The Impact of Evidence Lineups on Fingerprint Expert Decisions

Jeff Kukucka[1], Itiel E. Dror[2], Melissa Yu[2], Lisa Hall[3], & Ruth M. Morgan[2]

[1]Towson University

[2]University College London

[3]Metropolitan Police, London, United Kingdom

Abstract

Forensic examiners routinely compare a crime-relevant mark of unknown origin against a single

suspect's sample, which may create an expectation that the two will match. We tested how

embedding the suspect's sample among known-innocent fillers (i.e., an *evidence lineup*) affects

expert decision-making. Experienced fingerprint examiners (*N* = 43) compared crime-relevant

marks against either individual suspect fingerprints (i.e., the standard procedure) or arrays of

fingerprints (i.e., evidence lineups), with a matching fingerprint either present or absent.

Evidence lineups promoted conservative decision-making, as evidenced by fewer correct IDs and

a higher rate of inconclusive judgments. Though errors were rare, evidence lineups also

occasionally revealed errors that would have otherwise gone undetected. Our findings thus

support arguments that evidence lineups can expose fraud, identify flawed methodologies, and

curb overconfidence. The potential benefits and challenges of implementing evidence lineups in

forensic laboratories are discussed.

Keywords: *forensic science; cognitive bias; expert decision-making; fingerprints; evidence*

lineups.

The Impact of Evidence Lineups on Fingerprint Expert Decisions

In criminal investigations, forensic examiners are often asked to compare a pattern or impression found at a crime scene (e.g., a fingermark, signature, bullet cartridge, or shoemark) against a sample from a suspect and opine as to whether they originated from the same source (i.e., "match"). However, recent studies have overwhelmingly demonstrated that contextual factors can influence examiners' judgments of forensic science evidence in ways that undermine their probative value and increase the risk of costly errors (i.e., *forensic confirmation bias*; Kassin, Dror, & Kukucka, 2013; see also Cooper & Meterko, 2019; Dror, 2018; Kukucka, 2018).

Sources of Bias

While most prior research on forensic confirmation bias has tested the impact of extraneous case information (e.g., knowledge of a suspect's confession or race; see Cooper & Meterko, 2019), bias can also arise from other sources—some of which are inherent to the evidence or procedure itself. Dror (2020) developed a taxonomy of eight sources of bias that can impact forensic examiners' judgments. At Level 1, an examiner may be biased by the content of the evidence itself (e.g., a handwriting sample that contains suggestive language, a bitemark that betrays the nature of the crime). At Level 2, the act of comparing two stimuli may be inherently biasing in that knowledge of the known suspect sample may bias the analysis of the unknown mark from the crime scene—that is, rather than the evidence data driving the comparison, the suspect sample guides the comparison. At Level 3, examiners can be biased by knowledge of task-irrelevant case information (e.g., the suspect's criminal history or confession). These three levels all pertain to factors that are unique to the specific case at hand.

The next five levels in the taxonomy relate to wider factors, not specific to the case at hand. At Level 4, examiners may hold base-rate expectations, derived from their experience and/or from the analytic procedure itself, which predispose them to expect a particular outcome. At Level 5, examiners are subject to organizational and cultural influences (e.g., pressure to solve a case, allegiance to the party that hired them; Morgan, 2017; Murrie, Boccaccini, Guarnera, & Rufino, 2013). At Level 6, the examiner's education and training—and at Level 7, personal factors such as the need for closure, motivation, and ideology (e.g., Dunning & Balcetis, 2013)—may influence their judgment. Finally, at Level 8, basic psychological mechanisms, such as conformity from social psychology and top-down processes from cognitive psychology, may influence an examiner's judgment.

In a demonstration of Level 2 bias, Dror et al. (2011) found that experienced fingerprint examiners identified fewer minutiae (i.e., features) in the same fingerprint if it was presented alongside a suspect's fingerprint than if it was presented alone, suggesting that the mere presence of a comparison print caused examiners to disregard features that they would have otherwise deemed relevant. To mitigate Level 2 bias, Dror et al. (2015) proposed a protocol called Linear Sequential Unmasking (LSU), whereby examiners first examine the crime scene evidence in isolation and document their initial judgments before viewing the suspect's sample. Recent work suggests that LSU can easily be implemented in many forensic science disciplines, such as forensic entomology (Archer & Wallman, 2016), forensic document examination (Found & Ganas, 2013), firearms examination (Mattijssen, Kerkhoff, Berger, Dror, & Stoel, 2016), and even DNA analysis (Jeanguenat, Budowle, & Dror, 2017).

Most studies of forensic confirmation bias have examined Level 3 bias by investigating how task-irrelevant information impacts forensic decision-making. In one of the earliest studies, Dror and Charlton (2006) presented experienced fingerprint examiners with pairs of fingerprints that, unbeknownst to them, they had analyzed earlier in their careers. When reshown these same prints along with extraneous case information that implied the suspect's innocence or guilt, examiners changed 17% of their own previous decisions on the same prints.

This effect has since been replicated with respect to fingerprints (Stevenage & Bennett, 2017) and similar effects have been found in many other forensic disciplines—including arson investigation (Bieber, 2012), forensic anthropology (Nakhaeizadeh, Dror, & Morgan, 2014), crime scene investigation (van den Eeden, de Poot, & van Koppen, 2016), bloodstain pattern analysis (Osborne, Taylor, Healey, & Zajac, 2016), and complex DNA analysis (Dror & Hampikian, 2011), among others (for reviews, see Cooper & Meterko, 2019; Kukucka, 2018).

Accordingly, many have advised that forensic examiners be blinded to any information that is potentially biasing and irrelevant to the task at hand (e.g., Dror, 2018; Kukucka, Kassin, Zapf, & Dror, 2017; President's Council of Advisors on Science and Technology [PCAST], 2016). Nevertheless, the forensic science community has often been loath to recognize the existence and importance of cognitive bias (Buckleton, Evett, & Weir, 2014; Champod, 2014; Kukucka et al., 2017; Oliver, 2018).

Evidence Lineups

Of central importance to the current study, some have argued that the standard practice of asking a forensic examiner to compare a crime scene mark of unknown origin against a sample from a single known suspect is inherently suggestive of guilt (Risinger, Saks, Thompson, &

Rosenthal, 2002; Whitman & Koppl, 2010)—a form of Level 4 (procedure-driven) bias. That is to say, because the police do not identify suspects at random, their decision to submit *this* suspect's sample for analysis implies that other incriminating evidence exists and thus predisposes the examiner to believe that the two stimuli will match. This practice is analogous to an eyewitness showup, where an eyewitness views a single suspect and reports whether they believe this person was the culprit (Dysart & Lindsay, 2007). For decades, eyewitness showups have been criticized as inherently suggestive, and research has indicated that showups increase the risk of misidentifying an innocent suspect as guilty (e.g., Neuschatz et al., 2016; Steblay, Dysart, Fulero, & Lindsay, 2003; Wells et al., 1998; Wetmore et al., 2015). As such, best practice guidelines for eyewitness identification now discourage the use of showups unless absolutely necessary (National Academy of Sciences [NAS], 2014; Wells et al., 2020).

To counteract this bias, researchers have proposed that forensic examiners instead follow a procedure analogous to an eyewitness lineup by comparing the crime scene mark against an *evidence lineup*—i.e., an array that embeds the suspect's sample among several known-innocent filler samples. Under this procedure, the examiner must decide *which, if any,* of the lineup samples matches (not whether *this suspect's* sample matches) the mark from the crime scene (Kassin et al., 2013; Risinger et al., 2002; Wells, Wilford, & Smalarz, 2013). In theory, evidence lineups should protect against misidentifications by siphoning errors away from innocent suspects and onto filler samples—just as eyewitness lineups do (Smith, Wells, Lindsay, & Penrod, 2017). Accordingly, evidence lineups should quickly expose incompetence or fraud by virtue of their ability to produce known errors (i.e., filler identifications; Wells et al., 2013). Although forensic identification is a perceptual—not memory-based—task, advocates have argued that

evidence lineup procedures could nonetheless be modeled after empirically-based best practices for eyewitness identification (NAS, 2014; Wells et al., 2020).

Few studies have examined how evidence lineups impact forensic identification relative to the standard procedure. Miller (1987) asked trained students to compare a crime scene hair against either a single suspect hair or a lineup of five hairs—none of which truly matched the crime scene hair. When shown only one suspect hair, 30% of students misjudged it as a match, but when shown a lineup, only 4% misjudged any of the five lineup hairs as a match. Though groundbreaking and encouraging, Miller's study was limited in several ways—including its use of only non-matching hairs, which left it unknown whether evidence lineups might likewise decrease correct identifications of samples that truly match. An unpublished study by Thompson, Tangen, Treloar, and Ivison (2010) found that students were equally accurate in identifying fingerprints from an evidence lineup whether it was presented simultaneously (i.e., all lineup prints shown at once) versus sequentially (i.e., lineups prints shown one-at-a-time)—but they did not compare evidence lineups against the standard "showup" procedure.

Most recently, Quigley-McBride and Wells (2018) trained undergraduates in fingerprint analysis and had each perform eight fingerprint comparisons using either the standard procedure or an evidence lineup, including four comparisons where a matching print was present and four where it was not. They found that evidence lineups decreased correct identifications of the matching print relative to the standard procedure, but also decreased misidentifications of the innocent suspect's print—and to a greater degree. Evidence lineups also produced fewer correct rejections (i.e., correct decisions that no matching print was present) than the standard procedure, indicating that evidence lineups siphoned errors away from the innocent suspect and

onto filler samples. Notably, both of these findings are common to eyewitness identification studies (e.g., Neuschatz et al., 2016; Smith et al., 2017; Wells et al., 1998), suggesting that lineup presentations influence eyewitness and forensic identification similarly. Consistent with this idea, Smith, Mackovichova, Jalava, and Pozzulo (2020) recently demonstrated that the advantage of lineups over showups also extends to non-facial stimuli (i.e., vehicles).

Novice versus Expert Decision-Making

The aforementioned studies found that evidence lineups benefitted students' judgments of forensic stimuli relative to the standard procedure—but to date, no study has examined how evidence lineups affect the decision-making of forensic *experts* (i.e., experienced professional examiners). One would naturally expect professional examiners to be more accurate than novices; for example, studies have found that experienced fingerprint examiners are undeniably more accurate than novice students and new trainees (Tangen, Thompson, & McCarthy, 2011; Thompson, Tangen, & McCarthy, 2014). These studies have also found that examiners tend to make more conservative judgments than novices, such that examiners committed more false negative errors (i.e., failing to implicate a guilty person) than false positive errors (i.e., implicating an innocent person; Thompson & Tangen, 2014; Thompson et al., 2014).

However, experience, training, and accuracy do not imply immunity from cognitive bias—although many examiners appear to believe otherwise (Kukucka et al., 2017). Conversely, some have argued that experts' greater reliance on top-down processing, heuristics, and other cognitive shortcuts (e.g., schemas, selective attention, chunking) may actually render them *more* vulnerable to bias (Chi, 2006; Dror, 2011; Walther, Fiedler, & Nickel, 2003). While there is ample evidence that forensic experts are vulnerable to bias (see Cooper & Meterko, 2019; Kukucka,

2018), only one study has directly compared the biasability of forensic novices and experts: van den Eeden, de Poot, and van Koppen (2019) found that students' and professionals' investigations of a mock death scene were equally impacted by biasing contextual information. Given that contextual (i.e., Level 3) bias has been shown to affect both novices and experts, and evidence lineups have been shown to improve novices' discriminability, it stands to reason that evidence lineups may likewise benefit expert decision-making.

The Current Studies

In sum, researchers have argued that the standard forensic identification procedure is inherently suggestive (i.e., Level 4 bias) and proposed that examiners instead make identification decisions from evidence lineups that embed the suspect's sample among known-innocent fillers. A few studies have found that students who viewed evidence lineups exhibited advantages similar to those commonly seen in eyewitness identification research, but no study has tested how evidence lineups impact forensic *expert* decision-making.

To address this question, we recruited a sample of experienced professional fingerprint examiners and asked them to make several context-free identification decisions using either the standard procedure (i.e., one crime scene mark and one suspect print) or a sequential evidence lineup (i.e., one crime scene mark and five comparison samples, shown one-at-a-time). Some trials included a matching pair of prints while others did not, allowing us to investigate how presentation format affected both correct and false identifications and non-identifications. We also explored whether presentation format impacted examiners' conservatism (i.e., tendency to believe that a matching print was present or judge the prints as inconclusive) and/or response latency (i.e., how quickly they made their identification decisions).

Method

Participants and Design

After obtaining approval from the management of a major forensic laboratory in the United Kingdom, they agreed to send recruitment e-mails on our behalf to practicing fingerprint examiners in the UK. Each e-mail explained that the study aimed to better understand factors that impact forensic decision-making. By random assignment, each e-mail contained a link to either the showup or lineup condition, which were hosted on separate webpages. Ultimately, 43 examiners volunteered to participate in either the showup (n = 15) or lineup (n = 28) condition.[1] The average examiner was 44.07 years old (SD = 7.60; Med = 42; range = 32–58) and had considerable experience as a fingerprint examiner (M = 18.14 years, SD = 8.20; Med = 16; range = 9–40). Most examiners worked in large laboratories (M = 59.64 fellow examiners, SD = 33.69; Med = 72.5; range = 0–100). No other demographic information was collected.

Procedure

Each examiner viewed four fingerprint displays online, all of which followed either the standard procedure (i.e., one crime scene mark and one suspect print; *showup* condition) or sequential evidence lineup procedure (i.e., one crime scene mark and five comparison prints shown one-at-a-time; *lineup* condition). Three of these four displays were "match-present," meaning that the suspect's print (showup condition) or one of the comparison prints (lineup condition) did in fact match the crime scene mark. One display was "match-absent," meaning that neither the suspect's print (showup condition) nor any of the comparison prints (lineup condition) actually matched the crime scene mark. For the suspect print (showup condition) or for each comparison print (lineup condition), examiners reported whether they believed that

print did or did not match the crime scene mark or if they were unsure (i.e., "inconclusive"), and

our website covertly recorded how quickly they made each judgment. We later coded these

judgments as identification decisions in the manner described below and reported in Table 1.

Upon completing the study, examiners were instructed not to discuss the study with their peers.

Materials

       Instructions. After giving consent, examiners read instructions that explained the nature

of the impending fingerprint comparison task. In accordance with best practices for eyewitness

identification, these instructions reminded examiners that each display may or may not contain

a matching print. Examiners in the lineup condition were also told that each lineup would include

five comparison prints; although "back-loading" is the recommended practice for sequential

eyewitness lineups, examiners in our lineup condition each viewed four lineups and therefore

would have inferred the size of each lineup after completing the first lineup.

       Fingerprint stimuli. Each examiner viewed three match-present displays and one match-

absent display; we opted to include more match-present trials because these are the most critical

in forensic work, and we wanted to ensure that we collected sufficient data relevant to these

decisions. For the three match-present displays, we used images of latent fingermarks and the

corresponding fingerprints that were a match; these were generated for use in the current

studies (including a record of which latent marks were made by which finger) and therefore

ground truth was known. For the match-absent display and lineup fillers, we used a large

database (i.e., Automated Fingerprint Identification System [AFIS]) to generate non-matching

prints that were similar to the latent fingermark (Dror & Mnookin, 2010). Figure 1 shows

examples of a crime scene mark, matching print, and filler (i.e., non-matching) print that were

used. A qualified expert (i.e., the fourth author) oversaw the creation of the four fingerprint

displays to ensure that their difficulty approximated that of real-world casework.

   Showup condition. Examiners in the showup condition viewed four fingerprint showups,

each consisting of one crime scene mark (always shown on the left) and one suspect print (always

shown on the right) side-by-side on a computer screen. For each showup, examiners indicated

whether they believed that the mark and print matched, with options of "match," "non-match,"

and "inconclusive." Once they provided a judgment, the next showup automatically appeared.

Three showups were match-present (i.e., the suspect print did in fact match the crime scene

mark). One showup, which was always presented third, was match-absent (i.e., the suspect print

did not match the crime scene mark).

   For match-present showups, responses were coded as correct if the examiner judged the

suspect's print as a match to the crime scene mark, and coded as incorrect non-identifications if

they misjudged the suspect's print as a non-match to the crime scene mark. For the match-absent

showup, responses were coded as false identifications if the examiner misjudged the suspect's

print as a match to the crime scene mark, and coded as correct non-identifications if they judged

the suspect's print as a non-match to the crime scene mark. For all showups, inconclusive

judgments were coded as their own separate category (for a detailed discussion of inconclusive

decisions, see Dror & Langenburg, 2019).

   Lineup condition. Examiners in the lineup condition viewed four sequential evidence

lineups, each consisting of one crime scene mark (shown on the left side of the screen) and five

comparison prints (shown one-at-a-time on the right side of the screen). Examiners were

required to provide a judgment of each comparison print (i.e., match, non-match, or

inconclusive) before viewing the next comparison print. A "continuation rule" was used (Wells, Steblay, & Dysart, 2015a), such that examiners judged all five comparison prints for each lineup even if they had judged one of the earlier comparison prints as a match. Each examiner therefore made 20 total judgments (i.e., four lineups times five comparison prints).

Three lineups were match-present (i.e., one of the five comparison prints matched the crime scene mark and the other four did not). To account for serial position effects, we varied the ordinal position of the matching print across lineups, such that it always appeared third, fourth, and second, respectively, in the three match-present lineups. One lineup, which was always the third of the four lineups, was match-absent (i.e., none of the comparison prints matched the crime scene mark). One of the five comparison prints in the match-absent lineup— which always appeared second—was the same non-matching suspect print used in the match-absent showup; for coding purposes, this print was designated as belonging to an innocent suspect and the other four comparison prints were designated as fillers.

For match-present lineups, responses were coded as correct if the examiner judged *only* the matching print as a match to the crime scene mark, and coded as incorrect non-identifications if they misjudged the matching print as a non-match and did not misjudge any of the fillers as a match. For the match-absent lineup, responses were coded as false identifications if they misjudged *only* the innocent suspect's print as a match to the crime scene mark, and coded as correct non-identifications if they judged the innocent suspect's print as a non-match and did not misjudge any of the fillers as a match. For all lineups, responses were coded as filler identifications if they misjudged any of the fillers as a match,[2] and coded as "inconclusive" if they

judged the suspect's print (i.e., either the matching print or the innocent suspect's print) as inconclusive and did not misjudge any of the fillers as a match.

Results

Examiners made 172 total identification decisions, including 129 from match-present displays and 43 from match-absent displays. Table 1 presents examiners' identification decisions as a function of condition (showup vs. lineup) and match presence.

For each examiner, we calculated their total number of positive identifications (i.e., judgments that a matching sample was present, irrespective of accuracy) and their total number of "inconclusive" judgments—both out of a possible four. For the three match-present displays, we calculated each examiner's total number of correct identifications and incorrect non-identifications—both out of a possible three. Finally, we noted whether each examiner made a false identification or a correct non-identification from the lone match-absent display.

Positive Identifications and Inconclusive Judgments

Overall, examiners made a positive identification for 37.8% of displays, and they judged 25.6% of displays as inconclusive. Examiners in the showup ($M = 1.73$, $SD = 0.80$) and lineup ($M = 1.39$, $SD = 0.73$) conditions made the same number of positive identifications, $t(41) = 1.40$, $p = .168$, $d = 0.45$ [95% CI: -0.19, 1.08]. However, examiners in the lineup condition made significantly more inconclusive judgments ($M = 1.29$, $SD = 0.71$) than those in the showup condition ($M = 0.53$, $SD = 0.83$), $t(41) = 3.11$, $p = .003$, $d = 1.00$ [95% CI: 0.33, 1.66].

Correct and False Identifications

Examiners in the showup condition made more correct identifications ($M = 2.60$, $SD = 0.63$) than those in the lineup condition ($M = 2.14$, $SD = 0.45$), $t(41) = 2.76$, $p = .009$, $d = 0.88$ [95%

CI: 0.23, 1.54]. False identifications were almost nonexistent: One examiner in the lineup condition (3.6%) and none in the showup condition made a false identification; a Fisher's exact test indicated that this difference was not significant, p = .651.

Correct and Incorrect Non-Identifications

Overall, 95.3% of examiners correctly rejected the match-absent display, and correct non-identifications were equally common in the showup (100%) and lineup (92.9%) conditions, $\chi^2(1)$ = 1.12, p = .289, ɸ = .16, OR = undefined. To be exact, examiners who viewed a match-absent lineup collectively produced two incorrect match judgments (7.1%)—one of which was a filler identification (i.e., a known error)—while none of the examiners who viewed a match-absent showup made an incorrect judgment that would implicate an innocent suspect.

Conversely, examiners made incorrect non-identifications for 17.1% of match-present displays. Examiners in the lineup condition made somewhat fewer incorrect non-identifications (M = 0.39, SD = 0.50) than those in the showup condition (M = 0.73, SD = 0.70)—but this difference was not statistically significant, $t(41)$ = 1.85, p = .072, d = 0.59 [95% CI: -0.05, 1.23].

Response Latency

Examiners made 620 discrete judgments (i.e., 60 in the showup condition and 560 in the lineup condition); for each, we recorded how quickly (in seconds) they made their judgment. We later identified and removed five extreme outliers, whose scores were more than 24 IQRs above Q_3 (i.e., the 75th percentile) and more than 10 IQRs above the next-highest score; we presume that these scores reflect participants who became distracted mid-study.

With these scores excluded, examiners spent an average of 57.84 seconds (SD = 88.70; Med = 29.78) on each judgment—but the distribution remained positively skewed, Shapiro-Wilk

W = .53, p < .001. We therefore used non-parametric tests (i.e., Mann-Whitney U tests) to compare response latencies between conditions.

Overall, examiners in the lineup condition (*Med* = 25.66) made their judgments faster than those in the showup condition (*Med* = 87.55), Z = -8.43, p < .001. For the four prints that were included in both conditions (i.e., the three matching prints and the one innocent suspect print), examiners in the lineup condition judged those same prints faster (*Med* = 44.42) than those in the showup condition (*Med* = 87.55), Z = -4.68, p < .001.

Discussion

In light of growing concern over error and bias in forensic science decision-making (e.g., Dror, 2018; NAS, 2009; PCAST, 2016), researchers have suggested structuring the task of forensic identification similarly to that of eyewitness identification (Kassin et al., 2013; Risinger et al., 2002). Yet, only a few studies have examined how evidence lineups affect decision-making relative to the standard practice of analyzing only the suspect's sample, and none have used professional forensic examiners. While novice participants in eyewitness identification studies provide a fitting parallel to real-world eyewitnesses, student participants do not similarly represent forensic examiners, who are highly skilled experts in their domain. The current study provides the first data on how evidence lineups impact forensic experts' decision-making. Below, we discuss how this study adds to our understanding of evidence lineups, and we consider the benefits and challenges of implementing such procedures in the real world.

Wells et al. (2013) argued that a critical benefit of evidence lineups is their ability to quickly expose incompetence or fraud while allowing skilled examiners to demonstrate their expertise. Our findings underscore these claims. With respect to the former, evidence lineups

revealed rare but potentially important errors (i.e., filler identifications) that would have gone undetected in the standard procedure; in real-world situations, these errors can draw attention to invalid techniques or examiners who are dishonest or inadequately trained. With respect to the latter, examiners who viewed evidence lineups still produced highly accurate judgments, thus clearly demonstrating their expertise (Tangen et al., 2011; Ulery, Hicklin, Buscaglia, & Roberts, 2011). Importantly, our examiners also committed fewer false positive errors than false negative errors, which may reflect a response bias toward judging the prints as a "non-match" when uncertain (Thompson & Tangen, 2014; Thompson et al., 2014) and/or examiners' skill at detecting and excluding non-matching prints.

As in Quigley-McBride and Wells' (2018) study of novices, as well as myriad studies of eyewitness identification (Neuschatz et al., 2016), our examiners were less likely to make a correct identification (ID) from a match-present evidence lineup (43%) than a match-present showup (58%). Conversely, whereas prior work suggests that evidence lineups should likewise decrease false IDs, false IDs in our study were exceedingly rare: Only one examiner (in the showup condition) misidentified the innocent suspect's sample as a match to the crime scene mark. This finding could indicate that evidence lineups do not affect the rate of false IDs—or it could be that our materials did not allow for such an effect to reveal itself. Although our materials were designed to mimic the difficulty of real-world casework, the fact that false IDs were almost non-existent suggests that our target-absent displays were quite easy for examiners, such that they precluded a proper test of how evidence lineups impact identification accuracy. Perhaps our manipulation would have produced a noticeable effect on false IDs had we included more

challenging materials or more trials. As such, we refrain from drawing any strong conclusions as to whether and how evidence lineups impact accuracy.

That said, examiners in the current study were not 100% accurate, and a closer look at their inaccuracies suggests an important trend. Collectively, examiners who viewed evidence lineups made three errors, which included one false ID and two filler IDs. In other words, two of the three errors committed by examiners in this study would be known errors in the real world. This pattern is reminiscent of the phenomenon of filler siphoning in eyewitness identification, whereby lineups improve discriminability by siphoning false identifications away from the innocent suspect onto known-innocent fillers (Smith et al., 2017). While fingerprint examiners do tend to be highly accurate, our findings serve as a reminder that errors do occur, and they suggest that evidence lineups can make errors more transparent—and perhaps more preventable—by helping to identify and refine/abandon techniques that produce filler identifications.

While our manipulation did not affect the rate of positive identifications (i.e., "choosing"), evidence lineups elicited more inconclusive judgments (32%) than did the standard procedure (13%). An inconclusive judgment is difficult to interpret: Does it indicate that the examiner *cannot* decide or that they do not *want to* commit to a categorical decision—and when, if ever, should inconclusive judgments be considered erroneous or correct? (see Dror & Langenburg, 2019). In any case, our findings indicate that evidence lineups prompted examiners to be more conservative in their decision-making. While we reserve judgment as to whether this is a desirable outcome, we note that fingerprint experts tend to be more fearful of committing false positive errors than false negative errors (Charlton, Fraser-Mackenzie, & Dror, 2010). Our findings

suggest that evidence lineups, insofar as they promote conservative decisions, may provide examiners with an extra layer of protection against the errors they fear most.

Relatedly, Wells et al. (2013) speculated that evidence lineups should discourage overconfidence by making examiners more aware of their own mistakes and limitations, such that their confidence should become better calibrated to their true ability over time. Consistent with this argument, a recent meta-analysis of eyewitness identification studies found that confidence was a better predictor of accuracy for witnesses who viewed a lineup as opposed to a showup (Neuschatz et al., 2016). Because we did not measure confidence, we can only speculate to as whether viewing an evidence lineup might attenuate confidence. For example, the observed difference in inconclusive judgments raises the question of whether examiners might judge the same sample as a match less confidently if it is embedded among fillers than if it is not—which is to say, the standard procedure may artificially inflate examiners' confidence. Future research should assess how evidence lineups impact confidence in one's opinion and the confidence-accuracy relationship—perhaps by asking participants to report judgments probabilistically rather than categorically, which may be a more realistic and/or preferred method in some forensic domains (Thompson, Vuille, Taroni, & Biedermann, 2018).

While evidence lineups produced more inconclusive judgments, they also produced a slight—albeit not statistically significant—decrease in incorrect nonidentifications (13%) relative to the standard procedure (24%). Similarly, a meta-analysis by Steblay, Dysart, Fulero, and Lindsay (2003) found that target-present eyewitness lineups elicited fewer false rejections than target-present showups. This pattern suggests that evidence lineups prompted conservatism, but they did not create a general reluctance to make positive identifications, as examiners who

viewed a lineup were actually somewhat *less* likely to "miss" a matching sample when one was present. Thus, by increasing the rate of inconclusive judgments, evidence lineups sometimes caused a failure to inculpate a guilty suspect (i.e., fewer correct IDs), but sometimes prevented examiners from incorrectly exculpating a guilty suspect (i.e., fewer incorrect non-IDs)—but again, our conclusions vis-à-vis accuracy are tentative and warrant replication.

Lastly, we found that examiners who viewed a lineup judged fingerprint pairs much faster than those who saw only a single pair (i.e., showup). While this difference may be partly explained by examiners quickly ruling out some lineup fillers, examiners judged the same pair in half the time (i.e., 44 vs. 88 seconds) when that pair was part of a lineup—yet this increased speed did not affect rates of false IDs or correct non-IDs, and it actually produced somewhat *fewer* misses. These values are difficult to interpret without knowing how the observed reaction times compare to examiners' real world reaction times, but one possibility is that the lineup forced examiners to work more efficiently by focusing their attention on features that are most likely to be diagnostic of a match or non-match. In the eyewitness arena, decision time has received mixed support as a predictor of accuracy (see, e.g., Brewer & Weber, 2008). Future work should likewise explore whether the speed of a forensic identification decision predicts its accuracy.

Implications for Forensic Science Practice

Quigley-McBride and Wells (2018) found that evidence lineups produced similar effects on novices' decision-making as those commonly seen in eyewitness identification research. In the current study, evidence lineups produced a significant decrease in correct identifications and a non-significant decrease in incorrect non-identifications relative to the standard "showup" procedure—both of which are also seen in meta-analyses of eyewitness identification studies

(Neuschatz et al., 2016; Steblay et al., 2003). Thus, the current study provides some additional

evidence that evidence lineups and eyewitness lineups produce effects on decision-making that

are similar in kind (though perhaps not in magnitude).

If that is the case, then best practices for eyewitness identification (NAS, 2014; National

Institute of Justice, 2017) may help facilitate the implementation and optimization of evidence

lineup procedures. For example, examiners could receive unbiased instructions (Malpass &

Devine, 1981) and document their confidence before receiving feedback (e.g., verification by a

fellow examiner; Douglass & Steblay, 2006). Many factors that hinder eyewitnesses would be

moot for forensic examiners—including both *estimator* (e.g., intoxication, retention interval) and

system variables (e.g., double-blind administration, if evidence lineups are self-administered;

Wells, 1984), but other "estimator variables" could hinder examiners, such as the quality of the

available evidence (e.g., the clarity and/or completeness of a latent print).

As a practical matter, some have questioned how to generate and select filler samples for

evidence lineups (Reese, 2012; Whitman & Koppl, 2010). For disciplines with large databases of

samples, it should be relatively simple. For example, fingerprint experts often use a computerized

database (i.e., AFIS) to identify potential matches (Dror & Mnookin, 2010; Dror, Wertheim,

Fraser-Mackenzie, & Walajtys, 2012), which could be used to generate fillers in actual cases just

as in the current study. Whereas using highly similar faces can impede eyewitnesses (Wells,

Memon, & Penrod, 2006), using highly similar evidence lineup fillers should not impede skilled

experts—though this remains an open empirical question. In disciplines without such a database,

evidence lineups could be constructed by a fellow examiner (e.g., a case manager; Dror, 2013;

Thompson, 2011) using similar fillers that originate from known-innocent individuals (e.g.,

individuals who were deceased or incarcerated when the crime occurred). For example, the American Board of Forensic Odontology now recommends that bitemark examiners "engage another dentist to produce a 'dental lineup' of dentition evidence… includ[ing] evidence from the person or persons of interest and from other individuals as foils" (ABFO, 2018), though they provide no guidance on how to identify fillers that are appropriately similar. Future research should compare various methods for selecting fillers and examine the unique challenges of constructing evidence lineups in disciplines other than fingerprint analysis.

While some examiners and disciplines have been receptive to the idea of using evidence lineups, many others have opposed this and other debiasing measures on pragmatic grounds, including financial cost and time (Champod, 2014; Charlton, 2013; Langenburg, 2017). If conducting an evidence lineup proves to be considerably more onerous and/or expensive than the standard procedure, laboratories might instead consider a triage approach, whereby evidence lineups are used only in situations where the risk of error or bias is greatest (e.g., cases where exposure to task-irrelevant information is inevitable; situations where a second examiner is asked to independently verify another examiner's judgment; Dror, 2013). In any case, laboratories that adopt evidence lineups should involve psychologists in the implementation process; this collaboration will be critical in ensuring that the procedures are based on the best-available psychological evidence but also mindful of the idiosyncratic needs and challenges of the forensic discipline in question.

Limitations and Future Directions

The current study the first to examine how evidence lineups affect the decision-making of professional forensic examiners. However, unlike in some prior studies of forensic experts

(e.g., Dror & Charlton, 2006; Dror et al., 2012), our examiners knew that they were participating in a study, which could have led them to behave differently than they do in their normal casework. In one study, for example, fingerprint examiners who knew they were taking part in a study produced an unusually high rate of inconclusive judgments, which the researchers attributed to suspicion of being evaluated (Langenburg et al., 2009). Our examiners likewise produced many inconclusive judgments, which may reflect evaluation apprehension, but this would not explain why inconclusive judgments were significantly more common in the lineup condition. Regardless, the risk of reactivity (Orne, 1962) can be reduced by integrating research efforts into examiners' normal casework such that they do not realize they are being measured (Dror et al., 2012; see also Thompson, Black, Jain, & Kadane, 2017). Future studies of forensic expert decision-making should aspire to this ideal. Unfortunately, it is difficult—if not impossible—to do this when testing new methods such as the evidence lineup, as forensic examiners do not use this method in their normal work.

As noted above, the low observed rate of false identifications suggests that the materials used in the current study were not challenging for examiners. Evidence lineups are thought to mitigate a bias toward positive identifications (Whitman & Koppl, 2010), and it is well-established that judgments of ambiguous stimuli are more vulnerable to bias (e.g. Kunda, 1990). Consistent with this, Quigley-McBride and Wells (2018) found that evidence lineups had a stronger effect on judgments of high ambiguity fingerprints. Moreover, because examiners knew that their judgments carried no consequences, the standard procedure in our study may have been less suggestive than it is in the real world. Indeed, examiners judged only 58% of match-present showups as matches, implying that the procedure was not highly suggestive. Therefore, the

relative ease and inconsequential nature of examiners' judgments in our study may have led us

to underestimate the true impact of evidence lineups.

Our experts also evaluated fingerprint evidence that was entirely devoid of context, yet

real-world examiners often receive other case information (which may be task-relevant or task-

irrelevant) in addition to the forensic evidence in question (e.g., Gardner, Kelley, Murrie, &

Blaisdell, 2019). Perhaps evidence lineups affect decision-making differently in the presence of

biasing contextual information. To test this possibility, Quigley-McBride and Wells (2018) had

some students read a case report that was "highly suggestive of guilt" before analyzing

fingerprint evidence, but they found little indication that this report moderated the effect of

using an evidence lineup. Even so, future research should further explore this possibility by

manipulating examiners' a priori knowledge of ostensibly authentic case information.

Finally, given the multitude and diversity of forensic science disciplines, researchers

should exercise some caution in generalizing results across disciplines that may vary in their

vulnerability to bias. For example, at least two studies failed to find confirmation bias effects

among professional shoeprint (Kerstholt, Paashuis, & Sjerps, 2007) and firearms (Kerstholt et al.,

2010) examiners in the Netherlands. The authors of these studies speculated that these domains

experience less risk of bias because they utilize a highly standardized methodology (though there

are other possible explanations for these null effects; see Kukucka, 2018). In any case, the costs

and benefits of using evidence lineups may well differ across disciplines.

Conclusion

Researchers (e.g., Neuschatz et al., 2016; Wetmore et al., 2015) and policy makers (e.g.,

NAS, 2014; National Institute of Justice, 2017) have unequivocally rejected eyewitness showups,

which increase the risk of misidentifying an innocent suspect as guilty, yet forensic science examiners routinely use an analogous procedure. The current studies replicate and extend prior research suggesting that evidence lineups affect decision-making similarly, such that the forensic sciences may benefit from modeling their own procedures after best practices for eyewitness identification. However, as discussed above, the implementation of evidence lineups is likely to pose logistical challenges that may differ between forensic disciplines. While an exhaustive cost/benefit analysis is beyond the scope of this paper, we hope that these studies stimulate additional efforts to understand the benefits and challenges of using evidence lineups, toward the superordinate goal of maximizing the value of forensic science evidence.

References

American Board of Forensic Odontology. (2018). *Standards and guidelines for evaluating*

bitemarks. Retrieved from http://abfo.org/wp-content/uploads/2012/08/ABFO-

Standards-Guidelines-for-Evaluating-Bitemarks-Feb-2018.pdf.

Archer, M. S., & Wallman, J. F. (2016). Context effects in forensic entomology and use of

sequential unmasking in casework. *Journal of Forensic Sciences*, *61*, 1270-1277.

doi:10.1111/1556-4029.13139

Bieber, P. (2012). *Measuring the impact of cognitive bias in fire investigation.* Proceedings of

the International Symposium on Fire Investigation, Science and Technology, 3-17.

Available online at http://truthinjustice.org/Cognative_Bias_ARP.pdf

Brewer, N., & Weber, N. (2008). Eyewitness confidence and latency: Indices of memory

processes not just markers of accuracy. *Applied Cognitive Psychology*, *22*, 827-840.

doi:10.1002/acp.1486

Buckelton, J., Evett, I., & Weir, B. (2014). Letter to editor in response to editorial by Risinger et

al. *Science and Justice*, *54*, 510. doi:10.1016/j.scijus.2014.07.003

Champod, C. (2014). Research focused mainly on bias will paralyse forensic science. *Science and

Justice*, *54*, 107-109. doi:10.1016/j.scijus.2014.02.004

Charlton, D. (2013). Standards to avoid bias in fingerprint examination: Are such standards

doomed to be based on fiscal expediency? *Journal of Applied Research in Memory and

Cognition, 2*, 56–57. doi:10.1016/j.jarmac.2013.01.009

Charlton, D., Fraser-Mackenzie, P. A., & Dror, I. E. (2010). Emotional experiences and motivating

    factors associated with fingerprint analysis. *Journal of Forensic Sciences*, *55*, 385-393.

    doi:10.1111/j.1556-4029.2009.01295.x

Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In K. A. Ericsson, N.

    Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise*

    and expert performance (pp. 21-30). New York: Cambridge University Press.

Cooper, G. S., & Meterko, V. (2019). Cognitive bias research in forensic science: A systematic

    review. *Forensic Science International*, *297*, 35-46. doi:10.1016/j.forsciint.2019.01.016

Douglass, A. B., & Steblay, N. (2006). Memory distortion in eyewitnesses: A meta-analysis of the

    post-identification feedback effect. *Applied Cognitive Psychology*, *20*, 859-869.

    doi:10.1002/acp.1237

Dror, I. E. (2011). The paradox of human expertise: Why experts get it wrong. In N. Kapur (Ed.),

    The paradoxical brain (pp. 177-188). Cambridge, UK: Cambridge University Press.

Dror, I. E. (2013). Practical solutions to cognitive and human factor challenges in forensic

    science. *Forensic Science Policy & Management, 4,* 105-113.

    doi:10.1080/19409044.2014.901437

Dror, I. E. (2018). Biases in forensic experts. *Science, 360,* 243. doi:10.1126/science.aat8443

Dror, I. E. (2020). Cognitive and human factors in expert decision making: Six fallacies and the

    eight sources of bias. *Analytical Chemistry.* Advance online publication.

Dror, I. E., Champod, C., Langenburg, G., Charlton, D., Hunt, H., & Rosenthal, R. (2011).

    Cognitive issues in fingerprint analysis: Inter-and intra-expert consistency and the effect

of a 'target' comparison. *Forensic Science International*, *208*, 10-17.

doi:10.1016/j.forsciint.2010.10.013

Dror, I. E., & Charlton, D. (2006). Why experts make errors. *Journal of Forensic Identification*, *56*, 600-616.

Dror, I. E., & Hampikian, G. (2011). Subjectivity and bias in forensic DNA mixture interpretation. *Science & Justice*, *51*, 204-208. doi:10.1016/j.scijus.2011.08.004

Dror, I.E., & Langenburg, G. (2019). "Cannot decide": The fine line between appropriate inconclusive determinations vs. unjustifiably deciding not to decide. *Journal of Forensic Sciences*, *64*, 1-15. doi: 10.1111/1556-4029.13854

Dror, I. E., & Mnookin, J. (2010). The use of technology in human expert domains: Challenges and risks arising from the use of automated fingerprint identification systems in forensics. *Law, Probability and Risk, 9,* 47-67.

Dror, I. E., Thompson, W. C., Meissner, C. A., Kornfield, I., Krane, D., Saks, M., & Risinger, M. (2015). Context management toolbox: A linear sequential unmasking (LSU) approach for minimizing cognitive bias in forensic decision making. *Journal of Forensic Sciences*, *60*, 1111-1112. doi:10.1111/1556-4029.12805

Dror, I. E., Wertheim, K., Fraser-Mackenzie, P., & Walajtys, J. (2012). The impact of human-technology cooperation and distributed cognition in forensic science: Biasing effects of AFIS contextual information on human experts. *Journal of Forensic Sciences*, *57*, 343-352. doi:10.1111/j.1556-4029.2011.02013.x

Dunning, D., & Balcetis, E. (2013). Wishful seeing: How preferences shape visual perception. *Current Directions in Psychological Science*, *22*, 33-37. doi:10.1177/0963721412463693

Dysart, J. E., & Lindsay, R. C. L. (2007). Show-up identifications: Suggestive technique or reliable

    method? In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of*

    eyewitness psychology, Volume II: Memory for people (pp. 137-154). Mahwah, NJ:

    Lawrence Erlbaum Associates.

Found, B., & Ganas, J. (2013). The management of domain irrelevant context information in

    forensic handwriting examination casework. *Science & Justice*, *53*, 154-158.

    doi:10.1016/j.scijus.2012.10.004

Gardner, B. O., Kelley, S., Murrie, D. C., & Blaisdell, K. N. (2019). Do evidence submission forms

    expose latent print examiners to task-irrelevant information? *Forensic Science*

    International, *297*, 236-242. doi:10.1016/j.forsciint.2019.01.048

Jeanguenat, A. M., Budowle, B., & Dror, I. E. (2017). Strengthening forensic DNA decision

    making through a better understanding of the influence of cognitive bias. *Science and*

    Justice, 57, 415-420. doi:10.1016/j.scijus.2017.07.005

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems,

    perspectives, and proposed solutions. *Journal of Applied Research in Memory and*

    Cognition, *2*, 42-52. doi:10.1016/j.jarmac.2013.01.001

Kerstholt, J., Eikelboom, A., Dijkman, T., Stoel, R., Hermsen, R., & van Leuven, B. (2010). Does

    suggestive information cause a confirmation bias in bullet comparisons? *Forensic*

    Science International, *198*, 138-142. doi:10.1016/j.forsciint.2010.02.007

Kerstholt, J., Paashuis, R., & Sjerps, M. (2007). Shoe print examinations: Effects of expectation,

    complexity and experience. *Forensic Science International*, *165*, 30-34.

    doi:10.1016/j.forsciint.2006.02.039

Kukucka, J. (2018). Confirmation bias in the forensic sciences: Causes, consequences, and

    countermeasures. In W. J. Koen & C. M. Bowers (Eds.), *The psychology and sociology of*

    wrongful convictions: Forensic science reform (pp. 223-245). New York: Elsevier.

Kukucka, J., Kassin, S. M., Zapf, P. A., & Dror, I. E. (2017). Cognitive bias and blindness: A global

    survey of forensic science examiners. *Journal of Applied Research in Memory and*

    Cognition, *6*, 452-459. doi:10.1016/j.jarmac.2017.09.001

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480-498.

    doi:10.1037/0033-2909.108.3.480

Langenburg, G. (2017). Addressing potential observer effects in forensic science: A perspective

    from a forensic scientist who uses linear sequential unmasking techniques. *Australian*

    Journal of Forensic Sciences, *49*, 548-563. doi:10.1080/00450618.2016.1259433

Langenburg, G., Champod, C., & Wertheim, P. (2008). Testing for potential contextual bias

    effects during the verification stage of the ACE-V methodology when conducting

    fingerprint comparisons. *Journal of Forensic Sciences*, *54*, 571-582. doi:10.1111/j.1556-

    4029.2009.01025.x

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the

    absence of the offender. *Journal of Applied Psychology*, *66*, 482-489. doi:10.1037/0021-

    9010.66.4.482

Mattijssen, E. J. A. T., Kerkhoff, W., Berger, C. E. H., Dror, I. E., and Stoel, R. D. (2016).

    Implementing context management in forensic casework: Minimizing contextual bias in

    firearms examination. *Science and Justice, 56,* 113-122. doi:10.1016/j.scijus.2015.11.004

Miller, L. S. (1987). Procedural bias in forensic science examinations of human hair. *Law and*

    Human Behavior, *11*, 157-163. doi:10.1007/BF01040448

Morgan, R. M. (2017). Conceptualising forensic science and forensic reconstruction. Part II: The

    critical interaction between research, policy/law and practice. *Science & Justice*, *57*, 460-

    467. doi:10.1016/j.scijus.2017.06.003

Murrie, D. C., Boccaccini, M. T., Guarnera, L. A., & Rufino, K. A. (2013). Are forensic experts

    biased by the side that retained them? *Psychological Science*, *24*, 1889-1897.

    doi:10.1177/0956797613481812

Nakhaeizadeh, S., Dror, I. E., & Morgan, R. (2014). Cognitive bias in forensic anthropology:

    Visual assessment of skeletal remains is susceptible to confirmation bias. *Science &*

    Justice, *54*, 208-214. doi:10.1016/j.scijus.2013.11.003

NAS (2009). *Strengthening forensic science in the United States: A path forward*. Washington,

    DC: National Academies Press.

NAS (2014). *Identifying the culprit: Assessing eyewitness identification.* Washington, DC:

    National Academies Press.

National Institute of Justice (2017). *Eyewitness Identification: Procedures for conducting photo*

    arrays. Washington, DC: Department of Justice.

Neuschatz, J. S., Wetmore, S. A., Key, K. N., Cash, D. K., Gronlund, S. D., & Goodsell, C. A. (2016).

    A comprehensive evaluation of showups. In M. K. Miller & B. H. Bornstein (Eds.),

    Advances in psychology and law (pp. 43-69). Springer.

Oliver, W. R. (2018). Comment on Dror, Kukucka, Kassin, and Zapf (2018), "When expert

      decision making goes wrong." *Journal of Applied Research in Memory and Cognition*, *7*,

      314-315. doi:10.1016/j.jarmac.2018.01.010

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular

      reference to demand characteristics and their implications. *American Psychologist*, *17*,

      776-783. doi:10.1037/h0043424

Osborne, N. K. P., Taylor, M. C., Healey, M., & Zajac, R. (in press). Bloodstain pattern

      classification: Accuracy, effect of contextual information and the role of analyst

      characteristics. *Science & Justice*. doi: 10.1016/j.scijus.2015.12.005

President's Council of Advisors on Science and Technology. (2016). *Forensic science in criminal*

      courts: Ensuring scientific validity of feature-comparison methods.

Quigley-McBride, A., & Wells, G. L. (2018). Fillers can help control for contextual bias in forensic

      comparison tasks. *Law and Human Behavior*, *42*, 295-305. doi:10.1037/lhb0000295

Reese, E. J. (2012). Techniques for mitigating cognitive biases in fingerprint identification. *UCLA*

      Law Review, *59*, 1252-1290.

Risinger, D. M. (2009). The NAS/NRC report on forensic science: A glass nine-tenths full (this is

      about the other tenth). *Jurimetrics*, *50*, 21-34.

Risinger, D. M., Saks, M. J., Thompson, W. C., & Rosenthal, R. (2002). The Daubert/Kumho

      implications of observer effects in forensic science: Hidden problems of expectation and

      suggestion. *California Law Review*, *90*, 1-56.

Smith, A. M., Mackovichova, S., Jalava, S. T., & Pozzulo, J. (2019). Fair forensic-object lineups are

superior to forensic-object showups. *Journal of Applied Research in Memory and

Cognition*, *9*, 68-82. doi:10.1016/j.jarmac.2019.11.001

Smith, A. M., Wells, G. L., Lindsay, R. C. L., & Penrod, S. D. (2017). Fair lineups are better than

biased lineups and showups, but not because they increase underlying discriminability.

Law and Human Behavior, *41*, 127-145. doi:10.1037/lhb0000219

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L., (2003). Eyewitness accuracy rates in police

showup and lineup presentations: A meta-analytic comparison. *Law and Human

Behavior*, *27*, 523-540. doi:10.1023/A:1025438223608

Stevenage, S. V., & Bennett, A. (2017). A biased opinion: Demonstration of cognitive bias on a

fingerprint matching task through knowledge of DNA test results. *Forensic Science

International*, *276*, 93-106. doi:10.1016/j.forsciint.2017.04.009

Tangen, J. M., Thompson, M. B., & McCarthy, D. J. (2011). Identifying fingerprint

expertise. *Psychological Science*, *22*, 995-997. doi:10.1177/0956797611414729

Thompson, M. B., & Tangen, J. M. (2014). The nature of expertise in fingerprint matching:

Experts can do a lot with a little. *PLoS One*, *9*, e114759.

doi:10.1371/journal.pone.0114759

Thompson, M. B., Tangen, J. M., & McCarthy, D. J. (2014). Human matching performance of

genuine crime scene latent fingerprints. *Law and Human Behavior*, *38*, 84-93.

doi:10.1037/lhb0000051

Thompson, M. B., Tangen, J. M., Treloar, R., & Ivison, K. (2010, September). Humans matching

fingerprints: Sequence and size. In *Proceedings of the Human Factors and Ergonomics

Society Annual Meeting* (Vol. 54, No. 4, pp. 478-481). Los Angeles, CA: Sage.

Thompson, W. C. (2011). What role should investigative facts play in the evaluation of scientific

evidence? *Australian Journal of Forensic Sciences*, *43*, 123-134.

doi:10.1080/00450618.2010.541499

Thompson, W., Black, J., Jain, A., & Kadane, J. (2017). *Forensic science assessments: a quality

and gap analysis–latent fingerprint examination*. Washington, DC: American Association

for the Advancement of Science.

Thompson, W. C., Vuille, J., Taroni, F., & Biedermann, A. (2018). After uniqueness: The evidence

of forensic science opinions. *Judicature*, *102*, 18-27.

Thompson, M. B., Tangen, J. M., & McCarthy, D. J. (2014). Human matching performance of

genuine crime scene latent fingerprints. *Law and Human Behavior*, *38*, 84-93.

doi:10.1037/lhb0000051

Ulery, B. T., Hicklin, R. A., Buscaglia, J., & Roberts, M. A. (2011). Accuracy and reliability of

forensic latent fingerprint decisions. *Proceedings of the National Academy of Sciences*,

108, 7733-7738. doi:10.1073/pnas.1018707108

van den Eeden, C. A. J., de Poot, C. J., & van Koppen, P. J. (2016). Forensic expectations:

Investigating a crime scene with prior information. *Science & Justice*, *56*, 475-481.

doi:10.1016/j.scijus.2016.08.003

van den Eeden, C. A. J., de Poot, C. J., & van Koppen, P. J. (2019). The forensic confirmation bias:

A comparison between experts and novices. *Journal of Forensic Sciences*, *64*, 120-126.

doi:10.1111/1556-4029.13817

Walther, E., Fiedler, K., & Nickel, S. (2003). The influence of prior knowledge on constructive

biases. *Swiss Journal of Psychology*, *62*, 219-231. doi:10.1024/1421-0185.62.4.219

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social

Psychology*, *14*, 89-103. doi:10.1111/j.1559-1816.1984.tb02223.x

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., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness

nonidentifications. *Psychological Bulletin*, *88*, 776-784. doi:10.1037/0033-2909.88.3.776

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative

value. *Psychological Science in the Public Interest*, *7*, 45-75. doi:10.1111/j.1529-

1006.2006.00027.x

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., Small, M., Penrod, S., Malpass, R. S., Fulero, S., & Brimacombe, C. A. E. (1998).

Eyewitness identification procedures: Recommendations for lineups and photospreads.

Law and Human Behavior, *22*, 603-647. doi:10.1023/A:1025750605807

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., & Turtle, J. W. (1986). Eyewitness identification: The importance of lineup models.

      Psychological Bulletin, *99*, 320-329. doi:10.1037/0033-2909.99.3.320

Wells, G. L., Wilford, M. M., & Smalarz, L. (2013). Forensic science testing: The forensic filler-

      control method for controlling contextual bias, estimating error rates, and calibrating

      analysts' reports. *Journal of Applied Research in Memory and Cognition*, *2*, 53-55.

      doi:10.1016/j.jarmac.2013.01.004

Wetmore, S. A., Neuschatz, J. S., Gronlund, S. D., Wooten, A., Goodsell, C. A., & Carlson, C. A.

      (2015). Effect of retention interval on showup and lineup performance. *Journal of

      Applied Research in Memory and Cognition*, *4*, 8-14. doi:10.1016/j.jarmac.2014.07.003

Whitman, G., & Koppl, R. (2010). Rational bias in forensic science. *Law, Probability, & Risk*, *9*,

   69-90. doi:10.1093/lpr/mgp028

Table 1

Results of 172 Identification Decisions (129 MP and 43 MA) by Examiners (N = 43)

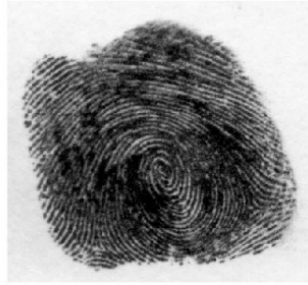|		Identification Decision				
		Hit	False ID	Filler ID	Non-ID	Inconclusive
Showup	MP	57.8% (26)	N/A	N/A	24.4% (11)	17.8% (8)
	MA	N/A	0	N/A	100% (15)	0
Lineup	MP	42.9% (36)	N/A	1.2% (1)	13.1% (11)	42.9% (36)
	MA	N/A	3.6% (1)	3.6% (1)	92.9% (26)	0

Note. MP = Match-Present; MA = Match-Absent.

Figure 1

Examples of a crime scene mark, matching print, and filler print used in the current study
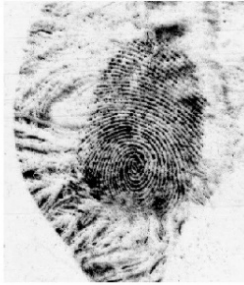
Crime scene mark   Matching print   Filler print

Note. The resolution of these images was much greater in our study; we have lowered the

resolution here in order to protect these individuals' anonymity.

Footnotes

[1] Because we used random assignment, we sent a comparable number of recruitment e-mails containing either of the two study links (i.e., showup or lineup). These e-mails were otherwise identical in content, and each examiner received only one of the two links. The unusual difference in sample size between the showup and lineup conditions is due to a higher rate of volunteering among those who received the link to the lineup condition; we can think of no other explanation for this discrepancy aside from random chance.

[2] If an eyewitness makes multiple positive identifications from a sequential lineup, they are sometimes allowed to resolve these multiple selections by viewing the photos again (Wells et al., 2015). We did not give examiners this option. On two occasions (one match-present and one match-absent), an examiner identified the suspect's print *and* one or more of the filler prints from the same lineup as a match, and we coded these as filler identifications (see Table 1). Coding these two judgments as 'inconclusive' rather than as filler identifications would not change the observed pattern of results.