Paper:
Horry, R., Brewer, N., Weber, N. & Palmer, M. (in press). The Effects of Allowing a Second Sequential Lineup Lap on Choosing and Probative Value.. *Psychology, Public Policy, and Law*

http://dx.doi.org/10.1037/law0000041

The Effects of Allowing a Second Sequential Lineup Lap on Choosing and Probative Value

Ruth Horry*, Neil Brewer, and Nathan Weber

Flinders University

Matthew A. Palmer

University of Tasmania

Accepted for publication in *Psychology, Public Policy, and Law.* This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.

Author note

Ruth Horry, School of Psychology, Flinders University; Neil Brewer, School of Psychology, Flinders University; Nathan Weber, School of Psychology, Flinders University; Matthew A. Palmer, Psychology, School of Medicine, University of Tasmania.

*Present address: Department of Psychology, College of Human and Health Sciences, Swansea University, Swansea, SA2 8 PP, Wales, UK.

This research was funded by ARC Discovery Grant DP1093210 to N. Brewer, N. Weber, and D. S. Lindsay. We thank Timothy Gillespie, Monique Jenkins, Jessica Shen, Emma Bowen, Anita Bunton, Elizabeth Dos Santos, Sophia Izzo, Annabel MacLachlan, Samantha Wakelin, and Nicola Guerin for their assistance with data collection. We also thank Carmen Lucas, Tomoko Nishizawa, and Nicole Reid for their assistance with preparation of experimental materials, data entry, and coding.

Correspondence concerning this article should be addressed to Ruth Horry, Department of Psychology, College of Human and Health Sciences, Swansea University, Swansea, SA2 8 PP, Wales, UK Email: R.Horry@swansea.ac.uk

Abstract

When presented with a sequential lineup, witnesses see each member of the lineup individually, essentially making a yes/no decision for each person shown. An important policy question is whether witnesses should be allowed to see an additional lap of a sequential lineup. We investigated the impact of a second lap on eyewitness decision-making and on the probative value of suspect identifications. We recruited a large community sample of participants (*N* = 393), each of whom viewed a target person before seeing a sequential lineup that did or did not include the target. A second lap was either required or optional. The group of participants who accepted the second lap were less able to discriminate between the target and the fillers and responded more conservatively in lap 1 than the group of witnesses who declined the second lap. Responding became more lenient from lap 1 to lap 2. Of the participants who saw a second lap, roughly 40% changed their response, most frequently from a non-identification to an identification. Both culprit identifications and filler identifications increased from lap 1 to lap 2. The probative value of suspect identifications was not significantly different whether witnesses were allowed two laps or one. However, the observed effects may be moderated by a number of system and estimator variables. Further, even small changes in probative value can have very different consequences depending upon the target-absent base rate.

Keywords: Eyewitness identification; sequential lineup; response bias.

The Effects of Allowing a Second Sequential Lineup Lap on Choosing and Probative Value

The sequential lineup was developed by Lindsay and Wells (1985) to reduce the likelihood that innocent suspects are chosen from lineups. The fundamental element of the sequential procedure is the presentation of lineup members one at a time so that a witness effectively decides whether each lineup member is or is not the culprit. There have been a number of variations on the procedure, including allowing witnesses to see the entire sequence of photographs in the lineup more than once (Lindsay, Lea, & Fulford, 1991; MacLin & Phelan, 2007; Steblay, Dietrich, Ryan, Raczynski, & James, 2011). These modifications have often been made in response to concerns that strict sequential presentation (i.e., a single presentation of each lineup member) may lead to cautious witnesses missing the culprit (Clark, 2012; McQuiston-Surrett, Malpass, & Tredoux, 2006). Many jurisdictions that have adopted the sequential lineup have also included a provision to allow witnesses to see additional lineup laps (Wells, 2014). However, empirical evaluations of the sequential lap procedure are scarce. Here we present empirical evidence on the impact of an additional sequential lineup lap on identification decisions and probative value of suspect identifications. First, however, we review the extant research on sequential lineup laps and detail current uses of the sequential lap procedure and its variants in the field.

Empirical data on sequential lineup laps

In Lindsay and Wells’ (1985) original sequential lineup study, participants viewed six photographs, recording a yes/no response for each one before moving on to the next. The participants could not return to a previously seen face and the lineup ended when all six photographs had been seen. This procedure differed from the standard simultaneous procedure in several ways, including: i) sequential presentation of images; ii) misrepresentation of the number of photographs in the lineup such that the witness believed he or she would see more images than were actually included; and iii) the requirement for multiple decisions from the participant. The sequential lineup, therefore, is best thought of as a suite of techniques rather than as a simple change in presentation format (Lindsay, Mansour, Beaudry, Leach, & Bertrand, 2009).

In the subsequent decades, modifications have been made to the sequential procedure. While the core aspects of the procedure have largely been maintained (e.g., requiring multiple decisions; misrepresentation, or alternatively, non-disclosure of the number of images to be seen; see Malpass et al., 2008), other aspects of the procedure have been modified. Some researchers, for example, have imposed a stopping rule, whereby the lineup is terminated when the participant chooses a lineup member (e.g., Horry, Palmer, & Brewer, 2012; Kneller, Memon, & Stevenage, 2001); others have allowed the participant to continue through the whole lineup, though the participants are told that only the first ‘yes’ counts (e.g., Carlson, Gronlund, & Clark, 2008; Mickes, Flowe, & Wixted, 2012). A different modification is the use of a “not sure” response option in addition to the standard yes/no response options (Steblay & Phillips, 2011).

The modification on which we focus is the use of a second lineup lap. Three published papers have explored this issue in the laboratory. In the first, Lindsay et al. (1991, Experiment 2) presented a small number of participants (*N* = 32) with a sequential, target-absent lineup. After the first lap of the lineup, half of the participants saw a second lap of the sequential lineup. Only one of these participants changed their response (from a lineup rejection to an innocent suspect identification). The results of this study are difficult to interpret, however, due to the very small sample size (only 16 participants saw two sequential lineup laps), the inclusion of only a target-absent lineup, and the use of a biased lineup (the fillers were deliberately selected to be poor, and the innocent suspect was wearing the same shirt as that worn by the culprit during the crime).

In the second paper, MacLin and Phelan (2007) allocated their participants to one of three conditions: a simultaneous lineup condition; a sequential lineup with the option of a second lap; or a sequential lineup with the option of a subsequent simultaneous lineup. Within the sequential lap condition, correct and incorrect choices increased from lap 1 to lap 2. However, these results must be interpreted cautiously for two reasons. First, the authors did not indicate how many participants saw a second lap, making the lap 2 data difficult to interpret. Second, the data suggested that the target-absent lineup was biased toward the innocent suspect. Fifty per cent of the witnesses who saw a target-absent simultaneous lineup chose someone, with 100% of those choices of the innocent suspect (as opposed to the 17% that would be expected in an unbiased, six-person lineup).

The most comprehensive study of sequential lineup laps to date involves two experiments reported by Steblay, Dietrich, et al. (2011). In Experiment 1, participants watched a videotaped staged crime, and then saw a sequential or simultaneous lineup (either target-present or target-absent). Participants in the sequential condition were offered a second lap of the lineup upon conclusion of the first lap. The key findings were that witnesses who identified no-one from lap 1 were more likely to accept the offer of a second lap than witnesses who did identify someone. Of those witnesses who accepted the second lap, roughly one third switched their response (usually from a non-identification to an identification). Most of these changes, however, were to an erroneous decision (i.e., a filler identification), and this applied to both target-present and target-absent lineups.

In Experiment 2, Steblay, Dietrich, et al. (2011) assigned their participants to one of three sequential lineup lap conditions: single lap, required second lap, or optional second lap. The required second lap condition allowed the authors to ask whether seeing a second lap, in itself, is sufficient to increase incorrect choosing rates. For the optional lap participants, the results were similar to Experiment 1: non-choosers were more likely than choosers to accept a second lap and, when a change in response was made, it was often erroneous. The results were less marked for the participants who were required to see two laps. There was a trend towards an increase in correct suspect identification rates, with correct identifications 1.5 times more likely in lap 2 than in lap 1, as well as a trend towards an increase in filler identifications from target-absent lineups, with filler identifications 1.44 times more likely in lap 2 than in lap 1. Of course, one would expect the effect size to be smaller in the required condition, as this group would have included some participants who would have accepted a second lap as well as some who would have declined a second lap.

The results of Steblay, Dietrich, et al. (2011) suggest that participants shifted to a more lenient decision criterion from a first lap to a second lap. The result was an increase in culprit identifications, which was accompanied by an increase in filler identifications. Thus, Steblay et al.'s results suggested that lapping may not be a particularly good way to ameliorate the impact of the sequential lineup on correct identifications.

In this study we re-investigated the impact of sequential lineup laps on eyewitness decision-making. To ensure that our results were robust and generalizable, we used a field experiment methodology to recruit a large sample of 393 participants, of a wide variety of ages and backgrounds (see also Horry, Palmer, et al., 2012; Lindsay, Semmler, Weber, Brewer, & Lindsay, 2008; Palmer, Brewer, Weber, & Nagesh, 2013). We also used seven different targets, as patterns of identification decisions can vary widely from one lineup to the next, influenced by factors such as target distinctiveness, target-innocent suspect similarity, lineup fairness, and quality of the encoding conditions (see the arguments of Wells & Windschitl, 1999).

The sequential lap procedure in practice

Despite ongoing debate in the scientific literature about whether simultaneous or sequential procedures should be preferred (e.g., Carlson et al., 2008; Gronlund, Wixted, &

Mickes, 2014; Mickes et al., 2012), uptake of the sequential lineup procedure is increasing. Smith and Cutler (2013) examined recommendations for procedural reform from the US, Canada and the UK. Some of these were national policies (for example, the UK's Police and Criminal Evidence Act, 1984, updated in 2011), while others were at state levels (e.g., New Jersey, North Carolina) or local levels (e.g., Northampton, MA; Santa Clara, CA). Approximately 62% of these reforms included some form of sequential lineup presentation. However, a major policy issue in implementing the sequential lineup has been that both correct and false identifications tend to be lower than when the simultaneous lineup is used (e.g., Clark, 2012; McQuiston-Surrett et al., 2006; Palmer & Brewer, 2012). Allowing additional lineup laps offers one potential technique for ameliorating the effects of sequential presentation on correct identification rates, and has therefore appeared in many protocols in the field. Wells (2014) stated that 'every U.S. jurisdiction that has adopted the sequential procedure has included a proviso that the witness can view the sequence a second time if the witness explicitly requests to see the lineup again' (p. 14). In the UK, witnesses are actually required to see two laps of the lineup *before* making a decision; they may then request additional views of the lineup (which do not necessarily have to be full laps – the witness can request to return to specific images; see Horry, Memon, Wright, & Milne, 2012, for a detailed description of UK procedures).

Evidence from the field is scant, but there are at least some suggestions that sequential laps increase choosing rates with potentially negative effects on accuracy. In a large field experiment of double-blind sequential (vs. simultaneous) lineups with witnesses to real crimes, Wells, Steblay, and Dysart (in press) allowed the sequential witnesses to see a second lap, if requested. Thirty-seven witnesses (of 235 who viewed a sequential lineup) viewed a second lap. Suspect identification rates increased from 23.4% to 27.5% from lap 1 to final response, and filler identifications increased from 11.1% to 12.3%. Klobuchar, Steblay, &

Caligiuri (2006) reported a significant difference in filler identification rates between witnesses who viewed one lap (3%) and witnesses who viewed more than one lap (29%). However, this comparison must be viewed cautiously, as data concerning the number of laps seen by witnesses were recorded for only 46% of lineups conducted. Significant increases in filler identifications have been found in archival studies of eyewitness identification in the UK, with witnesses who requested to see one or more faces from the lineup again choosing fillers at more than double the rate of those who made no such requests (Horry, Memon, et al., 2012). However, the procedural differences between the UK and the US (including requiring two laps before a decision is made; allowing returns to specific faces in the lineup) limit the extent to which we can generalize these findings to the sequential lineups used in the US and in most laboratory studies.

Our data allowed us to consider two sequential lineup methods: a sequential lap procedure in which additional laps were offered to all witnesses, and a no-lap procedure, in which all witnesses were restricted to a single lap. However, policy makers have a third option to consider, which is offering a lap only if the witness makes an explicit request for an additional lap. The latter procedure appears to be the status quo, presumably based on the assumption that the extra laps will catch some correct identifications that would have been missed with the no-lap procedure. The laps-by-request procedure is difficult to emulate in the laboratory, as participants rarely spontaneously request additional laps. We return to this issue in the discussion, when we consider the policy implications of our findings. Nonetheless, this research provides much needed empirical data to test the assumption that lapping increases hit rates. We also examined whether there was any observable change in the probative value of suspect identifications between sequential lap and no-lap lineups. Like Steblay, Dietrich, et al. (2011), we included a condition in which participants were required to see two laps, so that we could assess whether mere exposure to a second lap influenced decision making, or whether the effects were driven by those witnesses who chose to see a second lap.

The study

Given the now widespread use of some form of sequential lineup administration in police jurisdictions throughout the world (see Smith & Cutler, 2013), identifying the most efficacious mode of sequential delivery is important from the perspective of providing evidence-based policy guidelines. Moreover, independent replications are vitally important if this objective is to be realized. In this study, we attempted to replicate and extend Steblay, Dietrich, et al.'s (2011) findings. Participants were randomly assigned to one of two sequential lineup conditions. In the first condition, the participants were required to view two lineup laps. In the second condition, participants were offered a second lap, which they could accept or decline.

This research was motivated by three main aims. Our primary aims were to investigate the effect of sequential lineup laps on eyewitness identification outcomes and to assess the impact of additional laps on the probative value of suspect identifications. Our third aim was to investigate the cognitive mechanisms that drive any observed lapping effects on decision outcomes. To this end, we used a signal detection approach (e.g., Green & Swets, 1966; Macmillan & Creelman, 1991) that allowed us to estimate independent parameters for discriminability and response bias. Discriminability reflects a witness's ability to distinguish culprits from fillers, and is indexed by the parameter d'. A d' value of zero indicates no ability to distinguish culprits from fillers, and higher values indicate better ability. Response bias reflects a witness's tendency to choose from or reject a lineup (i.e., respond "not present"). Response bias is indexed by the parameter c, where a value of zero indicates unbiased responding (i.e., no systematic tendency to favor choosing from or rejecting the lineup).

Negative values indicate lenient responding (a tendency to choose from the lineup), and positive values indicate conservative responding (a tendency to reject the lineup).

Steblay, Dietrich, et al. (2011) reported that witnesses who chose no-one at lap 1 were more likely to accept a second lap than witnesses who did choose someone. This finding suggests that witnesses who set a strict decision criterion at lap 1 will be most likely to accept a second lap. Thus, our first prediction was that we would observe a significantly more conservative lap 1 *c* for witnesses who accepted a second lap than for witnesses who declined a second lap. Of course, the witnesses who accepted a second lap may also have had poorer discrimination abilities than the participants who declined a second lap, which would be indicated by a lower d' estimate.

For those participants who accepted a second lap, we predicted that we would observe a shift to a more lenient response bias from lap 1 to lap 2, which would manifest as a significant difference in *c*. Consequently, we expected to find a decrease in the proportion of non-identifications from lap 1 to lap 2. Importantly, a shift to a more lenient response bias increases the chances of each face in the lineup being identified; consequently, we expected an increase in both the proportion of suspect identifications and in filler identifications (as also found by Steblay, Dietrich, et al., 2011). We did not expect a second lap to improve discriminability (though see Carlson et al., 2008, for an argument that discrimination improves over the course of a sequential lineup).

From a policy perspective, the crucial question is this: How trustworthy are suspect identifications if laps are allowed *versus* disallowed? The most relevant statistic for policy makers and the courts (Aitken et al., 2011) is probative value (the ratio of correct to incorrect suspect identifications), which provides a measure of by how much one should adjust their belief in the guilt of the suspect, given that he was identified (Wells, 2014). Probative value is affected by both discriminability and response bias, such that it will generally increase as discriminability increases but decrease as responding becomes more lenient (Wixted & Mickes, 2012). The interplay between discriminability and response bias makes it difficult to form firm predictions regarding probative value; however, based upon previous research (e.g., Steblay, Dietrich, et al., 2011), we expected that probative value would be similar or higher if a single lap was allowed than if two laps were allowed (i.e., we did not expect probative value to suffer as a consequence of allowing only one lap).

Our inclusion of a condition in which all participants were required to see two laps allowed us to examine whether any effects observed in the optional condition were merely a result of exposure to a second lap, or whether they were driven by the specific sub-group of witnesses who chose to see a second lap (see also Steblay, Dietrich, et al., 2011). The required condition would have included some participants who would have accepted and some who would have declined a second lap, had that lap been optional rather than mandatory. If mere exposure to a second lap causes participants to respond more leniently, we would expect to see effects of a similar magnitude in the *required-second-lap* group as in the *accepted-second-lap* group. Alternatively, if the effects were driven by those witnesses who choose to see a second lap, we would expect to observe effects of a similar magnitude in the required and optional conditions (assuming that a similar proportion of witnesses in the required condition would have chosen to see a second lap as in the optional condition).

Method

Participants and Design

Participants (N = 449) were recruited in public places around Adelaide, Australia. Participants not wearing their normal corrective eyewear were excluded (N = 39), as were 17 participants in the optional condition who cited lack of time as a reason for declining a second lap. The final sample included 393 participants with a mean age of 32.5 years (SD =

16.5 years; range 18 to 78; three participants did not disclose their age). The sample included 259 women (65.90%; two participants did not disclose their gender).

The experiment followed a 2 (Target presence: target-present vs. target-absent) × 2 (Lap condition: required vs. optional) design. The optional group included an additional non-manipulated variable, as witnesses could accept or decline the second lap. To ensure adequate numbers in each cell, approximately twice as many participants were allocated to the optional condition as to the required condition.

The dependent variables were: 1) decision type: Suspect identification (target-present only), filler identification, non-identification, and multiple identification. Note that, rather than applying an arbitrary rule to resolve multiple identifications, we chose to code them as a separate category of response; 2) signal detection estimates of discriminability (d') and response bias (c); and 3) probative value (the ratio of correct suspect identifications to estimated incorrect suspect identifications).

Materials

Seven undergraduates (six female; six Caucasian, one Asian) volunteered as culprits for the study. A head-and-shoulders photograph of each culprit was taken several weeks prior to the start of data collection. The culprits were all instructed not to wear the same clothing during data collection that they were wearing in their lineup photographs.

Modal descriptions were generated for each culprit by combining the descriptions of three participants who had no further involvement in the study. These descriptions were then used to search several databases for potential fillers. The final selections of six fillers were based on visual similarity to the culprit.

Lineup fairness was assessed using the mock witness procedure (Doob & Kirshenbaum, 1973). Fifty participants who did not participate in the experiment completed the mock witness task. For each of the seven lineups, the participants read the modal description and were presented with the target and six fillers on a computer screen. The participants were asked to select the best match to the description. The descriptions were quite detailed (e.g., “Caucasian female, late teens to early 20s, with long dark brown hair, light skin tone, square-shaped face, round nose, and dark eyebrows”). The order of the lineups was randomized for each participant, and the position of the suspects within the lineup was counterbalanced across targets. Across the seven targets, the mean Tredoux’s *E* (Tredoux, 1998) was 3.00 (*SD* = 1.24), with a range from 1.76 to 4.93. Note that the level of detail included in our descriptions may have increased the rate at which the culprits were selected, reducing the *E* estimates.

The images were cropped to 720 × 585 mm and printed onto laminated cards. Six of the lineups were printed in color; the remaining lineup was printed in grayscale because it proved impossible to find suitable fillers with a similar hair color.

Procedure

Potential participants were approached in public locations around Adelaide, Australia. Each participant provided informed consent before participating. The experimenter directed the participant’s attention to a designated location 10 m away. The culprit stepped into view for 10 s, before stepping back out of view. Demographic information was then collected, as well as information about normal or corrected vision.

The experimenter read aloud the following instructions to all participants:

> “Remember the person I asked you to look at a few minutes ago? I have here a series of photos, which I am going to show you. These photos may or may not include the person you just saw. The photos are in no particular order. I will show you the photos one at a time. For each photo, please indicate whether that photo is of the person you just saw, by saying “yes” or “no”. Take as long as you need to view each photograph. Even if you identify someone, you will

> still be shown all of the photos in the series. You'll see a card that says "stop" when you have seen all of the photos."

A second experimenter then held up a stack of cards (including the six lineup photographs, a "stop" card, and a number of blank cards) so that the participant could only see the top card. The order of the fillers was randomized for each participant, and the position of the target (in target-present lineups) was counterbalanced between positions 2 and 5.

After lap 1, participants in the required condition were told that they would need to see the lineup again. To minimize the possibility that the participants would infer that their first decision must have been incorrect, we informed them that *all* participants were required to see the lineup again. Specifically, the instructions were as follows:

> "Now I would like you to look at the photos again. The photos are in the same order. Once again, I would like you to say "yes" if you think this is the person you saw, or "no" if you do not think it is the person you saw. Please note that all participants will be asked to view the photos twice – it does not mean that you got it wrong the first time."

Participants in the optional condition were offered the chance to see the lineup again. Once again, to minimize the possibility that the participants would infer that their first decision was incorrect, they were told "We are offering everybody the chance to look at the photos once again. Would you like to see them once more?" Following a 'yes' response, the participant was told "Here are the photos again in the same order. Again, please say "yes" if you think this is the person you saw, or "no" if you do not think it is the person you saw". The second lap followed the same procedure as the first lap.

After the final lap, confidence judgments were elicited, on a 0 to 100% scale. Choosers were asked how confident they were that they chose the correct photograph, and non-choosers were asked how confident they were that the person they saw was not in the photographs. All two-lap participants were then asked to think back to their lap 1 decision, and to provide a confidence judgment for that decision.[1] Finally, participants in the optional condition were asked why they did or did not choose to see a second lap. Responses to this question were recorded verbatim.[2]

Maintaining experimenter blindness

We designed our protocol to ensure experimenter blindness concerning lap condition (until the end of lap 1), target-presence, and suspect position. The culprit consulted a condition allocation sheet then shuffled the six fillers; for target-present lineups, one filler was discarded at random and the culprit photograph was placed in position 2 or 5.The culprit added the “stop” and blank cards to the stack and placed the photographs into an envelope. The culprit then placed the appropriate response sheet onto a clipboard. Page 1 of the response sheet was identical for all participants. Thus, the experimenter was blind to the participant’s lap condition until the sheet was turned over after lap 1.Throughout the lineup, the two experimenters stood side by side. One experimenter held the photographs at chest height, away from the body, out of line of sight of both experimenters.

Results

Discriminability and response bias

Estimating *d'* and *c* values for eyewitness identification tasks is not straightforward. Eyewitness identification decisions are compound recognition decisions, and include two components: a detection component (equivalent to an *old/new* recognition decision) and an identification component (equivalent to a multiple-alternative forced-choice recognition decision). In a lineup response, the witness must consider whether the culprit is in the lineup somewhere (a detection decision) and, if so, which lineup member to choose (the identification component). As a result, methods for calculating *d'* and *c* for simple recognition decisions yield biased estimates of these parameters when applied to compound decisions (for a more comprehensive discussion of this topic, see Palmer & Brewer, 2012, or Palmer, Brewer, & Weber, 2010).

To address this issue, we used a model designed to estimate discriminability and response bias in compound decision tasks: SDT-CD (Signal Detection Theory – Compound Decisions; Duncan, 2006; Palmer & Brewer, 2012; Palmer et al., 2010). This allowed us to estimate d' and c parameters for each condition. We also calculated 95% confidence intervals (CIs) around these d' and c values, using a jackknife procedure (Mosteller & Tukey, 1968; for examples, see, Koriat, Lichtenstein, & Fischhoff, 1980; Weber & Brewer, 2006). Details of model fitting using SDT-CD, and of the jackknife procedure used to generate 95% CIs, can be found in the Supplemental Material available online.

Table 1 shows the estimated discriminability and decision criterion parameters, with 95% CIs. To make inferential comparisons between these estimates, we used Cumming's (2009) inference-by-eye method of calculating the proportion of overlap (*POL*) between the CIs. This method compares the lower arm of the CI of the highest mean with the upper arm of the CI of the lowest mean. If the *POL* is less than .50, the difference is significant at $p < .05$. First, we compared discriminability between the groups. Discrimination was significantly poorer for participants who accepted a second lap (1.07, [0.63, 1.51]) than for participants who declined a second lap (2.16 [1.67, 2.65]), *POL* = -0.16, $p < .01$. The required group (1.61, [0.99, 2.23]) did not significantly differ from the *accepted-second-lap* group, *POL* = 1.01, $p > .20$, or the *declined-second-lap* group, *POL* = 0.98, $p > .10$.

Next, we examined changes in discriminability from lap 1 to lap 2 within the *required-second-lap* and *accepted-second-lap* groups. Because these parameters were based on paired data, we could not apply Cumming's (2009) method of inferential comparison. Instead, we calculated the 95% CI of the difference between the lap 1 and lap 2 proportions, again using Mosteller and Tukey's (1968) jackknifing method (see Supplemental Materials for further details). If the CI of the difference excluded 0, we could conclude that the difference was statistically significant at $p < .05$. Discriminability did not significantly improve from lap 1 to lap 2, for either the *accepted-second-lap* participants (difference = 0.43, 95% CI [-0.06, 0.92]), or the *required-second-lap* participants (difference = 0.06, 95% CI [-0.49, 0.61]).

With respect to response bias (see lower half of Table 1), at lap 1, participants who accepted a second lap were more conservative (0.05, [-0.29, 0.39]) than participants who declined a second lap (-0.87, [-1.14, -0.60]), *POL* = -1.02, $p < .001$. The *required-second-lap* group (-0.43, [-0.76, -0.10]) did not significantly differ from the *accepted-second-lap* group, *POL* = 0.53, $p > .05$, or the *declined-second-lap* group, *POL* = 0.57, $p > .05$. *The accepted-second-lap* group became more lenient from lap 1 to lap 2 (difference = 0.99, 95% CI [0.66, 1.32]). The *required-second-lap* group also became more lenient from lap 1 to lap 2, though the effect was smaller (difference = 0.36, 95% CI [0.12, 0.60]).

The signal detection analyses indicated that the group of participants who chose to see a second lap were less able to discriminate between the target and fillers, and were more conservative responders, than the participants who declined to see the second lap. Participants who saw two laps became more lenient from lap 1 to lap 2, and this criterion shift was largest for the *accepted-second lap* group. This latter difference in effect size suggests that mere exposure to a second lap does not cause participants to become more lenient; rather, the effect is driven by those participants who choose to see (or who would have chosen to see) a second lap. There was no evidence that discriminability improved from lap 1 to lap 2.

Lap 1 Responses

Response data were analyzed using multilevel logistic regression (see Wright & London, 2009). We included culprit number as a random effect in all analyses, which allowed the intercepts to vary across the lineups (Horry, Palmer, et al., 2012). Suspect, filler, multiple, and non-identifications were analyzed in separate regression models using the lme4 package for R (Bates, Maechler, & Bolker, 2011). We also analyzed target-present and target-absent lineups in separate models. As a measure of effect size, we report the risk ratio (*RR*), which can be expressed as *P*(outcome in Condition A)/*P*(outcome in Condition B). An *RR* of 2, for example, indicates that outcome *x* occurred twice as frequently in Condition A than in Condition B. The null value of the *RR* is 1; if the 95% Confidence Interval includes 1, one cannot conclude that the outcome probability differs across conditions. Risk ratios were calculated from the raw data. We calculated 95% CIs around the log-transformed *RR*s, which we then back-transformed for ease of interpretation (see formula provided by Tredoux, 1998). We use the *RR* instead of the odds ratio due to its more intuitive interpretation (A'Court, Stevens, & Heneghan, 2012; Holcomb, Chaiworapongsa, Luke, & Burgdorf, 2001). Importantly, "risk ratio" is a statistical term for comparing the probability of an event in two conditions. Though common usage of the term "risk" would imply a negative event, this does not have to be the case. One can, for example, compute the risk ratio for correct identifications (a positive outcome), by dividing the probability of a correct identification in Condition A by the probability of a correct identification in Condition B. In such a case, a number significantly greater than 1 would indicate that the probability of a correct identification was higher in Condition A than in Condition B.

Table 2 shows lap 1 decision frequencies by target presence and lap condition. To ensure that experimenter blindness was successfully maintained, we compared lap 1 decisions for required participants and optional participants, regardless of the decision to accept or decline the second lap (Column 2 of Table 2). To do so, we compared a baseline regression model that included only the random effect of target with a model that included condition (required vs. optional). We created separate models for each decision outcome from target-present and target-absent lineups. In no case did adding condition to the model improve model fit, indicating that the proportions of the various decisions were similar across the two conditions (maximum χ^2 (1) = 0.47, p = .49, *RR* = 1.42, 95% CI [0.52, 3.90]).

First, we asked whether lap 1 decisions differed between participants who later declined or accepted a second lap. The method of analysis was similar to that described above, only with group (*accepted-second-lap*, *declined-second-lap*, *required-second-lap*) as a predictor instead of condition. Looking first at target-present lineups, Table 2 shows that the proportion of correct identifications was higher for the *declined-second-lap* group than for the *accepted-second-lap* group, with the *required-second-lap* group falling in between. This observation was confirmed by a significant effect of participant group, χ^2 (2, n = 222) = 19.77, p < .001. The proportion of culprit identifications was higher for the *declined-second-lap* group than for the *required-second-lap* group, z = 2.29, p = .02, *RR* = 1.47, 95% CI [1.06, 2.03], and for the *required-second-lap* group than for the *accepted-second-lap* group, z = 2.41, p = .02, *RR* = 1.90, 95% CI [1.10, 3.28]. The proportion of non-identifications also differed significantly across the groups, χ^2 (2, n = 222) = 8.07, p = .004, being higher for the *accepted-second-lap* group than for the *declined-second-lap* group, z = 2.80, p = .005, *RR* = 1.82, 95% CI [1.19, 2.80]. Filler identifications were not significantly different across the groups, χ^2 (2, n = 222) = 0.17, p = .92, maximum *RR* = 1.23, 95% CI [0.46, 3.29]. However, there was a non-significant trend for multiple identifications, χ^2 (2, n = 222) = 5.65, p = .06, with a higher proportion of multiple identifications for the *accepted-second-lap* group than for the *declined-second-lap* group, z = 2.18, p = .03, *RR* = 3.87, 95% CI [1.11, 13.44].

For target-absent lineups, the proportion of non-identifications varied across the groups, χ^2 (2, n = 171) = 6.25, p = .04, with a higher proportion of non-identifications for the *accepted-second-lap* group than for the *declined-second-lap* group, z = 2.41, p = .02, *RR* = 1.43, 95% CI [1.06, 1.93]. The proportion of filler identifications, χ^2 (2, n = 171) = 2.31, p = .31, did not significantly vary across the groups (maximum *RR* = 1.54, 95% CI = 0.80, 2.96]).

The difference in multiple identifications again did not reach statistical significance, χ^2 (2, n = 171) = 4.95, p = .06. Despite this marginal difference, none of the pairwise comparisons between the groups were statistically significant (maximum z = 1.50, p = .13, RR = 4.01, 95% CI [0.91, 17.67]).

Changes in response from lap 1 to lap 2

We approached response changes in two ways. First, we examined whether certain types of lap 1 decision were changed more frequently than others. Second, we compared the difference in the proportion of each decision type in lap 1 and lap 2. Only participants who saw two laps were included in these analyses (i.e., *accepted-second-lap* and *required-second-lap* participants).

First, we asked whether certain types of responses were more likely to be changed than others. We coded response change as a binary outcome variable (response changed vs. response not changed), which we predicted from group (*accepted-second-lap* vs. *required-second-lap*) and Lap-1-decision (suspect identification, filler identification, multiple identification, or non-identification) in two separate mixed-effects logistic regressions (one for target-present lineups and one for target-absent lineups, with culprit number as a random effect). As the Group × Lap-1-decision interaction was not statistically significant for either the target-present lineups, χ^2 (3, n = 150) = 0.79, p = .86, or the target-absent lineups, χ^2 (2, n = 116) = 2.52, p = .28, we examined the main effects independent of the interaction terms.

To assess the main effects, we used an approach analogous to the Type-II approach in ANOVA. Specifically, we examined the effect of each predictor after controlling for the main effects of the other predictors in the analysis[3]. To accomplish this, we first created two models (one for target-present lineups and one for target-absent lineups) that included both predictors: Group and Lap-1-decision (model 1). The main effect of Group was then examined by comparing model 1 with a model that included only Lap-1-decision as a predictor. For target-present lineups, the difference in model fit was significant, indicating a significant main effect of Group, χ^2 (1, *n* = 150) = 4.51, *p* = .03, *RR* = 1.79, 95% CI [1.21, 2.63]. The *accepted-second-lap* group changed a higher proportion of their responses (54.84%) than the *required-second-lap* group (30.68%). For target-absent lineups, the difference in model fit was not statistically significant, χ^2 (1, *n* = 116) = 2.61, *p* = .11, *RR* = 1.31, 95% CI [0.86, 2.00].

The main effect of Lap-1-decision was then assessed by comparing the fit of model 1 with a model including only Group (model 3). For target-present lineups, the difference in model fit was significant, indicating a significant main effect of Lap-1-decision, χ^2 (3, *n* = 150) = 24.50, *p* < .001. Table 3 shows the frequency with which responses were changed as a function of lap 1 decision, with unchanged responses printed in bold. As we were examining only the main effect of decision type, the frequencies in Table 3 are from both groups combined. For target-present lineups, the proportion of culprit identifications that were changed (14.29%) was lower than the proportions of non-identifications (57.97%), *z* = -4.28, *p* < .001, *RR* = 0.25, 95% CI [0.11, 0.53] and multiple identifications (55.56%), *z* = -3.05, *p* = .002, *RR* = 0.26, 95% CI [0.11, 0.60] that were changed; the difference between culprit identifications and filler identifications was not significant, *RR* = 0.38, 95% CI [0.13, 1.06]. No other pairwise differences were statistically significant (maximum *RR* = 1.74, 95% CI [0.83, 3.66]). For target-absent lineups, the change in model fit was not statistically significant, indicating that response changes were not significantly associated with lap 1 decision, χ^2 (2, *n* = 116) = 4.94, *p* = .08 (maximum pairwise *RR* = 1.50, 95% CI [0.63, 3.58]).

Our second approach was to compare the net change in response frequencies between lap 1 and lap 2. Because each participant contributed both a lap 1 and a lap 2 response, the lap 1 and lap 2 data were paired. To compare them, we calculated the 95% CI around the difference between the lap 1 and lap 2 proportions. A difference was considered statistically significant if the CI of the difference excluded 0. Table 4 shows the proportions of each decision type at lap 1 and at lap 2, along with the 95% CI of the difference. The right-hand column of Table 4 shows the overall changes in response, collapsed across participants who accepted a second lap and those who were required to see a second lap. For target-present lineups, the proportion of correct suspect identifications significantly increased from lap 1 to lap 2, 95% CI of the difference [.10, .26], *RR* = 1.56[4], and this difference was significant for both the *accepted-second-lap* group, 95% CI of the difference [.13, .42], *RR* = 2.30, and for the *required-second-lap* group, 95% CI of the difference [.02, .21], *RR* = 1.28. The proportion of target-present filler identifications also increased from lap 1 to lap 2, 95% CI of the difference [.02, .14], *RR* = 1.80; however, the increase was statistically significant for the *accepted-second-lap* group, 95% CI of the difference [.02, .27], *RR* = 2.49, but not for the *required-second-lap* group, 95% CI of the difference [-.02, .09], *RR* = 1.33. The proportion of non-identifications from target-present lineups decreased from lap 1 to lap 2, 95% CI of the difference [-.34, -.17], *RR* = 0.46, and the decrease was statistically significant for both the *accepted-second-lap* group, 95% CI of the difference [-.50, -.21], *RR* = 0.34, and for the *required-second-lap* group, 95% CI of the difference [-.28, -.08], *RR* = 0.58. The proportion of multiple identifications from target-present lineups did not significantly change from lap 1 to lap 2, 95% CI of the difference [-.07, .04], *RR* = 0.89.

For target-absent lineups, the proportion of filler identifications increased from lap 1 to lap 2, 95% CI of the difference [.04, .25], *RR* = 1.55. The increase in filler identifications was statistically significant for the *accepted-second-lap* group, 95% CI of the difference [.03, .37], *RR* = 2.00, but not for the *required-second-lap* group, 95% CI of the difference [-.02, .23], *RR* = 1.34. The proportion of non-identifications significantly decreased from lap 1 to lap 2, 95% CI of the difference [-.32, -.13], *RR* = 0.67, and this decrease was significant for both the *accepted-second-lap* group, 95% CI of the difference [-.45, -.13], *RR* = 0.62, and for the *required-second-lap* group, 95% CI of the difference [-.29, -.07], *RR* = 0.71. Finally, there was a significant increase in multiple identifications from lap 1 to lap 2, 95% CI of the difference [.01, .15], *RR* = 2.30. However, the increase was not statistically significant for either group: *accepted-second-lap*, 95% CI of the difference [-.03, .19], *RR* = 2.98; *declined-second-lap*, 95% CI of the difference [-.01, .16], *RR* = 1.99.

Comparing final decisions across groups

Next, we examined differences in final decisions across the groups. The frequencies are shown in Table 5. Mixed-effect logistic regressions were run for each decision type, separately for target-present and target-absent lineups, with Group as the predictor variable. For target-present lineups, the proportion of filler identifications significantly varied across the groups, χ^2 (1, n = 222) = 6.35, p = .04. The *accepted-second-lap* group identified a higher proportion of fillers than the *declined-second-lap* group, *RR* = 2.90, 95% CI [1.20, 7.03]. No other effects were significant for target-present lineups (maximum χ^2 (1, n = 222) = 3.06, p = .22). For target-absent lineups, the proportion of filler identifications did not significantly vary across the groups, χ^2 (1, n = 171) = 2.21, p = .33 (the risk ratio for the *accepted-second-lap* group versus the *declined-second-lap* group was 1.32 (95% CI [0.79, 2.22])), and no other effects were significant, (maximum χ^2 (1, n = 171) = 1.81, p = .40).

Differences between lineups

Our seven lineups likely differed from each other in many ways, just as real police lineups would. Indeed, fairness as measured by Tredoux's *E* varied considerably across our seven lineups. In the preceding analyses, we allowed the intercepts to vary across the seven lineups, accounting for baseline differences in decision outcomes across the lineups. However, within a mixed-effects model, we can also allow the *slopes* to vary across lineups. This relaxes the assumption that the size of any effect will be the same across all of the targets, but it does so at the expense of parsimony, as it adds degrees of freedom to the model.

To determine whether the magnitude of the effects observed here varied significantly across the lineups, we created an additional model for each of the analyses reported above, in which the slopes were allowed to vary. The fit of the random-slopes model was then compared to the fit of the corresponding random-intercepts model; a significant change in model fit would indicate that the main effect of interest varied in magnitude across the lineups. In no case did allowing the slopes to vary improve significantly the fit of the model (maximum χ^2 (5) = 8.67, p = .12).

Probative value and information gain

With regards to lineup laps, there are (at least) two possible policies: allow all witnesses to see a second lap of the lineup (policy A) or constrain all witnesses to a single lap (policy B). From a policy maker's perspective, it is necessary to know whether policies A and B produce differences in: 1) the proportion of correct identifications from target-present lineups; 2) the proportion of false identifications from target-absent lineups; and 3) the ratio of correct to false identifications. The data from our optional-second-lap condition can directly inform each of these considerations. We collapsed our data across the *accepted-second-lap* and *declined-second-lap* groups to compare the lap 1 decisions (i.e., what would have happened under policy B) and the final decisions (i.e., what would have happened under policy A) of all participants in the optional condition.

Because these data were paired (i.e., each participant provided both a first and final response), we could not compare them using conventional tests. Instead, we calculated the 95% CI of the difference; if the CI excluded 0, we could conclude that the difference was statistically significant. For target-present lineups, the proportion of correct identifications increased from 41% at first decision to 54% at final decision, 95% CI of difference [6%, 19%]; *RR* = 1.32. For target-absent lineups, the proportion of filler identifications increased from first decision (26%), to final decision (36%), 95% CI of the difference [2%, 18%]; *RR* =

1.38. Thus, allowing a second lap increased both the proportion of correct identifications and the proportion of target-absent filler identifications. Note that the CIs of the differences were almost entirely overlapping, suggesting that the effect size was similar for the two types of decision. Though not directly relevant to probative value, the proportion of target-present filler identifications also increased, from 9% at first decision to 16% at final decision, 95% CI of difference [1%, 12%]; $RR = 1.78$.

Dividing the proportion of correct identifications by the proportion of false identifications provides a measure of probative value. As we did not designate an innocent suspect, we divided the target-absent filler identification rate by six (the nominal size of the lineups) to create a proxy of false identifications. A probative value of 1 would indicate that the ratio of guilty to innocent identifications was equal, and suspect identifications were therefore uninformative. Probative value is asymmetric, as it can vary from 0 (with values between 0 and 1 indicating exonerating evidence) to infinity (with values greater than 1 indicating incriminating evidence). Shown in Table 6 is the proportion of correct identifications, the estimated proportion of false identifications, and the probative value estimates (with 95% CIs) for lap 1 and final decisions across the entire optional group. Using the formula suggested by Tredoux (1998), we compared the probative value of the lap 1 and final decisions. The difference was not statistically significant, $\chi^2(1) = 0.01$, $p > .90$.[5]

Though probative value did not significantly differ for lap 1 and final responses across the optional group as a whole, the signal detection analyses showed that the *accepted-second-lap group* had poorer discrimination than the *declined-second-lap group*. An interesting question is whether an investigator could have used this information as a cue to accuracy. In other words, were the final decisions of participants who decline a second lap more probative than the final decisions of participants who accept a second lap? Probative value was estimated at 11.33, 95% CI [3.58 – 35.79], for the *declined-second-lap* group and

7.11, 95% CI [2.45 – 20.69], for the *accepted-second-lap* group. The difference between these estimates was not statistically significant, $\chi^2(1) = 0.34$, $p > .10$. Thus, lapping may not provide a reliable cue for assessing witness reliability.

Discussion

This study investigated the effect of a second sequential lineup lap on eyewitness identification accuracy. Participants were either required to see, or given the option to see, a second lap of the lineup. The group of participants who chose to see a second lap were less able to discriminate targets from fillers, and were more conservative responders at lap 1 than the group of participants who declined to see a second lap. However, from lap 1 to lap 2, these participants became more lenient in their responding. Non-identifications were frequently changed, leading to increases in both culprit identifications and filler identifications. The effect on probative value was minimal, as the proportional increases in correct and incorrect identifications were similar. Below we discuss how our results compare with those of Steblay, Dietrich, et al. (2011). We then discuss the potential causes of the sequential lap effect. Finally, we consider the policy implications of our findings.

In many ways, our results converge with those of Steblay, Dietrich, et al. (2011). For participants who chose a second lap, we found a significant increase in culprit identifications from lap 1 to lap 2, from 21% to 48%, producing a risk ratio of 2.31. Steblay, Dietrich, et al. (2011, Experiment 2) reported a similar increase, from 21% to 50%, which gives a risk ratio of 2.34. Of course, an increase in suspect identifications is only desirable if those suspect identifications are probative of guilt. Thus, one must also consider errors from target-absent lineups. For participants who chose a second lap, target-absent filler identifications increased from 20.4% to 40.8%, producing a risk ratio of 2.00. In Steblay, Dietrich, et al.'s Experiment 2, the comparable increase was from 52.9% to 88.2%, which gives a risk ratio of 1.67. Note that, although the proportion of errors was much higher in Steblay, Dietrich, et al. than in our study, the effect sizes were quite similar.

Of course, not all incorrect choices from target-absent lineups will implicate the suspect. To estimate the increase in false identifications, we must rely on a proxy of false identifications – the proportion of filler identifications divided by the nominal size of the lineup. For those participants who chose a second lap, estimated false identifications in our study increased from 3% to 7%; in Steblay, Dietrich, et al., the comparable increase was from 8% to 15%. These estimates must be interpreted cautiously, as they assume that the lineups were completely unbiased, and that the suspect was no more or less likely to be chosen than any of the fillers. Consequently, these difference estimates will be conservative for lineups in which the innocent suspect is chosen more frequently than would be expected by chance, and they will be liberal for lineups in which the innocent suspect is chosen less frequently than would be expected by chance.

Our signal detection analyses were consistent with Steblay, Dietrich, et al.'s (2011) assertion that the sequential lap effect is best described as a criterion shift leading to an increase in choosing from lap 1 to lap 2. Furthermore, the group that accepted the second lap when it was optional, as a whole, had a more conservative response bias in lap 1 than the group that declined the second lap. Importantly, they were also less able to discriminate between the target and the fillers in lap 1, and this ability did not significantly improve from lap 1 to lap 2. The sequential lapping effect seems to be driven by a group of cautious witnesses, with relatively poor discrimination ability, who take the opportunity to adjust their response bias when it is offered. Importantly, a criterion shift does not guide the witness to any one lineup member. Rather, the odds that *each face* will be chosen increase, though the proportional increase in choosing for each face will depend upon the shape of and distance between the underlying memorial distributions (Flowe & Bessemer, 2011). Thus, criterion shifts inevitably increase both correct and incorrect identifications.

The data from the *required-second-lap* group allowed us to ask whether the effects observed in the *accepted-second-lap* group resulted merely from exposure to a second lap, or whether they were related to the specific subset of witnesses who accepted a second lap. These possibilities are not straightforward to tease apart, as the required condition would have included some participants who would have accepted, and some who would have declined, a second lap. However, the first and final decisions of the optional and required groups were strikingly similar, and when we compared changes in response between the required and optional groups, we found similar effect sizes for all responses[6]. Furthermore, the criterion shift from lap 1 to lap 2 was larger for the *accepted-second-lap* group than for the *required-second-lap* group, with no overlap in the confidence intervals of the differences. These data strongly suggest that mere exposure to a second lap is neither harmful nor beneficial. Rather, the participants in the required condition who changed their responses were likely those participants who would have opted to see a second lap anyway, had the second lap been optional instead of mandatory.

Though the probative value of final suspect identifications was numerically higher for the *declined-second-lap* group than for the *accepted-second-lap* group, the difference was not statistically significant; in fact, to detect a significant difference given the effect size, an 11-fold increase in sample size would have been required (totalling 2,618 participants in the optional condition). Furthermore, by calculating probative value for lap 1 and lap 2 decisions, we can see that, if anything, probative value for the *accepted-second-lap* group increased from lap 1 (7.53) to lap 2 (12.84), though this difference is not likely to be statistically significant. We note that Steblay, Dietrich, et al. (2011) also documented a small increase in probative value from lap 1 (2.43) to lap 2 (3.40). These diagnosticity indices reflect the observation that the proportional increase in correct identifications was slightly larger than the proportional increase in false identifications in both studies.

Policy implications

The sequential lineup is becoming increasingly common in the field (Smith & Cutler, 2013). When using the sequential procedure, policy makers and investigators must consider whether witnesses should be allowed to see a second lap. The sequential lap procedure appears to be the status quo, with all US jurisdictions that have introduced the sequential lineup allowing laps (Wells, 2014), and with procedures in the UK mandating second laps and allowing additional subsequent laps (see Horry, Memon, et al., 2012). To make an informed decision on the lapping issue, policy makers need to know how lapping affects correct identifications, errors, and the probative value of suspect identifications. The present study provides much needed empirical data on each of these points. Specifically, across multiple target persons, and with a large sample of the general public, we found that allowing additional laps: a) significantly increased correct identifications; b) significantly increased filler identifications from target-absent lineups; and c) had little effect on probative value. Though the sequential lap procedure had little effect on probative value, we draw the reader's attention to the following points: 1) small increases in false identifications are not necessarily non-trivial; 2) there may be key moderator variables that could change the pattern of results from that observed here; and 3) despite the use of multiple target persons, these data represent a relatively small sample of lineups.

First, though the change in probative value observed here was not statistically significant, it would be premature to conclude that lapping does not affect the quality of the identification evidence. Even a small increase in false identifications could have potentially damaging consequences given the large numbers of identification parades that are conducted across the world. Furthermore, the impact of a small increase in false identifications will vary hugely depending upon the base rates with which innocent suspects are placed in lineups (see Wells & Olson, 2002). To demonstrate the importance of base rates, we will draw an analogy with the medical screening literature.

Let us consider a hypothetical scenario in which we wished to screen for a particular disease. Two tests are available: Test A has a correct detection rate of 95% and a false detection rate of 0.1%; Test B has a higher correct detection rate of 99%, but it also has a higher false detection rate of 0.3%. Let us now assume that we screened two different populations, each consisting of 100,000 people. Population X has a high disease prevalence rate of 20%; Population Y has a much lower prevalence rate of 1%. Table 7 shows the number of correct diagnoses, the number of false diagnoses, and the probability of having the disease given a positive diagnosis. Within Population X, Test B produces an extra 800 correct diagnoses at a cost of 160 false diagnoses. An individual from this population who received a positive result would have a 99.58% likelihood of having the disease under Test A and a 98.80% likelihood of having the disease under Test B. Within Population Y, however, Test B produces only 40 extra correct diagnoses, but 198 extra false diagnoses. An individual from Population Y who received a positive result would have a 90.56% likelihood of having the disease under Test A but only a 76.92% likelihood of having the disease under Test B. Thus, the difference in posterior odds between Test A and Test B is markedly larger when the base rate of disease prevalence is lower.

Turning back to lineup identifications, the target-present base rate is analogous to the disease prevalence rate. As a hypothetical example, let us imagine two jurisdictions – one with a very high target-present base rate of 90% (Jurisdiction X), and one with a very low target-present base rate of 20% (Jurisdiction Y). If we ran 100,000 lineups in each of these jurisdictions, how many extra culprits would we catch, and how many innocent suspects would be implicated, under the lap procedure? Given our observed correct identification rates

(41% and 54% for no-lap and lap lineups, respectively) and false identification rates (4% and 6% for no-lap and lap lineups, respectively), we would catch 11,430 extra culprits at a cost of 160 false identifications in Jurisdiction X by allowing a second lap. In Jurisdiction Y, we would catch 2,540 extra culprits at a cost of 1,280 false identifications. Though the posterior odds change little whether lapping is allowed or disallowed (98.84% to 98.79% in Jurisdiction X; 70.44% to 69.47% in Jurisdiction Y), the cost in terms of numbers of falsely implicated suspects is influenced hugely by the base rate. This example is illustrative only. We do not intend to imply that our observed data provide definitive estimates of the correct and false identification rates that would be observed in the real world. Our estimates are subject to sampling and measurement error, just as in every empirical study. However, no matter what numbers one plugs into the equation, the outcome would be the same: if one test produces a higher false alarm rate than another, the consequences of that increase will be larger as the target-present base rate decreases.

Our second point is that the sequential lapping effect may be moderated by a wide range of system and estimator variables (Wells, 1978). For example, conditions that reduce memory strength, such as long retention intervals (Sauer, Brewer, Zweck, & Weber, 2010), long viewing distances (Lindsay et al., 2008), and brief exposure durations (Palmer et al., 2013), could conceivably increase the effect of laps on false identifications. Non-blind lineup administration is one obvious example of a system variable that could moderate the lapping effect. For example, non-blind administrators might abuse the sequential lap procedure to encourage a witness to get the ‘right’ answer. The dangers of non-blind administration have been discussed in depth elsewhere (Clark, Brower, Rosenthal, Hicks, & Moreland, 2013; Wells & Bradfield, 1998; Wright, Carlucci, Evans, & Compo, 2010); suffice to say here that the lap procedure provides an additional route for undue influence. A full understanding of how the sequential lap effect operates, and of the impact it is likely to have in real identifications, will require systematic investigation across a range of these forensically relevant variables.

Third, though we included multiple target persons (*N* = 7) in this study, our study nonetheless represents a small sample of the population of all potential targets (Wells & Windschitl, 1999). Lineups will vary greatly in a number of factors that are likely to influence witness decision-making, including: a) the similarity between the perpetrator's appearance at the time of the crime and in the lineup photograph; b) the similarity between the culprit and the innocent suspect; c) the similarity between the culprit and the fillers; and d) the similarity between the innocent suspect and the fillers. Each of these factors could plausibly moderate the lapping effect. For example, an innocent suspect who bears a strong resemblance to the culprit and who stands out from the fillers may attract a large proportion of lap 2 decisions, thus creating a large effect of lapping on false identifications. Systematic study of these relationships would enable us to better predict when lapping may be beneficial and when it may be harmful, as well as improving our understanding of the cognitive processes that drive the effects.

One final point is that the lap procedure tested here and in previous research (e.g., Steblay, Dietrich, et al., 2011) is different to the one commonly used in the field, in which laps are permitted only if the witness initiates the request. The impact of a sequential lap effect in the field will obviously depend on how frequently witnesses make a request to view additional laps. It is difficult to know how often real witnesses make such a request. Wells et al. (in press) reported that 15.7% of their sequential lineup witnesses requested a second lap, whereas the comparable figure from Klobuchar et al. (2006) was 46.9%, at least for those lineups in which the details were appropriately recorded. From UK field data, Horry, Memon, et al. (2012) reported that 47.1% of their witnesses requested an additional lap, even after their mandatory second lap. To add a further complication, there are likely to be a host of social factors (such as the perceived friendliness or authoritarianism of the lineup administrator) that influence a witness's willingness to request a second lap, such that the proportion of witnesses who see extra laps may vary widely from jurisdiction to jurisdiction, and even within a single force. Perhaps, if only a small percentage of witnesses request additional laps, we have overestimated the impact of laps on real identifications. However, it is possible that witnesses who spontaneously initiate a request for an additional lap are those with the poorest memories, which may render any second-lap decisions particularly unreliable. The lapping-by-request procedure used in the field will likely prove very difficult to study in the laboratory, as the sample sizes needed to ensure an adequate number of two-lap participants would be very large. For now, we must acknowledge that this is a limitation with all of the extant research on lapping, including this study.

Conclusions

The sequential lap procedure produces a higher proportion of correct identifications than the no-lap procedure; however, it also produces a higher proportion of target-absent filler identifications. For policy makers, this research suggests that the sequential lap procedure will help to identify more offenders at the expense of identifying more innocent suspects. The impact of any change in false identifications will vary markedly with changes in target-present base rate, which policy makers should bear in mind when evaluating the sequential lap procedure. Future research should include a wide range of forensically relevant variables that may moderate the effect. The impact of similarity relationships between lineup members should also be tested, as the sequential lap effect may vary quite substantially over fluctuations in target-filler similarity, target-innocent suspect similarity, and innocent suspect-filler similarity.

References

A'Court, C., Stevens, R., & Heneghan, C. (2012). Against all odds? Improving the understanding of risk reporting. *British Journal of General Practice, 62,* e220-e223. doi: 10.3399/bjgp12X630223

Aitken, C., Berger, C. E. H., Buckleton, J. S., Champod, C., Curran, J., Dawid, A. P., … Zadora, G. (2011). Expressing evaluative opinions: A position statement. *Science & Justice, 51,* 1-2. doi: 10.1016/j.scijus.2011.01.002

Bates, D., Maechler, M., & Bolker, B. (2011). *lme4: Linear mixed-effects models using s4 classes.* Available at www.r-project.org

Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence-accuracy relationship in eyewitness identification: The effects of reflection and disconfirmation on correlation and calibration. *Journal of Experimental Psychology: Applied, 8,* 44-56. doi:10.1037//1076-898X.8.1.44

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

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform: Psychological science and public policy. *Perspectives on Psychological Science, 7,* 238-259. doi: 10.1177/1745691612439584

Clark, S. E., Brower, G. L., Rosenthal, R., Hicks, J. M., & Moreland, M. B. (2013). Lineup administrator influences on eyewitness identification and eyewitness confidence. *Journal of Applied Research in Memory and Cognition, 2,* 158-165. doi: 10.1016/j.jarmac.2013.06.003

Cumming, G. (2009). Inference by eye: Reading the overlap of independent confidence intervals. *Statistics in Medicine, 28*, 205-220. doi: 10.1002/sim.3471

Doob, A. N., & Kirshenbaum, H. M. (1973). Bias in police lineups – partial remembering. *Journal of Police Science and Administration, 1*, 287-293. doi: 10.1016/0022-1031(73)90062-0

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

Flowe, H., & Bessemer, A. (2011). The effects of target discriminability and criterion placement on accuracy rates in sequential and simultaneous target-present lineups. *Psychology, Crime, & Law, 17,* 587-610. doi: 10.1080/10683160903397540

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley & Sons.

Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating eyewitness identification procedures using receiver operating characteristic analysis. *Current Directions in Psychological Science, 23,* 3-10. doi: 10.1177/0963721413498891

Holcomb, W. L., Chaiworapongsa, T., Luke, D. A., & Burgdorf, K. D. (2001). An odd measure of risk: Use and misuse of the odds ratio. *Obstetrics & Gynecology, 98,* 685-688. doi: 10.1016/S0029-7844(01)01488-0

Horry, R., Memon, A., Wright, D. B., & Milne, R. (2012). Predictors of eyewitness identification decisions from video lineups in England: A field study. *Law & Human Behavior, 36,* 257-265. doi: 10.1007/s10979-011-9279-z

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

Klobuchar, A., Steblay, N. K. M., & Caligiuri, H. L. (2006). Improving eyewitness identifications: Hennepin County's blind sequential lineup pilot project. *Cardazo Public Law, Policy, and Ethics Journal, 2,* 381-414.

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

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 107-118. doi:10.1037//0278-7393.6.2.107

Lindsay, R. C. L., Lea, J. A., & Fulford, J. A. (1991). Sequential lineup presentation: Technique matters. *Journal of Applied Psychology, 76,* 741-745. doi: 10.1037/0021-9010.76.5.741

Lindsay, R. C. L., Mansour, J. K., Beaudry, J. L., Leach, A.-M., & Bertrand, M. I. (2009). Beyond sequential presentation: Misconceptions and misrepresentations of sequential lineups. *Legal & Criminological Psychology, 14,* 31-34. doi: 10.1348/135532508X382104

Lindsay, R. C. L., Semmler, C., Weber, N., Brewer, N., & Lindsay, M. R. (2008). How variations in distance affect eyewitness reports and identification accuracy. *Law & Human Behavior, 32*, 526-535. doi: 10.1007/s10979-008-9128-x

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

MacLin, O. H., & Phelan, C. M. (2007). PC_Eyewitness: Evaluating the New Jersey method. *Behavior Research Methods, 39*, 242-247. doi: 10.3758/BF03193154

Macmillan, N., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press

Malpass, R. S., Tredoux, C. G,. Compo, N. S., McQuiston-Surrett, D., MacLin, O. H., Zimmerman, L. A., & Topp, L. D. (2008). Study space analysis for policy development. *Applied Cognitive Psychology, 22,* 789-801. doi: 10.1002/acp.1483

McQuiston-Surrett, D., Malpass, R. S., & Tredoux, C. G. (2006). Sequential vs. simultaneous lineups: A review of methods, data, and theory. *Psychology, Public Policy, & Law,* 12, 137-169. doi: 10.1037/1076-8971.12.2.137

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,* 361-376. doi: 10.1037/a0030609

Mosteller, F., & Tukey, J. W. (1968). Data analysis including statistics. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (pp. 80-203). Reading, MA: Addison-Wesley.

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

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

Palmer, M. A., Brewer, N., Weber, N., & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied, 19,* 55-71. doi: 10.1037/a0031602

Police and Criminal Evidence Act (1984), Code D: Code of Practice for the Identification of Persons by Police Officers. London, UK: Home Office.

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

Smith, A. M., & Cutler, B. L. (2013). Identification test reforms. In B. L. Cutler (Ed.), *Reform of eyewitness identification procedures* (pp203-219). Washington, DC: American Psychological Association.

Steblay, N. K., Dietrich, H. L., Ryan, S. L., Raczynski, J. L., & James, K. A. (2011). Sequential lineup laps and eyewitness accuracy. *Law & Human Behavior,35*, 262-274. doi: 10.1007/S10979-010-9236-2

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

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

Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: Confidence, accuracy, and response latency. *Applied Cognitive Psychology, 20*, 17-31. doi:10.1002/acp.1166

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality & Social Psychology,* 36, 1546-1557. doi: 10.1037/0022-3514.36.12.1546

Wells, G. L. (2014). Eyewitness identification: Probative value, criterion shifts, and policy regarding the sequential lineup. *Current Directions in Psychological Science, 23*, 11-16. doi: 10.1177/0963721413504781

Wells, G. L., & Bradfield, A. L. (1998). “Good, you identified the suspect”: Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology, 83,* 360-376. doi: 10.1037//0021-9010.83.3.360

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

Wells, G. L., Steblay, N. K., & Dysart, J. E. (in press). Double-blind photo-lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior*.

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

Wixted, J. T., & Mickes, L. (2012). The field of eyewitness memory should abandon probative value and embrace Receiver Operating Characteristic analysis. *Perspectives on Psychological Science, 7,* 275-278. doi: 10.1177/1745691612442906

Wright, D. B., Carlucci, M. E., Evans, J. R., & Compo, N. S. (2010). Turning a blind eye to double-blind line-ups. *Applied Cognitive Psychology, 24,* 849-867. doi: 10.1002/acp.1592

Wright, D. B., & London, K. (2009). Multilevel modelling: Beyond the basic applications. *British Journal of Mathematical & Statistical Psychology,* 62, 439-456. doi: 10.1348/000711008X327632

Footnotes

[1]We recorded confidence ratings retrospectively because we were faced with the possibility that asking for confidence judgments might influence participants' decisions to accept or decline a second lap. However, these retrospective confidence judgments may differ from the judgments that would have been made at the time of the identification (see, for example, Brewer, Keast, & Rishworth, 2002). We did not report the confidence judgments here, because we were concerned about their validity. However, interested readers may contact the first author for further information.

[2]In the interests of brevity, we did not report these responses here. Please contact the first author directly for further information concerning these data.

[3]In the absence of significant interactions, the Type II approach is more powerful than the Type III approach, which controls for interactions and main effects.

[4]We could not calculate CIs around the risk ratios, as the data were paired rather than independent.

[5]An important caveat is that this formula is for independent samples comparisons, whereas the lap 1 and final probative values were paired. Consequently, this test is too conservative, and so it must be interpreted with caution. That said, the very small chi square value makes it unlikely that the difference would be significant even if we had been able to take the correlation between the lap 1 and final responses into account.

[6]Please contact the first author for full details of these comparisons.

Table 1.

Signal detection estimates of discriminability (d′) and response bias (c) with 95% confidence intervals (CIs).

Parameter	Decision	Required group		Optional group (overall)		Accepted-second-la	
		Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
d′	1	1.61	[0.99, 2.23]	1.69	[1.45, 1.93]	1.07	[0.63, 1.5
	2	1.65	[1.40, 1.90]	---		1.50	[1.09, 1.9
	Final	1.65	[1.40, 1.90]	1.80	[1.42, 2.18]	1.50	[1.09, 1.9
c	1	-0.43	[-0.76, -0.10]	-0.44	[-0.63, -0.25]	0.05	[-0.29, 0.3
	2	-0.93	[-1.18, -0.68]	---		-0.94	[-1.16, -0.
	Final	-0.93	[-1.18, -0.68]	-0.89	[-1.11, -0.67]	-0.94	[-1.16, -0.

Note: Est. = Estimate.

Table 2.

Lap 1 responses by lap condition and target presence.

	Required condition	Optional condition		
Response	Whole group	Whole group	accepted-second-lap	declined-second-lap
Target-present				
Suspect ID	39.8% (35)	41.1% (55)	21.0% (13)	58.3% (42)
Filler ID	10.2% (9)	9.0% (12)	9.7% (6)	8.3% (6)
Non-ID	40.9% (36)	40.3% (54)	53.2% (33)	29.2% (21)
Multiple ID	9.1% (8)	9.7% (13)	16.1% (10)	4.2% (3)
Condition *N*	88	134	62	72
Target-absent				
Filler ID	31.3% (21)	26.0% (27)	20.4% (10)	30.9% (17)
Non-ID	61.2% (41)	63.5% (66)	75.5% (37)	52.7% (29)
Multiple ID	7.5% (5)	10.6% (11)	4.1% (2)	16.4% (9)
Condition *N*	67	104	49	55

Note: *n*s are shown in parentheses. ID = identification.

Table 3.

Lap 2 responses as a function of lap 1 response for accepted-second-lap and required-second-lap participants combined.

	Lap 1 Response			
Lap 2 Response	Suspect ID	Filler ID	Non-ID	Multiple ID
		Target-present		
Suspect ID	**42**	2	23	8
Filler ID	2	**10***	14	2
Non-ID	1	1	**29**	0
Multiple ID	3	2	3	**8**
		Target-absent		
Filler ID	-	**21**†	24	3
Non-ID	-	1	**50**	1
Multiple ID	-	9	4	**3**

Note: Unchanged responses are shown in bold typeface. ID = identification.

*All of these 10 witnesses identified the same filler in lap 2 as in lap 1. †7 of these 21 witnesses identified a different foil in lap 2 than in lap 1.

Table 4.

Comparison between Lap 1 and Lap 2 responses for accepted-second-lap and required-second-lap participa

Decision type	Required-second-lap			Accepted-second-lap		
	Lap 1	Lap 2	Difference [95% CI]	Lap 1	Lap 2	Difference [95% CI]
Target-present						
Suspect ID	39.8%	51.1%	.11 [.02, .21]	21.0%	48.4%	.27 [.13, .42]
Filler ID	10.2%	13.6%	.03 [-.02, .09]	9.7%	24.2%	.15 [.02, .27]
Non-ID	40.9%	23.9%	-.19 [-.28, -.08]	53.2%	17.7%	-.35 [-.50, -.21]
Multiple ID	9.1%	11.4%	.02 [-.04, .09]	16.1%	9.7%	-.06 [-.16, .04]
Target-absent						
Filler ID	31.3%	41.8%	.10 [-.02, .23]	20.4%	40.8%	.20 [.03, .37]
Non-ID	61.2%	43.3%	-.18 [-.29, -.07]	75.5%	46.9%	-.29 [-.45, -.13]
Multiple ID	7.5%	14.9%	.07 [-.01, .16]	4.1%	12.2%	.08 [-.03, .19]

Table 5.

Final responses by target presence and condition.

		Optional Condition	
Response	Required Condition	accepted-second-lap	declined-second-lap
Target present			
Suspect ID	51.1% (45)	48.4% (30)	58.3% (42)
Filler ID	13.6% (12)	24.2% (15)	8.3% (6)
No ID	23.9% (21)	17.7% (11)	29.2% (21)
Multiple ID	11.4% (10)	9.7% (6)	4.2% (3)
Total N	88	62	72
Target absent			
Filler ID	41.8% (28)	40.8% (20)	30.9% (17)
No ID	43.3% (29)	46.9%(23)	52.7% (29)
Multiple ID	14.9% (10)	12.2% (6)	16.4% (9)
Total N	67	49	55

Note: *n*s are shown in parentheses.

Table 6.

Correct identifications, estimated false identifications, and probative value (with 95% Confidence Intervals) of lap 1 and final decisions in the optional condition (collapsed across accepted-second-lap and declined-second-lap participants).

				95% CIs	
Decision	Correct ID rate	False ID rate	Probative value	Lower	Upper
Lap 1	41.0%	4.3%	9.46	3.76	23.95
Final	53.7%	5.9%	9.06	4.15	19.80

Table 7.

Hypothetical data from two medical screening tests (Tests A and B) in a high-prevalence population (Population A) and a low-prevalence population (Population B).

Population	Test	Per 100,000 tests N correct	N incorrect	Posterior odds
X	A	19,000	80	.996
X	B	19,800	240	.988
Y	A	950	99	.906
Y	B	990	297	.769

Note: Population X disease prevalence = 20%; Population Y disease prevalence = 1%. Test A correct detection rate = .95; Test A false detection rate = .001; Test B correct detection rate = .99; Test B false identification rate = .003.