ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research

Article submitted to journal

Subject Areas:
behaviour, cognition, psychology

Keywords:
eyewitness identification, US lineup, UK lineup, simultaneous lineup, sequential lineup

Author for correspondence:
L. Mickes
e-mail: laura.mickes@rhul.ac.uk

US Lineups Outperform UK Lineups

Travis M. Seale-Carlisle[1], Laura Mickes[1,2]

[1]Royal Holloway, University of London
[2]University of California, San Diego

In the US and the UK, many thousands of police suspects are identified by eyewitnesses every year. Unfortunately, many of those suspects are innocent, which becomes evident when they are exonerated by DNA testing, often after having been imprisoned for years [1]. It is therefore imperative to use identification procedures that best enable eyewitnesses to discriminate innocent from guilty suspects. Although police investigators in both countries often administer lineup procedures, the details of how lineups are presented are quite different and an important direct comparison has yet to be conducted. We investigated whether these two lineup procedures differ in terms of 1) discriminability (using receiver operating characteristic analysis) and 2) reliability (using confidence-accuracy characteristic analysis). A total of 2249 participants watched a video of a crime and were later tested using either a 6-person simultaneous photo lineup procedure (US) or a 9-person sequential video lineup procedure (UK). US lineup procedure yielded significantly higher discriminability and significantly higher reliability. The results do not pinpoint the reason for the observed difference between the two procedures, but they do suggest that there is much room for improvement with the UK lineup.

THE ROYAL SOCIETY PUBLISHING

1. Introduction

The US and the UK are like-minded nations with similarities that extend well beyond their common language. An institutional similarity is in legal systems (e.g., both nations operate under common law) and a cultural similarity is in crime rates [2]. Yet another similarity is that police investigators in both nations often administer a lineup procedure to an eyewitness during the course of a criminal investigation. However, the details of how lineups are presented in the two nations are quite different, and our goal was to determine if the diagnostic accuracy of the US lineup procedure differs from that of the UK lineup procedure.

A lineup consists of the police suspect, who is either innocent or guilty, and several other individuals, or fillers, who resemble the suspect and are known innocents. Although the US and UK lineup procedures share those general characteristics, they differ in several respects, including the number of lineup members presented (typically 6 in the US vs. 9 in the UK), the presentation of the lineup members' images (typically photographs in the US vs. video presentations in the UK), and the procedure used to present the lineup (simultaneous or sequential presentation of lineup members). In the US, the lineup procedure varies from jurisdiction to jurisdiction, but the most common procedure involves the simultaneous presentation of 6 static photographs, with each photograph showing a front view of a face. In the UK lineup procedure (standardised across England and Wales), videos of nine lineup members are sequentially presented, and each video shows an individual facing forward, and then turning to each side for profile views. Witnesses watch two rounds of each video before making a decision [3].

One difference from the current study and current practices is that statements of confidence are not routinely taken in the UK [3], but are often taken in the US [4]. We collected confidence (in accordance with a recommendation of the recent National Academy of Sciences committee on the state of eyewitness identification research [5]) and made use of that information to determine which procedure yields higher diagnostic accuracy.

In the US, sequential lineups were long thought to be superior in terms of discriminability (the ability of eyewitnesses to distinguish between innocent and guilty suspects) to simultaneous lineups because they often yield a lower false ID rate, a marginally lower correct ID rate, and, critically, a higher diagnosticity ratio (correct ID rate / false ID rate) [6–8]. The sequential superiority claim has resulted in up to 30% of US law enforcement agencies to change from the simultaneous to the sequential lineup [4]. Recently, however, it came to light that receiver operating characteristic (ROC) analysis is a more appropriate strategy when the goal is to measure discriminability [5,9–13] but see [14]. The fact that the diagnosticity ratio, a likelihood ratio, does not purely measure discriminability is fairly new to the field of eyewitness identification research, but has been known for decades in other applied fields, such as in diagnostic medicine [15]. When ROC analysis is used, the simultaneous lineup has often been found to outperform the sequential lineup [16–19].

Results of ROC analysis are important for policymakers deciding which type of lineup to use [16]. However, once a criminal case reaches a court of law, regardless of the procedure that was used during the investigation, and regardless of whether one procedure is shown to have greater discriminability than the other, judges and jurors need to know if identifications during the initial lineup procedure are reliable. That is, they need to know the positive predictive value of a suspect ID made with a particular level of confidence. ROC analysis does not provide that answer, but an analysis of the confidence-accuracy relationship does.

To measure this relationship, data are typically analysed using calibration analysis or confidence accuracy characteristic (CAC) analysis. There is consistently a strong confidence-accuracy relationship for individuals who make an identification from a lineup in the lab [20–23], and in the field [24,25]. Calibration analyses often involve plotting accuracy for those who identify suspects or fillers [26], but CAC analysis most directly supplies the answer to the question that judges and juries have about a testifying eyewitness who has identified a suspect: how accurate is that suspect identification likely to be given the level of confidence that was expressed?

Two experiments were conducted to compare discriminability and reliability of US and UK lineups, and they differed only slightly with regard to the UK lineup condition. In one of the experiments (but not the other), after lapping through the lineup twice, participants in the UK condition had the opportunity to view as many lineup members as often as desired before making their decision [3]. Because there were no important differences in the results, we combined the data and present them together (and present the frequency counts separately in Table 1). We report the results of both ROC analysis, which evaluates the level of discriminability supported by the US and UK lineup procedures, and CAC analysis, which measures the confidence-accuracy relationship associated with suspect IDs for the two procedures.

2. Material and methods

(a) Participants

Participants, undergraduate students from the University of California, San Diego, completed the experiment in exchange for course credit ($N = 2249$; 1551 female, 681 male, and 17 did not state; age in years: $M = 20.62$; $sd = 2.80$; ethnicity: Asian 56%, White 19%, Hispanic 15%, Black 1%, Other 6%, and did not state 2%). Participants were randomly assigned to the US lineup or UK lineup condition, and to a target-present lineup or a target-absent lineup. We determined that a sample size of 1000 (for both Experiment 1a and 1b) would yield sufficient power to detect an effect size as large as the one observed in previous research for simultaneous vs. sequential lineups [18]. Data collection continued until the term ended.

(b) Materials

(i) Video

In a 20 s video of a mock crime, a young White male stole several items from a vacated office. The front of the offender's face was clearly shown for 8 s.

(ii) Lineup Construction

An experienced London Metropolitan Police Officer with specialised training in eyewitness identification procedures filmed the actor according to PACE code specifications [3]. The Officer also selected nine fillers based on PACE code guidelines from the PROMAT database (the database used by the London Metropolitan Police Force for constructing lineups). No specific filler was designated as the innocent suspect in target-absent lineups. For the US lineup, 5 or 6 of the 9 fillers were randomly selected for target-present or target-absent lineups, respectively. For the UK lineup, 8 of the fillers were randomly selected for target-present lineups and all of the 9 fillers were used in target-absent lineups. Positions of the lineup members were randomly set for each participant. The same stimuli were used for both the UK and US lineup procedures.

(c) Procedure

The experiment took place online. After digitally consenting, participants entered demographic information (age, ethnicity, education level), watched the video, played a 5-minute game of Tetris as a distractor task, and then were tested on a lineup. They chose someone, or rejected the lineup, from a US or UK lineup (that was target-absent or target-present), rated their confidence on a 100-point scale (0 = just guessing and 100 = absolutely certain). They then answered several multiple-choice questions about the video, including a validation question ("What crime was committed?"), and were debriefed.

Because the experiment took place online there was no administrator influence. The experiment was programmed so that reloading, pausing, and returning to a previous page was disabled; and participants were prevented from participating more than once.

(i) US Lineup Presentation

In the US lineup condition, photographs of the front view of six lineup members (that were still images of the videos) were presented in a 3×2 matrix. The target's and fillers' positions were randomly determined for each participant.

(ii) UK Lineup Presentation

In the UK lineup condition, videos of nine lineup members were presented in sequential order that lapped through twice. The order for both laps was the same for each participant, but the target's and fillers' positions were randomly determined for each participant. The lineup took approximately 6 minutes (depending on internet connection speed) to complete.

3. Results

Participants who incorrectly answered the validation question were excluded from all analyses ($n = 44$). Of the 2205 remaining, 571 were in the US target-present condition, 554 were in the UK target-present condition, 577 were in the US target-absent condition, and 503 were in the UK target-absent condition. Response frequencies for every level of confidence are in Table 1.

Table 1. Frequency counts of suspect IDs, filler IDs, no IDs for target-present and target-absent lineups for every level of confidence for Experiment 1a and Experiment 1b. Note. ID = identification, SIDs = suspect IDs, FIDs = filler IDs

	US Condition					**UK Condition**				
	Target-present			**Target-absent**		**Target-present**			**Target-absent**	
Confidence	SIDs	FIDs	No IDs	FIDs	No IDs	SIDs	FIDs	No IDs	FIDs	No IDs
	Experiment 1A									
0	0	4	1	5	6	0	3	10	7	7
10	1	0	0	4	0	0	3	3	1	1
20	2	3	4	4	4	1	5	2	5	2
30	7	10	6	12	6	1	17	4	6	2
40	10	9	5	14	8	3	15	3	12	5
50	15	13	7	25	11	3	19	12	18	8
60	10	12	8	28	11	5	19	6	17	8
70	17	12	18	27	18	11	27	12	21	10
80	17	12	12	12	13	5	13	7	20	10
90	12	5	12	5	19	4	5	10	11	6
100	6	0	5	2	14	7	4	8	9	5
	Experiment 1B									
0	1	3	1	6	2	3	7	6	8	4
10	1	3	3	4	0	0	2	2	1	1
20	5	6	4	4	1	1	5	2	7	1
30	5	10	6	13	12	2	8	10	21	4
40	13	7	5	27	11	5	13	4	14	1
50	12	26	16	26	19	6	19	12	33	8
60	21	11	14	22	14	9	20	8	31	11
70	28	14	18	38	31	16	26	17	44	13
80	16	10	13	19	29	13	28	15	36	19
90	13	4	11	12	16	8	14	10	22	11
100	8	0	8	6	17	7	7	2	9	13

Table 2. Suspect IDs, Filler IDs and No IDs for target-present and target-absent lineups rates by level of confidence per condition. Note. ID = identification, SIDs = suspect IDs, FIDs = filler IDs

		US Condition			UK Condition		
	Confidence	SIDs	FIDs	No IDs	SIDs	FIDs	No IDs
Target-present	0	**0.39**	0.30		**0.20**	0.50	
	10	0.38	0.29		0.19	0.49	
	20	0.38	0.29		0.19	0.48	
	30	0.37	0.27		0.19	0.46	
	40	0.35	0.24		0.18	0.41	
	50	0.31	0.21	0.31	0.17	0.36	0.30
	60	0.26	0.14		0.15	0.29	
	70	0.20	0.10		0.13	0.22	
	80	0.13	0.05		0.08	0.13	
	90	0.07	0.02		0.05	0.05	
	100	0.02	0.00		0.03	0.02	
Target-absent	0	**0.09**	0.45		**0.08**	0.62	
	10	0.09	0.44		0.07	0.60	
	20	0.09	0.43		0.07	0.59	
	30	0.08	0.42		0.07	0.57	
	40	0.08	0.38		0.07	0.52	
	50	0.06	0.32	0.45	0.06	0.48	0.30
	60	0.05	0.25		0.05	0.39	
	70	0.03	0.17		0.04	0.30	
	80	0.02	0.08		0.02	0.19	
	90	0.01	0.04		0.01	0.09	
	100	0.00	0.01		0.00	0.03	

(a) ROC Analysis

The overall correct ID rate is the number of suspect IDs from target-present lineups divided by the number of target-present lineups presented. The overall false ID rate is the number of estimated suspect IDs from target-absent lineups divided by the number of target-absent lineups presented. Because there is no actual innocent suspect in a lab study, there are several different approaches to computing the false ID rate. First, one filler in a target-absent lineup can be randomly designated to serve as the innocent suspect (such that any ID of that suspect would count as an innocent suspect ID). Second, the filler who is most often misidentified can be designated as the innocent suspect. A third, now standard, practice [27] estimates the number of innocent suspect IDs from the number of filler IDs from target-absent lineups. This estimate is obtained by dividing the number of filler IDs by the number of lineup members (six for the US lineup and nine for the UK lineup). That estimated value is then divided by the number of target-absent lineups to estimate the false ID rate. All three approaches yielded the same conclusions in our study, so we report the results using the third method.

The suspect ID rates for target-present lineups (i.e., correct ID rates), suspect ID rates for target-absent lineups (i.e., false ID rates), and filler ID rates for both target-present and target-absent lineups are shown in Table 2. The red values were used to construct the ROC curves in Fig. 1. The bolded red values are the overall correct and false ID rates that have been traditionally analysed in an effort to determine lineup superiority. However, because both the correct ID rate and the false ID rate are lower for the UK procedure (and could therefore mean a shift in responding, not a difference in discriminability), an analysis of the full ROC provides a clearer picture of the discriminability associated with the two procedures.

Fig. 1 shows the ROC curves for both US and UK conditions, and it is apparent that those in the US lineup condition discriminated innocent from guilty suspects better than those in the UK lineup condition. Partial areas under the curve (pAUC) values were computed using a false ID cut-off of .078 (the rightmost point for the UK lineup) with the statistical package pROC [28]. The pAUC for US lineup condition (.017) was significantly greater than the pAUC for UK lineup condition (.010), $D = 2.74$, $p = .006$. Note that using the rightmost point for the US lineup did not change the conclusion ($p = .002$).

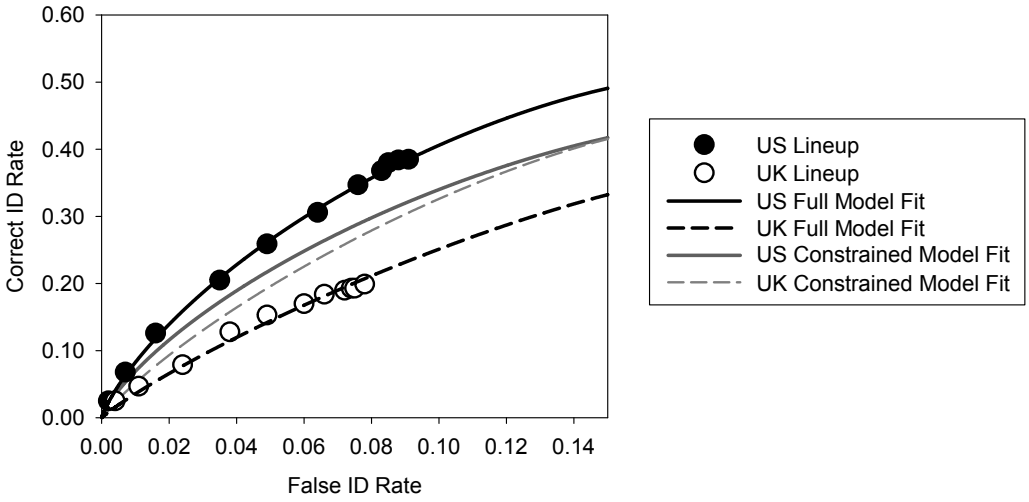

Figure 1. ROC data and curve fits for the US and UK lineup conditions. The solid line represents the fit of the signal detection model and the solid grey line represents the fit of the signal detection model to the US lineup data when d' was constrained to be equal for both conditions. The dashed black line represents the fit of the full signal detection model and the dashed grey line represents the fit of the signal detection model to the UK lineup data when d' was constrained to be equal for both conditions.

(b) Model Fits

It was recently argued that the results of ROC analysis based on an atheoretical measure like pAUC need not agree with results based on a theoretical measure like d' obtained by fitting a theoretical model to the same data [29]. Although it is theoretically possible for the two approaches to yield different conclusions about which procedure is diagnostically superior, in practice, this is likely to rarely occur. Here, we fit a theoretical model to the US and UK ROC data and find that, as expected, the atheoretical pAUC analysis and theoretical signal detection analysis agree on which procedure yields higher discriminability.

An equal variance signal detection model was fit to the data. In the model, memory strength values for innocent suspects (and fillers) and guilty suspects are distributed in two Gaussian, lure and target distributions, respectively, along a memory strength axis. The lure distribution is set to $\mu_{\text{lure}} = 0$, $\sigma_{\text{lure}} = 1$, and the corresponding mean for the target distribution (μ_{target}, which is the same as d' for the equal-variance model we used) was estimated by fitting the model to the ROC data. A fair target-absent lineup is conceptualized as 6 or 9 random draws (for US or UK target-absent lineups, respectively) from the lure distribution. A target-present lineup is conceptualized as 5 or 8 random draws (for US or UK target-present lineups, respectively) from the lure distribution and one random draw from the target distribution (for US and UK target-present lineups).

To keep the number of parameters down, the observed correct and false IDs were binned into low (ratings from 0-60), medium (ratings from 70-80) and high (ratings from 90-100) confidence levels, and were treated as different decision criteria. The full model estimates values of d', variance (fixed to 1 for both lures and targets), and low, medium, and high levels of criteria (c1, c2 and c3, respectively). We also found it necessary to include another parameter δ, such that c1, c2 and c3 represent the confidence criteria for target-absent lineups, and those values divided by δ represent the confidence criteria for target-present lineups. Allowing the criteria to differ in this way is mathematically equivalent to keeping them fixed and instead allowing the standard deviation of the filler distribution to differ for target-present and target-absent lineups (perhaps because fillers are processed differently when there is a familiar target in the lineup). No conclusions would change if this parameter was omitted from the analysis, but the overall fits would be worse.

The 8 parameters were adjusted until the difference between the observed and predicted values was minimised using a chi-square goodness-of-fit statistic. The fit was very good: The minimum chi-square goodness-of-fit statistic was not significant, $\chi^2(8) = 4.93$, $p = .765$. Fig. 1 shows the observed points from the data and the predicted curves generated from the full signal detection model from both conditions and, for comparison, also shows the curves generated from the signal detection model with the d' parameter constrained to be equal in both conditions. This resulted in a much poorer fit, $\chi^2(7) = 31.30$, $p = .001$, that is significantly worse than when the d' parameter is free to vary, $p < .001$. The fact that the fit is significantly worse when d' is constrained means that d' for the US and UK procedures differ significantly (in agreement with the results from pAUC analysis).

(c) Comparing Discriminability of Repeated Viewings

We measured whether discriminability for those participants who opted to view lineup members again (n = 128) differed from those who did not. To do so, we computed d' from the overall correct and false ID rates and compared them using the G statistic. We used this approach instead of ROC analysis because separating the data in this manner resulted in too few observations to perform a meaningful pAUC analysis [30]. Those who viewed lineup members more than the required two times had lower discriminability ($d' = .35$) than those who viewed the lineup members twice ($d' = .68$), but the difference was not significant, $G = 1.14$, $p = 0.253$.

(d) Analysis of the Confidence-Accuracy Relationship

Fig. 2 shows the CAC curves for the US and UK lineup conditions. The CAC dependent variable = (number of of correct suspect IDs) / (number of correct suspect IDs + number of incorrect suspect IDs) for every level of confidence, where incorrect suspect IDs refers to the estimated innocent suspect IDs obtained in the manner previously mentioned. Levels of confidence were collapsed into three bins because there were too few responses in certain bins (the same bins that were used for model fitting). The dependent measure is the positive predictive value (PPV), that is, it is the probability that a suspect who was identified by a witness truly is the perpetrator. Consistent with recent findings, PPV for both conditions increased with confidence [20]. PPV for the UK lineup condition was lower for each level of confidence than for the US lineup condition. The non-overlapping standard errors in Fig. 2 indicate that the suspect ID accuracy scores for the US condition were reliably higher than the corresponding values from the UK condition at each level of confidence.

Note that the PPV values in Fig. 2 correspond to the approximately 50% base rate of target-present lineups used in this experiment. In the real world, the base rate of target-present lineups is unknown. The PPV values in Fig. 2 would be higher for base rates greater than 50% and lower for base rates less than 50%, but the relative standing of the two procedures would not change so long as the base rates were the same for both procedures.

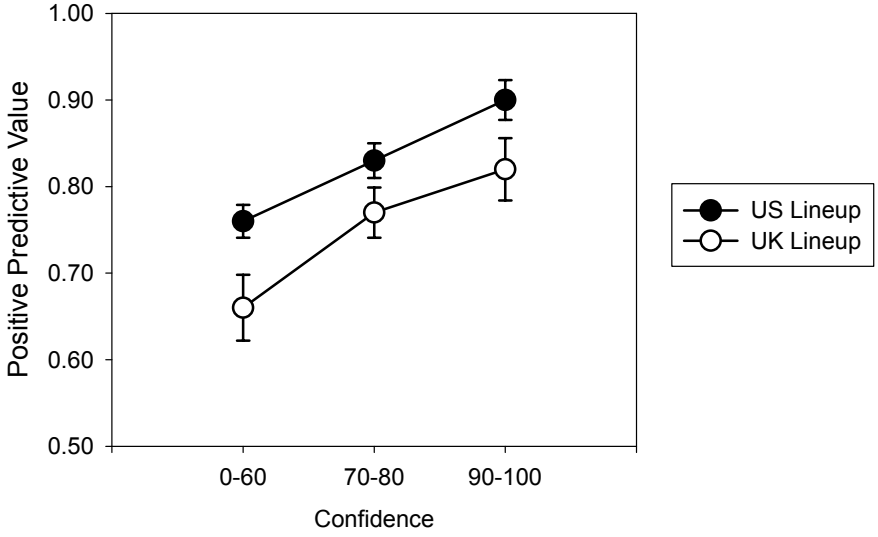

Figure 2. CAC plots for the US and UK lineup conditions. Bars represent standard errors bars estimated using a bootstrap procedure.

(e) Computing CAC Standard Errors

The standard errors associated with suspect ID accuracy scores cannot be directly computed and were therefore estimated using a 10,000-trial bootstrap procedure. On each trial, the observed data from target-present lineups were randomly sampled with replacement to obtain a bootstrap sample of suspect IDs for that trial. For example, for the observed TP data in the US condition, there were 150 high-confidence suspect IDs out of 500 lineups, so the observed high-confidence suspect ID hit rate = $150/500$ = .30. Thus, on each bootstrap trial, a high-confidence suspect ID was registered with probability .30 for each of 500 lineups (i.e., a high-confidence suspect ID would be registered approximately every third lineup, on average). The first bootstrap trial might yield 157 suspect IDs, the next bootstrap trial might yield 141 suspect IDs, and so on. Similarly, on each bootstrap trial, the observed data from target-absent lineups were randomly sampled with replacement to obtain a bootstrap sample of filler IDs for that trial. For example, for the observed TA data in the US condition, there were 100 high-confidence filler IDs out of 500 lineups, so the observed high-confidence filler ID hit rate = $100/500$ = .20. Thus, on each bootstrap trial, a high-confidence filler ID was registered with probability .20 for each of 500 lineups (i.e., approximately every fifth lineup yielded a high-confidence filler ID). The first bootstrap trial might yield 94 filler IDs, the next bootstrap trial might yield 101 filler IDs, and so on. After obtaining a bootstrap sample of suspect IDs and filler IDs on a given bootstrap trial, a suspect ID accuracy score was computed in exactly the same manner it was computed for the observed data. Thus, for example, if there were 157 suspect IDs and 94 filler IDs on the first bootstrap trial, then suspect ID accuracy for the first bootstrap trial = $157/(157 + 94/6)$ = .909. Note that the bootstrap sample of 94 filler IDs was divided by lineup size (6) to estimate innocent suspect IDs from target-absent lineups. Similarly, if there were 141 suspect IDs and 101 filler IDs on the second bootstrap trial, then suspect ID accuracy for the second bootstrap trial = $141/(141 + 101/6)$ = .893. This process was repeated for 10,000 bootstrap trials, and the standard deviation of the 10,000 bootstrap suspect ID scores provided the estimated standard error. The same procedure was followed for each confidence level separately in the US condition and for each confidence level separately in the UK condition.

4. Discussion

Lineup procedures used in the US and UK were not developed by scientists and then implemented in the field, but were instead developed by law enforcement agencies who have no objective basis for preferring one procedure to another. The best way to determine which procedure is diagnostically superior is to use ROC analysis [16]. In the first direct comparison of US and UK lineup procedures using ROC and CAC analysis, we found that the US lineup yielded significantly higher discriminability and significantly higher accuracy at each level of confidence.

The two procedures differ in several ways (e.g., dynamic vs. static images, nominal lineup size) so it is not possible to say why the US procedure outperformed the UK procedure. Our findings may be another example of the often-replicated difference between US sequential photo lineups and US simultaneous photo lineups, which generally favour the latter. Alternatively, participants in the UK condition (and with sequential lineups more generally), but not in the US condition (or with simultaneous lineups more generally), may lose attention during the course of the protracted lineup procedure. The UK lineup takes about six minutes before a decision could be made whereas a decision could be made within seconds after presentation of the US lineup. Could this difference, that is inherent to the procedures, explain the results? Retention interval can be construed as time from end of the video presented during the study phase to the average time the target is first presented during the test phase. For the US procedure, it is approximately 5 minutes (i.e., the duration of the distractor task). For the UK procedure, it is approximately 5 minutes in addition to the time to get to position 4 or 5 in the lineup (the average position of the target). Each lineup member's video lasts approximately 15 seconds, thus, on average, the UK procedure would add about an extra minute to the retention interval. Such a small difference in retention interval is unlikely to account for the differences, nonetheless it is a possibility worth considering.

Whatever the reason for the difference, these findings underscore the importance of directly comparing lineup procedures in terms of their ability to discriminate innocent from guilty suspects. Using a lineup constructed by an experienced police officer for one set of stimuli and one exposure duration, we found that the US procedure unambiguously outperformed the UK procedure. Although it seems unlikely that our results are specific to the testing conditions used here, future work should investigate a wide range of stimuli and conditions (to more definitively answer the applied question of which procedure is superior), and it should also investigate the specific source of the difference between the diagnostic accuracy of the two procedures (to facilitate theory development). Given how many innocent and guilty suspects are tested using lineup procedures in both the US and the UK, such work should be an urgent priority.

Ethics statement

The study was approved by the University of California, San Diego ethics committee IRB Project Number: 121186.

Data accessibility

The entire dataset is accessible http://dx.doi.org/10.5061/dryad.mp08g/1

Competing interests

The authors have no conflicts of interest.

Authors' contributions

LM developed the study concept and design, and performed the model fits; TMS-C programmed the experiments; both authors created the stimuli, collected data, conducted data analysis, and wrote the manuscript. Both authors approved the final version of the manuscript for submission.

Acknowledgment

We thank John T. Wixted for discussions about this research. We also thank the London Metropolitan Police Officers who assisted with the materials used in the experiments.

Funding statement

This work was supported in part by the Economic and Social Research Council [ES/L012642/1] to Laura Mickes.

References

1. Innocence Project Understand the causes: the causes of wrongful conviction. New York: Innocence Project; 2015.
   Accessed: 2016-04-18.
   http://www.innocenceproject.org/causes-wrongful-conviction.
2. United Nations Office on Drugs and Crime; 2008.
   Accessed: 2016-04-18.
   http://www.unodc.org/unodc/en/data-and- analysis/Tenth-United-Nations-Survey-on-Crime- Trends-and-the-Operations-of-Criminal-Justice-Systems.html.
3. Police and Criminal Evidence Act (1984) Codes of Practice, Code D; 2011.
   Accessed: 2016-04-18.
   https://www.gov.uk/government/publications/pace-code-d-2011.
4. Police Executive Research Forum. A National Survey of Eyewitness Identification Procedures in Law Enforcement Agencies; 2013.
   http://www.policeforum.org/.
5. National Research Council.
   Identifying the Culprit: Assessing Eyewitness Identification.
   Washington, DC: The National Academies Press; 2014.
   Available from: http://www.nap.edu/catalog/18891/identifying-the-culprit-assessing-eyewitnes
6. Steblay NK, Dysart JE, Wells GL.
   Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion.
   Psych Public Policy Law. 2011;17(1):99–139.
7. Steblay N, Dysart J, Fulero S, Lindsay RCL.
   Eyewitness Accuracy Rates in Sequential and Simultaneous Lineup Presentations: A Meta-Analytic Comparison.
   Law Hum Behav. 2001;25(5):459–473.
   Available from: http://dx.doi.org/10.1023/A:1012888715007.
8. Lindsay RC, Wells GL.
   Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation.
   J Appl Psychol. 1985;70(3):556–564.
9. Wixted JT, Mickes L.
   The Field of Eyewitness Memory Should Abandon Probative Value and Embrace Receiver Operating Characteristic Analysis.
   Perspect Psychol Sci. 2012;7(3):275–278.
   Available from: http://pps.sagepub.com/content/7/3/275.abstract.
10. Wixted JT, Mickes L.
    Evaluating eyewitness identification procedures: ROC analysis and its misconceptions.
    J Appl Res Mem Cogn. 2015;4(4):318 – 323.

Available from: http://www.sciencedirect.com/science/article/pii/S2211368115000510.
11. Wixted JT, Mickes L.
ROC analysis measures objective discriminability for any eyewitness identification procedure.
J Appl Res Mem Cogn. 2015;4(4):329 – 334.
Available from: http://www.sciencedirect.com/science/article/pii/S2211368115000492.
12. Gronlund SD, Wixted JT, Mickes L.
Evaluating Eyewitness Identification Procedures Using Receiver Operating Characteristic Analysis.
Curr Dir Psychol Sci. 2014;23(1):3–10.
Available from: http://cdp.sagepub.com/content/23/1/3.abstract.
13. Wixted JT, Mickes L.
A signal-detection-based diagnostic-feature-detection model of eyewitness identification.
Psychol Rev. 2014;121(2):262–276.
14. Wells GL, Smalarz L, Smith AM.
ROC analysis of lineups does not measure underlying discriminability and has limited value.
J Appl Res Mem Cogn. 2015;4(4):313 – 317.
Available from: http://www.sciencedirect.com/science/article/pii/S2211368115000509.
15. Zweig MH, Campbell G.
Receiver operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine.
Clin Chem. 1993;39(8):561–577.
16. Mickes L.
Receiver operating characteristic analysis and confidence accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory.
J Appl Res Mem Cogn. 2015;4(2):93 – 102.
Available from: http://www.sciencedirect.com/science/article/pii/S2211368115000169.
17. Dobolyi DG, Dodson CS.
Eyewitness confidence in simultaneous and sequential lineups: A criterion shift account for sequential mistaken identification overconfidence.
J Exp Psychol Appl. 2013;19(4):345–357.
18. Mickes L, Flowe HD, Wixted JT.
Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous versus sequential lineups.
J Exp Psychol Appl. 2012;18(4):361–376.
19. Gronlund SD, Carlson CA, Neuschatz JS, Goodsell CA, Wetmore SA, Wooten A, et al.
Showups versus lineups: An evaluation using ROC analysis.
J Appl Res Mem Cogn. 2012;1(4):221 – 228.
Available from: http://www.sciencedirect.com/science/article/pii/S2211368112000927.
20. Wixted JT, Mickes L, Clark SE, Gronlund SD, Roediger III HL.
Initial eyewitness confidence reliably predicts eyewitness identification accuracy.
Am Psychol. 2015;70(6):515–526.
21. Horry R, Palmer MA, Brewer N.
Backloading in the sequential lineup prevents within-lineup criterion shifts that undermine eyewitness identification performance.
J Exp Psychol Appl. 2012;18(4):346–360.
22. Sauer J, Brewer N, Zweck T, Weber N.
The Effect of Retention Interval on the Confidence–Accuracy Relationship for Eyewitness Identification.
Law Hum Behav. 2009;34(4):337–347.
Available from: http://dx.doi.org/10.1007/s10979-009-9192-x.
23. Brewer N, Wells GL.
The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates.
J Exp Psychol Appl. 2006;12(1):11–30.
24. Wixted JT, Mickes L, Dunn JC, Clark SE, Wells W.
Estimating the reliability of eyewitness identifications from police lineups.
Proc Natl Acad Sci. 2016;113(2):304–309.
Available from: http://www.pnas.org/content/113/2/304.abstract.

25. Behrman BW, Davey SL.
Eyewitness Identification in Actual Criminal Cases: An Archival Analysis.
Law Hum Behav. 2001;25(5):475–491.
Available from: http://dx.doi.org/10.1023/A:1012840831846.
26. Juslin P, Olsson N, Winman A.
Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidenceâĂŞaccuracy correlation.
Journal of Experimental Psychology: Learning, Memory, and Cognition. 1996;22(5):1304–1316.
27. Palmer MA, Brewer N, Weber N, Nagesh A.
The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention.
J Exp Psychol Appl. 2013;19(1):55–71.
28. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al.
pROC: an open-source package for R and S+ to analyze and compare ROC curves.
BMC Bioinform. 2011;12(1):1–8.
Available from: http://dx.doi.org/10.1186/1471-2105-12-77.
29. Lampinen JM.
ROC analyses in eyewitness identification research.
J Appl Res Mem Cogn. 2016;5(1):21 – 33.
Available from: http://www.sciencedirect.com/science/article/pii/S2211368115000480.
30. Mickes L, Moreland MB, Clark SE, Wixted JT.
Missing the information needed to perform ROC analysis? Then compute d', not the diagnosticity ratio.
J Appl Res Mem Cogn. 2014;3(2):58 – 62.
Available from: http://www.sciencedirect.com/science/article/pii/S2211368114000308.