"Only your first yes will count": The impact of pre-lineup instructions on sequential lineup decisions

Ruth Horry[a*], Ryan J. Fitzgerald[b], and Jamal K. Mansour[c]

[a]Department of Psychology, College of Human and Health Sciences, Swansea University

[b]Department of Psychology, Simon Fraser University

[c] Memory Research Group; Centre for Applied Social Sciences; Division of Psychology, Sociology and Education, Queen Margaret University

Accepted for publication in the *Journal of Experimental Psychology: Applied.*

© American Psychological Association, 2020. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. Please do not copy or cite without the author's permission.

*Corresponding author. Ruth Horry, Department of Psychology, Vivian Tower, Singleton Park Campus, Swansea, SA2 8PP, United Kingdom. Tel: +44 1792 295927. Email: R.horry@swansea.ac.uk

Author note: Data are openly available on the project's OSF page: https://osf.io/cwfth/. The data were presented at the International Meeting of the Psychonomic Society, in Amsterdam, the Netherlands, May 2018, and at the January 2020 meeting of the Experimental Psychology Society in London, UK. This research was supported in part by a grant to the third-author from the Centre for Applies Social Sciences at Queen Margaret University.

The authors acknowledge Jacqueline Ackermann, Harry Anderson, Sophie Berryman, Grant Chapman, Rebecca Clark, Lucy Collins, Rachel Cullinane, Roberta Donatone, Jeri Gosling, Emily Gray, Monica Gurung, Madress Henderson, Matt Higginson, Freya Lenton, Georgia Maullin, Anna McTurk, Vincent Nittel, Tatiana Opletina, Sharyne Patel, Chrissie Petane, Joshua Quiza-Perez, Key Pugh, Toni Ralph, Catriona Rennie, Mhairi Robertson, Amy Roohi, Katharina Schlimm, Kim Schneider, Amelia Snape, and Amelia Thomas for their assistance with data collection.

Abstract

When administering sequential lineups, researchers often inform their participants that only their first yes response will count. This instruction differs from the original sequential lineup protocol and from how sequential lineups are conducted in practice. Participants (*N* = 896) viewed a videotaped mock crime and viewed a simultaneous lineup, a sequential lineup with a *first-yes-counts* instruction, or a sequential control lineup (with no *first-yes-counts* instruction); the lineup was either target-present or target-absent. Participants in the *first-yes-counts* condition were less likely to identify the suspect and more likely to reject the lineup than participants in the simultaneous and sequential control conditions, suggesting a conservative criterion shift. The diagnostic value of suspect identifications, as measured by partial Area Under the Curve, was lower in the *first-yes-counts* lineup than in the simultaneous lineup. Results were qualitatively similar for other metrics of diagnosticity, though the differences were not statistically significant. Differences between the simultaneous and sequential control lineups were negligible on all outcomes. The *first-yes-counts* instruction undermines sequential lineup performance and produces an artefactual simultaneous lineup advantage. Researchers should adhere to sequential lineup protocols that maximize diagnosticity and that would feasibly be implemented in practice, allowing them to draw more generalizable conclusions from their data.

Keywords: Eyewitness identification, sequential lineup, instructions, criterion shifts

Public Significance Statement: When participants viewing sequential lineups were told that only their first "yes" response to a lineup member would count, their lineup performance was poorer than if they did not receive that instruction or if they viewed a simultaneous lineup. Studies using the *first-yes-counts* instruction may underestimate the potential of sequential lineups.

"Only your first yes will count": The impact of pre-lineup instructions on sequential lineup decisions

Eyewitness identification evidence is often used in criminal cases to support prosecutions of suspects. Typically, an eyewitness is presented with a lineup which includes the police suspect and several known-innocent fillers who have been selected to match the suspect's appearance and/or the eyewitness's description of the culprit. However, eyewitness identification errors have been implicated in hundreds of wrongful convictions (www.innocenceproject.org). It is crucial, therefore, to identify lineup protocols that allow investigators to most effectively diagnose whether their suspect is likely to be guilty or innocent, based on the decisions made by a witness; we refer to this as the *diagnosticity* of the lineup.

Researchers have identified and tested a variety of variables that impact lineup decisions which have led to several best-practice guidelines and influenced practice in many jurisdictions (e.g., National Institute of Justice, 1999; Wells et al., 2020). However, an ongoing issue of contention is whether the images in a lineup should be presented simultaneously or sequentially. In a simultaneous lineup, the eyewitness is presented with all lineup members at the same time and can either choose to identify a lineup member or choose no one. Lindsay and Wells (1985) developed the sequential lineup, in which each lineup member is seen individually, with the eyewitness making a yes/no decision for each lineup member before moving on to the next. Their initial investigation found that sequential presentation reduced the rate at which an innocent suspect was misidentified when the true perpetrator was absent from the lineup, whereas correct identifications of the perpetrator fell only slightly. Consequently, the ratio of correct to false identifications (commonly termed the Diagnosticity Ratio) favored the sequential lineup (often referred to as the *sequential superiority effect*). This basic pattern of findings was replicated in several independent

studies, which were summarized in meta-analyses in 2001 and 2011 (Steblay, Dysart, Fulero, & Lindsay, 2001; Steblay, Dysart, & Wells, 2011). A subsequent meta-analysis, applying a signal-detection model to the data, suggested that this basic pattern was best explained by a criterion shift, with participants responding more conservatively to sequential lineups than to simultaneous lineups (Palmer & Brewer, 2012; see also Meissner, Tredoux, Parker, & MacLin, 2005).

More recently, several large-sample experiments comparing simultaneous and sequential presentation have yielded quite different results (e.g., Carlson & Carlson, 2014; Gronlund, Carlson, Dailey, & Goodsell, 2009; Mickes, Flowe, & Wixted, 2012). These studies have generally reported sizeable decreases in correct identification rates from culprit-present sequential lineups, accompanied by much smaller decreases in the false identification rate from culprit-absent sequential lineups (cf. simultaneous lineups). Unsurprisingly, therefore, the authors of these studies have tended to conclude that decisions from simultaneous lineups are as diagnostic, and perhaps more diagnostic, than decisions from sequential lineups.

Some authors have argued that this shift in the general pattern of findings regarding the sequential-simultaneous comparison can be explained by a change in analytic strategy—principally, the application of Receiver Operating Characteristic (ROC) analysis to lineup data (Gronlund, Mickes, Wixted, & Clark, 2015; Mickes et al., 2012). However, an alternative possibility is that these discrepant findings have been driven, at least in part, by the sequential lineup protocols used in these studies. Indeed, when it comes to sequential lineup presentation, there are many decisions that researchers must make about how to administer the lineup, including: where the suspect will be placed in the lineup. whether the number of images in the lineup should be concealed from the eyewitness, whether the eyewitness should be allowed to make more than one identification, how to resolve multiple

identifications if they occur, and whether the eyewitness should be allowed to see any of the lineup images more than once. These decisions provide many degrees of freedom, both in how sequential lineups are administered, and in how the data from them are coded. We note that these degrees of freedom also exist in practice, with considerable variability in how sequential lineups are implemented by law enforcement professionals (Police Executive Research Forum, 2013; Bertrand et al., 2018).

Some of these protocol variations have been manipulated experimentally. Horry, Palmer, and Brewer (2012) investigated how knowledge of the number of to-be-seen lineup images in a sequential lineup affects decision making. Participants were either correctly informed of the number of images included or were misled to expect that the lineup contained more images than it actually did (i.e., the lineup was backloaded). When the number of images was known and the suspect appeared late in the lineup, both correct and false identifications increased substantially, undermining the diagnosticity of these decisions. Furthermore, participants responded increasingly conservatively as the expected number of images increased. This criterion shift was replicated by Carlson, Carlson, Weatherford, Tucker, and Bednarz (2016) when comparing conditions in which the number of images in the lineup was disclosed versus undisclosed to participants.

In another test of sequential lineup protocol variations, Steblay, Dietrich, Ryan, Raczynski, and James (2011) investigated the effect of allowing a second sequential lineup lap on identification decisions. They found that second-lap decisions were generally less accurate than first-lap decisions, such that the diagnosticity of identification decisions was generally lower after a second lap. Horry, Brewer, Weber, and Palmer (2015) found that participants who elected to see a second lap generally shifted to a more liberal response criterion in that second lap, though diagnosticity remained about the same.

These examples are informative in that they make it clear that protocol variations in the sequential lineup can have non-trivial effects on eyewitness identification decisions, with downstream implications for the diagnosticity of the test. Yet there are many protocol variations that exist within the research literature that have not yet been experimentally manipulated. We focus on one such variation here, the *first-yes-counts* instruction. This instruction comes in several flavors. Sometimes, participants are informed that they will be required to view all images in the lineup, but that only their first "yes" response will be counted (e.g., Carlson, 2011; Carlson et al., 2016; Carlson, Gronlund, & Clark, 2008; Gronlund et al., 2009; Mickes et al., 2012); sometimes, participants are informed that if they say "yes" to an image, they will not be able to respond "yes" to any subsequent images (e.g., Horry et al., 2012; Cutler & Penrod, 1988); sometimes, participants are instructed to make only one "yes" response (e.g., Elphick, Pike, & Hole, 2019); and in other studies, participants are informed that the lineup will terminate following a "yes" response (e.g., Dobolyi & Dodson, 2013; Douglass & McQuiston-Surrett, 2006; Humphries & Flowe, 2015; Morgan et al., 2004). In many studies that have used termination rules, pre-lineup instructions are not reported clearly enough to determine whether the participants were aware that the lineup would terminate following a "yes" response. We return to this point in the General Discussion.

Regardless of the exact form that they take, *first-yes-counts* instructions share the core feature of informing the eyewitness that only their first positive response will be counted, and that any subsequent positive responses (should they be permitted) will be discounted. It is worth highlighting, at this point, that the *first-yes-counts* instruction was not part of the original sequential lineup protocol (Lindsay & Wells, 1985).

Applied considerations regarding the first-yes-counts instruction

In practice, a *first-yes-counts* instruction is unlikely to be used. Indeed, where sequential procedures have been adopted by police departments, they have tended to include the following features: i) the eyewitness is required to view all lineup members, even following a positive response; and ii) any multiple identifications are resolved by the eyewitness (Wells, Steblay, & Dysart, 2015). To understand the logic of these features, let us consider three hypothetical eyewitnesses viewing a sequential lineup. Eyewitness A views a lineup with the suspect placed in position 2, and makes a positive identification. If the lineup were terminated at this point, a defense attorney could argue that the eyewitness was only shown a two-person lineup, which may then undermine the admissibility and perceived probative value of the evidence. Eyewitness B views a lineup with the suspect placed in position 4 but makes a positive identification of the filler in position 3. If the lineup is terminated following the filler identification, this would create the rather bizarre situation in which the eyewitness is never actually given the opportunity to identify the police suspect—clearly an unacceptable scenario for any investigator. Finally, eyewitness C makes a tentative identification of the filler in position 3, before then making an identification of the suspect, who is in position 4. Eyewitness C is much more sure of their identification of the suspect, and strongly indicates that she wishes to make that her final decision. It is hard to imagine that an investigator would be willing to discount that suspect identification in favour of the preceding filler identification. Indeed, when conducting their field experiment of simultaneous versus sequential lineup presentation, Wells et al. (2015) argued that their police partners would never have allowed them to use many of the protocols that dominate the experimental literature, including the use of a *first-yes-counts* instruction.

These protocol differences between the sequential lineup as used in practice and the sequential lineup as studied in the lab have potentially significant implications. Whether

simultaneous or sequential lineups are more diagnostic of a suspect's guilt or innocence is an important applied question. We cannot hope to answer that question adequately if the protocols that dominate the experimental literature differ markedly from the protocols that are used in practice. In particular, our ability to draw meaningful conclusions about the comparative diagnostic value of simultaneous and sequential lineups may have been confounded by the presence, in some studies, of the *first-yes-counts* instruction. This possibility was discussed by Steblay et al. (2011) in their meta-analysis, though they were unable to draw any meaningful conclusions due to confounds across studies that could not be disentangled.

Seale-Carlisle, Wetmore, Flowe, and Wixted (2019) recently summarized the results of 11 experiments from seven published papers and two unpublished studies which had applied ROC analysis to comparisons of sequential and simultaneous lineup performance. We summarize two key protocol aspects (usage of a *first-yes-counts* instruction and implementation of a termination rule, whereby the lineup is terminated following a "yes" response) of the eight relevant experiments in Table 1. Where the information was not readily available in the public domain (i.e., the information was not included in the published report, or the report was unpublished), we contacted the authors to request the information. To summarize, three of the eight experiments used a *first-yes-counts* instruction and all used a termination rule. The remaining three experiments summarized by Seale-Carlisle et al. (2019) (Seale-Carlise and Mickes, 2016; Seale-Carlisle et al., 2019, Experiments 1 and 5) compared simultaneous lineups with UK lineups, which differ markedly from the sequential lineups used in other parts of the world (e.g., by requiring that the eyewitness withholds a decision until the entire lineup has been seen twice). In other words, in not one of the experiments considered by Seale-Carlisle et al. (2019) did the sequential lineup use the protocol originally outlined by Lindsay and Wells (1985) or a protocol used by law enforcement in the US. We

note that this issue is not specific to this particular set of studies; indeed, similar protocol variation is evident in the studies summarized in the meta-analysis by Steblay et al. (2011).

Theoretical considerations regarding the first-yes-counts instruction

Theoretically, we posit that there are two primary ways in which the *first-yes-counts* protocol might influence identification outcomes. The first is purely mechanistic—by considering the first "yes" response to be final, the *first-yes-counts* protocol prohibits changes of mind. This essentially changes the decision rule from *best-over-criterion* to *first-over-criterion* (see Wilson, Donnelly, Christenfeld, & Wixted, 2019, for a detailed discussion of decision rules in sequential lineups). If a filler that precedes the suspect is identified, then the participant will be unable to rectify that error, even if they have a strong recognition experience when viewing the suspect. We call this the *termination mechanism*. The second mechanism is a psychological one: the *first-yes-counts* instruction may also encourage conservative responding, as participants become wary of 'spending' their only positive response in case a better alternative comes along later in the sequence. The result of a conservative criterion shift would be to lower the overall likelihood of a suspect identification, as well as to increase the likelihood that the lineup will be rejected.

The criterion shifts induced by the *first-yes-counts* instruction may come in two distinct varieties. The first, most straightforward possibility, is that the witness adopts a more conservative criterion at the beginning of the lineup (cf. a witness who did not receive the *first-yes-counts* instruction), which is then maintained throughout the lineup. We will call this the *static-conservative-shift* mechanism. Alternatively, the participant may begin by adopting a more conservative decision criterion, which is gradually relaxed over the course of the

lineup, as the witness begins to feel that he or she is running out of options to identify the perpetrator.[1] We call this the *dynamic-conservative-shift* mechanism.

Regardless of whether criterion shifts are static or dynamic, we would expect them to lower the overall probability of a suspect identification and increase the overall probability of a lineup rejection. However, these criterion shifts may also induce position effects, such that the probabilities of different types of responses vary depending on where the suspect is placed. We consider this possibility next.

Suspect position effects

Several studies have reported position effects in sequential lineups, wherein the likelihood of a suspect being identified changes depending on where within the sequence he is placed (e.g., Carlson, Gronlund, & Clark, 2008; Clark & Davey, 2005; Horry et al., 2012; Meisters et al., 2018; Wilson, Donnelly, Christenfeld, & Wixted, 2019), although the extant literature varies in terms of whether effects are found with target-present and/or target-absent lineups, or neither (e.g., Carlson & Carlson, 2014; Dobolyi & Dodson, 2013; Flowe et al., 2016; Kneller, Memon, & Stevenage, 2001; Lindsay & Wells, 1985; Wells et al., 2011; Sporer, 1993). One explanation for such position effects is that the eyewitness's decision criterion may become increasingly liberal as the lineup progresses and the eyewitness begins to feel that she is running out of opportunities to make a positive identification (i.e., a dynamic-conservative shift; Goodsell, Gronlund, & Carlson, 2010; Horry et al., 2012; Meisters et al., 2018; see also Smith, Wells, Lindsay, & Myerson, 2018). If this is the case, then we might expect the *first-yes-counts* protocol to exacerbate these position effects.

[1] Note that, for this to occur, the eyewitness does not need to know how many images are in the lineup; she merely has to think that or wonder if the end of the lineup is nearing. We know of no studies that have asked participants ahead of a lineup how many images they expect to see.

Below, we consider how the *termination* mechanism, the *static-conservative-shift* mechanism, and the *dynamic-conservative-shift* mechanism may create position effects.

First, let us consider the *termination* mechanism in isolation. The probability that at least one filler will exceed the decision threshold increases with every additional filler that is shown. Recall that if a termination rule is implied, the first item to exceed the participant's decision threshold will be chosen, and that decision will be final. Consequently, under a termination rule, the probability that the target will be chosen decreases with each additional filler that precedes it in the lineup. The *first-yes-counts* instruction should, therefore, decrease the probability of a suspect identification to a greater extent when the suspect appears late in the lineup than when the suspect appears early.

If a *static-conservative-shift* is added to the termination mechanism, then we would expect to see qualitatively similar position effects. However, we would also expect that they would be smaller in magnitude, as fillers would be less likely to exceed the more stringent decision criterion. In essence, the more stringent decision criterion would be expected to buffer the impact of the termination rule on suspect identifications, at least to some extent.

The *dynamic-criterion-shift* mechanism, however, could produce qualitatively different suspect position effects. If the participant begins with a conservative decision criterion which becomes increasingly liberal, then the probability of any individual item exceeding that threshold will increase as the lineup progresses. Suspects who appear early will have a lower likelihood of exceeding the decision threshold than suspects who appear late, and who are thus compared to a more liberal criterion. Consequently, the *first-yes-counts* instruction should decrease the probability of a suspect identification to a greater extent if the suspect appears early than if the suspect appears late[2].

[2] Simulations of these mechanisms can be found on the project's OSF page, and are described in the Supplementary Materials.

The Present Study

The primary aims of our study were to determine whether, and how, the *first-yes-counts* instruction affects eyewitness identification decisions from sequential lineups, and whether the *first-yes-counts* instruction produces artefactual differences between sequential and simultaneous lineups. Participants viewed a mock-crime video and then a six-person target-present or -absent lineup. The suspect either appeared in position 2 or position 5. Participants were randomly allocated to one of three lineup conditions: a simultaneous lineup, a sequential lineup with a *first-yes-counts* instruction, or a sequential control lineup (with no *first-yes-counts* instruction).

Our first research question concerned the impact of the *first-yes-counts* instruction on patterns of identification responses. We hypothesised that the *first-yes-counts* instruction would influence the pattern of decisions via two key mechanisms: by mechanistically prohibiting witnesses from changing their minds and by inducing conservative criterion shifts. Without the *first-yes-counts* instruction and with the freedom for witnesses to change their minds, we expected that identification responses would be similar in the sequential control and simultaneous lineups. This led to the following predictions about how the *first-yes-counts* lineup would compare to both the simultaneous and sequential control lineups:

H1: The sequential *first-yes-counts* condition would be associated with a lower likelihood of suspect identifications than the simultaneous and sequential control conditions.

H2: The sequential *first-yes-counts* condition would be associated with a higher likelihood of non-identifications than the simultaneous and sequential control conditions.

We had no firm predictions about the impact of the *first-yes-counts* instruction on filler identifications. On the one hand, if participants are responding more conservatively, we

might expect the likelihood of a filler identification to be lower in the *first-yes-counts* condition than in the sequential control and simultaneous conditions. However, this effect might be counteracted by the mechanistic action of the termination rule, which disallows participants from changing erroneous filler identifications that occur earlier in the lineup to suspect identifications.

Our second research question concerned the impact of the *first-yes-counts* instruction on position effects in suspect identification rates. We made no directional hypotheses about position effects, as we would expect to see different patterns depending on the nature of any criterion shifts that are associated with the *first-yes-counts* protocol. If the *first-yes-counts* instruction does not induce any criterion shifts and exerts its influence only through the termination mechanism, then we would expect suspect identifications to be disproportionately reduced when the suspect appears late in the lineup. Therefore, we would expect the magnitude of the difference in suspect identifications between the *first-yes-counts* and sequential control/simultaneous conditions to be larger when the suspect appears in position 5 than when the suspect appears in position 2. We would expect to see a qualitatively similar pattern (though smaller in magnitude) if *the first-yes-counts* instruction induces a static conservative criterion shift. If, however, the *first-yes-counts* instruction induces a dynamic criterion shift, then we would expect suspect identifications to be disproportionately reduced when the suspect appears early in the lineup. Consequently, the magnitude of the difference between the *first-yes-counts* and sequential control/simultaneous conditions would be larger when the suspect appears in position 2 than when the suspect appears in position 5.

Our third research question concerned the impact of the *first-yes-counts* instruction on diagnosticity. We focus predominantly on pAUC, as this is the measure that has informed recent claims of simultaneous lineup advantages (Seale-Carlisle et al., 2019). However, to examine the robustness of our conclusions, we also explored the extent to which our findings

converged across alternative measures of diagnosticity, including *d'* on suspect identifications (Mickes, Moreland, Clark, & Wixted, 2014) and the Diagnosticity Ratio; these analyses are reported in the Supplementary Materials. We made no directional predictions about the impact of the *first-yes-counts* instruction on diagnosticity, as we had hypothesised that it would primarily affect response bias.

Method

Participants. Participants were recruited online and in the laboratory. Table 2 provides a breakdown of participants by source. For the MTurk sample, eligibility criteria were that workers had to have a HIT (Human Intelligence Task, i.e., previous completed tasks on MTurk) approval rate greater than 75% and more than 100 HITs approved. Another group of online participants were recruited through the Qualtrics Online Sample service (https://www.qualtrics.com/uk/online-sample/). We commissioned Qualtrics to recruit 250 complete responses that met the following eligibility criteria: Aged 18 or over, fluent English speaker, and passed at least one attention check[3]. This study was reviewed and approved by the Institutional Review boards at each of the three universities where data were collected.

A total of 2260 participants clicked through to the study, and a total of 1159 participants completed the study, producing an attrition rate of 48.7%, which is broadly consistent with attrition rates reported in other online experiments in behavioural science (Zhou & Fishbach, 2016). Many of these participants clicked through, but got no further than the information and consent screens. Attrition almost always occurred prior to allocation to experimental conditions, ruling out differential attrition as potential explanation of our findings (Zhou & Fishbach, 2016). No laboratory participants were lost to attrition.

[3] We had intended to only exclude participants who failed both attention checks. However, the Qualtrics Online Sample team excluded participants if they failed either of the attention checks. This led to a higher proportion of exclusions due to attention checks from the Qualtrics Online Sample than from the MTurk and social media samples

In accordance with our pre-registered analysis plan (https://osf.io/c4dve/), 278 participants were excluded for failing attention checks that pertained to the content of the mock-crime videos (14 laboratory participants, 264 online participants). A further 58 participants were excluded for completing the study more quickly than would have been feasible had they been attending seriously to the task and instructions. The Qualtrics Online Sample excluded participants who did not provide a valid age (n = 13) or indicated that they were not fluent in English (n = 26). In addition, MTurk participants were excluded if participants had duplicate IP addresses (n = 6). After exclusions, the final sample consisted of 896 participants (327 laboratory participants, 569 online participants).

Design. The experiment followed a 3 (Lineup Type: simultaneous, sequential *first-yes-counts*, sequential control) × 2 (Target Presence: target-present, target-absent) × 2 (Suspect Position: 2, 5) between-subjects design. Participants were randomly allocated to one of the twelve cells of the design. The mean cell size was 74.7 (*Range* = 60-89). In addition, participants were randomly assigned to view one of four different targets in the mock-crime video (Target Person), and to view one of two different filler orders (A-E or E-A).

Materials.

Stimuli. To increase generalizability, we used four mock-crime videos. These videos came from the stimulus pool used by Mansour, Lindsay, Brewer, and Munhall (2009), although not all of the stimuli used here appeared in Mansour et al. Each video depicted the same events but with a different perpetrator. The videos lasted between 37 and 50 seconds. In each video, the perpetrator walked into the shot and approached a counter, where he informed a woman (off-camera) that he had come to pick up a VCR to take it to the shop. The woman said that she would go to look for the VCR in the back, leaving the perpetrator alone. The perpetrator looked around before stealing cash from a handbag on the counter. The woman returned to say that she could not find the VCR and requested that the perpetrator return later.

He agreed and walked away. The perpetrator's face is visible throughout the video, including approximately three to five seconds of a close-up, frontal view.

The four perpetrators were all of similar appearance—White males with short, dark hair, slim to average builds, and in their early 20s. Fillers were selected using a match to description approach from the database of fillers maintained by the third author. The lineup photographs were cropped at the neck and were pasted onto a coloured background. All lineup photos in the same lineup shared a background colour.

We used Oriet and Fitzgerald's (2018) single lineup paradigm to manipulate target presence. Specifically, for each perpetrator, we created a lineup that consisted of the perpetrator and five fillers. When Lineup A was paired with the video featuring Perpetrator A, it served as a target-present lineup. When Lineup A was paired with a video featuring Perpetrator B, it served as a target-absent lineup. Lineups A and B were paired together in this way, as were lineups C and D. The advantage of this approach is that target-present and -absent lineups are composed of identical images; what has changed is the identity of the perpetrator. This approach also naturally creates an innocent suspect to whom the fillers have been matched.

We created eight versions of each lineup, in which we varied: i) the position of the suspect (position 2 or position 5); ii) the order of the fillers (order A-E or E-A); and iii) simultaneous or sequential presentation. For simultaneous lineups, the images were presented in a 2 × 3 array with a number from 1 to 6 beneath each image. For sequential control and sequential *first-yes-counts* lineups, the images were presented individually without a number.

Filler task. We created a five minute filler task. A portion of a "Where's Waldo?"[4] image was displayed on the centre of the screen and participants were asked a series of questions about the image (e.g., "How many closed umbrellas are in the picture?", "What

[4] TM & © 2008 Entertainment Rights Distribution Limited. All rights reserved.

color is the inflatable mattress in the scene?”). Participants typed their responses into a text box. Participants were allowed up to 10 seconds to answer each question, and we asked 34 questions in total.

Lineup protocols. Participants in all conditions read the following pre-lineup instructions: *“Based on the video you just watched, you are now about to view a photo lineup. You will be asked to identify the perpetrator of the crime in the video. This person may or may not be present in this lineup. Following this decision you will be asked to rate your confidence”.*

Separate lineup instructions were developed for the three lineup types. The instructions were based on those used by Norwood Police Department, MA (http://www.norwoodma.gov/departments/police/mptc_training_material.php, accessed 10th March 2020).

In the simultaneous lineup condition, participants received the following instructions:

“1. You are being asked to view a set of photographs.

2. You will be viewing all photographs at the same time.

3. Please look at all photographs carefully and take time before making a decision.

4. The person you saw may or may not be in the set of photographs you are about to view.

5. You should remember that it is just as important to clear innocent persons from suspicion as to identify the guilty.”

In the sequential control condition, participants received the following instructions:

“1. You are being asked to view a set of photographs.

2. You will be viewing all photographs one at a time and in a random order.

3. Please look at all of them. I am required to show you the entire series.

4. Please make a decision about each photograph before moving onto the next one.

5. The person you saw may or may not be in the set of photographs you are about to view.

6. You should remember that it is just as important to clear innocent persons from suspicion as to identify the guilty." The participants in the were not informed of how many lineup members they would see.

The *first-yes-counts* instructions were the same as in the sequential control condition with the following additional point between points 4 and 5: *"If you respond "yes" to a photo, you will not be able to change that decision, and you will not be able to respond "yes" to any later photos".*

Participants in the simultaneous condition viewed the six lineup members in a 3 × 2 array and were asked *"Is the culprit included in any of the images shown?"* They could respond by selecting a number from 1 to 6 or by selecting *"Not present"*. If the participant identified a lineup member, he/she was asked *"How confident are you in your decision?"*. If the participant selected the Not present option, he/she was asked *"How confident are you that none of the people in the lineup were the culprit?"* Participants selected their confidence rating using a slider labelled *"% confidence"*, which extended from 0 to 100. Markers were shown on the scale at 10-point intervals, though the participant could select any integer between 0 and 100. The anchors *"Not at all confident"* and *"Certain"* were shown at the 0- and 100-points, respectively.

The sequential control and *first-yes-counts* lineup protocols were identical, except for the differences in instructions described above. For participants in both sequential conditions, the first lineup image appeared in the centre of the screen. Below the image was the question *"Is this the culprit?"* with two options: *Yes* or *No*. If the participant responded "Yes", he/she was asked *"How confident are you in your decision?"* using the confidence slider described above. The participant continued through all six lineup images, making a "Yes" or "No" decision for each. Participants could say "Yes" to multiple images, regardless of whether they were in the sequential control condition or whether they were in the *first-yes-counts*

condition; that is, even if the participant had been instructed that they could only make a single "Yes" decision, they were, in fact, able to say "Yes" to any number of images. If the participant reached the end of the lineup having said "No" to each item, he/she was asked "*How confident are you that none of the people in the lineup were the culprit?"*, which was answered using the same confidence scale described above.

If a participant said "Yes" to more than one image, he/she saw a second lap of the lineup. Before the second lap, the participant was instructed *"You are about to see all of the images again, in the same order. You are being asked to view the lineup again because you said "yes" to two different images the first time. This time, the lineup will end after your first "yes" response, so please make sure you only say "yes" to the face you wish to identify. Of course, you should still bear in mind that the culprit may or may not be present, so you should not feel like you must make an identification."*

To be clear, the sequential control and sequential *first-yes-counts* protocols differed only in their pre-lineup instructions. Regardless of what they had been informed, there was nothing to stop participants in either condition from responding affirmatively to more than one image. This allowed us to conduct exploratory analyses of the frequency of multiple "yes" decisions among participants in the *first-yes-counts* condition, and to examine the accuracy of resolved multiple "yes" decisions made by these participants. Links to demonstration versions of the protocols are available on the project's OSF page (https://osf.io/79bnc/).

Procedure. The experiment was programmed and administered in Qualtrics. Participants signed up for a study on "Perception of events in a video". After providing informed consent via a consent screen, participants were informed that they were about to watch a short video, and that they would need to pay attention as they would be asked about

their perception of the event later. Participants were not informed that a crime would take place or that they would be asked to identify the perpetrator later in the experiment.

Participants were then randomly assigned to view one of the four mock-crime videos. Participants were unable to advance to the next screen until enough time to watch the video in its entirety had elapsed. After watching the video, participants completed the 5-minute filler task. Participants then read the standard pre-lineup instructions, were randomly allocated to the simultaneous, sequential *first-yes-counts*, or sequential control condition, and were presented with the appropriate set of instructions. To increase the likelihood that the instructions would be attended to carefully, participants first listened to audio-recorded instructions. They then read the instructions, which were presented as a bulleted list. Participants then proceeded through the lineup, following the protocols outlined in the Materials.

Following the lineup, the participants were asked two multiple-choice questions to ensure that they had been paying attention to the video: 1) What did the thief steal from? A handbag (correct response); A cash register; A box on the desk; A charity collection tin. 2) Why did the thief say he was there? To take part in a study; To apply for a job; To look for a missing item; To pick up a VCR (correct response).

The participants were then asked to recall, in as much detail as possible, the instructions that they were given before they viewed the lineup. After they submitted their response, they were asked five yes/no questions about the pre-lineup instructions. *1) Were you told before the photo lineup that only your first "yes" would count? 2) Were you told before the photo lineup that the culprit's appearance may have changed since the video? 3) Were you told before the photo lineup that the thief may or may not be present in the lineup? 4) Were you told before the photo lineup that it would be just as important to clear innocent persons from suspicion as it would be to identify the guilty? 5) Were you told before the*

photo lineup that you would have a certain amount of time to make a decision? The correct answer to questions 3 and 4 was always "yes"; the correct answer to questions 2 and 5 was always "no"; the correct answer to question 1 depended on the condition to which the participant was assigned. As per our preregistration, participants were not excluded if they answered these questions incorrectly.

After answering these questions, participants were thanked and debriefed.

Measures. We recorded identification decision (suspect identification, filler identification, non-identification) and decision confidence. These were used to calculate our focal measure of diagnosticity (pAUC), as well as convergent measures of diagnosticity (d' on suspect identifications and the Diagnosticity Ratio).

Researchers must decide how to code multiple identifications from sequential lineups. For *first-yes-counts* lineups, we coded the first "yes" decision as the participant's identification decision, as this reflects the most common practice in the published studies that have used a *first-yes-counts* instruction. Recall, however, that any eyewitnesses who made multiple identifications were required to view a second lap in which they could provide, at most, one identification. For sequential control lineups, we coded the eyewitness's final decision as their identification decision.

Results

Manipulation checks. Overall, participants in the sequential conditions accurately remembered whether a *first-yes-counts* instruction had been presented. Accuracy rates in response to the forced-choice question about the pre-lineup instructions did not significantly differ between the sequential control (88.98%) and *first-yes-counts* (84.26%) conditions, χ^2 (1, N = 559) = 3.11, p = .082.[5] Participants were more likely to respond "yes" to more than

[5] As per our pre-registration, to improve comparability with previous studies, we did not exclude participants who answered the instructional manipulation checks incorrectly. However, we conducted additional exploratory analyses to check the robustness of our

one image in the sequential control condition (17.32%) than in the *first-yes-counts* condition (7.54%), χ^2 (1, N = 559) = 12.57, p < .001. Taken together, these results suggest that most participants remembered the presence/absence of the *first-yes-counts* instruction.[6]

Confirmatory analyses. Our first set of analyses were designed to address our first research questions: what is the impact of the *first-yes-counts* instruction on patterns of identification responses? And does the *first-yes-counts* lineup induce position effects for suspect identification rates? Table 3 shows the frequencies of each decision type, broken down by lineup type, target presence, and suspect position. The lower rows of Table 3 also show frequencies collapsed across suspect position. In accordance with our pre-registered analysis plan (https://osf.io/c4dve/), we analyzed decision frequencies in a series of mixed effects logistic regressions using the package lme4 for R (Bates, Maechler, Bolker, & Walker, 2015; R Core Team, 2019). We created separate regression models for predicting suspect identifications, filler identifications, and non-identifications. The fixed effects were Target Presence (0 = TA, 1 = TP), Lineup Condition (0 = *first-yes-counts*, 1 = Simultaneous, 2 = Sequential control), Suspect Position (0 = Position 2, 1 = Position 5), and their interactions. We built each regression model in three steps: 1) Main effects only; 2) Main effects plus two-way interactions; 3) Main effects, two-way interactions, and the three-way interaction. Random intercepts for perpetrator identity were included in the analyses. The variance for random intercepts for sample type (lab vs. online), and the variances of the random slopes for perpetrator identity and sample type, were estimated to be close to zero

findings if these participants were excluded. These additional analyses showed that removing participants in the sequential control and *first-yes-counts* conditions who responded incorrectly to this question made very little difference to identification response patterns.

[6] For descriptive statistics pertaining to memory for other aspects of the pre-lineup instructions, see the Supplementary Materials, https://osf.io/kwqh5/

and produced a singular model fit and/or caused the models to fail to converge; these random effects were therefore not included in the models.[7]

Our first hypothesis (H1) was that the likelihood of a suspect identification would be lower for *first-yes-counts* lineups than for either simultaneous or sequential control lineups. The main effects model for suspect identifications was a significant improvement over the random effects model, χ^2 (4, N = 896) = 304.68, p < .001. In support of H1, suspect identifications were significantly more likely to occur in simultaneous lineups (32.64%) than in *first-yes-counts* lineups (25.57%), z = 2.64, p = .008, OR = 1.73, 95% CI [1.15, 2.61]. Suspect identifications were also more likely in sequential control lineups (29.13%) than in *first-yes-counts* lineups (25.57%), but the difference was not statistically significant, z = 1.67, p = .095, OR = 1.46, 95% CI [0.94, 2.27]. Suspect identification rates did not significantly differ between simultaneous and sequential control lineups, z = 0.78, p = .43, OR = 1.19, 95% CI [0.77, 1.83]. Unsurprisingly, guilty suspects were more likely to be identified (53.91%) than innocent suspects (4.68%), as evidenced by a statistically significant main effect of Target Presence, z = 13.11, p < .001, OR = 25.10, 95% CI [15.50, 40.64]. The main effect of suspect position was not statistically significant, z = 0.80, p = .42, OR = 1.15, 95% CI [0.81, 1.63]. The two-way interactions model for suspect identifications was not a statistically significant improvement over the main effects model, χ^2 (5, N = 896) = 6.66, p = .25, and the three-way interaction model was not a statistically significant improvement over the two-way interactions model, χ^2 (2, N = 896) = 0.02, p = .99. Consequently, our pre-registered confirmatory analyses were inconclusive regarding our second research question.

[7] We also conducted exploratory analyses with Sample Type (Lab vs Online) as a fixed factor. These analyses revealed that online participants were more likely than lab participants to identify fillers from sequential control lineups, z = 3.37, p < .001, OR = 3.38, 95% CI [1.66, 6.86]. No other effects including Sample Type were statistically significant. Full details of the analyses can be found in the Supplementary Materials: https://osf.io/kwqh5/

Our second hypothesis (H2) was that the likelihood of a non-identification would be higher for the *first-yes-counts* lineup than for simultaneous or sequential control lineups. The main effects model for non-identifications was a statistically significant improvement over the random effects model, χ^2 (4, N = 896) = 111.85, p < .001. As predicted, non-identifications were less likely to occur from simultaneous lineups (43.92%) than from *first-yes-counts* lineups (53.44%), z = 2.75, p = .006, OR = 0.63, 95% CI [0.45, 0.87]. Non-identifications were also significantly less likely to occur from sequential control lineups (42.91%) than from *first-yes-counts* lineups (53.44%), z = 3.01, p = .003, OR = 0.58, 95% CI [0.40, 0.82]. Non-identification rates did not significantly differ between simultaneous and sequential control lineups, z = -0.47, p = .64, OR = 0.92, 95% CI [0.65, 1.30]. Unsurprisingly, Non-identifications were less likely to occur when the target was present (30.43%) than when the target was absent (63.25%), as revealed by the main effect of Target Presence, z = 9.80, p < .001, OR = 0.24, 95% CI [0.18, 0.32]. The main effect of suspect position on non-identifications was not statistically significant, z = 0.92, p = .36, OR = 0.88, 95% CI [0.66, 1.16]. The two-way interactions model did not provide a statistically significant improvement in model fit over the main effects model, χ^2 (5, N = 896) = 6.32, p = .28. The three-way interactions model did not provide a statistically significant improvement in model fit over the two-way interactions model, χ^2 (2, N = 896) = 1.65, p = .44.

We had no directional predictions about the effect of the first-yes-counts instruction on filler identifications. However, our pre-registered analysis plan included an exploratory analysis of filler identification rates using the same linear mixed model approach applied to suspect and non-identifications. The main effects model was a statistically significant improvement over the random effects model, χ^2 (4, N = 896) = 37.00, p < .001. The likelihood of a filler identification did not significantly differ between the *first-yes-counts* (20.98%) and simultaneous lineups (23.44%), z = 0.65, p = .52, OR = 1.13, 95% CI [0.78,

1.66], or between the *first-yes-counts* and sequential control lineups (27.95%), $z = 1.77$, $p = .08$, $OR = 1.43$, 95% CI [0.96, 2.13]. Fillers were less likely to be identified on target-present trials (15.66%) than on target-absent trials (32.07%), $z = -5.27$, $p < .001$, $OR = 0.40$, 95% CI [0.29, 0.55]. The two-way interactions model for filler identifications was not a statistically significant improvement over the main effects model, χ^2 (5, $N = 896$) = 5.80, $p = .33$, and the three-way interaction model was not a statistically significant improvement over the two-way interaction model, χ^2 (2, $N = 896$) = 0.17, $p = .92$.

Our third research question concerned the effect of the *first-yes-counts* instruction on the diagnosticity of the sequential lineup. For these analyses, we collapsed across suspect position, thereby increasing statistical power for detecting differences between lineup conditions. We focus primarily on pAUC, as this has been the focal measure of lineup diagnosticity used in the recent large-sample studies summarized by Seale-Carlisle et al. (2019). Qualitatively similar findings emerged when we considered alternative measures of diagnosticity (d' on suspect identifications and the Diagnosticity Ratio); for brevity, these are reported in the Supplementary Materials.

We plotted ROC curves for the simultaneous, sequential control, and *first-yes-counts* lineups (Figure 1A). As per our pre-registration, the ROC curves were plotted using the estimated false identification rate (number of incorrect selections / 6). To increase the stability of the curves, we collapsed the 101-point confidence scale into five confidence bins (0-29, 30-49, 50-69, 70-89, 90-100). To avoid extrapolating any of the curves beyond the observed data, we calculated pAUC in the false alarm rate region (.05) shared by all curves (i.e., the lowest observed false alarm rate). The pAUC values were calculated and compared using the pROC package for R (Robin et al., 2011). As Figure 1A shows, the curve for the *first-yes-counts* lineup was lowest at each operating point. However, the area under the curve for the *first-yes-counts* lineup (.014, 95% CI [.008, .021]) was not significantly different from

the area under the curve for the simultaneous lineup (.020, 95% CI [.012, .027]), $D = 1.01$, $p = .31$ or the sequential control lineup, (.017, 95% CI [.009, .026]), $D = 0.51$, $p = .61$. The simultaneous and sequential control lineups were also not significantly different, $D = 0.44$, $p = .66$[8].

Exploratory analyses

Although none of the interaction terms including suspect position were statistically significant, the descriptive statistics in Table 3 suggest that when the culprit was present in the lineup and the culprit appeared early, *the first-yes-counts* instruction may have had a disproportionate impact on the correct identification rate. We therefore conducted additional exploratory analyses, predicting correct identifications from target-present trials as a function of lineup type. We created separate mixed effects logistic regression models for suspect position 2 and suspect position 5. Random intercepts were included for Target Person. When the suspect appeared in position 2, the main effect of Lineup Type was statistically significant, χ^2 (2, $N = 226$) = 11.93, $p = .003$. The probability of a correct identification was higher for simultaneous lineups (64.10%) than for sequential *first-yes-counts* lineups (38.37%), $z = 3.25$, $p = .001$, $OR = 2.90$, 95% CI [1.53, 5.53]. The probability of a correct identification was also higher for sequential control lineups (58.06%) than for sequential *first-yes-counts* lineups, $z = 3.25$, $p = .001$, $OR = 2.26$, 95% CI [1.14, 4.46]. The probability of a correct identification did not significantly differ between simultaneous and sequential control lineups, $z = -0.71$, $p = .48$, $OR = 0.78$, 95% CI [0.39, 1.56]. When the suspect appeared in position 5, however, the main effect of Lineup Type was not statistically significant, χ^2 (2, $N = 221$) = 0.31, $p = .85$. Thus, though the interaction term was not statistically significant in our pre-registered analyses, we have tentative evidence that the

[8] The results are qualitatively similar if the ROC curves are extrapolated to be equivalent to the maximum observed false alarm rate.

first-yes-counts instruction has a disproportionate effect on correct identifications when the culprit appears early in the lineup.

Our confirmatory analyses of diagnosticity used the estimated innocent suspect identification rate (i.e., all incorrect identifications divided by the nominal size of the lineup) rather than the observed innocent suspect identification rate. However, estimating the innocent suspect identification rate is potentially problematic for two key reasons: First, this method smooths out incorrect identifications across all lineup members. Given that we have theoretical grounds to expect that suspect position will influence suspect identification rates in first-yes-counts lineups (a hypothesis which received tentative support in the above exploratory analysis), estimating the false identification rate could minimize a true difference in false identifications between conditions. Second, estimating the false identification rate rests on the implicit assumption that the innocent suspect is chosen no more frequently than the average filler. In practice, this is unlikely to be true, even when considerable efforts are made to produce fair lineups (see, for example, Sauer, Palmer, & Brewer, 2019). We therefore conducted additional exploratory analyses, estimating diagnosticity using observed false identifications of the designated innocent suspect. The ROC curves are shown in Figure 1B, and the d' and c analyses are reported in the Supplementary Materials.

We calculated pAUC in the false alarm rate region (.04) shared by all curves (i.e., the lowest observed false alarm rate). The area under the curve for the *first-yes-counts* lineup was significantly smaller (.012, 95% CI [.008, .018]) than the area under the curve for the simultaneous lineup (.020, 95% CI [.013, .025]), $D = 2.13$, $p = .03$. Thus, when a *first-yes-counts i*nstruction was used, a simultaneous lineup advantage emerged. However, when no *first-yes-counts* instruction was used, and participants were given no reason not to believe that they were not free to make and resolve multiple identifications, the difference in pAUC between the simultaneous and sequential (control) lineup (.017, 95% CI [.013, .023]) was not

statistically significant, $D = 0.79$, $p = .43$. Taken together, these results suggest that the *first-yes-counts* protocol can induce an artefactual simultaneous lineup advantage. Finally, the difference between the sequential control and *first-yes-counts* lineups did not significantly differ, $D = 1.39$, $p = .16$.

Since completing our preregistration, several other approaches to assessing lineup diagnosticity and utility have been published, including Yang, Smalarz, Moody, Cabell, and Copp's (2019) expected cost model, Lee and Penrod's (2019) Multi-*d'* model, and Smith, Lampinen, Wells, Smalarz, and Mackovichova's (2018) Deviation from Perfect Performance (DPP). To provide a more complete picture of our findings and to assess the robustness of our conclusions, we also applied these approaches to our data. However, in the interests of brevity, we report these analyses in the Supplementary Materials (https://osf.io/kwqh5/). These measures converged on similar conclusions—that the *first-yes-counts* procedure is associated with identification decisions that are less diagnostic and of less utility than simultaneous and sequential control lineups. One result of particular note is that Yang et al.'s (2019) expected cost model demonstrated that, across the range of parameters we considered, there was never a circumstance in which the *first-yes-counts* procedure was preferable to the sequential control lineup. The expected cost model allows consideration of costs across the full range of hypothetical target-absent base-rates (i.e., a range of prior guilt probabilities). This model showed that the relative costs of using a simultaneous versus a sequential control lineup depended on assumptions about prior probability of guilt. Specifically, when the prior probability of guilt was assumed to be higher than around .50, the simultaneous lineup was associated with lower costs; when the prior probability of guilt was lower than around .50, the sequential control lineup was associated with lower costs.

Discussion

In an experimental study with 896 participants, we set out to address three research questions. First, how does the *first-yes-counts* instruction affect patterns of eyewitness identification outcomes (suspect identifications, non-identifications, and filler identifications)? Participants who completed a sequential lineup with a *first-yes-counts* instruction were less likely to identify the suspect than participants who viewed a simultaneous lineup; the difference between the sequential *first-yes-counts* and sequential control lineups was not statistically significant. Participants in the *first-yes-counts* lineup were more likely to reject the lineup than participants who saw a simultaneous lineup or who completed a sequential lineup with no *first-yes-counts* instruction. These results strongly suggest that the *first-yes-counts* instruction induces a conservative criterion shift.

Second, does the *first-yes-counts* instruction produce order effects in sequential lineups? We observed tentative evidence that the *first-yes-counts* instruction had a larger effect on correct identifications when the suspect appeared in 2 than in position 5. These results suggest that the *first-yes-counts* creates a dynamic criterion shift, whereby the witness initially adopts a conservative criterion, which is then relaxed over the course of the lineup. However, these analyses were exploratory, and we note that the hypothesised interaction terms in our pre-registered analyses were not statistically significant.

Third, does the *first-yes-counts* instruction affect the diagnostic value of identification decisions? Inferential analyses on our focal measure, pAUC, indicated that a simultaneous lineup advantage may be an artefact of the combination of a *first-yes-counts* instruction and a termination rule; however, the critical difference between the simultaneous and *first-yes-counts* lineups was only statistically significant when we used observed false identifications of the designated innocent suspect (as opposed to the estimating false identifications by dividing all target-absent lineup identifications by the nominal size of the lineup). Taken together, our results suggest that the *first-yes-counts* instruction undermines sequential lineup

decision making by inducing conservative criterion shifts and prohibiting changes-of-mind, with potentially negative downstream consequences on diagnosticity.

Applied Implications

A central goal of eyewitness identification research is to determine which procedures produce evidence that is most diagnostic of a suspect's guilt or innocence. There has been considerable debate surrounding the relative diagnostic value of identification decisions from simultaneous and sequential lineups (e.g., Mickes et al., 2012). Several recent experiments have reported that simultaneous lineups are more diagnostic than sequential lineups (see Seale-Carlisle et al., 2019, for a review)—a finding commonly attributed to the analytic strategy used in these studies (Gronlund et al., 2015). These findings have been used to argue that simultaneous lineups should be the preferred method of conducting lineups (Gronlund et al., 2015) and have laid the foundation for the development of new theoretical models of eyewitness identification decision making (e.g., Wixted & Mickes, 2014; Wixted, Vul, Mickes, & Wilson, 2018). Our results, however, suggest that the simultaneous lineup advantage observed in these studies may be at least partially an artefact of the sequential lineup protocols used. Notably, in a meta-analytic comparison of simultaneous and sequential lineups (Seale-Carlisle et al., 2019), the weighted mean difference in pAUC values was small—just 0.0103, 95% CI [0.0064 to 0.0142]. This small difference could readily be accounted for by the use of *first-yes-counts* instructions (Mickes et al., 2012; Gronlund et al., 2012) and termination rules (Andersen et al., 2014; Carlson & Carlson, 2014; Meisters et al., 2018).

Our results experimentally demonstrate that a *first-yes-counts* instruction with sequential lineups can create a simultaneous lineup advantage in measures of diagnosticity such as pAUC. However, when eyewitnesses who received a sequential lineup were *not* given a *first-yes-counts* instruction and were given no reason to believe that they could not

make and resolve multiple ‘yes’ decisions, we found negligible differences between simultaneous and sequential lineups in decision outcomes and in the diagnostic value of those outcomes. We can also tentatively suggest that the *first-yes-counts* instruction can induce problematic order effects wherein the likelihood that a suspect will be identified depends upon where in the lineup he is placed.

The lack of any meaningful differences between the simultaneous and sequential control conditions in this study may come as a surprise to many readers, given Steblay et al. (2011)’s meta-analytic findings. They found that sequential lineups were associated with lower correct identification rates, higher correct rejection rates, and higher Diagnosticity Ratios than simultaneous lineups. However, the meta-analytic effect sizes (expressed as Cohen’s *h*) for the differences in correct identification rates and correct rejection rates were .16 and .45, respectively in Steblay et al. The average sample size (reported by Palmer & Brewer, 2012, in their re-analysis of the same corpus of studies) was approximately 45 participants per cell. Using the pwr package for R (Champley, 2018), we calculated that the statistical power of the average study was just 56% for the difference in correct rejections and 12% for the difference in correct identifications. Where a meta-analysis is built on underpowered studies, it becomes difficult to state with any degree of certainty whether the true effects being estimated are actually different from zero (Luke, 2019).

To make matters worse, the meta-analytic effect sizes reported by Steblay et al. (2011) were likely to be overestimates. There has been growing awareness that meta-analyses routinely overestimate effect sizes, and that this occurs because of systemic problems in the research studies that comprise the meta-analysis, including undisclosed flexibility in design, analysis, and reporting; low statistical power; and publication bias (Kvarven, Strømland, & Johannesson, 2019). If the meta-analytic effects reported by Steblay et al. are overestimated by a factor of two, then the true effects may be closer to Cohen’s h = .08 and .23. This would

mean that the power of the average study in Steblay et al.'s meta-analysis was actually closer to 19% for the difference in correct rejections and 7% for the difference in correct identifications. Our results are inconclusive as to whether optimal sequential lineups are superior (or inferior) to simultaneous lineups. However, we would argue that the entire literature is currently inconclusive. High-powered, high-quality comparison tests, using appropriate sequential lineup protocols, are sorely needed to answer this question.

Theoretical implications

Our results suggest that the *first-yes-counts* protocol reduces suspect identifications and increases non-identifications. This pattern is inconsistent with what we would expect if the *first-yes-counts* protocol left participants' response bias unaffected and exerted its influence in a purely mechanistic manner by disallowing changes-of-mind. In that case, the lost suspect identifications would be displaced onto fillers, leaving the non-identification rate unchanged. Rather, our results suggest that the *first-yes-counts* instruction exerts a *psychological* effect on participants, causing them to shift their decision criteria to a more conservative position.

We considered two different types of criterion shift—a static shift to a more conservative position, which is then maintained constant across all lineup members, and a dynamic shift, which is gradually relaxed as the lineup progresses. These two types of criterion shifts can be disentangled by examining position effects. The *static-criterion-shift* account predicts that the magnitude of the *first-yes-counts* effect will be stronger when the suspect appears *late* in the lineup, whereas the *dynamic-criterion-shift* account makes the opposite prediction. Our findings are qualitatively more similar to the predictions of the *dynamic-criterion-shift* account: the magnitude of the decrease in correct identifications associated with the *first-yes-counts* instruction was larger when the suspect appeared in position 2 than when the suspect appeared in position 5. We acknowledge that our evidence

for dynamic criterion shifts is tentative, as the interaction between Suspect Position and Lineup Type was not statistically significant. Confirmatory research with larger samples is needed to better test this hypothesis. However, we note that evidence for dynamic criterion shifts in sequential lineups has been observed in prior studies, and has been taken to suggest that participants respond increasingly liberally if they feel that they are running out of opportunities to make a positive identification (Carlson et al., 2016; Horry et al., 2012; Meisters et al., 2018; see also Smith et al., 2015).

Disentangling the *first-yes-counts* instruction from the termination rule

In our main analyses, we deliberately confounded the *first-yes-counts* instruction with the application of a termination rule. We did so for good reason; to the best of our knowledge, every study that has used a *first-yes-counts* instruction has also employed a termination rule. In contrast, while termination rules are sometimes applied in the absence of a *first-yes-counts* instruction (e.g., Andersen et al., 2014; Carlson & Carlson, 2014), there are also many instances of studies where eyewitnesses are allowed to change their minds. Indeed, this latter procedure maps far more closely onto police procedures (see Wells et al., 2015). Therefore, by comparing *first-yes-counts* lineups with a termination rule and control lineups without a termination rule, we were able to consider meaningfully how the protocols that have come to dominate in the laboratory differ from those that would be employed in practice. However, our data do allow us to separate the effects of the *first-yes-counts* instruction and the termination rule, as participants in both sequential conditions proceeded through the entire lineup and were given the opportunity to resolve multiple identifications. We consider these issues descriptively below.

First, we can explore what would happen if a termination rule was imposed on the sequential control data, mimicking studies in which sequential lineups are terminated following a positive response, but participants are unaware that this will happen (e.g.,

Andersen et al., 2014; Carlson & Carlson, 2014; Meisters et al., 2018). When the target was present in the lineup, applying a termination rule reduced the overall correct identification rate from 56.6% to 50.0% and increased the filler identification rate from 18.9% to 25.4%. Thus, applying a termination rule—even when participants are unaware of it—decreases the correct identification rate and displaces those lost identifications onto fillers.

Splitting the data by suspect position makes clear that the termination rule exerts its influence more strongly when the suspect is placed later in the lineup. When the target was present and in position 5, applying a termination rule reduced the correct identification rate from 55.0% to 43.3%, with a corresponding increase in the filler identification rate from 21.7% to 33.3%. When the target was present in position 2, applying a termination rule reduced the correct identification rate only slightly, from 58.1% to 56.5%; the corresponding increase in filler identifications was from 16.1% to 17.7%. In other words, if we had considered the first positive decision as final, the overall correct identification rate would have dropped and suspect position effects would have emerged in correct identifications—in this case, a 13 percentage point difference between positions 2 and 5. Thus, even if participants are not provided with a *first-yes-counts* instruction, imposing a termination rule or counting only the first positive decision artefactually reduces the correct identification rate, increases the target-present filler identification rate, and introduces problematic order effects.

Importantly, the reduction in the correct identification rate (cf. the false identification rate) is disproportionate—and this is especially true in designs with no designated innocent suspect (see, for example, Mickes et al., 2012, Experiment 1). Our data demonstrates that the application of a termination rule has no effect on *whether* a lineup member is identified; what changes is *who* is identified. If there is no designated innocent suspect, and the false identification rate is therefore estimated by dividing the number of filler identifications by the nominal size of the lineup, then the termination rule will have no effect on the estimated false

identification rate. The combination of a lower correct identification rate and an unchanged false identification rate is necessarily a reduction in diagnosticity, however it is measured.

We can also isolate the effect of the *first-yes-counts* instruction by coding the final, resolved decisions of participants in the *first-yes-counts* condition that, despite the instruction, made more than one "yes" response. Because few participants in the *first-yes-counts* condition made multiple responses (n = 23), these effects are small. However, allowing these changes-of-mind would have increased the overall correct identification rate from *first-yes-counts* lineups slightly from 44.9% to 48.1%. This figure is still lower than in the sequential control condition (56.6%). Clearly, then, the *first-yes-counts* instruction reduces correct identifications from sequential lineups, even if a termination rule is not then imposed. In other words, the *first-yes-counts* instruction exerts a *psychological* influence on responding, not just a *statistical* one.

Observed vs estimated innocent suspect identification rate

When calculating measures of diagnosticity, it is necessary to derive both a correct identification rate and a false identification rate (of the innocent suspect) from the data. Two distinct approaches to deriving the false identification rate exist in the literature: estimating the false identification rate by dividing all correct choices by the nominal size of the lineup, and using the observed identification rate of a designated innocent suspect. The choice of which of these two procedures is necessarily constrained in cases with no designated innocent suspect. However, assuming that an experiment has been conducted with a designated innocent suspect, it is possible to apply either approach. Our findings demonstrate that the two approaches will not always produce the same conclusions. Specifically, our observed effect sizes were smaller, and our contrasts non-significant, when we used the estimated innocent suspect rate than when we used identifications of the designated identification suspect.

One drawback of the estimated false identification rate approach is that false identifications are distributed across all members of the lineup, which is likely to reduce differences in the false identification rate between conditions. As a consequence, statistical power for comparisons that include the false identification rate will be lower. Furthermore, if there is some theoretical reason to expect that the position of the suspect in the lineup will influence decision-making, then the estimated false identification rate approach will likely mask those effects in the target-absent data.

On the other hand, one could argue that the observed false identification rate is open to being buffeted about by noisy data—a few additional chance identifications of the innocent suspect could have a disproportionate effect on measures of diagnosticity. To some extent, larger samples will buffer this concern. However, even when thousands of participants are tested, fair target-absent lineups still produce a relatively small number of innocent suspect identifications.

These two false identification rates offer different information that can be informative to policy-makers. The estimated false identification rate assumes that the target-absent lineup is perfectly fair, and so would generalize most readily to fair lineups. However, studies of real-world lineups suggest that they are unlikely to be perfectly fair (e.g., Valentine, Harris, Colom Piera, & Darling, 2003; Valentine & Heaton, 1999). Furthermore, evidence from the laboratory suggests that even when considerable efforts are made to produce fair lineups, some degree of lineup bias is likely to be evident (Sauer et al., 2019). Using the observed innocent suspect identification rate does not introduce the same assumptions about lineup fairness and arguably generalizes more readily to a wider range of real-world scenarios, including those in which the innocent suspect has a higher probability of being selected than the fillers.

Often, researchers must make decisions about how to analyze their data from multiple, defensible alternatives. This creates numerous potential analysis pipelines, which can lead to quite different outcomes. Recently, techniques such as 'multiverse analysis' have been published, which make transparent the impact of analytic choices on study outcomes (Steegan, Tuerlinckx, Gelman, & Vanpaemel, 2016). In a similar vein, we would encourage researchers to present diagnosticity estimates using both techniques for deriving false identifications, which would increase transparency and also inform considerations of generalizability.

Broader Implications

Sequential lineups provide considerable room for methodological and analytical flexibility. Indeed, when scrutinizing the methods and results sections of published studies, we were struck by the variability in how sequential lineups have been administered. We were further struck by how difficult it often was to determine exactly what had been done—particularly concerning what eyewitnesses had been told prior to the lineup. The present findings, together with prior research (e.g., Carlson et al., 2016; Horry et al., 2012; Horry et al., 2015; Steblay et al., 2011), demonstrate how relatively minor changes to sequential lineup protocols can impact decision outcomes, with downstream consequences for the diagnostic value of the identification evidence. We recommend that, as a field, we adopt gold standard protocols for sequential lineups, and that these gold standards are informed by a careful consideration of protocols that are feasible for practice. Though further research will be needed to determine the optimal protocol, we suggest that it might look something like the protocol used by Wells et al. (2015) in their field experiment: 1) eyewitness are informed before the lineup is shown that they will be required to view all of the lineup members (to prevent eyewitnesses from inferring that they made a mistake because the lineup procedure continued after an identification); 2) eyewitnesses are not informed of the number of images

that are included in the lineup; 3) eyewitnesses are not instructed that they can only make a single yes response or provided with any instruction that implies that only their first yes will count; 4) eyewitnesses are shown all lineup members regardless of the decisions made (i.e., a termination rule should not be applied); and 5) eyewitnesses who make multiple yes responses are given the opportunity to resolve their decision.

In terms of reporting guidelines, each of the five points above should be explicitly described in the manuscript. Pre-lineup instructions should be reproduced in the manuscript or supplementary materials verbatim. Where it is feasible to do so, we would encourage researchers to make their protocols openly available through repositories such as the OSF (www.osf.io). For example, where lineups are computer administered, experimental scripts could be deposited on the OSF and/or links to demonstration versions could be embedded in manuscripts. Where lineups are administered manually, an example experimental session (with a mock participant) could be video recorded and shared through the OSF or a transcript provided.

There are additional sequential protocol aspects for which the data are less clear, such as the use of a "don't know" response option (e.g., Steblay & Phillips, 2011; Weber & Perfect, 2012), and the possibility of viewing multiple laps (e.g., Horry et al., 2015; Steblay et al., 2011). If a goal of research is to investigate the relative diagnosticity of procedures that are in widespread usage, then it is worth looking to how sequential procedures are implemented in practice. Bertrand et al. (2018), for example, found that approximately half of police officers surveyed allowed witnesses to see more than one lap of the lineup, indicating that lapping is fairly common in practice. Further, detailed surveys covering jurisdictions of North America could inform our laboratory protocols, enabling us to maximise the external validity of our research.

Limitations and Future Directions

As previously noted, despite a sample approaching 900 participants, the variability in the data meant that we lacked statistical power to detect all but large differences in our convergent diagnosticity metrics and to detect statistically significant interactions between Suspect Position and Lineup Type on identification decisions. Consequently, our conclusions regarding these facets of the results remain tentative and await confirmation from replications with larger samples. Such studies may enable us to more firmly distinguish between dynamic and static criterion shifts, and to more accurately determine how the *first-yes-counts* instruction impacts different metrics of diagnosticity.

There is considerable variability in the wording of *first-yes-counts* instructions in the literature. We modelled our instructions on those used by Horry et al. (2012) which incorporated two components: 1) that only the first "yes" would count; and 2) that the participant would not be able to say "yes" to any subsequent images. While some studies do explicitly inform their participants that a "yes" response will prohibit any further opportunities to say "yes" – potentially because the lineup will be terminated (e.g., Dobolyi & Dodson, 2013; Douglass & McQuiston-Surrett, 2006; Humphries & Flowe, 2015) – other studies do not provide any explicit instructions about what will happen following a "yes" response (e.g., Carlson et al., 2016; Mickes et al., 2012). It is reasonable to assume that participants likely infer that they will not be able to say "yes" to any subsequent images—or that, even if it is possible to say "yes" to a subsequent decision, there is little point in doing so, as this decision will be disregarded. However, it remains an open question as to whether both elements of this *first-yes-counts* instruction are necessary to create the effects we observed here.

Though we modelled the protocol in our control condition on those used in the real world (e.g., Norwood Police Department, MA; Wells et al., 2015), there are, of course, important aspects of the eyewitness's experience that we could not capture. For example, in

practice, eyewitnesses may request (and be granted) a second lineup lap regardless of the decisions that they made previously. It is difficult to emulate this laps-on-request protocol in the lab (especially so online), but it is worth considering how lapping might interact with the *first-yes-counts* instruction. Previous studies have found that responding tends to become more liberal from a first to a second lap; participants appear to recalibrate their decision criterion to make up for lost identifications the first time around (Horry et al., 2015; Steblay et al., 2011). Perhaps *first-yes-counts* participants would particularly benefit from a second lap, as they would be able to recoup some of the lost perpetrator identifications that resulted from setting their decision criteria conservatively. Effectively, a protocol that allows lapping may at least partially offset the damaging impact of the *first-yes-counts* instruction—though lapping is also typically accompanied by an increase in false identifications (Steblay, Dietrich, et al., 2011).

On the other hand, the lab experience may soften the impact of the *first-yes-counts* instruction compared to the real-world experience of an actual eyewitness. In the laboratory, witnesses are aware that their errors carry no consequences. If they happen to pick a filler and a better match comes along later, then so what? Contrast this with the experience of a real eyewitness who fears that he/she will lose their only opportunity to identify the perpetrator if they happen to pick a filler earlier in the sequence. It is reasonable to expect that the *first-yes-counts* instruction will create stronger criterion shifts in real eyewitnesses than in laboratory participants (see Eisen, Smith, Olaguez, & Skerrit-Perta, 2017, for a compelling demonstration of the impact of participant beliefs about decision consequences on showup outcomes). Consequently, the impact of the *first-yes-counts* instruction may be underestimated here. It will be useful for moving forward in our understanding of the sequential procedure to ask eyewitnesses what they expect to happen when they view sequential lineups. Perhaps many participants anticipate a *first-yes-counts* approach and this

is why they use a more conservative response criterion than with simultaneous lineups. If so, then informing participants that they will be able to resolve multiple identifications may counteract this conservative shift, at least to some extent.

In any study, concerns about ecological validity must be balanced against pragmatic and ethical considerations. In this instance, to increase the ease of participant recruitment, and to minimise attrition, we used a non-violent mock crime and a short retention interval, allowing participants to complete the study within a single experimental session. Of course, real witnesses may witness violent, distressing events and will also experience much longer retention intervals than we have used here. Both stress and retention interval have been shown to decrease the accuracy of eyewitness identification decisions (e.g., Deffenbacher, Bornstein, Penrod, & McGorty, 2004; Sauer, Brewer, Zweck, & Weber, 2010). How such factors might interact with the use of a *first-yes-counts* instruction is unknown. However, we cannot think of a compelling theoretical or applied reason to use a *first-yes-counts* instruction under any circumstances.

Conclusions

Informing participants that only their first yes will count undermines decision-making from sequential lineups. Specifically, the *first-yes-counts* instruction reduces the proportion of suspect identifications and increases the proportion of non-identifications. These results are inconsistent with an assumption that the impact of the *first-yes-counts* instruction is purely a statistical one that results from the application of a termination rule. Rather, the results strongly suggest that the *first-yes-counts* instruction has psychological consequences for eyewitnesses—that it produces conservative criterion shifts. Somewhat more tentatively, our results suggest that these criterion shifts may be dynamic, with participants relaxing their decision criteria as the lineup progresses. Further research will be required to confirm this

hypothesis. Our results also provide preliminary evidence that the *first-yes-counts* instruction reduces the diagnostic value of identification decisions, measured by metrics such as pAUC.

The debate concerning simultaneous and sequential lineups has reignited recently, as several large-sample studies have reported that simultaneous lineups produce identification evidence that is more diagnostic than identification evidence from sequential lineups (see Seale-Carlisle et al., 2019, for a review). Our results suggest that the simultaneous advantage that has emerged from those studies may be at least partly attributable to artefacts of the sequential lineup protocols used in these studies, such as the *first-yes-counts* instruction and the termination rule. Going forward, it is imperative that we are conscientious about reporting how lineups are administered and how identification decisions are coded. Deviations from how sequential lineups are administered in practice and from the original protocol developed by Lindsay and Wells (1985), should be explicitly justified and the implications of those deviations should be carefully considered and clearly communicated. By reducing inconsistencies in protocols between labs, and by ensuring that our protocols accurately represent those used in practice, we can more precisely answer pressing questions regarding lineup administration that practitioners and policymakers wish to know.

References

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

Andersen, S. M., Carlson, C. A., Carlson, M. A., & Gronlund, S. D. (2014). Individual differences predict eyewitness identification performance. *Personality and Individual Differences, 60,* 36-40. doi: 10.1016/j.paid.2013.12.011

Bertrand, M. I., Lindsay, R. C. L., Mansour, J. K., Kalmet, N., & Melsom, E. I. (2018). Examining how lineup practices of Canadian and U.S. police officers adhere to their national best practice recommendations. *Manitoba Law Journal, 41,* 1-47.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1-48. doi: 10.18637/jss.v067.i01

Canty, A., & Ripley, B. (2019). boot: Bootstrap R (S-Plus) Functions. R packge version 1.3-22.

Carlson, C. A. (2011). Influence of a perpetrator's distinctive facial feature on eyewitness identification from simultaneous versus sequential lineups. *Applied Psychology in Criminal Justice, 7,* 77-92.

Carlson, C. A., & Carlson, M. A. (2014). An evaluation of lineup presentation, weapon presence, and a distinctive feature using ROC analysis. *Journal of Applied Research in Memory and Cognition, 3,* 45-53. doi: 10.1016/j.jarmac.2014.03.004

Carlson, C. A., Carlson, M. A., Weatherford, D. R., Tucker, A., & Bednarz, J. (2016). The effect of backloading instructions on eyewitness identification from simultaneous and sequential lineups. *Applied Cognitive Psychology, 30,* 1005-1013. doi: 10.1002/acp.3292

Carlson, C. A., Gronlund, S. D., & Clark, S. E. (2008). Lineup composition, suspect position, and the sequential lineup advantage. *Journal of Experimental Psychology: Applied, 14,* 118-128. doi: 10.1037/1076-898X.14.2.118

Carlson, C. A., Jones, A. R., Goodsell, C. A., Carlson, M. A., Weatherford, D. R., Whittington, J. E., & Lockamyeir, R. F. (2019). A method for increasing empirical discriminability and eliminating top-row preference in photo arrays. *Applied Cognitive Psychology.* doi: 10.1002/acp.3551

Champley, S., (2018). pwr: Basic functions for power analysis. R package version 1.2-2. https://CRAN.R-project.org/package=pwr

Clark, S. E., & Davey, S. L. (2005). The target-to-foils shift in simultaneous and sequential lineups. *Law and Human Behavior, 29,* 151-172. doi: 10.1007/s10979-005-2418-7

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

Clark, S. E., & Wells, G. L. (2008). On the diagnosticity of multiple-witness identifications. *Law and Human Behavior, 32,* 406-422. doi: 10.1007/s10979-007-9115-7

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60,* 170-180. doi: 10.1037/0003-066X.60.2.170

Cutler, B. L., & Penrod, S. D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. *Journal of Applied Psychology, 73,* 281-290. doi: 10.1037/0021-9010.73.2.281

Deffenbacher, K. A., Bornstein, B. A., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior, 28,* 687-706. doi: 10.1007/s10979-004-0565-x

Dobolyi, D. G., & Dodson, C. S. (2013). Eyewitness confidence in simultaneous and sequential lineups: A criterion shift account for sequential mistaken identification overconfidence. *Journal of Experimental Psychology: Applied, 19,* 345-357. doi: 10.1037/a0034596

Douglass, A. B., & McQuiston-Surrett, D. (2006). Post-identification feedback: Exploring the effects of sequential photospreads and eyewitness' awareness of the identification task. *Applied Cognitive Psychology, 20,* 991-1007. doi: 10.1002/acp.1253

Eisen, M. L., Smith, A. M., Olaguez, A. P., & Skerrit-Perta, A. S. (2017). An examination of showups conducted by law enforcement using a field-simulation paradigm. *Psychology, Public Policy, and Law, 23,* 1-22. doi: 10.1037/law0000115

Elphick, C. E. J., Pike, G. E., & Hole, G. K. (2019). You can believe your eyes: Measuring implicit recognition in a lineup with pupillometry. *Psychology, Crime & Law.* doi: 10.1080/1068316X.2019.1634196

Flowe, H. D., Smith, H. M. J., Karoğlu, N., Onweugbusi, T. C., & Rai, L. (2016). Configural and component processing in simultaneous and sequential lineup procedures. *Memory, 24,* 306-314. doi: 10.1080/09658211.2015.1004350

Goldfield, K. G. (2019). simstudy 0.1.14. https://CRAN.R-project.org/package=simstudy

Goodsell, C. A., Gronlund, S. D., & Carlson, C. A. (2010). Exploring the sequential lineup advantage using WITNESS. *Law and Human Behavior, 34,* 445-459. doi: 10.1007/s10979-009-9215-7

Gronlund, S. D., Carlson, C. A., Dailey, S. B., & Goodsell, C. A. (2009). Robustness of the sequential lineup advantage. *Journal of Experimental Psychology: Applied, 15,* 140-152. doi: 10.1037/a0015082

Gronlund, S. D., Mickes, L., Wixted, J. T., & Clark, S. E. (2015). Conducting an eyewitness lineup: How the research got it wrong. In B. H. Ross (Ed.) *Psychology of Learning and Motivation, Volume 63.* Waltham, MA: Academic Press.

Horry, R., Brewer, N., Weber, N., & Palmer, M. A. (2015). The effects of allowing a second sequential lineup lap on choosing and probative value. *Psychology, Public Policy, and Law, 2,* 121-133. doi: 10.1037/law0000041

Horry, R., Palmer, M. A., & Brewer, N. (2012). Backloading in the sequential lineup prevents within-lineup criterion shifts that undermine eyewitness identification performance. *Journal of Experimental Psychology: Applied, 18,* 346-360. doi:10.1037/a0029779

Humphries, J. E., & Flowe, H. (2015). Receiver operating characteristic analysis of age-related changes in lineup performance. *Journal of Experimental Child Psychology, 132,* 189-204. doi: 10.1016/jecp.2014.12.009

Kneller, W., Memon, A., & Stevenage, S. (2001). Simultaneous and sequential lineups: Decision processes of accurate and inaccurate eyewitnesses. *Applied Cognitive Psychology, 15,* 659-671. doi: 10.1002/acp.739

Kvarven, A., Strømland, E., & Johannesson, M. (2019). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour.* doi: 10.1038/s41562-019-0787-z

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

Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identification decisions from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70,* 556-564. doi: 10.1037/0021-9010.70.3.556

Luke, T. J. (2019). Lessons from Pinocchio: Cues to deception may be highly exaggerated. *Perspectives on Psychological Science, 14,* 646-671. doi: 10.1177/1745691619838258

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

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

Meisters, J., Diedenhofen, B., & Musch, J. (2018). Eyewitness identification in simultaneous and sequential lineups: An investigation of position effects using receiver operating characteristics. *Memory, 26,* 1297-1309. doi: 10.1080/09658211.2018.1464581

Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous versus sequential lineups. *Journal of Experimental Psychology: Applied, 18, 4,* 361-376. doi: 10.1037/a0030609

Mickes, L., Moreland, M. B., Clark, S. E., & Wixted, J. T. (2014). Missing the information needed 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

Morgan, C. A., Hazlett, G., Doran, A., Garrett, S., Hoyt, G., Thomas, P., Baranoski, M., & Southwick, S. M. (2004). Accuracy of eyewitness memory for persons encountered during exposure to highly intense stress. *International Journal of Law and Psychiatry, 27,* 265-279. doi: 10.1016/j/ijlp.2004.03.004

Oriet, C., & Fitzgerald, R. (2018). The single lineup paradigm: A new way to manipulate target presence in eyewitness identification experiments. *Law and Human Behavior, 42,* 1-12. doi: 10.1037/lhb000272

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

Police Executive Research Forum (2013). A national survey of eyewitness identification procedures in law enforcement agencies. *National Institute of Justice.* Washington, DC.

R Core Team, (2019). R: A language for statistical computing. R foundation for Statistical Computing, Vienna, Austria.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyse and compare ROC curves. *BMC Bioinformatics, 12,* 77. Doi: 10.1186/1471-2105-12-77

Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence-accuracy relationship for eyewitness identification. *Law and Human Behavior, 34,* 337-347. doi: 10.1007/s10979-009-9192-x

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

Seale-Carlisle, T. M., & Mickes, L. (2016). US line-ups outperform UK line-ups. *Royal Society Open Science, 3:* 160300. doi: 10.1098/rsos.160300

Seale-Carlisle, T. M., Wetmore, S. A., Flowe, H. D., & Mickes, L. (2019). Designing police lineups to maximize memory performance. *Journal of Experimental Psychology: Applied, 25,* 410-430. doi: 10.1037/xap0000222

Smith, A. M., Lampinen, J. M., Wells, G. L., Smalarz, L., & Mackovichova, S. (2018). Deviation from perfect performance measures the diagnostic utility of eyewitness lineups but partial area under the curve does not. *Journal of Applied Research in Memory and Cognition, 8,* 50-59.

Smith, A. R., Wells, G. L., Lindsay, R. C. L., & Myerson, T. (2018). Eyewitness identification performance on showups improves with an additional-opportunities instruction: Evidence for present-absent criteria discrepancy. *Law and Human Behavior, 42,* 215-226. doi: 10.1037/lhb000284

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology, 78,* 22-33.

Steblay, N. K., Dietrich, H. L., Ryan, S. L., Raczynski, J. L., & James, K. A. (2011). Sequential lineup laps and eyewitness accuracy. *Law and Human Behavior, 35,* 262-274. doi: 10.1007/s10979-010-9236-2

Steblay, N. K., Dysart, J. E., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 25,* 459-473. doi: 10.1023/A:1012888715007

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

Steblay, N. K., & Phillips, J. D. (2011). The not-sure response option in sequential lineup practice. *Applied Cognitive Psychology, 25,* 768-774. doi: 10.1002/acp.1755

Steegan, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science, 11,* 702-712. doi: 10.1177/1745691616658637

Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis significance tests. *Psychological Methods, 6,* 371-386.

Valentine, T., Harris, N., Colom Piera, A., & Darling, S. (2003). Are police video identifications fair to African-Caribbean suspects? *Applied Cognitive Psychology, 17,* 459-476. doi: 10.1002/acp.880

Valentine, T., & Heaton, P. (1999). An evaluation of the fairness of police line-ups and video identifications. *Applied Cognitive Psychology, 13,* S59-S72. doi:10.1002/(SICI)1099-0720(199911)13:1+<S59::AID-ACP679>3.0.CO;2-Y

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

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

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

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

Wilcock, R. A., Bull, R., & Vrij, A. (2005). Aiding the performance of older eyewitnesses: Enhanced non-biased line-up instructions and line-up presentation. *Psychiatry, Psychology, and Law, 12,* 129-140. doi: 10.1375/pplt.2005.12.1.129

Wilson, B. M., Donnelly, K., Christenfeld, N., & Wixted, J. T. (2019). Making sense of sequential lineups: An experimental and theoretical analysis of position effects. *Journal of Memory and Language, 104,* 108-125. doi: 10.1016/j.jml.2018.10.002

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

Wixted, J. T., Vul, E., Mickes, L., & Wilson, B. M. (2018). Models of lineup memory. *Cognitive Psychology, 105,* 81-114. doi: 10.1016/j.cogpsych.208.06.001

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

Zhou, H., & Fishbach, A., (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology, 111,* 493-504. doi: 10.1037/pspa0000056

Table 1. Use of *first-yes-counts* instructions and termination rules in studies summarized by Seale-Carlisle et al. (2019)

Study	First-yes-counts instruction	Termination rule
Mickes, Flowe, & Wixted (2012) Experiment 1A	Yes	Yes
Mickes, Flowe, & Wixted (2012) Experiment 1B	Yes	Yes
Gronlund et al. (2012)	Yes	Yes
Andersen, Carlson, Carlson, & Gronlund (2014)	No	Yes
Carlson & Carlson (2014)	No	Yes
Meisters, Diedenhofen, & Musch (2018)	No	Yes
Willing, Diedenhofen, & Musch (unpublished)	No	Yes
Goodsell (unpublished)	No	Yes

Note: This table excludes studies that tested UK sequential procedures.

Table 2. Participant numbers by source

Participant source	N before attrition	N after attrition	N after exclusions
Lab – institution 1	115	115	115
Lab – institution 2	114	114	109
Lab – institution 3	112	112	103
Online – Mturk	321	274	255
Online – Qualtrics	1445	598	257
Online Sample			
Online – social media	153	58	57
Total	2260	1159	896

Table 3. Identification decisions by lineup type and suspect position.

		Target –present			Target-absent		
Lineup type	Suspect position	Target ID	Filler ID	Non-ID	Innocent suspect ID	Filler ID	Non-ID
Simultaneous	2	64.10%	11.54%	24.36%	4.71%	37.65%	57.65%
		[53.02, 73.85]	[6.19, 20.50]	[16.19, 34.94]	[1.85, 11.49]	[28.09, 48.27]	[47.04, 67.6]
		50/78	9/78	19/78	4/85	32/85	49/85
	5	57.30%	13.48%	29.21%	5.88%	30.59%	63.53%
		[46.93, 67.06]	[7.88, 22.10]	[20.78, 39.36]	[2.54, 13.04]	[21.81, 41.05]	[52.92, 72.97]
		51/89	12/89	26/89	5/85	26/85	54/85
Sequential Control	2	58.06%	16.13%	25.81%	3.03%	34.85%	62.12%
		[45.66, 69.52]	[9.00, 27.21]	[16.56, 37.88]	[0.83, 10.39]	[24.48, 46.89]	[50.06, 72.85]
		36/62	10/62	16/62	2/66	23/66	41/66
	5	55.00%	21.67%	23.33%	4.54%	37.88%	57.58%
		[42.49, 66.91]	[13.13, 33.62]	[14.44, 35.43]	[1.56, 12.54]	[27.15, 49.94]	[32.63, 55.92]
		33/60	13/60	14/60	3/66	25/66	38/66
Sequential *first-yes-counts*	2	38.37%	13.95%	47.67%	2.74%	27.40%	69.86%
		[28.80, 48.93]	[8.16, 22.82]	[37.44, 58.10]	[0.75, 12.54]	[18.49, 38.57]	[58.56, 79.17]
		33/86	12/86	41/86	2/73	20/73	51/73
	5	52.78%	19.44%	27.78%	6.76%	24.32%	68.92%
		[41.40, 63.88]	[11.95, 30.03]	[18.76, 39.05]	[2.92, 14.87]	[15.97, 35.20]	[44.10, 66.19]

		38/72	14/72	20/72	5/74	18/74	51/74
				Collapsed across suspect position			
Simultaneous	-	60.48%	12.57%	26.95%	5.29%	34.12%	60.59%
		[52.91, 67.58]	[8.37, 18.45]	[20.79, 34.14]	[2.81, 9.75]	[27.41, 41.53]	[53.09, 67.62]
		101/167	21/167	45/167	9/170	58/170	103/170
Sequential	-	56.56%	18.85%	24.59%	3.79%	36.36%	59.85%
control		[47.70, 65.02]	[12.90, 26.70]	[17.80, 32.93]	[1.63, 8.56]	[28.65, 44.84]	[51.32, 67.82]
		69/122	23/122	30/122	5/132	48/132	79/132
Sequential	-	44.94%	16.46%	38.61%	4.76%	25.85%	69.39%
first-yes-		[37.40, 52.72]	[11.49, 23.02]	[31.08, 46.71%	[2.32, 9.50]	[19.45, 33.48]	[61.52, 76.27]
counts		71/158	26/158	61/158	7/147	38/147	102/147

Note. Values in square brackets are 95% confidence intervals.
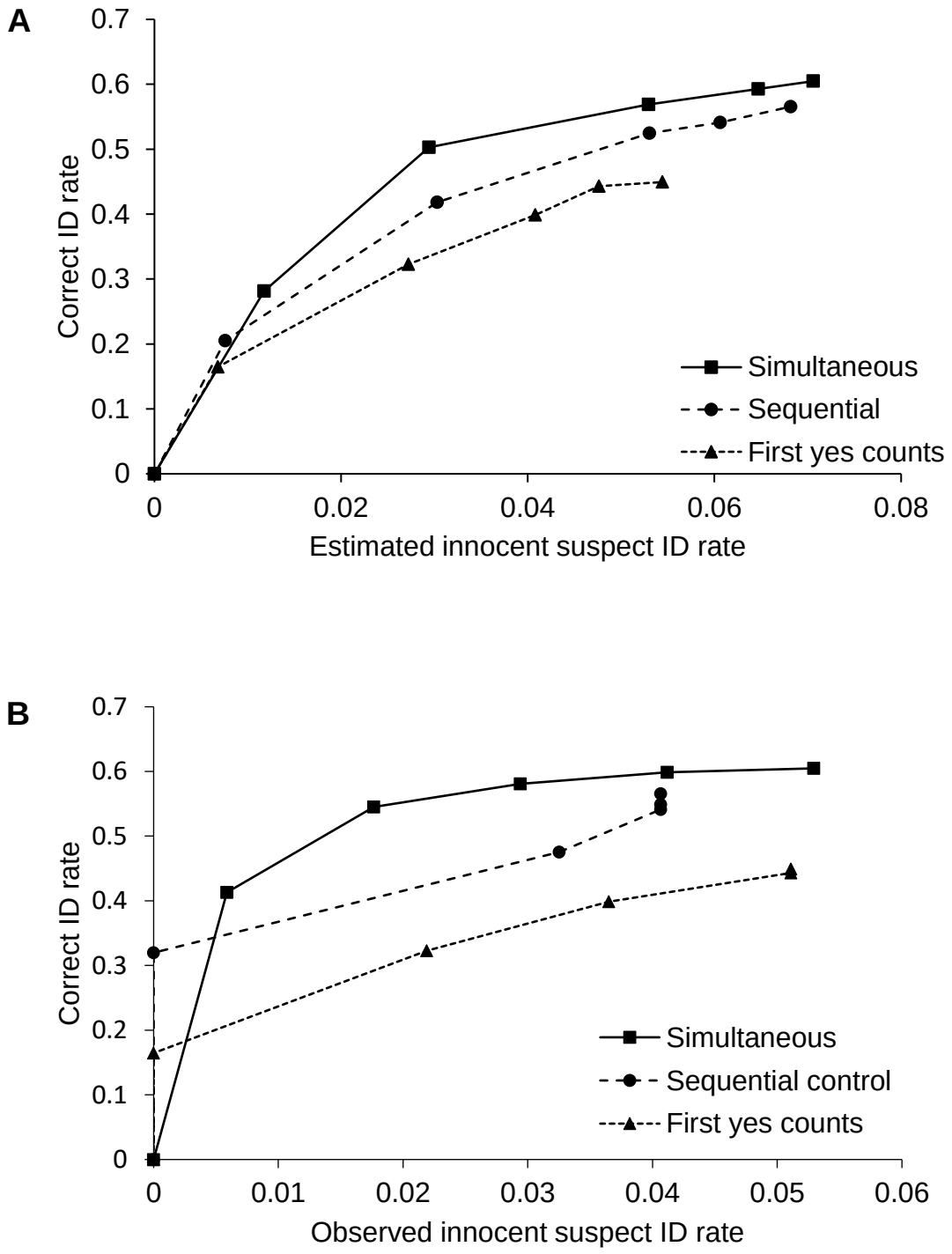
Figure 1. ROC curves for the simultaneous, sequential control, and sequential *first-yes-counts* lineups. Panel A uses the estimated innocent suspect identification rate (filler identifications / 6), and Panel B uses the designated innocent suspect identification rate