The European Journal of Psychology Applied to Legal Context

A Test of Combi-Lineups: A Lineup Procedure for Multiple Perpetrator Cases in the Netherlands

Melanie Sauerland[1], Micol Iannuzzi[1], Nina Tupper[1], Adri G. van Amelsvoort[2], and Colin G. Tredoux[3]

[1]Maastricht University, Maastricht, The Netherlands; [2]National Facial Biometrics Working Group, The Netherlands; [3]University of Cape Town, Cape Town, South Africa

ARTICLE INFO

ABSTRACT

Background: International best-practice guidelines for lineup administration typically address single perpetrator crimes; few guidelines exist for cases involving multiple perpetrators. In the Netherlands, lineups testing the possible presence of several perpetrators are administered in a single intermixed sequence. This *combi-lineup* has the advantage that the witness does not know how many people are in the lineup and that it is cognitively less demanding for witnesses. It may also enhance memory performance through cueing effects. *Method:* In a laboratory (N = 158) and an online experiment (N = 213), we compared identification performance and the confidence-accuracy relationship for combi-lineups versus separate standard lineups. Participants viewed a mock-theft video depicting two thieves, and completed lineup tests for both. They either made identification decisions from a 12-person combi-lineup or from two separate 6-person lineups. *Results:* Discriminability, confidence-accuracy characteristic curves, and Bayesian information gain curves as a function of target presence base rates were very similar across both procedures. *Conclusion:* These findings provide evidence that the combi-lineup leads to results very similar to those obtained with multiple, separate standard lineups.

Puesta a prueba de ruedas de reconocimiento combinadas: un procedimiento de ruedas de reconocimiento para los casos de múltiples sospechosos en los Países Bajos

RESUMEN

Antecedentes: Las directrices internacionales de buenas prácticas relativas a las ruedas de reconocimiento normalmente se centran en los delitos con un único sospechoso. Hay pocas normas referidas a múltiples sospechosos. En los Países Bajos, las ruedas de reconocimiento con varios sospechosos se administran en una única rueda con todos ellos. Esta rueda combinada tiene las ventajas de que el testigo no sabe cuántos sospechosos están presentes en la rueda y de que es menos exigente cognitivamente para los testigos. También puede mejorar el rendimiento de la memoria gracias a los efectos de las señales. *Método:* En un experimento de laboratorio (N = 158) y en un experimento online (N = 213) comparamos el desempeño en la identificación en ruedas combinadas en comparación con ruedas de un único sospechoso. Los participantes vieron el vídeo de un robo simulado con dos autores y y se les solicitó la identificación de ambos, bien en una rueda combinada de 12 miembros con los dos sospechosos, bien en 2 ruedas distintas de 6 miembros con un sospechoso cada una. *Resultados:* Los resultados mostraron que la capacidad de discriminación, las curvas características de confianza-exactitud y las curvas bayesianas de ganancia de información como función de las tasas base de presencia del estímulo objetivo (sospechosos) eran muy parecidas en ambos tipos de ruedas. *Conclusiones:* Los resultados demuestran que las ruedas combinadas tienen resultados muy parecidos a los obtenidos con ruedas separadas para cada sospechoso.

Many crimes are committed by more than one perpetrator at a time. While international best practice guidelines for lineups specify the procedure for lineup construction and administration (e.g., Wells et al., 2020), those guidelines are usually silent about lineup procedures in cases with multiple perpetrators. Dutch lineup guidelines do address multiple perpetrator cases, though, and recommend a mixed-lineup procedure (van Amelsvoort, 2023). During this so-called combi-lineup, multiple suspects and foils are presented in one intermixed sequence. Although Dutch police have used combi-lineups since 1999 (van Amelsvoort, 1999), implications for eyewitness performance have not been tested empirically. In this study, we compared lineup performance across combi-lineups with two perpetrators and two separate lineups, in two experiments. We also assessed the confidence-accuracy relationship in each type of lineup.

Surveys of police investigators and lineup administrators across Europe suggest that between 10 and 47% of crimes are committed by more than one perpetrator (Hobson et al., 2012; Tupper et al., 2019). When asked to indicate their practices when faced with a two-perpetrator crime, a third of Dutch, Belgian, and Swedish lineup administrators ($N = 51$) indicated that they would place two suspects into the same lineup (Tupper et al., 2019). The situation in the UK is different: in line with the Police and Criminal Evidence Act (PACE, 1984), 100% of police officers in a UK survey ($N = 29$) reported administering separate lineups in cases of multiple suspects (Hobson et al., 2012). Yet, many officers reported adapting the standard instructions to clarify that each lineup contained only one suspect and that the different lineups tested identification of different perpetrators. In a South-African survey ($N = 75$), many police officers (66%) reported placing multiple suspects together in a single lineup (Nortje et al., 2020). This is in line with the South-African guideline that additional suspects can be added to the parade so long as the suspects are physically similar and more foils are added to the parade, per suspect (Du Toit et al., 1987). These findings illustrate that lineups are in some places already conducted with more than one suspect, even though guidelines based on empirical evidence are missing.

Lineup procedures for multiple perpetrator crimes are also relevant from a memory perspective. Increasing the number of faces during encoding reduces recognition performance (Megreya & Burton, 2006). In eyewitness identification research, witnessing a crime with multiple perpetrators, compared to one perpetrator, affects identification performance for target-present lineups (Clifford & Hollin, 1981; Lockamyeir et al., 2021; Megreya & Bindemann, 2012; Nortje 2018; Yarmey, 1982). Such negative effects can be attributed to divided attention and increased perceptual load during encoding (Bindemann et al., 2005; Megreya & Bindemann, 2012; Murphy & Greene, 2016).

However, there seems to be no effect of multiple perpetrators on lineup rejections (or false alarms) in target-absent conditions (Lockamyeir et al., 2021; Megreya & Bindemann, 2012; Nortje 2018). This could suggest that people who witnessed a multiple perpetrator crime adopt a stricter criterion, increasing the likelihood of a rejection (Nortje, 2018). The negative effect of the presence of multiple perpetrators at a crime scene on hit rates raises the question of whether certain lineup procedures might support lineup performance of people who witnessed a multiple perpetrator crime.

Multiple suspects, if more than one is a perpetrator, may provide cues for each other during lineup administration, and this could assist memory retrieval in a lineup procedure. This is because associative information and binding independent pieces of information is intrinsic to memory functioning (Aue et al., 2012). Recognition rates for a previously studied face are often higher when the face is shown with a previously associated face as a cue, compared to being presented alone or with a different face (Tupper, Sauer et al., 2018, Experiment 2; Watkins et al., 1976; Winograd et al., 1977). For lineups, cueing has been tested in studies by presenting two perpetrator lineups side-by-side (Wells & Pozzulo, 2006), by presenting a photo of an accomplice next to the lineup containing the suspect (Dempsey, 2012), or by presenting separate lineups after another, twice, before allowing lineup decisions (Hobson & Wilcock, 2011).

The success of these attempts was mixed, with an advantage of the cued procedure for target-present but not target-absent lineups in two studies (Dempsey, 2012; Hobson & Wilcock, 2011) or non-significant effects in another (Wells & Pozzulo, 2006). Relevant limitations of cueing effects include that they rely on a strong association between the two stimuli (cf. Tupper, Sauer et al., 2018) and that they might unfold only when memory for the perpetrator is relatively weak (Smith & Vela, 2001). In summary, witnesses of multiple perpetrator crimes may benefit from retrieval support during recognition under certain conditions.

Lineup Construction and Administration in the Netherlands

The Dutch handbook for identification procedures (van Amelsvoort, 2023; Handbook henceforth) details protocols for the construction and administration of lineups in the Netherlands since 1994. For example, according to the Handbook, lineups can include between five and eleven fillers; witnesses may view the lineup simultaneously or sequentially; and lineups can be administered live, or using photos or videos. In practice, sequential photo lineups come into use most frequently, whereas live lineups are rare due to the high costs (for more details on Dutch lineup construction and administration, see Sauerland et al., 2023).

For multiple perpetrator lineups, the Handbook (van Amelsvoort, 2023) proposes a sequential or a simultaneous lineup procedure. Here, we focus on the sequential procedure: witnesses see lineup members referring to several perpetrators intermixed in one large *combi-lineup*. The rationale for this adapted sequential procedure is that the witness should not know the size of the lineup. If witnesses did know, there would be a risk that the witness would use an increasingly lenient decision criterion when approaching the last lineup members. In case of multiple perpetrator crimes, the first of several sequential lineups would reveal the (approximate) number of lineup members in subsequent lineups, increasing the pressure to identify someone as the witness progresses from one lineup to the next (Horry et al., 2012; Lindsay et al., 1991). Additionally, going through multiple separate lineups can put strain on witnesses' cognitive and mental resources, and this can be reduced by using a single, combined procedure.

The Confidence-Accuracy Relationship as a Function of Lineup Procedure

Apart from adjusting lineup procedures or adding control variables, diagnosticity of lineup decisions can be improved by considering confidence judgments. In recent years, consensus has emerged that post-decision confidence and identification accuracy are strongly related for positive identification decisions (Sauerland & Sporer, 2009; Sporer et al., 1995; Weber & Brewer, 2006; Wixted et al., 2016) – under pristine conditions (Wixted & Wells, 2017) and when the confidence judgment was collected immediately following the identification decision (Douglass & Steblay, 2006).

For the Dutch combi-lineup, the question arises whether the specific conditions of this procedure affect the confidence-accuracy relationship. These conditions include witnessing a crime with multiple perpetrators rather than one perpetrator and viewing a somewhat larger rather than a smaller lineup[1]. As we have seen, witnessing a multiple perpetrator crime leads to poorer identification performance. Yet, witnesses may be able to account for such performance decreases when making their confidence judgment (Palmer et al., 2013; Semmler et al., 2018). One experiment comparing the confidence-accuracy relationship for a single vs. multiple-perpetrator crime supports this idea (Lockamyeir et al., 2021): despite a detrimental effect of multiple perpetrators on identification accuracy, confidence was still reliably associated with accuracy and, importantly, high confidence, on average, indicated high accuracy.[2]

Another line of reasoning could be that large lineups might suppress performance and reduce the strength of the confidence-accuracy relationship. To our knowledge, no work to date has tested the effect of lineup size for multiple perpetrator-crimes. For single perpetrator-crimes, effects of lineup size on diagnosticity or the confidence-accuracy relationship are mixed. Some work has not

found differences in discriminability between smaller and larger lineups with ROC curves (Akan et al., 2021, Experiment 1 and 2; Seale-Carlisle et al., 2019; Wooten et al., 2020). Others have found an advantage of 6-person over 2-person lineups under low filler-similarity conditions (Lam & Wixted, 2024, Experiment 1; Shen et al., 2024, Experiment 2). Experiments that reported calibration analyses or confidence-accuracy characteristic curves found no differences as a function of lineup size (Akan et al., 2021) or a small advantage for larger over smaller lineups (Seale-Carlisle et al., 2019; Wooten et al., 2020). Overall, these findings suggest that larger lineup size does not affect lineup performance or the confidence-accuracy relationship. Here, we put the confidence-accuracy relationship for combi- and standard lineups to another test.

The Present Study

In a laboratory experiment (Experiment 1) and an online replication of that experiment (Experiment 2), we assessed identification performance and the confidence-accuracy relationship for a two-perpetrator crime, either with two separate, standard lineups or the Dutch combi-lineup procedure. Participants witnessed a simulated theft perpetrated by two thieves. Half of the participants made two separate lineup decisions from sequential lineups containing one suspect each. The other half of the participants made a decision in a task containing a combi-lineup that presented the suspects and foils from thief lineups in an intermixed order. Based on the idea that in a combi-lineup, one target can provide cues regarding the identity of the other target (Tulving & Thompson, 1973; Tupper, Sauer et al., 2018), we expected better identification performance in the combi-lineup condition than the standard lineup condition. We did not expect differences in the confidence-accuracy relationship between the two procedures.

Method

The experiment received ethical approval by the Ethics Review Committee of the faculty (approval code OZL_231_140_12_2020_S4). The data are available here: https://doi.org/10.17605/OSF.IO/23EUN. We cannot share the lineup images because we do not have permission from the individuals.

Participants

The data collection procedures were in accordance with the Declaration of Helsinki. Power analyses for Generalized Estimating Equations (GEE) rely on estimates of parameters that are not obvious. To inform the required sample sizes, we therefore relied on power analyses for two separate 2 (identification accuracy: accurate vs. inaccurate) x 2 (lineup procedure: combi-lineup vs. standard) chi-squared tests for target-present and target-absent lineups. For a chi-squared test with a moderate effect size phi = .30, p = .05, and power = .80, G*Power (Faul et al., 2007) returns a sample size of n = 88. For two separate chi-squared tests for target-present and target-absent lineups, this means a sample of 2 x 88 = 176 is required.

We started data collection for Experiment 1 in the fall of 2019. Due to the COVID-19 pandemic, data collection was halted in the spring of 2020. We then collected as many participants as possible by the end of the academic year 2020/21. In total, N = 178 people participated in Experiment 1. We excluded 9 participants because they differed in ethnicity from the targets in the stimulus film, so as to reduce other-group bias (Meissner & Brigham, 2001). We further excluded five participants because they previously knew a lineup member, three participants due to technical errors, two participants because their confidence was not recorded through error, and one participant because their phone rang during the lineup administration.

The final Experiment 1 sample consisted of N = 158 participants, (23 men, 135 women; M_{age} = 20.63 years; age range = 17-32 years; SD = 2.7; Mdn = 20). Participants were students (75.9%), had completed higher vocational education (20.3%), university education (7.5%), medium vocational education (0.6%), or provided no answer (3.2%). Participants received either 0.5 participation credits or a 5€ gift voucher.

In Experiment 2, N = 265 people participated. We excluded 22 participants because they differed in origin from the targets in the stimulus film to reduce other-group bias (Meissner & Brigham, 2001). We further excluded 20 participants because they answered both attention check questions incorrectly, four participants because their identification decisions were not recorded, three participants because they previously knew a lineup member, two participants due to technical error during the administration of the experiment, and one participant because they indicated they were distracted in a follow-up question.

The final Experiment 2 sample consisted of N = 213 participants, (78 men, 133 women, 2 non-binary). Participants indicated their age as under 18, 18-24, 25-34, …, 75-84 years. Participants' age ranged from under 18 to 75-84 years, with most participants belonging to the group of 18-24-year-olds. Most participants were students (69.0%). We recruited participants through the participation platform of the faculty, word of mouth, social media (Facebook and Instagram), and SurveyCircle. Student participants received course credit in return for participation; other participants did not receive reimbursement.

Design

We used a 2 (lineup procedure: combi-lineup vs. standard) x 2 (target presence: present vs. absent) between-subject design in both experiments. Participants were randomly assigned to the conditions. In the standard lineup condition, participants made two separate lineup decisions from two sequential lineups. In the combi-lineup condition, participants received one combined sequential lineup that contained the suspects and foils from both lineups in the standard lineup condition in an intermixed order. All combinations of target presence were possible for the two thieves: both target-absent, one of the two target-absent and one target-present, or both target-present.

We created three versions of target-absent and target-present lineups to counterbalance the position of the innocent suspect(s) and perpetrator(s). We also counterbalanced the order of the two thief lineups in the standard lineup condition.

We considered identification decisions as accurate (hits, correct rejections) or inaccurate (filler or replacement selections, false rejections). We also measured participants' post-decision confidence on a scale from 0 to 100 for both lineup decisions.

Materials

Stimulus Film

Participants viewed a 1:37 min stimulus film that depicted a nonviolent multi-perpetrator mock-theft. In the film, a man, the victim, is unloading his car. In the background, two thieves (a man and a woman) gesture towards the victim. Once the victim carries a box inside his home nearby, the male thief follows and distracts him by asking for directions when the victim attempts to walk back to his car. At the same time, the female thief steals various objects from the victim's car and walks away. The two thieves were white, and 25 and 27 years old, respectively. The female thief appeared for 60 seconds on screen with 18 seconds of close-up view shots of her face and the male accomplice appeared for 58 seconds with 28 seconds of close-up shots of his face.

Lineups

Lineups Construction. We constructed the lineups in line with Dutch police protocol (van Amelsvoort, 2018). For each target, we created a person description that included information about age, body shape, hair color, and hair length. Following Dutch police procedure, we presented each target and a selection of possible fillers to four mock-witnesses (two men, two women) who were similar in age and origin to our target population (i.e., students). They were informed that a theft had occurred and received a general perpetrator description prepared by women (for female test observers) or men (for male test observers), respectively (e.g., She is between 22 and 27 years old, has long, blond hair and a normal figure). It was their task to point out anyone who stood out (if anyone), for whatever reason. Test observers pointed out several fillers who were subsequently dropped or replaced. The final selection of six fillers (one serving as replacement for target-absent conditions) was established following four rounds with test observers.

We established the effective lineup size by means of the mock-witness paradigm (Doob & Kirshenbaum, 1973). Independent samples of mock-witnesses (ns between 30 and 38) who had not seen the stimulus event, read a target description similar to the descriptions used for the Dutch police protocol described above and viewed a standard lineup. Mock-witnesses then selected the person from the lineup who matched the description best. E (i.e., the effective lineup size) ranged from E = 3.6 to 4.7 (of a possible 6), thereby marking them as a fair selection (Tredoux, 1998, 1999).

Lineup Composition. The standard lineups included one suspect (innocent or guilty) and five fillers. The combi-lineup included two suspects (innocent or guilty) and ten fillers. Lineup photographs showed each lineup member from the shoulders up in front of a white background; including one portrait photo and a 45° angle profile picture for each lineup member. The two photos of each lineup member were presented simultaneously; the lineup members were presented sequentially.

Lineup Administration

Experiment 1. We administered the lineups in line with Dutch police protocol (van Amelsvoort, 2018). Prior to viewing the lineups, participants were informed that they a) would see a sequential lineup, b) would see the selection once, and c) should immediately indicate if they saw the person who committed the theft. A leaflet emphasized that 1) the lineup administrator did not know the lineup members nor who was suspected of the theft and that the fillers were innocent citizens, 2) it was difficult to recognize a person from a photograph, 3) if in doubt, they should not identify anyone, and 4) it was possible that the person(s) who committed the theft was (were) not present in the selection.

We presented the lineups sequentially, using PowerPoint. Following Dutch procedure, each lineup member appeared on screen for 4 seconds followed by 3 seconds of a black screen. Lineup members were numbered 1-6. Lineups could only be viewed once and could be paused on the black screen. Participants first viewed two test photographs with the same timing as the actual lineup photographs. Just before participants saw the actual lineup, they were reminded that it was possible that the person(s) wanted was (were) not part of the selection and that they should not point out anyone if they were in doubt. These instructions were simultaneously given both orally and in writing, as part of the PowerPoint presentation (Kerstholt et al., 1998).

The lineup administrator documented the number of the selected photograph if participants identified a lineup member. In line with Dutch procedure, the lineup then continued to the last lineup member. Following an identification, lineup administrators asked the participant how confident they were in the decision they had just made on a scale from 0 to 100%. If the participant did not make an identification, administrators asked participants how confident they

were that the target was not present in the lineup at the end of the lineup.

Blinding Administrators. Blinding of administrators was accomplished in two ways. The lineup was presented on two screens – one for the participant and one for the lineup administrator. The administrator's screen was covered during lineup presentation such that only the number of the presented lineup member was visible. Additionally, and to avoid lineup administrators learning the likely position of the target over time, suspect positions varied between three different positions for each target. Lineups were created by the supervisor of the experiment, who was the only person aware of the suspect positions and counterbalancing scheme and who provided the lineup as a blinded file.

Experiment 2. We mirrored the procedures for Dutch police protocol as much as possible in the online testing setting. Prior to the lineup presentation, participants were informed that they would view lineup(s) of possible suspects and that they should fetch something for notetaking, so they could write down the photograph numbers of anyone they might recognize from the theft. Analogous to Experiment 1, participants then viewed two test photographs and were asked to adjust their screen or seating position so that they could see the photographs clearly. Next, participants learnt that they would see sequential lineup(s) and should take note of the referring lineup member number(s) if they saw the person(s) who committed the theft. Before the lineup presentation commenced, participants were warned that the person(s) wanted may not be present and that they should not select anyone if they had doubts. As in Experiment 1, these instructions were simultaneously given both orally and in writing (Kerstholt et al., 1998).

Unlike Experiment 1, participants then viewed *all* lineup members sequentially before indicating any recognition. As in Experiment 1, each lineup member appeared on screen for 4 seconds followed by 3 seconds of a black screen. Lineup members were numbered 1-6 (standard lineups) or 1-12 (combi-lineup), respectively. After participants had viewed all lineup members, they could indicate the lineup member (standard lineup) or members (combi-lineup) they had recognized or indicate that the perpetrator(s) were not present in the lineup. Participants indicated their confidence on a scale from 0-100% following each lineup decision. In the standard lineup condition, participants repeated this procedure twice, once for each lineup.

Procedure

Experiment 1

Participants who took part in Experiment 1 had already participated in a different experiment in the same session. There was a brief break between the two studies. In the first study (not reported here), participants viewed a short crime video and then made identification decisions for a thief and a victim. This study investigated an unrelated research question, namely the post-decision confidence-accuracy relationship for Dutch lineup protocols and is published elsewhere (Sauerland et al., 2023). Here, we report the results of the second study. Participants learned about the experiment through the University's participant recruitment platform (SONA), online advertisements, and flyers at various university faculties. The experiment was advertised as a study where participants would find out what it was like to be a police witness. The advertisement did not mention eyewitness identification specifically.

Participants were tested individually and the identification procedure was video recorded. After signing the written informed consent form, participants viewed the stimulus film. During a 10-minute retention interval, they provided demographic information and completed several unrelated filler tasks on the

computer. Next, they viewed the lineup(s). After each lineup decision, participants indicated their confidence and what role that person had played in the crime. Finally, participants were asked what they had thought the purpose of the experiment was prior to participating. Participants furthermore indicated whether they knew any of the lineup members. Before leaving, participants were asked not to talk about the details of the experiment to avoid bias in future participants. Participants then received the debriefing and reimbursement.

Experiment 2

Participants received a Qualtrics link to complete the experiment online. After providing informed consent, participants were instructed to pay attention to the video they were about to watch, because they would be asked questions about it later. To check if participants had watched the whole video, a still image of a blue star appeared at the end of the video. Participants received the cue to "Pay attention to this blue star" because they would be asked about it later. During a 10-minute retention interval, participants played a few short games (Pacman, Mario Kart, Flappy Bird, visual search tests). Before viewing the lineups, participants answered two attention check questions about the shape and the color of the blue star at the end of the stimulus film. Most participants answered both (n = 241) or one (n = 4) of the two questions correctly. We excluded participants who answered both questions incorrectly (n = 20). Then, participants viewed the lineup(s). After each lineup decision, participants indicated what role that person had played in the crime. Finally, participants provided demographic information and indicated what they had thought the purpose of the experiment was prior to participating and whether they personally knew any of the persons in the lineup(s). The final question was about the environment while completing the experiment (e.g., alone, surrounded by others). Participants then received the debriefing.

Data Analyses

Effect of Lineup Procedure on Lineup Performance

We used GEE to assess the effect of lineup procedure and target presence on lineup performance. GEE allows the specification of a regression model with dichotomous outcome measures and any combination of within- and between-subjects factors as predictors, with the specification of a so-called working correlation matrix (i.e., unstructured) as a nuisance parameter to accommodate the correlated residuals that result from repeated measurements (Hanley et al., 2003). To ensure the generalizability to a broader population and to avoid overfitting of the data, the minimum number of observations per possible outcome (identification accuracy: accurate vs. inaccurate) should be at least 10 times the number of predictors (Peduzzi et al., 1996). Our models included three predictors (two main effects and one interaction term). That means, we needed at least 30 accurate and 30 inaccurate responses per model. All of our GEE models met this condition.

The initial analyses included both main effects (lineup procedure, target presence) and the resulting two-way interaction in the equation. The interaction was non-significant for both experiments, p_{E1} = .991, p_{E2} = .055[3]. We therefore excluded the interaction from the model to assess the two main effects. The exponent of the b value is equal to the partial odds ratio. For binary predictors, this can be interpreted as effect size.

To test the effect of cueing on identification performance, we computed another GEE with effects of target presence and cueing (correct cue vs. incorrect cue) on identification performance. We coded the cue for the female thief as correct if the male thief

was present in the lineup. Likewise, we coded the cue for the male thief as correct if the female thief was present. The initial analyses included both main effects (cueing, target presence) and the resulting two-way interaction in the equation. The interaction was non-significant for both experiments, p_{E1} = .304, p_{E2} = .642. We therefore excluded the interaction from the model to assess the two main effects.

We also constructed the ROC curves for standard and combi-lineups and computed the area under the curve (Gronlund et al., 2014; Mickes et al., 2012; Wixted & Mickes, 2018; Wixted et al., 2017). A ROC curve is constructed by plotting the rate of correct identifications against the rate of false identifications for separate levels of confidence (rated by participants on a scale from 0 to 100, transformed to an 11-point Likert scale). ROC curves require large sample sizes to yield reliable estimates. We therefore collapsed data across both experiments. We deemed this approach justified, given similar patterns of results in identification performance and confidence-accuracy characteristic (CAC) curves.

Effect of Lineup Procedure on Confidence-Accuracy Relationship

We used two techniques for mapping accuracy of participants who made a selection from the lineup (i.e., choosers) across varying levels of confidence: CAC curves and Bayesian information gain curves as a function of target presence base rates. These methods require large sample sizes to yield reliable estimates. Following other work (Dodson & Dobolyi, 2016; Sauer et al., 2010; Sauerland et al., 2025; Tupper et al., 2023), we therefore collapsed data for both targets in these analyses.

Confidence-Accuracy Characteristic Curves. We constructed separate CAC curves for standard and combi-lineups. Innocent suspect selections in target-absent lineups were computed as number of filler selections divided by number of lineup members (Lockamyeir et al., 2021; Palmer et al., 2013; Sauerland et al., 2013). To allow for the construction of stable curves and following earlier work (Flowe et al., 2017; Sauerland et al., 2016; Wixted et al., 2015), we collapsed confidence categories into low, medium, and high confidence ([0, 54%], [55, 84%], [85, 100%]).

Confidence-Accuracy Relationship as a Function of Target Presence Base Rates. To provide insights into the confidence-accuracy relationship for standard vs. combi-lineups as a function of different prior base rates of target presence, we conducted Bayesian information gain analyses (Wells et al., 2015; Wixted & Wells, 2017). Eyewitness identification research often presents 50% of participants with target-present lineups and 50% with target-absent lineups (i.e., a base rate of target presence of 50%). For different legislations and police stations, however, the base rate probability of perpetrator presence is unknown and variable. We constructed Bayesian information gain curves for standard and combi-lineups separately for high and lower levels of confidence ([85, 100%], [0, 84%]).

Results

Effect of Lineup Procedure on Lineup Performance

Table 1 shows an overview of hits, filler selections, and false and correct rejections for the different lineup conditions and Table 2 shows the corresponding identification accuracy rates. In Experiment 1, identification accuracy for standard lineups was 84.5% [78.4, 89.9] and for combi-lineups it was 83.9% [78.0, 89.3]. For target-present lineups, correct performance in both lineup conditions was very similar (standard: 75.3% [66.2, 83.8], combi: 74.4% [65.2, 83.2]). Likewise, for target-absent lineups, correct performance in the two

lineup conditions was very similar (standard: 94.4% [88.3, 98.7], combi: 93.9% [88.2, 98.8]).

Table 1. Frequency (and Proportion, in parenthesis) of Decision Outcomes as a Function of Lineup Procedure and Target Presence (Experiment 1 and 2)

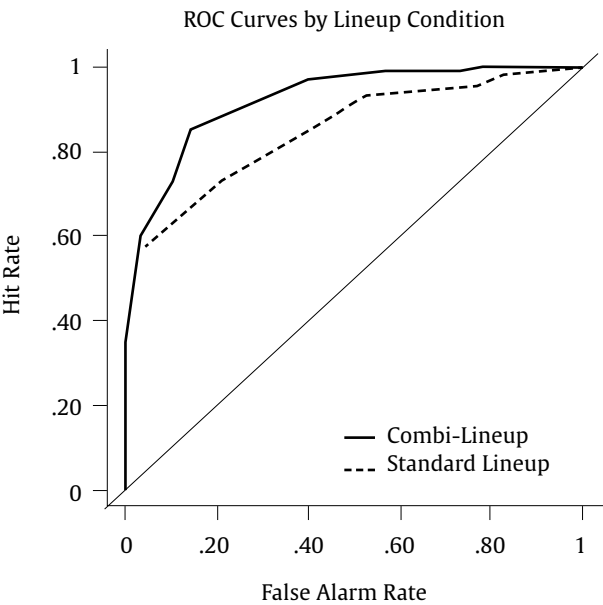|	Standard lineups		Combi-lineups	
Experiment 1				
Target-present lineups				
Hits	58	(.75)	64	(.74)
Filler selections	2	(.03)	0	(0)
False rejections	17	(.22)	22	(.26)
Target-absent lineups				
Correct rejections	67	(.94)	77	(.94)
Filler selections	4	(.06)	5	(.06)
Experiment 2				
Target-present lineups				
Hits	70	(.49)	41	(.55)
Filler selections	48	(.34)	16	(.21)
False rejections	24	(.17)	18	(.24)
Target-absent lineups				
Correct rejections	108	(.78)	48	(.68)
Filler selections	30	(.22)	23	(.32)

In Experiment 2, identification accuracy for standard lineups was 63.6% [57.5, 68.6] and for combi-lineups it was 61.0% [53.4, 69.2]. For target-present lineups, correct performance was 49.3% [41.7, 57.9] for standard lineups and 54.7% [43.5, 65.6] for the combi-lineup. For target-absent lineups, correct performance was 78.3% [71.0, 85.1] for standard lineups and 67.6% [56.5, 77.9] for the combi-lineup.

In Experiment 1, in line with the descriptives, performance did not vary significantly as a function of lineup procedure, Wald $\chi^2(1) = 0.04$, $p = .850$, Exp(b) = 0.94. However, performance did vary as a function of target presence, with better performance in target-absent lineups (94%) than target-present lineups (75%), Wald $\chi^2(1) = 17.68$, $p < .001$, Exp(b) = 5.38.

In Experiment 2, performance again did not vary significantly as a function of lineup procedure, Wald $\chi^2(1) = 0.18$, $p = .671$, Exp(b) = 1.11. Again, performance in target-absent lineups (75%) was superior to performance in target-present lineups (51%), Wald $\chi^2(1) = 25.02$, $p < .001$, Exp(b) = 2.76.

Figure 1 shows the ROC curves for standard and combi-lineups, together with a positive diagonal line that represents chance performance. Data points on this line would correspond to equal rates of correct and false identifications. A ROC curve has greater empirical discriminability when it rises closest to the upper-left area in the ROC graph. The different points that form the ROC curves represent pairs of correct and false identification rates at different levels of confidence. Looking at the ROC curve for the combi-lineup group, the

point located at the upper right of the curve shows the rates of correct and false identifications, determined over all confidence levels taken together. One data point further to the left and we find the rates of correct and false identifications computed over all confidence levels taken together, with the exception of the lowest level of confidence. As we move further towards the left, we ultimately end at the lower left-most point, where only responses that were rated with the highest possible confidence (100%) were included for the computation.

Figure 1. Receiver Operating Characteristic (ROC) Plots for Standard and Combi-lineups Combined for Experiment 1 and 2. The Diagonal Represents Chance Performance.
Note. False alarm rates represent all filler selections in target-absent line ups, i.e., we did not divide by the number of foils.

To assess a possible effect of lineup procedure on identification performance, we calculated the area under the curve for the relevant range of false identification rates. Using a bootstrap analysis via the pROC package in R (Robin et al., 2011), we found the pAUC for the ROC curve to be .85 (95% CI [.779, .906]) for standard lineups, and .92 (95% CI [.867, .967]) for combi-line-ups. DeLong's test indicated that the difference between the two AUCs was not statistically significant, D(290.78) = 1.81, $p = .071$. On this comparison, a post-hoc power analysis ($\alpha = .05$) showed that we had moderate power to detect an effect of this magnitude, 1-β = .77.

Effect of Cueing on Lineup Performance in Combi-Lineups

Table 2. Proportion (and Frequency, in parenthesis) of Correct Lineup Identification Decisions as a Function of Lineup Procedure and Target Presence for Experiment 1 and 2

	Target-Present Lineups		Target-Present Lineups		Across Target Presence	
	Proportion/ Frequency	95% CI	Proportion/ Frequency	95% CI	Proportion/ Frequency	95% CI
Experiment 1						
Standard lineups	.75 (58)	[.66, .84]	.94 (67)	[.88, .99]	.85 (125)	[.78, .90]
Combi-lineups	.74 (64)	[.65, .83]	.94 (77)	[.88, .99]	.84 (141)	[.78, .89]
Across lineup procedure	.75 (122)	[.68, .81]	.94 (144)	[.90, .98]	.84 (266)	[.80, .88]
Experiment 2						
Standard lineups	.49 (70)	[.42, .58]	.78 (108)	[.71, .85]	.64 (178)	[.58, .69]
Combi-lineups	.55 (41)	[.44, .66]	.68 (48)	[.57, .78]	.61 (89)	[.53, .69]
Across lineup procedure	.51 (111)	[.45, .58]	.75 (156)	[.69, .81]	.63 (267)	[.58, .67]

In neither experiment did cueing (accurate vs. inaccurate) have a significant effect on identification performance (Experiment 1: Wald $\chi^2(1)$ = 0.40, p = .526, Exp(b) = 1.34; Experiment 2: Wald $\chi^2(1)$ = 1.16, p = .282, Exp(b) = 1.48). In Experiment 1, the main effect of target presence was significant, as reported above, p = .002, Exp(b) = 5.45. In Experiment 2, the main effect of target presence was non-significant, p = .154, Exp(b) = 1.68. This is in line with the marginally significant interaction between target presence and lineup procedure in Experiment 2 that showed an effect of target presence for standard lineups but not combi-lineups (reported in footnote 2).

Effect of Lineup Procedure on Confidence-Accuracy Relationship

Confidence-Accuracy Characteristic Curves

Figure 2 displays the CAC curves for standard and combi-lineups. For both experiments, CAC curves for the two lineup procedures were very similar at high and moderate confidence levels. Generally, the curves displayed under-confidence, especially in the moderate-confidence bin. Highly confident choosers were highly accurate in both experiments.

Confidence-Accuracy Relationship as a Function of Target Presence Base Rates

To address the possibility that the confidence-accuracy relationship may depend on the target presence base rates, we plotted the probability that a suspect identification from portrait lineups was accurate across all possible target presence base rates (Figure 3). The identity line shows where the data would fall if an identification was non-diagnostic.
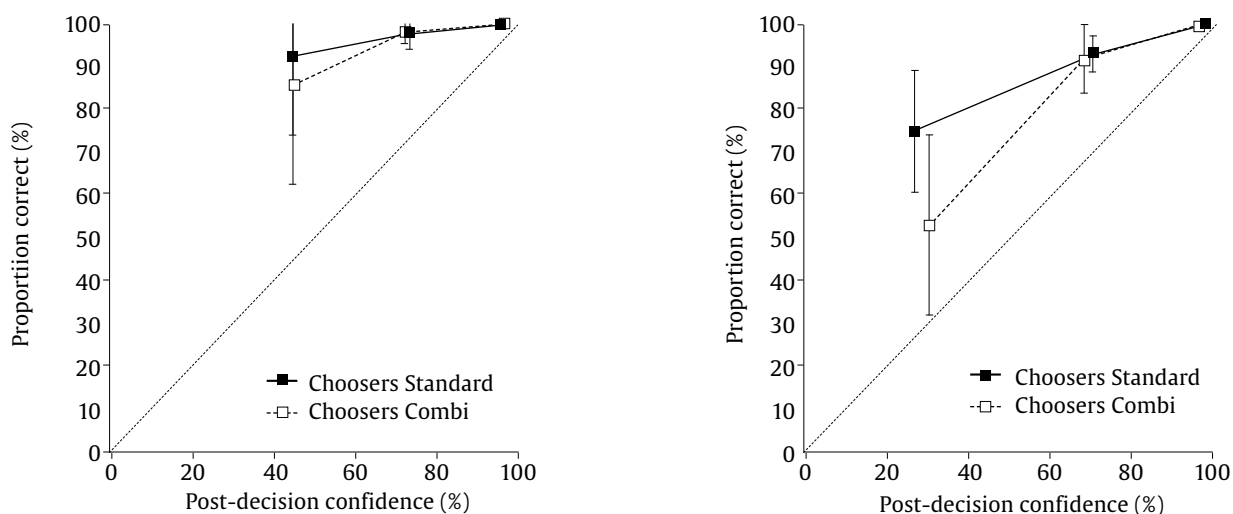
For both experiments, all curves were above the identity line, demonstrating that identifications were diagnostic of guilt. The heights of the curves indicate better discriminability for more confident participants at nearly the whole spectrum of target presence base rates than for less confident participants. The curves of highly confident participants in the two lineup procedure conditions were very similar. Decisions of less confident participants were

somewhat more diagnostic when they were in the combi-lineup condition than in the standard lineup condition in Experiment 1, but more diagnostic when they were in the standard lineup condition in Experiment 2.
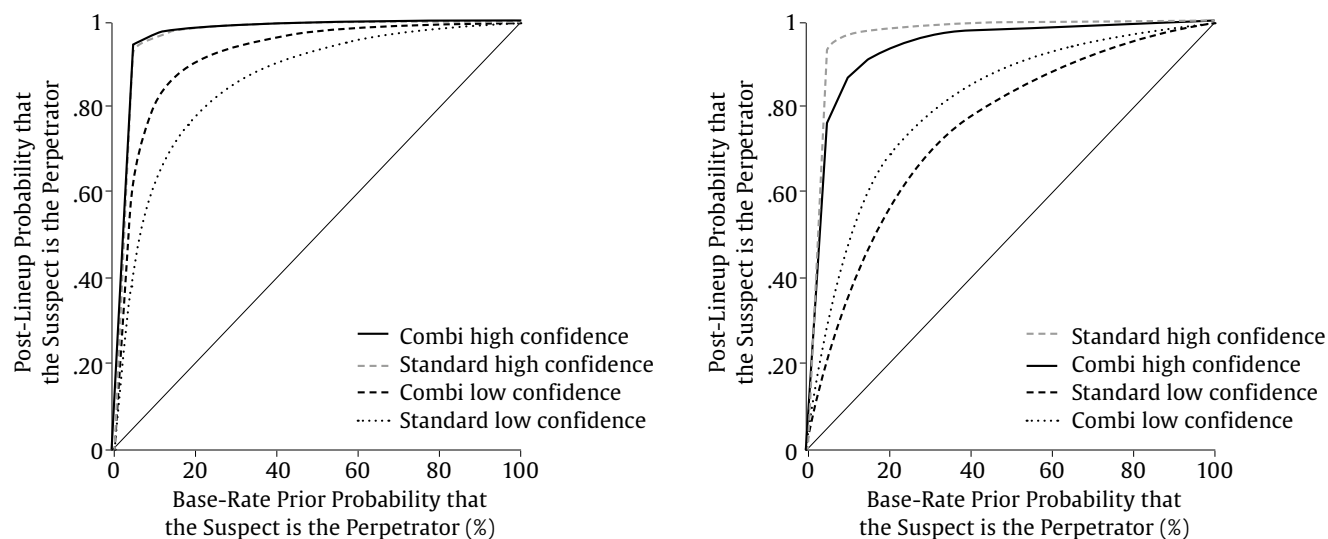
Discussion

Although many crimes are committed by more than one perpetrator, international best practice guidelines do not specify lineup procedures for multiple perpetrator cases. Dutch police have used a mixed-lineup procedure for multiple perpetrator cases since 1999 (van Amelsvoort, 1999). However, it has been unclear whether this combi-lineup procedure affects witness performance or the confidence-accuracy relationship. In two experiments, we compared performance and the confidence-accuracy relationship in combi-lineups and standard separate lineups. Unexpectedly, identification accuracy and discriminability were not significantly higher in combi-lineups than in standard lineups. As expected, the confidence-accuracy relationship was overall similar across procedures.

We expected better identification performance in combi-lineups, compared to standard lineups based on the idea that in a combi-lineup one suspect can serve as a cue to other suspects (Tulving & Thompson, 1973; Tupper, Sauer et al., 2018). Our null findings for both target-present and target-absent lineups are in line with earlier work (Wells & Pozzulo, 2006), but contrary to reported advantages of the cued procedure for target-present but not target-absent lineups (Dempsey, 2012; Hobson & Wilcock, 2011). The literature suggests that cuing effects depend on a strong association between the binding stimuli (Tupper, Sauer et al., 2018) and a weak memory for the perpetrator (Smith & Vela, 2001). Arguably, the association between the two thieves in the stimulus film was strong, as they were planning and carrying out a theft in one video sequence. The very high identification performance rates in Experiment 1 (see Table 2), however, suggest that memory for the perpetrator was strong and that a ceiling effect might have occurred. This could explain the non-significant effects of cueing. On the other hand, identification accuracy was more moderate in Experiment 2, especially for target-present lineups – the condition that showed cueing effects in some earlier work (Dempsey, 2012; Hobson & Wilcock, 2011). In summary, our findings align with a literature of inconsistent success in eliciting

Figure 2. Choosers' Confidence-Accuracy Characteristic Curves for Standard vs. Combi-Lineups for Low Confidence, Medium Confidence, and High Confidence ([0, 54%], [55 - 84%], [85, 100%]) in Experiment 1 (left panel) and Experiment 2 (right panel).

Figure 3. Bayesian Information Gain Curves: Post-Lineup Probability that the Suspect is the Perpetrator on Portrait Lineups for Experiment 1 (left) and Experiment 2 (right) as a Function of the Base Rate of Target Presence and Lineup Procedure.

Note. For some conditions, the number of innocent suspect selections was 0 (Experiment 1: combi-lineup; Experiment 2: standard condition). To be able to construct the curves, we replaced 0 with 0.5.

cueing effects (Dempsey, 2012; Hobson & Wilcock, 2011; Tupper, Sauer et al., 2018; Wells & Pozzulo, 2006) and cast some doubt on the cueing hypothesis.

Both our lineup procedures were designed to protect participants' confidence judgments from any external interference. Based on this, and also on earlier work comparing the confidence-accuracy relationship for a single vs. multiple-perpetrator crimes (Lockamyeir et al., 2021), and on evidence that witnesses can account for performance decreases when making their confidence judgment (Palmer et al., 2013; Semmler et al., 2018), we had no reason to believe that the confidence-accuracy relationship would differ between combi-lineups and standard lineups. Indeed, the CAC curves look very similar (see Figure 2), as do the Bayesian Information Gain Curves (see Figure 3). Overall, our findings suggest that confidence is reliably associated with accuracy in the Dutch combi-lineup procedure.

A limitation to this work is that participants in Experiment 1 participated in another lineup experiment in the same testing session. Potentially, the earlier study could affect the results of the current Experiment 1. Reassuringly, earlier work found that making multiple identification decisions was not harmful to eyewitness memory and had either no effect or a trivial effect on accuracy, choosing, or confidence (Mansour et al., 2017; Tupper, Sauerland et al., 2018). Additionally, the correspondence of the pattern of results in Experiments 1 and 2 gives us confidence in the findings of Experiment 1.

Policy Implications and Future Directions

We present empirical evidence that the Dutch combi-lineup procedure leads to very similar results as multiple, separate standard lineups. In two experiments, the combi-lineup and standard lineup resulted in comparable identification performance and confidence-accuracy relationship. Future studies might investigate how the combi-lineup fares with larger groups of perpetrators. Another motivation for using the combi-lineup is the idea that the combi-lineup might reduce the strain on witnesses, compared to multiple lineups. While such reasoning is intuitive, this has never been studied in an experimental setting. This is another avenue for future research.

To conclude, the current research provides evidence that the Dutch combi-lineup is a viable method for establishing the identity of perpetrators in cases with multiple perpetrators. It leads to similar performance and confidence-accuracy relationship as multiple, separate standard lineups. The procedure could serve as a model for police in other countries when dealing with multiple perpetrator crimes. Future research is to establish whether the combi-lineup reduces the strain on witnesses.

Conflict of Interest

The authors of this article declare no conflict of interest

Notes

[1]The number of lineup members is limited to six persons per suspect in Dutch combi-lineups.

[2]Other experiments that compared multiple and single perpetrator crimes did not collect confidence judgments (Megreya & Bindemann, 2012) or did not report the confidence-accuracy relationship as a function of number of perpetrators (Clifford & Hollin, 1981; Nortje 2018).

[3]We did compute the simple slopes for Experiment 2 (p = .055) for exploratory purposes. Performance did not vary as a function of lineup procedure in either target-present, p = .406, nor target-absent lineups, p = .103. Rather, performance in target-absent lineups (78%) was superior to performance in target-present lineups (49%) for standard lineups, p < .001, but not for combi-lineups, p = .168.

References

Akan, M., Robinson, M. M., Mickes, L., Wixted, J. T., & Benjamin, A. S. (2021). The effect of lineup size on eyewitness identification. *Journal of Experimental Psychology: Applied, 27*(2), 369-392. https://doi.org/10.1037/xap0000340

Aue, W. R., Criss, A. H., & Fischetti, N. W. (2012). Associative information in memory: Evidence from cued recall. *Journal of Memory and Language, 66*(1), 109-122. https://doi.org/10.1016/j.jml.2011.08.002

Bindemann, M., Burton, A. M., & Jenkins, R. (2005). Capacity limits for face processing. *Cognition, 98*(2), 177-197. https://doi.org/10.1016/j.cognition.2004.11.004

Clifford, B. R., & Hollin, C. R. (1981). Effects of the type of incident and the number of perpetrators on eyewitness memory. *Journal of Applied Psychology, 66*(3), 364-370. https://doi.org/10.1037/0021-9010.66.3.364

Dempsey, J. L. (2012). *An investigation of retrieval cues as a method to improve eyewitness identification accuracy of multiple perpetrators.* Doctoral dissertation. Carleton University, Ottawa, Canada.

Dodson, C. S., & Dobolyi, D. G. (2016). Confidence and eyewitness identifications: The cross-race effect, decision time and accuracy. *Applied Cognitive Psychology, 30*(1), 113-125. https://doi.org/10.1002/acp.3178

Doob, A. N., & Kirshenbaum, H. M. (1973). Bias in police lineups — partial remembering. *Journal of Police Science and Administration, 1,* 287-293.

Douglass, A. B., & Steblay, N. (2006). Memory distortion in eyewitnesses: A meta-analysis of the post-identification feedback effect. *Applied Cognitive Psychology, 20*(7), 859-869. https://doi.org/10.1002/acp.1237

Du Toit, E., de Jager, F. J., Paizes, A., Skeen, A., & van der Merwe, S. E. (1987). *Commentary on the Criminal Procedure Act.* Juta and Company (PTY) LTD.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. https://doi.org/10.3758/BF03193146

Flowe, H. D., Colloff, M. F., Karoğlu, N., Zelek, K., Ryder, H., Humphries, J. E., & Takarangi, M. K. (2017). The effects of alcohol intoxication on accuracy and the confidence–accuracy relationship in photographic simultaneous lineups. *Applied Cognitive Psychology, 31*(4), 379-391. https://doi.org/10.1002/acp.3332

Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating eyewitness identification procedures using receiver operating characteristic analysis. *Current Directions in Psychological Science, 23*(1), 3-10. https://doi.org/10.1177/0963721413498891

Hanley, J. A., Negassa, A., Edwardes, M. D. B., & Forrester, J. E. (2003). Statistical analysis of correlated data using generalized estimating equations: An orientation. *American Journal of Epidemiology, 157*(4), 364-375. https://doi.org/10.1093/aje/kwf215

Hobson, Z. J., & Wilcock, R. (2011). Eyewitness Identification of multiple perpetrators. *International Journal of Police Science & Management, 13*(4), 286-296. https://doi.org/10.1350/ijps.2011.13.4.253

Hobson, Z. J., Wilcock, R., & Valentine, T. (2012). Multiple suspect showing: A survey of police identification officers. *Policing, 7*(1), 79-87. https://doi.org/10.1093/police/pas021

Horry, R., Memon, A., Wright, D. B., & Milne, R. (2012). Predictors of eyewitness identification decisions from video lineups in England: A field study. *Law and Human Behavior, 36*(4), 257 -265. https://doi.org/10.1037/h0093959

Kerstholt, J. H., van Amelsvoort, A. G., & Koster, E. R. (1998). *Optimalisatie van daderherkenning met videoconfrontaties: Effecten van instructie en aanbiedingstijd.* (TNO-rapport TM-98-C069) [Optimization of offender recognition with video lineups: Effects of instruction and presentation time]. Ministerie van Justitie Wetenschappelijk Onderzoek- en Documentatiecentrum.

Lam, A. L., & Wixted, J. T. (2024). The effect of lineup size on discriminability is dependent on filler similarity and independent of encoding strength. *Memory & Cognition, 53*(4), 1272-1286. https://doi.org/10.3758/s13421-024-01649-x

Lindsay, R. C., Lea, J. A., & Fulford, J. A. (1991). Sequential lineup presentation: Technique matters. *Journal of Applied Psychology, 76*(5), 741-745. https://doi.org/10.1037/0021-9010.76.5.741

Lockamyeir, R. F., Carlson, C. A., Jones, A. R., Wooten, A. R., Carlson, M. A., & Hemby, J. A. (2021). One perpetrator, two perpetrators: The effect of multiple perpetrators on eyewitness identification. *Applied Cognitive Psychology, 35*(5), 1206-1223. https://doi.org/10.1002/acp.3853

Mansour, J. K., Beaudry, J. L., & Lindsay, R. C. L. (2017). Are multiple-trial experiments appropriate for eyewitness identification studies? Accuracy, choosing, and confidence across trials. *Behavior Research Methods, 49*, 2235–2254. https://doi.org/10.3758/s13428-017-0855-0

Megreya, A. M., & Bindemann, M. (2012). Identification accuracy for single- and double-perpetrator crimes: Does accomplice gender matter? *British Journal of Psychology, 103*(4), 439-453. https://doi.org/10.1111/j.2044-8295.2011.02084.x

Megreya, A. M., & Burton, A. M. (2006). Recognising faces seen alone or with others: When two heads are worse than one. *Applied Cognitive Psychology, 20*, 957-972. https://doi.org/10.1002/acp.1243

Meissner, C. A., & Brigham, J. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, & Law, 7*(1), 3-35. https://doi.org/10.1037/1076-8971.7.1.3

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

Murphy, G., & Greene, C. M. (2016). Perceptual load affects eyewitness accuracy and susceptibility to leading questions. *Frontiers in Psychology, 7,* Article 1322. https://doi.org/10.3389/fpsyg.2016.01322

Nortje, A. (2018). *The butcher, the baker, the candlestick maker: Investigating facial recognition for multiple-perpetrator crimes* (Doctoral dissertation, University of Cape Town). https://doi.org/hdl.handle.net/11427/29506

Nortje, A., Tredoux, C. G., & Vredeveldt, A. (2020). Eyewitness identification of multiple perpetrators. *South African Journal of Criminal Justice, 33*(2), 348-381. Retrieved from https://research.vu.nl/en/publications/eyewitness-identification-of-multiple-perpetrators/ on 27 November 2025

Palmer, M. A., Brewer, N., Weber, N., & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied, 19*(1), 55-71. https://doi.org/10.1037/a0031602

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology, 49*(12), 1373–1379. https://doi.org/10.1016/s0895-4356(96)00236-3

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

Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence–accuracy relationship for eyewitness identification. *Law and Human Behavior, 34*(4), 337-347. https://doi.org/10.1007/s10979-009-9192-x

Sauerland, M., Raymaekers, L. H. C., Otgaar, H., Memon, A., Waltjen, T. T., Nivo, M., Slegers, C., Broers, N. J., & Smeets, T. (2016). Stress, stress-induced cortisol responses, and eyewitness identification performance. *Behavioral Sciences and the Law, 34*(4), 580-594. https://doi.org/10.1002/bsl.2249

Sauerland, M., & Sporer, S. L. (2009). Fast and confident: Postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology: Applied, 15*(1), 46-62. https://doi.org/10.1037/a0014560

Sauerland, M., Stockmar, A. K., Sporer, S. L., & Broers, N. J. (2013). The reliability of identification evidence with multiple lineups. *The European Journal of Psychology Applied to Legal Context, 5*(1), 49-71. https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1889-18612013000100003

Sauerland, M., Tupper, N., Iannuzzi, M., & van Amelsvoort, A. G. (2023). The confidence-accuracy relationship for lineup decisions holds for the Dutch identification procedure. *PloS ONE, 8*(4), Article e0284205. https://doi.org/10.1371/journal.pone.0284205

Sauerland, M., Wiechert, S., Czarnojan, E., Deiman, E., Dörr, L., Broers, N. J., & Verschuere, B. (2025). Identification performance across the life span: Lineups and the reaction time-based Concealed Information Test. *Cognition, 254,* Article 105996. https://doi.org/10.1016/j.cognition.2024.105996

Seale-Carlisle, T. M., Wetmore, S. A., Flowe, H. D., & Mickes, L. (2019). Designing police lineups to maximize memory performance. *Journal of Experimental Psychology: Applied, 25*(3), 410-430. https://doi.org/10.1037/xap0000222

Semmler, C., Dunn, J. C., Mickes, L., & Wixted, J. T. (2018). The role of estimator variables in eyewitness identification. *Journal of Experimental Psychology: Applied, 24*(3), 400-415. https://doi.org/10.1037/xap0000157

Shen, K. J., Huang, J., Lam, A. L., & Wixted, J. T. (2024). The effects of filler similarity and lineup size on eyewitness identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 50*(9), 1444-1462. https://doi.org/10.1037/xlm0001342

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review, 8*(2), 203-220. https://doi.org/10.3758/BF03196157

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118*(3), 315-327. https://doi.org/10.1037/0033-2909.118.3.315

Tredoux, C. G. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior, 22*(2), 217-237. https://doi.org/10.1023/A:1025746220886

Tredoux, C. (1999). Statistical considerations when determining measures of lineup size and lineup bias. *Applied Cognitive Psychology, 13*(1), 9-26. https://doi.org/10.1002/(SICI)1099-0720(199911)13:1+<S9::AID-ACP634>3.0.CO;2-1

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352-373. https://doi.org/10.1037/h0020071

Tupper, N., Geisendörfer, A. K., Lorei, C., Sporer, S. L., Tredoux, C. G., & Sauerland, M. (2023). Police vs. laypeople: Little difference in identification performance and confidence-accuracy relationship for facial and body lineups. *Applied Cognitive Psychology, 37*(4), 845-860. https://doi.org/10.1002/acp.4085

Tupper, N., Sauer, J. D., Sauerland, M., Fu, I., & Hope, L. (2018) Face value: testing the utility of contextual face cues for face recognition, *Memory, 26*(10), 1436-1449. https://doi.org/10.1080/09658211.2018.1489968

Tupper, N., Sauerland, M., Sauer, J. D., Broers, N. J., Charman, S. D., & Hope, L. (2018). Showup identification decisions for multiple perpetrator crimes: Testing for sequential dependencies. *PLOS ONE, 13*(12), Article e0208403. https://doi.org/10.1371/journal.pone.0208403

Tupper, N., Sauerland, M., Sauer, J. D., & Hope, L. (2019). Eyewitness identification procedures for multiple perpetrator crimes: A survey of police in Sweden, Belgium, and the Netherlands. *Psychology, Crime & Law, 25*(10), 992-1007. https://doi.org/10.1080/1068316X.2019.1611828

van Amelsvoort, A. G. (1999). *Handleiding confrontatie* (3rd ed.) [Handbook eyewitness identification]. Stapel & De Koning.

van Amelsvoort, A. G. (2018). *Handleiding confrontatie* (10th ed.) [Handbook eyewitness identification]. Stapel & De Koning.

van Amelsvoort, A. G. (2023). *Handleiding confrontatie* (11th ed.) [Handbook eyewitness identification]. Sdu.

Watkins, M. J., Ho, E., & Tulving, E. (1976). Context effects in recognition memory for faces. *Journal of Verbal Learning and Verbal Behavior, 15*(5), 505-517. https://doi.org/10.1016/0022-5371(76)90045-1

Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: Confidence, accuracy, and response latency. *Applied Cognitive Psychology, 20*(1), 17-31. https://doi.org/10.1002/acp.1166

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*(1), 3–36. https://doi.org/10.1037/lhb0000359

Wells, E. C., & Pozzulo, J. D. (2006). Accuracy of eyewitnesses with a two-culprit crime: Testing a new identification procedure. *Psychology, Crime & Law, 12*(4), 417-427. https://doi.org/10.1080/10683160500050666

Wells, G. L., Yang, Y., & Smalarz, L. (2015). Eyewitness identification: Bayesian information gain, base-rate effect equivalency curves, and reasonable suspicion. *Law and Human Behavior, 39*(2), 99–122. https://doi.org/10.1037/lhb0000125

Winograd, E., & Rivers-Bulkeley, N. T. (1977). Effects of changing context on remembering faces. *Journal of Experimental Psychology: Human Learning and Memory, 3*(4), 397-405. https://doi.org/10.1037/0278-7393.3.4.397

Wixted, J. T., & Mickes, L. (2018). Theoretical vs. empirical discriminability: The application of ROC methods to eyewitness identification. *Cognitive Research: Principles and Implications, 3*(1), Article 9. https://doi.org/10.1186/s41235-018-0093-8

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. L. III. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist, 70*(6), 515-526. https://doi.org/10.1037/a0039510

Wixted, J. T., Mickes, L., Dunn, J. C., Clark, S. E., & Wells, W. (2016). Estimating the reliability of eyewitness identifications from police lineups. *Proceedings of the National Academy of Sciences, 113*(2), 304-309. https://doi.org/10.1073/pnas.1516814112

Wixted, J. T., Mickes, L., Wetmore, S. A., Gronlund, S. D., & Neuschatz, J. S. (2017). ROC analysis in theory and practice. *Journal of Applied Research in Memory and Cognition, 6*(3), 343-351. https://doi.org/10.1016/j.jarmac.2016.12.002

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest, 18*(1), 10-65. https://doi-org.mu.idm.oclc.org/10.1177/1529100616686966

Wooten, A. R., Carlson, C. A., Lockamyeir, R. F., Carlson, M. A., Jones, A. R., Dias, J. L., & Hemby, J. A. (2020). The number of fillers may not matter as long as they all match the description: The effect of simultaneous lineup size on eyewitness identification. *Applied Cognitive Psychology, 34*(3), 590-604. https://doi.org/10.1002/acp.3644

Yarmey, A. D. (1982). Eyewitness identification and stereotypes of criminals. In A. Trankell (Ed.), *Reconstructing the past: The role of psychologists in criminal trials* (pp. 205-225). Kluwer.