lineups, $z = 1.01$, $p = .31$, Cohen's $h = .18$. However, for target-absent lineups, children in the simultaneous condition were more accurate (.65) than children in the *clear* phase of the Interactive Simultaneous Procedure (.42), $z = 2.06$, $p = .04$, Cohen's $h = .47$. In the *fuzzy* phase, for target-present lineups, there was again no significant difference in accuracy between the *fuzzy* lineup (.58) and the simultaneous condition (.48), $z = 1.20$, $p = .24$, Cohen's $h = .20$. Similar to the target-absent lineups during the *clear* phase, children in the simultaneous condition were more accurate (.65) than children in the *fuzzy* phase (.41), $z = 2.90$, $p < .01$, Cohen's $h = .49$. Thus, it appears that use of the Interactive Simultaneous Procedure negatively impacted children's accuracy when the target was absent, but not when the target was present.

Which lineup navigation behaviors are related to target presence? Next, we examined the lineup navigation behaviors recorded while participants completed the *fuzzy* decision phase of the procedure. The following behaviors were included in subsequent analyses as postdictors (refer to Table 2 for descriptive information):

1. *Look time (any):* average time spent looking at any face.
2. *Looks at each face:* average number of looks at each face (averaged across all faces per participant).
3. *Filler looks*: average number of looks at any filler.
4. *Suspect looks*: number of looks at a suspect (guilty or innocent).
5. *Winner looks*: average number of looks at the winning face (face ultimately selected).
6. *Loser looks*: average number of looks at the losing faces (faces ultimately not selected).
7. *Filler look time*: average time spent looking at any filler.
8. *Suspect look time*: average time spent looking at suspect.
9. *Winner look time:* time spent looking at the winning face.

Table 1
Experiment 1 Final Identification (Clear Lineup) Decision Responding for Each Target

		Low similarity				High similarity			
Target	Set size	Suspect	Filler	Reject	*n*	Suspect	Filler	Reject	*n*
Male									
Target-absent	4	.06	.41	.53	32	.19	.44	.38	32
	6	.11	.43	.46	37	.12	.36	.52	33
	8	.09	.44	.47	32	.03	.62	.35	34
Target-present	4	.70	.15	.15	33	.54	.26	.20	35
	6	.39	.35	.26	31	.54	.26	.20	35
	8	.53	.27	.20	30	.45	.32	.23	31
Female									
Target-absent	4	.15	.24	.62	34	.09	.51	.40	35
	6	.15	.29	.56	34	.09	.47	.44	34
	8	.25	.41	.34	32	.04	.61	.36	28
Target-present	4	.58	.27	.15	33	.53	.19	.28	32
	6	.66	.09	.25	32	.53	.19	.28	36
	8	.63	.23	.13	30	.47	.32	.21	34
Total									
Target-absent	4	.11	.32	.58	66	.13	.48	.39	67
	6	.13	.37	.51	71	.10	.42	.48	67
	8	.17	.42	.41	64	.03	.61	.35	62
Target-present	4	.64	.21	.15	66	.54	.22	.24	67
	6	.52	.22	.25	63	.54	.23	.24	71
	8	.58	.25	.17	60	.46	.32	.22	65

Table 2
Experiment 1 Average Lineup Viewing Behavior as a Function of Stimulus Set Size and Similarity

	Low				High				Total			
Behavior	4	6	8	All	4	6	8	All	4	6	8	All
# Looks/face	1.39	1.23	1.28	1.30	1.43	1.33	1.30	1.35	1.41	1.28	1.29	1.33
Suspect looks	1.58	1.43	1.48	1.50	1.47	1.46	1.46	1.47	1.52	1.44	1.47	1.48
Filler looks	1.32	1.18	1.26	1.25	1.42	1.28	1.28	1.33	1.37	1.23	1.27	1.29
Winner looks	1.24	1.13	1.40	1.26	1.31	1.19	1.40	1.30	1.27	1.16	1.40	1.28
Loser looks	1.27	1.15	1.22	1.21	1.31	1.25	1.23	1.26	1.29	1.20	1.22	1.24
Winner look time	1.65	1.46	1.86	1.66	1.63	1.52	1.84	1.66	1.64	1.49	1.85	1.66
Loser look time	1.57	1.52	1.39	1.49	1.52	1.56	1.46	1.51	1.55	1.54	1.42	1.50
Look time (any)	2.50	2.00	1.98	2.16	2.45	2.29	2.09	2.28	2.47	2.15	2.04	2.22
Suspect look time	2.14	2.01	1.98	2.04	1.93	1.78	1.94	1.88	2.03	1.90	1.96	1.96
Filler look time	1.61	1.54	1.45	1.53	1.60	1.64	1.52	1.59	1.60	1.59	1.49	1.56

Note. There are multiple fillers and multiple losers but only one winning face. Thus, the data for the filler and loser variables are averaged across the averages for each participant. This table, broken down by age group is available in the online supplementary materials. Look times are measured in seconds.

10. *Loser look time*: average time spent looking at any losing face.

We also included participants' confidence ratings in their lineup decision as a postdictor. We explored which combination of these 11 variables could optimally predict target presence. Given that this is the first research to explore these combined behaviors, we examined all behaviors without imposing pre hoc assumptions. To do this, we developed a predictive model using a machine learning analysis in which a computer "learns" through progressively improving on its ability to predict target presence given the existing data. As discussed in the introduction, we used a specific type of machine learning called support vector machine. SVM is typically used for binary classification and regression purposes (Burges, 1998). Given the existing behavioral data, the SVM built a model to predict whether a lineup included the suspect or not.

We used a SVM pattern-recognition approach (Matlab v. 2019a), to determine which of the above lineup behaviors optimally discriminated target-present from target-absent lineups. We used a linear SVM classifier to find a "hyperplane" in a multidimensional space that maximally separated the data into the distinct features (i.e. target-present and target-absent lineups). Like previous research (see Chang & Lin, 2011; Zanette, Gao, Brunet, Bartlett, & Lee, 2016), we used a test/training approach to "learn" which lineup behaviors best predicted target presence. First, we systematically divided participants into two groups: a training and a testing set. Using the training set, the SVM learns the association between the lineup behavior and target presence. Then, using this knowledge, the SVM predicts which of the lineups include the suspect. Next, the SVM tests the accuracy of its prediction. Using a leave-one-out cross-validation approach, this process of learning and testing was repeated until data from each participant was used once as the testing set while all remaining participants were used as the training set.

To select the most parsimonious model with the highest classification performance accuracy, we used a sequential feature selection procedure. We started by including all lineup behaviors (plus confidence) in the initial model and, like Zanette et al. (2016), we eliminated the least important behavior one at a time (by removing the feature with the lowest weight in separating target-present from target-absent lineups in the model) until only one behavior remained. Model classification performance accuracy was recorded with each elimination. The final selected model was the most parsimonious model (i.e. the model with the highest classification accuracy using the fewest predictors). The SVM model with the best performance predicted target presence with 67% accuracy and contained five of the 11 inputted variables: filler look time, suspect look time, suspect looks, filler looks, and winner look time (see Figure 3, Panel A). We then checked to ensure that the final model's prediction accuracy levels were higher than chance. To do this, we constructed an empirical sampling distribution by generating 1,000 permutated instances of the data. That is, we calculated a chance model using all the original feature values (looking behaviors) and the original classifier (target presence or absence) values but randomly reordered (1,000 times) the classifications for each participant. Any prediction accuracy value outside the central 95% of these permuted values was deemed significantly statistically different from a chance level of model accuracy. The final model postdicted target presence at above the chance model's level (49.87%; 95% CI [32.87, 54.07]). This finding provides evidence that these five lineup behaviors can distinguish target-present from target-absent lineups.

Note that the online supplementary materials include results from an analysis of participants' response to the first lineup only (to account for multiple lineup decisions). This analysis resulted in inclusion of the same variables in the final model for both experiments (in Experiment 2, there were a few additional behaviors included, but the confidence intervals were much wider, likely as a result of the reduced statistical power associated with reduced sample size). As such, the results did not differ notably from results in which both identifications were combined.

To assess the model's predictive ability when taking into account set size and filler similarity manipulations, a 2 (target presence: present, absent) $\times$ 2 (actor: male, female) $\times$ 2 (lineup accuracy: accurate, inaccurate) $\times$ 2 (similarity: high, low) $\times$ 3 (stimulus set size: 4, 6, 8) hierarchical log-linear analysis (HILOG) was conducted with SVM model classification accuracy as the dependent variable. Odds ratios (*OR*) and associated 95% confidence intervals were computed as an effect size for significant differences. Confidence intervals that do not overlap with 1.00 represent a significant difference (α = .05). The highest-order

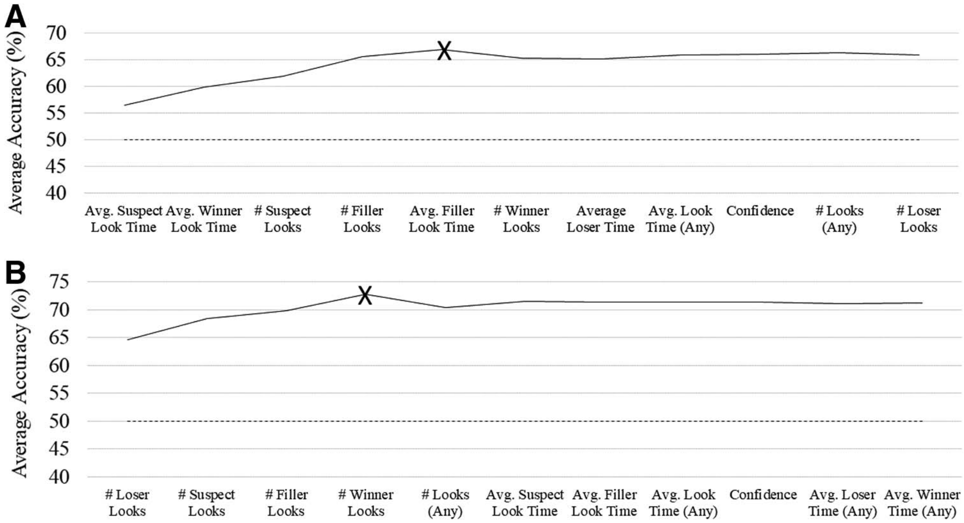

Figure 3. Panels A and B show SVM classification accuracy in predicting target presence (for children and adults, respectively). Each data point represents the average accuracy score when all behaviors listed to the left of the *x*-axis label were included in the model. Dashed line indicates chance distribution, and the X indicates the final model chosen.

effects were two-way effects, $\chi^2(88) = 293.39$, $p < .001$. The model classified target presence better when the target was absent (85% accuracy) than when the target was present (48% accuracy), $\chi^2(1) = 164.34$, $p < .001$, $OR = 6.12$, 95% CI [4.36, 8.61]. In addition, the model was better able to classify target presence in children who made a correct lineup decision (80% accuracy) than children who made an incorrect decision (53% accuracy), $\chi^2(1) = 105.76$, $p < .001$, $OR = 3.58$, 95% CI [2.60, 4.92]. There was no difference in model accuracy across different levels of similarity, stimulus set size, and the different actors (all p's $> .05$).

Further exploration of identified lineup navigation behaviors. To further explore the looking behaviors identified by the SVM models, we constructed a series of 2 (target presence: present, absent) × 2 (lineup accuracy: accurate, inaccurate) × 2 (similarity: high, low) × 3 (stimulus set size: 4, 6, 8) × age (continuous and centered around 0) models using each of the factors identified in the model as the dependent variables. Importantly, these subsequent analyses were conducted independent of the SVM analyses and, as such, describe the nature of the relationship between each behavior and the variables described above (e.g., age, accuracy, target presence, similarity).

We constructed two types of models to further explore the selected dependent variables (i.e. lineup behaviors). General linear models (GLM) were constructed for each continuous dependent variable (e.g., look time), and generalized linear models (GZLM) were constructed for each discrete dependent variable (i.e. count data, such as number of suspect views). All models initially included all main effects and all possible interactions and used only cases with no missing data among the postdictors and dependent variable for each specific model. We then applied a backward stepwise procedure using Akaike's information criterion (see Vrieze, 2012) as a metric for evaluating model parsimony to identify a best fitting model for each dependent variable. Unless otherwise indicated, all analyses used the complete sample ($n = 789$ identifications). All efforts were taken to address any violation of model assumptions. As a final step prior to interpreting results, the false discovery rate correction (FDR; Benjamini & Hochberg, 1995) was applied (set at $\alpha = .05$). See Table 3 for complete model results and online supplementary materials for detailed discussion of adjustments made to models due to violations of assumptions. Below, only significant results are reported.

Filler look time. Due to a technical error, one lineup from a single participant was excluded from the GLM. Accurate children showed a statistically significant decrease in filler looking times ($n = 393$, $M = 1.42$, $SD = 0.54$) relative to children who were inaccurate ($n = 395$, $M = 1.71$, $SD = 0 .85$).

Filler looks. Diagnostic tests suggested our GLM violated heterogeneity assumptions. We took a series of steps to account for this (i.e. applied a heteroscedasticity-consistent covariance matrix model; HCCM; Long & Ervin, 2000; Rosopa, Schaffer, & Schroeder, 2013; see online supplementary materials for full description of these steps). Children who viewed a target-present lineup had a lower average number of looks ($n = 391$, $M = 1.22$, $SD = 0.37$) compared to children who viewed a target-absent lineup ($n = 397$, $M = 1.36$, $SD = 0.50$). Accurate children had a lower average number of looks at any filler ($n = 393$, $M = 1.22$, $SD = 0.41$) relative to inaccurate children ($n = 395$, $M = 1.37$, $SD = 0.47$). There was also a statistically significant main effect of set size on the number of looks at any filler. Post hoc pairwise t tests (with FDR correction) revealed a statistically significant difference in the average number of filler looks between 4-image ($n = 266$, $M = 1.47$, $SD = .53$) and 6-image ($n = 271$, $M = 1.24$, $SD = .41$) set sizes, $t(785) = -3.41$, $p < .01$, and between 4-image and 8-image ($n = 251$, $M = 1.27$, $SD = .37$) set sizes, $t(785) = -2.65$, $p = .01$. Finally, there was a statistically significant main effect of centered age that indicated that as a participant's age increased, their average number of filler looks was expected to also increase, Pearson $r = .12$, $p < .01$.

Suspect look time. Children who viewed target-present lineups and made an accurate lineup decision spent the most time

Table 3
Experiment 1 Follow-up Model Results

Experiment 1	*df*	Filler look time[1]			# Filler looks[2]			Suspect look time			# Suspect looks[3]	
		F	*p*	η_p^2	*F*	*p*	η_p^2	*F*	*p*	η_p^2	χ^2	*p*
Presence	1	1.26	.78	.002	**18.61**	**<.001**	**.03**	**39.92**	**<.001**	**.05**	8.17	.07
Accuracy	1	**33.00**	**<.001**	**.042**	**32.43**	**<.001**	**.03**	6.71	.054	.009	5.63	.16
Similarity	1	0.40	.78	.001	2.60	.23	.006	2.21	.309	.003	0.03	.95
Set size (size)	2	2.92	.24	.008	**6.98**	**.005**	**.02**	0.39	.875	.001	0.94	.75
Age	1	4.17	.24	.006	**9.14**	**.01**	**.01**	2.01	.313	.003	5.16	.16
Presence × Accuracy	1	0.00	1.00	<.001	1.45	.41	.004	**24.65**	**<.001**	**.03**	**14.25**	**.005**
Presence × Similarity	1	1.11	.78	.001	−3.26	.57	.001	—	—	—	0.00	.99
Presence × Size	2	0.67	.78	.002	—	—	—	4.45	.054	.01	0.11	.98
Presence × Age	1	0.17	.85	<.001	0.44	.63	.001	—	—	—	0.48	.68
Accuracy × Similarity	1	0.20	.85	<.001	0.00	1.00	.001	0.86	.637	.001	3.87	.22
Accuracy × Size	2	0.78	.78	.002	0.07	.89	<.001	3.49	.112	.009	3.07	.56
Accuracy × Age	1	0.45	.78	.001	—	—	—	0.00	.978	<.001	0.26	.75
Similarity × Size	2	0.93	.78	.002	1.28	.45	.002	0.59	.771	.002	1.47	.68
Similarity × Age	1	0.47	.78	.001	0.05	.89	.001	0.11	.882	<.001	0.55	.68
Size × Age	2	0.94	.78	.002	0.85	.57	.002	0.70	.747	.002	1.43	.68
Presence × Accuracy × Similarity	1	3.62	.24	.005	2.85	.23	.004	—	—	—	0.84	.68
Presence × Accuracy × Size	2	0.63	.78	.002	—	—	—	—	—	—	1.09	.75
Presence × Accuracy × Age	1	—	—	—	2.47	.23	.003	—	—	—	5.00	.16
Presence × Similarity × Size	2	0.21	.90	.001	—	—	—	—	—	—	1.81	.68
Presence × Similarity × Age	1	6.18	.16	.008	—	—	—	—	—	—	1.55	.56
Presence × Size × Age	2	—	—	—	—	—	—	—	—	—	5.37	.24
Accuracy × Similarity × Size	2	0.15	.90	<.001	—	—	—	0.21	.91	.001	2.05	.68
Accuracy × Similarity × Age	1	0.11	.89	<.001	—	—	—	0.58	.73	.001	0.06	.93
Accuracy × Size × Age	2	0.56	.80	.001	—	—	—	3.17	.14	.008	0.20	.97
Similarity × Size × Age	2	0.19	.90	<.001	2.46	.233	.007	0.08	.98	<.001	4.70	.30
Presence × Accuracy × Similarity × Size	2	2.61	.26	.007	—	—	—	—	—	—	1.71	.68
Presence × Accuracy × Similarity × Age	1	—	—	—	—	—	—	—	—	—	1.06	.68
Presence × Accuracy × Size × Age	2	—	—	—	—	—	—	—	—	—	5.77	.22
Presence × Similarity × Size × Age	2	—	—	—	—	—	—	—	—	—	2.33	.68
Accuracy × Similarity × Size × Age	2	3.19	.24	.008	—	—	—	2.33	.25	.006	1.37	.68
Presence × Accuracy × Similarity × Size × Age	2	—	—	—	—	—	—	—	—	—	6.04	.22

Note. Significant effects are bolded. All *p* values are corrected.
[1] Technical error resulted in loss of one participant's data ($n = 788$). [2] Diagnostic tests suggested our data violated heterogeneity assumptions of GLM. We took a series of steps to account for this (i.e. applied a heteroscedasticity-consistent covariance matrix model; Long & Ervin, 2000; Rosopa, Schaffer, & Schroeder, 2013; see online supplementary materials for full description of these steps). [3] GLM: technical error resulted in loss of one participant's data ($n = 788$). The model was fit using a positive-Poisson distribution with a logarithmic link function (see online supplementary materials for details).

looking at the suspect ($n = 214$, $M = 2.69$, $SD = 2.13$), followed by children who viewed a target-present lineup and made an inaccurate lineup decision ($n = 178$, $M = 1.88$, $SD = 1.40$). Children who viewed a target-absent lineup and made an inaccurate lineup decision spent the next longest time looking at the suspect ($n = 217$, $M = 1.70$, $SD = 1.13$) followed by children who viewed target-absent lineups and made an accurate lineup decision ($n = 180$, $M = 1.49$, $SD = 0.85$).

Suspect looks. Children who made an accurate lineup decision from a target-present lineup had a higher number of suspect looks ($n = 213$, $M = 1.79$, $SD = 0.75$) than children who made an inaccurate decision from a target-present lineup ($n = 178$, $M = 1.36$, $SD = 0.62$). Children who made an accurate lineup decision from a target-absent lineup had a comparable number of suspect looks ($n = 180$, $M = 1.33$, $SD = 0.65$) with children who made an inaccurate lineup decision from a target-absent lineup ($n = 217$, $M = 1.41$, $SD = 0.67$).

Winner look time. A GLM was constructed to explore winner look time using only target present lineups in which a winner was selected ($n = 309$). The final model retained only a single term, set size, and it was not statistically significant, $F(2, 306) = 2.94$, $p = .06$, $\eta_p^2 = .019$.

Discussion

The Interactive Simultaneous Procedure allowed us to track 10 different behaviors while children completed the identification task. These behaviors, plus confidence ratings, were submitted to the machine learning analysis, which identified five behaviors that together significantly postdicted suspect guilt: filler look time, suspect look time, suspect looks, filler looks, and winner looks. Though we did not develop hypotheses particular to child witnesses due to the paucity of relevant literature, had we used the relevant adult literature to do so, our results would have been consistent with the hypotheses. Accurate classification of suspect guilt was related to fewer filler looks and shorter filler look time, more looks at the "winning" or selected face, more looks and longer look time at the suspect in target-present lineups, and shorter looks at the suspect in target-absent lineups. Fillers were viewed more often in target-absent relative to target-present lineups and more often in 4-person than 6- or 8-person lineups. In addition, with increasing age, children viewed fillers more frequently.

Examining the differences identified in the follow-up analyses when looking at the suspect relative to filler faces reveals an interesting story for children. Less look time and lower frequency of filler looks relates to higher accuracy. Fillers were also viewed more often in target-absent than target-present lineups. Further, increased suspect looks were indicative of suspect guilt only when children viewed a target present lineup. These patterns suggest that, for children, less looking at fillers and more looking at suspects may speak to the diagnostic value of a child's identification (i.e. may separate guilty from innocent suspects).

We were particularly interested in whether children's behavioral responses would change as a function of the complexity of the lineup (increased lineup size and filler similarity). There was little evidence that this was the case. Thus, the current data cannot contribute to understanding why increased stimulus set size may negatively impact children's lineup accuracy (e.g., Pozzulo & Lindsay, 1999; Price & Fitzgerald, 2016) nor why more similar fillers might increase lineup difficulty for children (e.g., Fitzgerald et al., 2014). Children appeared to take additional opportunities to view fillers in the smaller 4-person lineup than the 6- and 8-person lineups, but there was little else to suggest that children behaved differently across lineup conditions. The lack of differences between varying lineup sizes may be due to our set sizes not being different enough (i.e., 4-, 6-, or 8-person lineups may have been too similar perceptually to impact children's navigation) or to the restrictions we placed on viewing with the current procedure. We discuss this possibility in more detail in the General Discussion.

Children's general lack of behavioral adjustment to lineup conditions could be a result of at least two factors. One possibility is that children did not adjust to the increased complexity of the stimuli. That is, children may not have been aware of the potentially greater challenge posed by the increase in filler similarity and number of images. As a result, they did not adjust their search behavior. Alternatively, the focused nature of the task, which required children to look at only one photo at a time, may have provided the cognitive support children felt they needed to complete the task. Regardless of the reason for children's lack of adjusting their search strategy, we wondered if the more cognitively savvy adults would adjust their lineup navigation behavior when presented with lineups of varying complexity. Thus, in Experiment 2, we conducted a similar experiment as Experiment 1, but with adults.

Experiment 2

Method

The design of Experiment 2 was similar to Experiment 1, with four exceptions. First, there was no age condition because the sample comprised young adults (N = 342; M_{age} = 21.00, SD = 4.67; females = 75%). Second, to make target exposure appropriate for adults, participants viewed a video rather than live games (targets remained the same). Third, the video depicting two adults playing a game of catch was about 3 min in length and, as such, adults had a shorter exposure duration to the targets than children in Experiment 1. Last, a comparison sample was not included (e.g., the sample of 8-person, low-similarity, simultaneous lineup procedure in Experiment 1). All other aspects of the experiments were the same. Each adult made two identifications, for a total of 671 identifications. Thirteen identifications were not completed due to experimenter or technical error.

Results

The results of the lineup accuracy analyses are available in the online supplementary materials for interested readers, and selection rates across conditions are in Table 4. As with Experiment 1, results from an analysis of participants' first lineup response only are also available in online supplementary materials, which did not differ notably from results reported below.

Which lineup navigation behaviors indicate target presence? We used the same SVM (Matlab v. 2019a) analyses with leave-one-out cross-validation to identify a model to predict which behaviors (if any) optimally discriminated target presence. The accuracy score of the model is in Figure 3 (Panel B). A model with four behaviors classified target presence with 72.73% accuracy: suspect looks, filler looks, loser looks, and winner looks. This model classified at a level significantly higher than chance; Chance model = 49.64%; 95% CI [21.39, 54.55].

To assess the SVM model's classification ability under particular conditions, a HILOG (as described above) was conducted, and the FDR correction was applied to all results. There was a three-way effect between SVM, target, and target presence, $\chi^2(1) = 6.52$, $p = .04$. For the female actor, the SVM model was better able to classify in target-present conditions (82%) than target-absent conditions (65%), $\chi^2(1) = 12.31$, $p < .001$. No effect was found with the male target, $\chi^2(1) = 0.72$, $p = .40$. There was also a three-way effect between SVM, set size, and target presence, $\chi^2(1) = 16.03$, $p < .001$. When eight faces were shown, the SVM model was better able to classify the target-present (78%) than target-absent (60%) condition, $\chi^2(1) = 8.77$, $p < .01$. No differences were found with the 6-, $\chi^2(1) = 3.00$, $p = .08$ or 4-face conditions, $\chi^2(1) = 3.17$, $p = .08$. There was also a three-way effect between SVM, accuracy, and target presence, $\chi^2(1) = 14.00$, $p = .01$. When witnesses made a correct decision, the SVM model was better able to classify target-present (93%) than target-absent conditions (79%), $\chi^2(1) = 19.17$, $p < .001$. The opposite was true when witnesses made an incorrect decision; the SVM model was better able to classify in target-absent (61%) than target-present conditions (38%), $\chi^2(1) = 3.91$, $p < .05$.

Further exploration of identified lineup navigation behaviors. To further explore the navigation behaviors identified in the previous two SVMs, we constructed a series of 2 (target presence: present, absent) × 2 (lineup accuracy: accurate, inaccurate) × 2 (similarity: high, low) × 3 (stimulus set size: 4, 6, 8) × 2 (actor: male, female) models using each behavior as the dependent variables (see Table 5). As with Experiment 1, subsequent analyses are independent of SVM analyses, and two types of models were constructed to explore the relationships among the selected dependent variables and the postdictors (GLM, GZLM; see online supplementary materials for model notes). After model development, the FDR correction was applied (set at $\alpha = .05$). Only significant results are reported in-text. See Table 6 for complete model results and online supplementary materials for detailed discussion of ad-

Table 4
Experiment 2 Final Identification (Clear Lineup) Decision Responding for Each Target

Target	Set size	Low similarity				High similarity			
		Suspect	Filler	Reject	*n*	Suspect	Filler	Reject	*n*
Male									
Target-absent	4	.07	.22	.70	27	.16	.31	.53	32
	6	.03	.31	.66	29	.07	.25	.68	28
	8	.08	.38	.54	26	.04	.50	.46	26
Target-present	4	.75	.07	.18	28	.65	.12	.23	26
	6	.52	.15	.33	27	.55	.21	.24	29
	8	.55	.07	.38	29	.72	.07	.21	29
Female									
Target-absent	4	.19	.15	.67	27	.04	.20	.76	25
	6	.14	.04	.82	28	.00	.25	.75	28
	8	.17	.14	.69	29	.13	.13	.73	30
Target-present	4	.78	.00	.22	32	.76	.07	.17	29
	6	.60	.08	.32	25	.74	.03	.23	31
	8	.81	.15	.04	27	.83	.00	.17	24
Total									
Target-absent	4	.13	.19	.69	54	.11	.26	.63	57
	6	.09	.18	.74	57	.04	.25	.71	56
	8	.13	.25	.62	55	.09	.30	.61	56
Target-present	4	.77	.03	.20	60	.71	.09	.20	55
	6	.56	.12	.33	52	.65	.12	.23	60
	8	.68	.11	.21	56	.77	.04	.19	53

justments made to models due to violations of assumptions. Unless otherwise indicated, all analyses used the complete sample of identifications ($n = 671$).

Suspect looks. The final GZLM retained five three-way interactions and all lower order effects. After FDR correction, there were no statistically significant interactions or main effects.

Filler looks. The number of filler looks was higher in target-absent ($n = 335$, $M = 1.90$, $SD = .83$) than target-present ($n = 336$, $M = 1.45$, $SD = 0.62$) lineups, in high ($n = 337$, $M = 1.77$, $SD = .80$) than low similarity ($n = 334$, $M = 1.58$, $SD = 0.73$) lineups, and in male-target ($n = 336$, $M = 1.76$, $SD = 0.76$) than female-target ($n = 335$, $M = 1.58$, $SD = 0.77$) lineups. Further, accurate witnesses had a significantly lower number of filler looks ($n = 455$, $M = 1.54$, $SD = 0.67$) than inaccurate witnesses ($n = 216$, $M = 1.97$, $SD = 0.88$).

Winner looks. A zero-truncated GZLM was constructed to explore winner looks. No analyses were conducted for target-absent lineups or lineups in which no winner was selected, as no winner can be correctly selected in this condition. The backward stepwise procedure retained a three-way interaction between lineup accuracy, lineup similarity, and set size, though after applying the FDR correction, there were no statistically significant results in the model.

Loser looks. For target-absent lineups, there was a comparable average of loser looks for accurate ($n = 223$, $M = 2.18$, $SD = 0.86$) and inaccurate ($n = 112$, $M = 2.01$, $SD = 1.07$) decisions. However, for target-present lineups, there was a marked decrease in loser looks for accurate ($n = 232$, $M = 1.32$, $SD = 0.56$) versus inaccurate ($n = 104$, $M = 2.11$, $SD = 0.91$) decisions. Each of the remaining three main effects not involved in the interaction were statistically significant. There

Table 5
Experiment 2 Average Lineup Looking Behavior as a Function of Stimulus Set Size and Similarity

Behavior	Low				High				Total			
	4	6	8	All	4	6	8	All	4	6	8	All
# Looks/face	1.73	1.64	1.64	1.67	1.90	1.87	1.78	1.85	1.81	1.76	1.71	1.76
Suspect looks	2.10	2.10	2.05	2.08	2.22	2.21	2.33	2.25	2.16	2.16	2.19	2.17
Filler looks	1.61	1.55	1.58	1.58	1.79	1.81	1.70	1.77	1.70	1.68	1.64	1.67
Winner looks	1.24	1.10	1.62	1.32	1.50	1.38	1.54	1.47	1.37	1.24	1.58	1.40
Loser looks	1.89	1.75	1.64	1.76	2.03	1.97	1.81	1.94	1.96	1.86	1.73	1.85
Winner look time	1.73	1.67	1.30	1.57	1.92	1.82	1.69	1.81	1.82	1.75	1.49	1.69
Loser look time	1.60	1.51	1.63	1.58	1.62	1.55	1.56	1.58	1.61	1.53	1.59	1.58
Look time (any)	1.81	1.76	1.77	1.78	1.90	1.82	1.80	1.84	1.86	1.79	1.78	1.81
Suspect look time	2.42	2.42	2.15	2.33	2.57	2.51	2.32	2.47	2.50	2.46	2.23	2.40
Filler look time	0.42	0.31	0.34	0.36	0.36	0.33	0.35	0.34	0.39	0.32	0.34	0.35

Note. There are multiple fillers and multiple losers, but only one winning face. Thus, the data for the filler and loser variables are averaged across the averages for each participant.

Table 6
Experiment 2 Follow-up Model Results

		Winner looks[1]		Suspect looks		Filler looks			Loser looks		
Experiment 2	*df*	χ^2	*p*	χ^2	*p*	*F*	*p*	η_p^2	*F*	*p*	η_p^2
Presence	1	—	—	4.39	.17	**77.08**	**<.001**	**.10**	**133.23**	**<.001**	**.12**
Accuracy	1	5.31	.15	3.53	.17	**35.41**	**<.001**	**.08**	**18.06**	**<.001**	**.04**
Similarity	1	2.33	.44	2.37	.24	**18.21**	**<.001**	**.02**	**15.39**	**<.001**	**.01**
Size	2	0.37	.94	0.24	.89	2.49	.33	.002	**6.00**	**.02**	**.02**
Target	1	—	—	0.43	.66	**9.60**	**.02**	**.008**	**8.87**	**.02**	**.01**
Presence × Accuracy	1	—	—	5.75	.17	4.79	.18	.003	**39.06**	**<.001**	**.07**
Presence × Similarity	1	—	—	0.69	.79	1.54	.55	.001	0.66	.92	<.001
Presence × Size	2	—	—	0.18	.79	1.93	.50	.003	0.37	.99	.001
Presence × Target	1	—	—	4.39	.17	1.89	.50	<.001	1.09	.77	.001
Accuracy × Similarity	1	0.76	.86	3.52	.17	0.04	.98	.001	0.56	.94	.003
Accuracy × Size	2	0.29	.94	1.61	.65	0.03	.98	.001	0.15	.99	.001
Accuracy × Target	1	—	—	0.67	.65	0.00	.98	.001	0.00	.99	<.001
Similarity × Size	2	1.43	.86	0.55	.80	0.47	.97	.003	0.47	.99	.003
Similarity × Target	1	—	—	1.58	.36	0.31	.97	<.001	0.01	.99	<.001
Size × Target	2	—	—	1.30	.66	0.05	.98	.001	0.18	.99	.003
Presence × Accuracy × Similarity	1	—	—			0.05	.98	.001	0.00	.99	.002
Presence × Accuracy × Size	2	0.13	.94	5.70	.17	1.30	.65	.001	2.44	.34	.003
Presence × Accuracy × Target	1	—	—	5.84	.17	0.13	.98	<.001	0.09	.99	<.001
Presence × Similarity × Size	2	—	—			0.13	.98	<.001	0.01	.99	<.001
Presence × Similarity × Target	1	—	—			3.41	.33	.004	2.50	.39	.004
Presence × Size × Target	2	—	—	4.94	.18	0.16	.98	.001	0.29	.99	.003
Accuracy × Similarity × Size	2	0.13	.94	4.46	.17	2.53	.33	.009	1.47	.69	.005
Accuracy × Similarity × Target	1	—	—			1.03	.69	.003	0.94	.79	.002
Accuracy × Size × Target	2	—	—			0.15	.98	.007	0.43	.99	.006
Similarity × Size × Target	2	—	—	5.24	.17	0.98	.77	.005	1.41	.69	.008
Presence × Accuracy × Similarity × Size	2	—	—			0.82	.85	.001	0.39	.99	.001
Presence × Accuracy × Similarity × Target	1	—	—			0.01	.98	<.001	0.03	.99	<.001
Presence × Accuracy × Size × Target	2	—	—			0.61	.96	.007	0.55	.99	.005
Presence × Similarity × Size × Target	2	—	—			0.52	.97	.003	0.60	.99	.003
Accuracy × Similarity × Size × Target	2	—	—			0.22	.98	.002	0.18	.99	.001
Presence × Accuracy × Similarity × Size × Target	2	—	—			1.73	.50	.01	3.65	.12	.02

Note. Significant effects are bolded. Diagnostic tests for all four models suggested our data violated heterogeneity assumptions of GLM. We took a series of steps to account for this (i.e. applied a heteroscedasticity-consistent covariance matrix model; Long & Ervin, 2000; Rosopa et al., 2013; see online supplementary materials for full description of these steps). All *p* values are corrected.
[1] Model was fit using only target present lineups in which a winner was selected (n = 255).

was a significantly higher average number of loser looks in high (n = 337, M = 1.94, SD = 0.94) than low (n = 334, M = 1.76, SD = 0.87) similarity lineups, and in male (n = 336, M = 1.95, SD = 0.94) than female (n = 335, M = 1.75, SD = 0.86) lineups. Follow-up pairwise Welch t tests of the statistically significant main effect of set size revealed only a single significant difference across the three levels of set size. The smallest set size (four) resulted in a higher average number of loser looks (n = 226, M = 1.96, SD = 1.02) than a set size of eight (n = 220, M = 1.73, SD = 0.83), $t(429.58) = 2.67$, $p = .01$.

Discussion

For adult witnesses, the machine learning analyses identified four behaviors that together postdicted suspect guilt at above chance levels: suspect looks, filler looks, loser looks, and winner looks. The hypotheses we developed for adult witnesses, where differences were observed, were consistent with our findings. Like children, adults looked at filler images more frequently when viewing target-absent than target-present lineups, and filler looks were more common among inaccurate witnesses, compared to accurate witnesses. However, there was stronger evidence that adults adjusted their looking behavior as a function of lineup composition. In high similarity lineups, compared to low similarity lineups, adults looked more often at any face, at fillers, and at loser images. There was also minor evidence that lineup size impacted looking behavior, with the smallest lineup size resulting in a higher number of loser looks than the largest lineup size.

General Discussion

A primary aim of the present study was to determine if observable witness behavior could postdict whether or not the suspect was in the lineup. We explored these relations in the context of a procedure that scaffolded the witness experience by requiring witnesses to look at least once at each photo and ensuring they focused on one photo at a time during the lineup task. The present experiments measured 10 potential behavioral indices plus confidence ratings, and, using machine learning, identified those that emerged as most optimal postdictors of suspect guilt. Across two samples of participants, we found that inclusion of relatively few behaviors postdicted target presence (i.e., suspect guilt) quite well. For children (Experiment 1), filler look time, suspect look time, suspect looks, filler looks, and winner look time postdicted target presence with 67% accuracy. For adults (Experiment 2), suspect looks, filler looks, loser looks, and winner looks postdicted target presence with 73% accuracy. This novel approach provides robust evidence that such behavioral factors can be used to both pre- and postdict suspect guilt. These findings provide a springboard for future research using similar procedures, more advanced technology (e.g., eye trackers), or other witness behaviors (e.g., facial expression, body language, verbal reports) to explore reliable postdictors of suspect guilt.

The novel approach taken with the present work allowed us to identify previously undocumented behavioral cues to suspect guilt, but it also complements the extant literature. Perhaps most notably, filler looking behavior was a significant postdictor of suspect guilt in both experiments. For lineups that included the guilty suspect (target-present), both child and adult witnesses were less likely to look at fillers relative to lineups that did not include the guilty suspect (target-absent). Combined with the Charman and Cahill (2012) finding that strength of memory for fillers implies longer looking time at fillers, which implies a degree of indecision and thus, reduced decision automaticity, the present findings indicate that researchers would be well served to attend to witness interaction with fillers to better understand memory processes underlying identification decisions. In the future, both measures of self-reported automaticity (as per Dunning & Stern, 1994) that could be easily collected in the field and a test of memory for fillers (as per Charman & Cahill, 2012) could be assessed for concordance with the current behavioral measures.

These data may also add context to some existing beliefs about pre- and postdictors of identification accuracy. The more nuanced behavior information collected in the present experiments suggests that we should differentiate looking behavior for the suspect relative to the fillers. Further, the finding that children may need (or take) more views of a suspect face before making a correct versus an incorrect decision in a target-present lineup, suggests that greater cognitive effort may be needed for children to make a correct decision.

A Lack of Adaptation to Lineup Composition

Adults, but generally not children, made some adaptations to the composition of the lineup. One of the more interesting possibilities for this limited behavioral adjustment is that witnesses were unable to gauge task difficulty. Filler similarity is related to lineup difficulty (Fitzgerald et al., 2013), which might explain adults' behavioral adjustments. Though the evidence is less compelling regarding lineup size (with some adult witness research suggesting no decrease in correct identifications with lineups as large at 40; Levi, 2002), manipulations of set size may also contribute to lineup difficulty, particularly for children. In future research, more extreme manipulations of lineup complexity (e.g., lineups of 15 or 20 members) may help to clarify the nature of some of these differences. The possible lack of awareness of lineup composition, however, suggests that an explicit instruction about the challenge associated with lineup identification tasks may be worth exploring.

A second possibility is that limiting attention to one face at a time may have sufficiently supported, or led to the perception that it supported, witness decision-making. Uncertainty in decision-making is often accompanied by increases in information-seeking behavior (Desender, Boldt, & Yeung, 2018), and perhaps the perception of how difficult the task was (i.e. equally difficult across conditions) led to equivalent search behavior across lineup conditions. Alternatively, perceptions of this particular task's supportiveness may have made judgments of lineup complexity more challenging. By examining navigation behavior using different methods (e.g., eye trackers), we would be able to see whether children's and adults' viewing patterns adapt to increasing lineup difficulty under less focused conditions. Conversely, using eye tracking technology in combination with one of the newer procedures that uses a more focused stimulus set size, such as single or paired-presentation of faces with children (e.g., Bruer & Price, 2017; Price & Fitzgerald, 2016), would allow for comparison of even smaller set sizes and the potential for behavioral adjustment with these highly focused tasks.

Limitations and Future Directions

When compared to a traditional, simultaneous procedure, the Interactive Simultaneous Procedure produced lower accuracy in target-absent conditions for the child sample. Thus, if replicated, such a procedure may only be practical for researchers who are interested in learning more about decision mechanisms through monitoring search behavior. For such purposes, we believe this procedure is a practical alternative for researchers for whom using an eye tracker is not a feasible option. However, future research would benefit from more sophisticated technologies, such as eye trackers, to track children's navigation behaviors under varying conditions. An eye tracker can be applied directly to a traditional simultaneous lineup task, thus eliminating the potentially negative impact of the Interactive Simultaneous Procedure on accuracy. We encourage other researchers to consider tracking lineup navigation behavior and to include conditions of variable lineup composition and efforts to increase focus on individual lineup members to help understand children's search strategies.

The current procedure artificially restricted lineup navigation behavior by requiring witnesses to look at all images prior to making an identification decision and by occluding images that were not the focus of the witness' attention. This restriction may have impacted how a witness searched the lineup and/or the nature of relations between particular variables. Given the role of automaticity in eyewitness identification decision processes, our requirement that participants look at all images prior to making a decision could have minimized natural variation in lineup navigation behaviors. As suggested by an anonymous reviewer, though required image views that take place after an early target recognition are likely to be faster, the noise such required viewings add to the data likely decreased our power and ability to observe some differences. Thus, the behaviors identified in the present analyses as related to target presence may not be as strongly (or perhaps more strongly) related under more naturalistic searching conditions. Use of the occluded images without the requirement to look at all images would be a valuable extension.

Finally, in the present experiments, we sought to occlude the images to the point that recognition of the silhouette was possible, without drawing attention to specific facial features. The degree of occlusion and the degree to which the occluded and nonoccluded images match are factors that could have a potentially substantive impact on search behavior and should be explored in future research. Indeed, there is evidence that even a fully occluded silhouette that matches—or does not match—the general outline of a target can influence identification behavior (Zajac & Jack, 2016).

Conclusion

The present research takes an early step toward understanding witness search strategies using novel methodology and statistical analyses. The results of the present experiments indicate that further exploration of witness search behaviors has potential to provide context to identification decisions that is currently often restricted to witness confidence. As the body of literature exploring variables that complement a categorical identification decision grows, implications for better understanding how witness decisions are made and how to interpret such decisions is vastly improving.

References

Amini, P., Ahmadinia, H., Poorolajal, J., & Moqaddasi Amiri, M. (2016). Evaluating the high risk groups for suicide: A comparison of logistic regression, Support Vector Machine, Decision Tree and Artificial Neural Network. *Iranian Journal of Public Health, 45,* 1179–1187. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5149472/

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B, Methodological, 57,* 289–300. http://dx.doi.org/10.1111/j.2517-6161.1995.tb02031.x

Brewer, N., Keast, A., & Sauer, J. D. (2010). Children's eyewitness identification performance: Effects of a not sure response option and accuracy motivation. *Legal and Criminological Psychology, 15,* 261–277. http://dx.doi.org/10.1348/135532509X474822

Brewer, N., & Weber, N. (2008). Eyewitness confidence and latency: Indices of memory processes not just markers of accuracy. *Applied Cognitive Psychology, 22,* 827–840. http://dx.doi.org/10.1002/acp.1486

Bruer, K. C., Fitzgerald, R. J., Price, H. L., & Sauer, J. D. (2017). How sure are you that this is the man you saw? Child witnesses can use confidence judgments to identify a target. *Law and Human Behavior, 41,* 541–555. http://dx.doi.org/10.1037/lhb0000260

Bruer, K. C., & Price, H. L. (2017). A repeated forced-choice lineup procedure provides suspect bias information with no cost to accuracy for older children and adults. *Applied Cognitive Psychology, 31,* 448–466. http://dx.doi.org/10.1002/acp.3342

Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery, 2,* 121–167. http://dx.doi.org/10.1023/A:1009715923555

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods, 42,* 286–291. http://dx.doi.org/10.3758/BRM.42.1.286

Chang, C., & Lin, C. (2011). LIBSVM: A library for support vector machines. *ACM transactions on intelligent systems and technology, 2,* 1–27. http://dx.doi.org/10.1145/1961189.1961199

Charman, S. D., & Cahill, B. S. (2012). Witnesses' memories for lineup fillers postdicts their identification accuracy. *Journal of Applied Research in Memory & Cognition, 1,* 11–17. http://dx.doi.org/10.1016/j.jarmac.2011.08.001

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological Science, 29,* 761–778. http://dx.doi.org/10.1177/0956797617744771

Dunning, D., & Perretta, S. (2002). Automaticity and eyewitness accuracy: A 10- to 12-second rule for distinguishing accurate from inaccurate positive identifications. *Journal of Applied Psychology, 87,* 951–962. http://dx.doi.org/10.1037/0021-9010.87.5.951

Dunning, D., & Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal of Personality and Social Psychology, 67,* 818–835. http://dx.doi.org/10.1037/0022-3514.67.5.818

Fitzgerald, R. J., & Price, H. L. (2015). Eyewitness identification across the life span: A meta-analysis of age differences. *Psychological Bulletin, 141,* 1228–1265. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/10.1002/acp.3012

Flowe, H. (2011). An exploration of visual behaviour in eyewitness identification tests. *Applied Cognitive Psychology, 25,* 244–254. http://dx.doi.org/10.1002/acp.1670

Flowe, H. D., & Cottrell, G. W. (2011). An examination of simultaneous lineup identification decision processes using eye tracking. *Applied Cognitive Psychology, 25,* 443–451. http://dx.doi.org/10.1002/acp.1711

Gündüz, N., & Fokoué, E. (2015). *Robust classification of high dimension low sample size data*. arXiv preprint arXiv:1501.00592.

Levi, A. M. (2002). Up to forty: Lineup size, the modified sequential lineup, and the sequential lineup. *Cognitive Technology, 7,* 39–46.

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. http://dx.doi.org/10.1023/A:1024807202926

Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician, 54,* 217–224.

Mansour, J., Lindsay, R. C. L., Brewer, N., & Munhall, K. G. (2009). Characterizing visual behavior in a lineup task. *Applied Cognitive Psychology, 23,* 1012–1026. http://dx.doi.org/10.1002/acp.1570

Pozzulo, J. D., & Balfour, J. (2006). Children's and adults' eyewitness identification accuracy when a culprit changes his appearance: Comparing simultaneous and elimination lineup procedures. *Legal and Criminological Psychology, 11,* 25–34. http://dx.doi.org/10.1348/135532505X52626

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

Price, H. L., & Fitzgerald, R. J. (2016). Face-off: A new identification procedure for child eyewitnesses. *Journal of Experimental Psychology: Applied, 22,* 366–380. http://dx.doi.org/10.1037/xap0000091

Rosopa, P. J., Schaffer, M. M., & Schroeder, A. N. (2013). Managing heteroscedasticity in general linear models. *Psychological Methods, 18,* 335–351. http://dx.doi.org/10.1037/a0032553

Salazar, D. A., Vélez, J. I., & Salazar, J. C. (2012). Comparison between SVM and logistic regression: Which one is better to discriminate? *Revista Colombiana de Estadística Número especial en Bioestadística, 35,* 223–237. Retrieved from http://www.kurims.kyoto-u.ac.jp/EMIS/journals/RCE/V35/v35n2a03.pdf

Sauerland, M., Sagana, A., Sporer, S. L., & Wixted, J. T. (2018). Decision time and confidence predict choosers' identification performance in photographic showups. *PLoS ONE, 13*(1), e0190416. http://dx.doi.org/10.1371/journal.pone.0190416

Sporer, S. L. (1992). Post-dicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and non-choosers. *European Journal of Social Psychology, 22,* 157–180. http://dx.doi.org/10.1002/ejsp.2420220205

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods, 17,* 228–243. http://dx.doi.org/10.1037/a0027127

Weber, N., Brewer, N., Wells, G. L., Semmler, C., & Keast, A. (2004). Eyewitness identification accuracy and response latency: The unruly 10–12-second rule. *Journal of Experimental Psychology: Applied, 10,* 139–147. http://dx.doi.org/10.1037/1076-898X.10.3.139

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. http://dx.doi.org/10.1177/1529100616686966

Zajac, R., & Jack, F. (2016). Improving children's performance on photographic line-ups: Do the physical properties of the 'wildcard' make a difference? *Legal and Criminological Psychology, 21,* 358–371. http://dx.doi.org/10.1111/lcrp.12075

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

Zanette, S., Gao, X., Brunet, M., Bartlett, M. S., & Lee, K. (2016). Automated decoding of facial expressions reveals marked differences in children when telling antisocial versus prosocial lies. *Journal of Experimental Child Psychology, 150,* 165–179. http://dx.doi.org/10.1016/j.jecp.2016.05.007

Received February 11, 2019
Revision received January 6, 2020
Accepted January 9, 2020 ■