

1-1-2023

## Appearances can be deceiving: How naturalistic changes to target appearance impact on lineup-based decision-making

Dominic T. Jordan  
*Edith Cowan University*

Adrian J. Scott

Donald M. Thomson

Follow this and additional works at: <https://ro.ecu.edu.au/ecuworks2022-2026>



Part of the [Criminology Commons](#), and the [Psychology Commons](#)

---

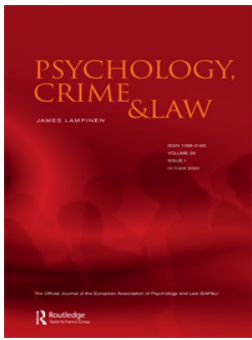
10.1080/1068316X.2023.2243001

Jordan, D. T., Scott, A. J., & Thomson, D. M. (2023). Appearances can be deceiving: How naturalistic changes to target appearance impact on lineup-based decision-making. *Psychology, Crime & Law*. Advance online publication.

<https://doi.org/10.1080/1068316X.2023.2243001>

This Journal Article is posted at Research Online.

<https://ro.ecu.edu.au/ecuworks2022-2026/2937>



## Appearances can be deceiving: how naturalistic changes to target appearance impact on lineup-based decision-making

Dominic T. Jordan, Adrian J. Scott & Donald M. Thomson

**To cite this article:** Dominic T. Jordan, Adrian J. Scott & Donald M. Thomson (2023): Appearances can be deceiving: how naturalistic changes to target appearance impact on lineup-based decision-making, *Psychology, Crime & Law*, DOI: [10.1080/1068316X.2023.2243001](https://doi.org/10.1080/1068316X.2023.2243001)

**To link to this article:** <https://doi.org/10.1080/1068316X.2023.2243001>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 03 Aug 2023.



[Submit your article to this journal](#)



Article views: 296



[View related articles](#)



[View Crossmark data](#)

# Appearances can be deceiving: how naturalistic changes to target appearance impact on lineup-based decision-making

Dominic T. Jordan <sup>a</sup>, Adrian J. Scott <sup>b</sup> and Donald M. Thomson<sup>c</sup>

<sup>a</sup>School of Arts and Humanities, Edith Cowan University, Joondalup, Australia; <sup>b</sup>Department of Psychology, Goldsmiths, University of London, London, UK; <sup>c</sup>Faculty of Health, Deakin University, Melbourne, Australia

## ABSTRACT

The present study examined the influence of appearance, procedure and position on identification decisions, post-decisional confidence ratings and estimates of discrimination and confidence-specific accuracy. Regarding appearance, the study examined the combined influence of three naturalistic changes that occur day-to-day (i.e. a reduction in cranial hair length, the removal of stubble, and a change of clothing), two of which have not been considered before in a lineup-decision context. Participants ( $N = 350$ ) completed four experimental lineups which involved: viewing a target person, completing a brief distractor task, and making an identification decision and a post-decisional confidence rating from a photographic lineup. Participants were randomly allocated to complete simultaneous or sequential lineups, with appearance (no change, change), position (early, late) and target (present, absent) systematically varied across the four trials. Appearance affected all dependent measures and was particularly influential in target-present lineups. Naturalistic changes to target appearance reliably decreased correct identification rates, confidence in correct identifications, discrimination accuracy, and confidence-specific accuracy. Procedure and position, by contrast, had a more limited impact. Of concern for the criminal justice system, neither procedure nor position manipulations offset any reductions in lineup-decision accuracy when target appearance changed.

## ARTICLE HISTORY

Received 7 November 2022


Accepted 21 July 2023

## KEYWORDS

Naturalistic appearance change; eyewitness identification; post-decisional confidence; discrimination accuracy; confidence-specific accuracy

More than 50 years of empirical and field research have shown that eyewitness identification decisions, obtained using lineup procedures, are highly fallible and often mistaken (Clark et al., 2008; Horry et al., 2014; Steblay et al., 2011). During this time, considerable attention has been afforded to better understanding eyewitnesses' lineup-based identification decisions, and how they might be improved (Wells et al., 2006). A primary objective of this research has been to identify influential factors that contribute to, or minimise, mistaken identifications. Such factors have been broadly categorised in the literature as

**CONTACT** Dominic T. Jordan  dtjordan.ac@gmail.com  School of Arts and Humanities, Edith Cowan University, 270 Joondalup Drive, Joondalup, WA 6027, Australia

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/1068316X.2023.2243001>.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

belonging to one of two groups: estimator or system variables (see Wells, 1978). Estimator variables refer to event and witness factors outside the criminal justice system's control, whereas system variables refer to procedural factors within the criminal justice system's control.

Perpetrator appearance change (i.e. appearance) is an estimator variable that describes the relationship between person-specific cues stored at encoding (i.e. physical attributes of a perpetrator as witnessed during a criminal event) and person-specific cues available at retrieval (i.e. the physical attributes of the suspect and lineup members viewed during a lineup). To date, eyewitness research examining appearance has primarily focused on the influence of distinct, disguise-related changes to appearance on lineup-based decision accuracy. However, research has rarely considered the influence of naturalistic appearance changes. As appearance characteristics vary day-to-day, and naturalistic changes can occur frequently during criminal investigations (Erickson et al., 2017; Pozzulo & Marciniak, 2006), it is important to examine how such changes to appearance influence eyewitness decision-making.

Procedure and position, which describe the mode of lineup administration and the placement of a suspect within a lineup respectively, are two system variables that have received considerable attention from eyewitness researchers. However, there is limited research examining how these two system-controlled variables interact with naturalistic changes to appearance in lineup-based identification contexts. As such, it is important to examine if and how system-controlled variables influence eyewitness decision-making under varied appearance conditions. The present study, therefore, sought to directly investigate the influence of appearance (estimator variable), procedure and position (system variables) on identification decisions, post-decisional confidence, and estimates of decision accuracy.

## **Appearance**

As part of a criminal investigation, police officers are sometimes unable to apprehend the perpetrator at the scene of a crime, resulting in a delay between the witnessed criminal event and subsequent lineup-based identification (Pozzulo & Balfour, 2006; Pozzulo & Marciniak, 2006). During this delay, which might span days, weeks, months or even years, the appearance characteristics of a perpetrator can change (Terry, 1994; Thomson, 1982). The appearance characteristics of a perpetrator play an important role in recognition, for in the absence of familiarity or prior knowledge of a perpetrator's identity, the person-specific cues associated with their appearance are the primary source of information available to witnesses when responding to a lineup (Levi & Jungman, 1995; Thomson, 1982, 1986).<sup>1</sup>

The types of changes made to perpetrator appearance during a criminal investigation can be classified as either disguise-related or naturalistic (Hope & Sauer, 2014). Disguise-related changes typically involve the addition or removal of paraphernalia worn to obscure salient parts of a perpetrator's underlying facial structure (Cutler et al., 1987; Mansour et al., 2012; O'Rourke et al., 1989). In lineup-based identification contexts, the addition and removal of eyewear (Mansour et al., 2012), headgear (Cutler et al., 1987; O'Rourke et al., 1989) and other disguise-related changes, such as wigs (Pozzulo & Balfour, 2006; Pozzulo & Marciniak, 2006; Yarmey, 2004) and masks (Davies & Flin, 1984;

Manley et al., 2021) have been found to have a significant deleterious effect on identification accuracy, particularly in target-present lineups. In target-absent lineups, the impact of disguise-related changes is less clear. Some studies have found that the addition of a wig at encoding has no effect on identification decisions (Pozzulo & Balfour, 2006; Pozzulo & Marciniak, 2006; Yarmey, 2004), whereas Mansour et al. observed that the addition of a toque (i.e. hat) at encoding significantly increased incorrect identifications made from target-absent lineups.

Naturalistic changes to perpetrator appearance occur with the passage of time, and typically involve alterations to cranial hair length, colour and style, facial hair length or presence (in male populations), body shape, weight, and age (Charman & Wells, 2007). To date, limited attention has been afforded to examining naturalistic appearance changes in lineup-based identification contexts (Molinaro, Arndorfer, & Charman, 2013). To the best of the authors' knowledge, only two studies have examined the impact of naturalistic changes to target appearance on identification (Gronlund et al., 2009; Memon & Gabbert, 2003). Memon and Gabbert, for example, manipulated the hair style of a female target so that their long hair was either worn down or pulled back at encoding. Gronlund et al., by comparison, manipulated the appearance of a male target, so that they either had or had not grown several days worth of facial hair, and altered their cranial hair slightly, at the point of retrieval. Consistent with research on disguise-related changes, both studies found that significantly fewer correct identifications were made in change, relative to no change conditions.

Theoretical accounts of recognition, such as signal detection theory, offer an explanation as to why correct identifications diminish when targets change their appearance. Signal detection accounts posit that the process of making an identification decision is dependent on the interaction between two internal cognitive processes: match-to-memory and decision criterion (Baranski & Petrusic, 1998; Sauer et al., 2008). Match-to-memory can be characterised as the subjective sense of similarity or feeling of familiarity a person experiences when comparing a present object with an object stored in memory (Levi & Jungman, 1995; Thomson, 1982; Tulving, 1981; Wixted, 2007). Decision criterion describes the degree of evidence required to trigger a positive identification (i.e. a member of the lineup is selected; Clark, 2011). In cases involving unknown perpetrators, appearance constitutes the main source of perceptual, and by extension memorial, information available to witnesses when responding to a lineup task (Charman & Wells, 2007). A changed perpetrator will, therefore, generate less match-to-memory evidence than an unchanged perpetrator when presented as part of a lineup, because fewer perceptual cues are available at retrieval that can act as reliable markers of recognition (Shapiro & Penrod, 1986; Thomson, 1986; Thomson et al., 1982). In signal detection terms, changing perpetrator appearance directly weakens match-to-memory evidence, making it less likely to exceed a witness's decision criterion and trigger a correct identification (Pozzulo & Balfour, 2006).

In actual criminal investigations, the potential impact of appearance change cannot be overstated. Of particular concern are naturalistic changes because they occur on a day-to-day basis and, therefore, could play a role in every criminal investigation and subsequent lineup task (Charman et al., 2022; Erickson et al., 2017). Furthermore, there is no way to account for the occurrence of naturalistic changes in real-world contexts. Recognising that a perpetrator has undergone a naturalistic change (or changes) is often conditional

on first recognising that the suspect is the perpetrator (Charman et al., 2022). From a criminal justice system perspective, it is not possible to determine whether differences between a witness description of a perpetrator and detained suspect, are the result of the perpetrator changing their appearance or an innocent suspect not matching the appearance of the perpetrator.

Given the relative paucity of research examining the effects of day-to-day naturalistic appearance changes on identification, more data is needed that replicates and extends the limited knowledge base. As Charman et al. (2022) stated, eyewitness research has tended to ignore the reality that perpetrators' appearance in real-world contexts often changes. As such, when making recommendations regarding the suitability of key system variables, researchers may have overestimated the accuracy and confidence with which witnesses make lineup-based decisions in real-world contexts. To date, the one study to have examined the influence of appearance, procedure, and position on identification, manipulated appearance by using additive changes (i.e. facial hair was added at retrieval; Gronlund et al., 2009). However, research must also consider the inverse affect, by manipulating appearance through removal-based changes (i.e. facial hair is removed at retrieval). This distinction is important, as several studies in the field of face recognition have found that additive and removal-based changes affect recognition accuracy differently (Righi et al., 2012; Terry, 1994). For example, Righi et al. and Terry found that removing eyeglasses from a target between encoding and retrieval reduced recognition accuracy, but adding eyeglasses had no effect. Further, Righi et al. found that removing a wig, and Terry found that adding a beard, contributed to larger respective declines in recognition accuracy than adding a wig and removing a beard.

### ***Procedure and position***

A plethora of research has sought to establish whether the simultaneous (i.e. lineup members are presented to the witness at the same time) or sequential (i.e. lineup members are presented to the witness one at a time) procedure encourages more accurate eyewitness responding (see Clark et al., 2008; Palmer & Brewer, 2012; Steblay et al., 2011 for meta-analyses). Most empirical evidence indicates that a trade-off in accuracy occurs across target present and absent lineups, whereby correct identifications occur more frequently from target-present simultaneous lineups and correct rejections occur more frequently from target-absent sequential lineups (Clark et al., 2008; Palmer & Brewer, 2012). In actual criminal investigations, the utility of this finding is limited, for investigators cannot establish with certainty whether a lineup contains the perpetrator or not (Wells & Olson, 2003). Accordingly, researchers have increasingly relied on analyses that collapse across target presence (e.g. diagnosticity ratios,  $d'$ , and pAUC) to form conclusions regarding procedural superiority (Smith et al., 2017). While much research initially favoured a sequential lineup advantage (see Steblay et al., 2011), more recent research using signal detection measures has almost unanimously favoured a simultaneous lineup advantage (Carlson & Carlson, 2014; Gronlund et al., 2014; Seale-Carlisle et al., 2019; Seale-Carlisle & Mickes, 2016).

The diagnostic feature detection hypothesis (DFD) was introduced to explain the simultaneous lineup advantage (Wixted & Mickes, 2014), which posits that simultaneous lineups facilitate improved discrimination relative to sequential lineups (i.e. the degree

to which witnesses can accurately categorise novel stimuli from previously encountered stimuli; Lee & Penrod, 2019). According to DFD, simultaneous lineups encourage witnesses to compare lineup members (Seale-Carlisle et al., 2019). This process of comparison allows witnesses to detect and immediately discount features shared by all lineup members as non-diagnostic; and instead focus on, and attribute weight solely to unique features which are diagnostic when making an identification decision. Sequential lineups, by contrast, deliberately restrict witnesses' capacity to compare faces. It is argued that by presenting lineup members in isolation witnesses can only establish which features are shared and unique as they progress through the lineup (Wetmore et al., 2017). As such, witnesses presented with a sequential lineup may attribute greater weight to non-diagnostic information when making an identification, decreasing discrimination accuracy (Wixted & Mickes, 2014).

The applications of DFD extend beyond procedure and can be useful for understanding how identifications obtained from sequential lineups are influenced by other system-controlled variables, like position (Wetmore et al., 2017). Interest in position effects, which describe the propensity for a participant to select a target or replacement as a function of where they are presented in a lineup has grown in the advent of research which found that position primarily influences identification decisions made within sequential, rather than simultaneous lineups (Carlson et al., 2008; Clark & Davey, 2005; Gronlund et al., 2009; Meisters et al., 2018; Wells et al., 2011). Research examining position effects within sequential lineups has generally found that identification performance increases when targets and replacements are presented late in a sequential lineup, rather than early (Carlson et al., 2008; Gronlund et al., 2009; Horry et al., 2012; Meisters et al., 2018). According to DFD, this is because witnesses become more familiar with which features are or are not diagnostic as a sequential lineup progresses, allowing them to better optimise their discrimination capabilities (Wetmore et al., 2017). It is worth noting, however, that some research has found the opposite effect, namely, that identification performance in target-present lineups increases when targets are presented early in a sequential lineup, rather than late (Carlson et al., 2016; Clark & Davey, 2005). Findings of an early position advantage are at odds with the assumptions of DFD, and directly contradict the proposition that witnesses optimise their discrimination capabilities over the course of a sequential lineup. Instead, it has been argued that witnesses are more likely to 'spend' their identifications on similar-looking foils preceding the target, when the latter is positioned late rather than early in a sequential array (Clark & Davey, 2005).

Position effects in sequential lineups are complex and the inconsistent nature of the relationship between procedure and position is not easily explained (Colloff & Wixted, 2020). Despite this, establishing the applied suitability of procedure and position factors in real-world contexts remains a priority for the criminal justice system (Horry et al., 2012). Given the propensity for estimator variables like appearance change (i.e. appearance) to occur in the real world, it is important to examine how appearance interacts with key system variables to impact identification (Charman et al., 2022). Currently, little is known about how procedure and position interact with appearance. In fact, the limited published research examining the relationship between appearance (including disguise) and procedure has produced inconsistent findings, with evidence of a sequential advantage (Gronlund et al., 2009), a simultaneous advantage (Memon & Gabbert,

2003) and no difference found (Cutler & Penrod, 1988; Mansour et al., 2012; Pozzulo & Marciniak, 2006) when target appearance was changed.

Conversely, the one study to examine appearance, procedure and position produced findings contrary to predictions (Gronlund et al., 2009). Gronlund et al., for example, hypothesised that simultaneous lineups would be less vulnerable to changes in appearance than sequential lineups because lineup members can be directly compared, and a changed target should still have the strongest resemblance to a witness's memory relative to other lineup members. However, the findings revealed that when target appearance was changed sequential lineups produced more accurate responding than simultaneous lineups. Regarding appearance, procedure and position, no significant interactions were found between the three variables. Yet position effects within sequential lineups were much more prominent in no change than change conditions, perhaps suggesting that positional advantages can be rendered weaker or even redundant in cases involving changed targets. In response, the authors called for more research, noting that it was unclear why the research findings were inconsistent with expectations.

### **Confidence**

Lineup-based decisions typically refer to categorical identifications (i.e. yes, that is the perpetrator; no, the perpetrator is not present), yet witnesses often provide investigators with additional information in the form of post-decisional confidence ratings (i.e. I am 90% certain that is the perpetrator). Interest in post-decisional confidence ratings (i.e. confidence) has grown in the advent of research which found that confidence and accuracy for positive identifications decisions are well calibrated and robustly associated (Brewer, 2006; Brewer & Wells, 2011; Juslin et al., 1996; Wixted & Wells, 2017). Empirical evidence supporting an association between confidence and decision accuracy has a strong theoretical basis (Baranski & Petrusic, 1998; Sauer et al., 2008; Van Zandt, 2000). For example, decisional locus theories, the most notable of which is signal detection theory (Macmillan & Creelman, 1991), posit that confidence and decision-making rely on the same information in memory, and thus confidence ratings and identifications are intrinsically related (Sauer et al., 2008). Confidence within a decisional locus framework reportedly indexes match-to-memory signal strength (Dobolyi & Dodson, 2013; Leippe et al., 2009) or the difference between match-to-memory signal strength and a witness's decision criterion (Brewer & Wells, 2011; Sauer & Brewer, 2015; Wixted et al., 2015). Regardless of which interpretation is favoured, it has been suggested that for positive identification decisions, confidence provides a relatively direct measure of the memorial evidence in favour of that decision (Sauer et al., 2008).

The small number of studies to have examined the influence of appearance, procedure or position on confidence, have reported inconsistent findings. Regarding appearance, some findings suggest that confidence in correct identifications decreases in change conditions (Mansour et al., 2012), while other findings suggest that confidence is completely unaffected by appearance (Cutler et al., 1987). Similar to appearance, the influence of procedure has varied considerably, as some findings suggest that confidence in positive identifications decreases in simultaneous compared to sequential lineups (Dobolyi & Dodson, 2013; Gronlund et al., 2009; Weber & Brewer, 2004, experiment 2), confidence in non-identifications decreases in simultaneous compared to sequential lineups



(Mansour et al., 2012), and confidence is unaffected by procedure (Lindsay & Wells, 1985; Weber & Brewer, 2004, experiment 3). No explanation has yet been offered to account for the discrepant findings regarding confidence, particularly as they relate to appearance and procedure effects. Unlike appearance and procedure, very little research has examined the influence of position on confidence. Although it is apparent that some research examining position effects within simultaneous and sequential lineups have obtained post-decisional confidence ratings, analyses relating to the influence of position on confidence were not reported (e.g. Carlson et al., 2016; Gronlund et al., 2009; Horry et al., 2012).

## Present study

There is a dearth of research examining the impact of day-to-day naturalistic appearance changes in a lineup-based identification context, with only two prior studies examining its influence in relation to eyewitness decision-making (Gronlund et al., 2009; Memon & Gabbert, 2003). Of these, Gronlund et al. examined the variables of interest in the present study (i.e. appearance, procedure and position), but their study involved additive-based featural transformations and focused only on identification decisions and post-decisional confidence ratings.

Continued examination of day-to-day naturalistic appearance changes is important in lineup-based identification contexts because the appearance characteristics of a perpetrator are the primary source of information available to witnesses, and changes that occur naturally likely influence most real-world lineup tasks (Charman et al., 2022; Erickson et al., 2017). It is also important to examine procedure and position to evaluate whether the manipulation of these system variables can be used to effectively mitigate the potential deleterious effects of appearance change. Establishing if or how procedure and position can influence participants' lineup decisions and maximise decision accuracy under variable appearance conditions will have implications for the criminal justice system.

The present study extends prior research by examining, for the first time, the influence of appearance, procedure and position on identification decisions, post-decisional confidence ratings, and estimates of discrimination and confidence-specific accuracy. In contrast to previous research, the appearance manipulation in the present study involved exclusively removal-based featural transformations; such that, in change conditions targets' facial and cranial hair were removed and less prominent in length respectively, during the lineup than at encoding. Distinguishing between additive and removal-based transformations in an identification context is important, given that research in the field of face recognition has found that these changes can affect recognition accuracy differently (Righi et al., 2012; Terry, 1994). It is also important to acknowledge, that targets in change conditions wore different clothing during encoding and the lineup.

Regarding procedure and position, the last decade has offered several important theoretical advancements regarding the influence of these variables on lineup-based decision accuracy, with recently introduced signal detection measures used to inform applied recommendations regarding the suitability of system variables in real-world contexts. Therefore, the present study extends prior research by adopting measures including multi-d' (Lee & Penrod, 2019) and confidence-accuracy characteristic analysis (Mickes, 2015) to better estimate the accuracy and reliability with which eyewitnesses make decisions from lineups in contexts where target appearance may have changed.

In examining the impact of appearance, the following hypothesis was proposed in light of signal detection theory and previous research findings:

- (1) It is hypothesised that correct identification decisions, confidence in correct identifications, discrimination and confidence-specific accuracy will be higher in no change than change conditions.

Regarding procedure and position, hypotheses were not deemed appropriate given the inconsistencies between theoretical predictions and empirical research findings. As such, two general research questions were proposed:

- (2) Does procedure influence the accuracy of identification decisions, post-decisional confidence ratings, and estimates of discrimination and confidence-specific accuracy?
- (3) Does position, within sequential lineups, influence the accuracy of identification decisions, post-decisional confidence ratings, and estimates of discrimination and confidence-specific accuracy?

Regarding the relationship between appearance, procedure and position, little is currently known about how they might interact, and the one prior study to have examined these factors produced findings contrary to predications. As such, a final research question is proposed:

- (4) Do appearance, procedure and position interact to influence the accuracy of identification decisions, post-decisional confidence ratings, and estimates of discrimination accuracy?

## Materials and methods

### Design

The study adopted a 2 (appearance)  $\times$  2 (procedure)  $\times$  2 (position)  $\times$  2 (target) mixed factorial design, with procedure (simultaneous, sequential) manipulated between-participant, and appearance (no change, change), position (early, late) and target (present, absent) manipulated within-participant. At the outset, participants were randomly allocated to one of two lineup conditions: simultaneous or sequential. Participants within each lineup condition were then randomly allocated to one of eight possible viewing conditions. Within each viewing condition, appearance, position, and target were varied across four lineup trials. It is important to note that participants viewed four as opposed to eight lineups, because appearance and position manipulations were perfectly conflated within each viewing condition (see the Appendix).

### Participants

A total of 350 participants with an average age of 30.99 years ( $SD = 12.21$ , range 17–71 years) completed the study. Most participants were female ( $n = 269$ , 78.4%) and Caucasian

( $n = 263$ , 76.7%); seven participants chose not to provide any demographic information. Approximately half of participants ( $n = 178$ , 50.9%) were community members, recruited via social media (Facebook, LinkedIn and Twitter), email invitation, posters, and word of mouth. Remaining participants ( $n = 172$ , 49.1%) were undergraduate students recruited via university research participation schemes at the authors' institutions. Community members were offered the opportunity to enter a prize draw to receive a \$25 e-gift card, and undergraduate students were offered the opportunity to receive course credit.

## **Materials**

Still, photographs from the multi-PIE face database (multi-PIE; Gross et al., 2010) were used both as encoding stimuli and in the construction of lineups. Multi-PIE is a resource which provides high-resolution passport-style images (i.e. head and shoulders) of faces of different ethnicities and genders. For this research, all photographs were of Caucasian males.

### **Target photographs**

In the present study, three different photographs of each target were selected from multi-PIE. One of two photographs of each target was presented to participants during encoding (i.e. the target stimulus photograph), depending upon the appearance condition (i.e. no change, change). The third photograph of the target was presented during retrieval (i.e. the target photograph included in the lineup) and remained the same regardless of which appearance condition participants were in.

Three person-specific characteristics (target facial hair presence, target cranial hair length, and target clothing) and one contextual characteristic (photograph background) were used to establish the suitability of potential target persons. In no change conditions, the target stimulus photograph and the target photograph included in the lineup only differed in relation to the one contextual characteristic (photograph background). The two photographs used in this condition were taken on the same day, thus the three person-specific characteristics were maintained between encoding and retrieval. By contrast, in change conditions, the target stimulus photograph and the target photograph included in the lineup were taken on different days and differed in relation to the three person-specific characteristics (target facial hair presence, target cranial hair length and target clothing) and the one contextual characteristic. Regarding target facial hair presence and target cranial hair length, changes were exclusively removal-based (see Terry, 1994). Thus, in change conditions, the target presented in the lineup always had less facial hair and slightly shorter cranial hair than the target presented to participants during encoding (see supplementary materials for an example of the appearance manipulation).

### **Lineups**

Participants viewed four lineup trials as part of this research, therefore, four experimental lineups were constructed (i.e. one lineup for each target). As multi-PIE is a modest size face database (i.e. includes < 200 Caucasian, male faces; see Bergold & Heaton, 2018) a variation of the resemblance-match method (see Tunnicliff & Clark, 2000) was used to select eight lineup members, including a target replacement (i.e. replacement), for each

target. The resemblance-match method adopted as part of this research occurred in three phases, the first two of which were undertaken by the authors and the third of which was completed by an independent sample of participants.

During the first phase, every Caucasian, male face in multi-PIE was placed into an attribute pool, based on their approximate age, hair colour and length. During the second phase, 15 faces adjudged to have the highest degree of resemblance to each target's no-change photograph were selected from the relevant attribute pools to form a 'target group' for each target. During the third phase, 34 participants ( $M$  age = 27.63  $SD$  = 9.59, range 20–60 years; female = 65.7%; Caucasian = 91.4%), selected and ranked eight faces from each target group that best resembled the respective target. The eight faces (not including the target) with the highest pooled mean resemblance rankings were selected as lineup members for each target, and the lineup members with the highest mean resemblance ranks became the replacements in each target-absent lineup (see supplemental materials).

### *Effective size and defendant bias*

Measures of Tredoux's (1998, 1999) effective size (Tredoux's  $E'$ ), and defendant bias (Malpass, 1981) were calculated for each target present and target absent lineup (see supplemental materials). Two independent groups of participants were recruited to assist in this process. The first group comprised 17 participants ( $M$  age = 25.71 years,  $SD$  = 7.65, range 20–53 years; female = 52.9%; Caucasian = 94.1%) who provided written descriptions of each target face. Descriptors mentioned by at least four participants were then collated to form a single modal description for each target. The second group comprised 58 participants ( $M$  age = 22.93,  $SD$  = 7.27, range 18–63 years; female = 74.6%; Caucasian = 69.8%) who engaged in an online mock witness task and evaluated each of the lineups by selecting the face that best fit the provided modal description.

No significant differences were observed in Tredoux's  $E'$  within any of the four experimental lineup pairs (i.e. target-present and target-absent variations for each lineup). Collapsed across target presence, measures of Tredoux's  $E'$  ranged from 2.17 to 4.75 ( $M$  = 3.60,  $SD$  = 0.91). Measures of defendant bias indicated that the proportion of mock witnesses who selected the target or replacement in Lineup 1 (1.46, 1.09, respectively) and Lineup 2 (0.68, 1.82, respectively) did not significantly differ from that expected by chance. By contrast, the proportion of mock witnesses who selected the target or replacement in Lineup 3 (2.82, 2.89, respectively) and Lineup 4 (5.04, 6.01 respectively) did differ significantly from that expected by chance.

### *Procedure*

Participants were directed to an online study, hosted by Qualtrics XM, via a weblink. The study began with an information letter and consent form. Once consent was obtained, participants were advised that they would engage with five lineup trials, the first of which was a practice to familiarise them with the procedure. Participants were informed that target persons may or may not be included in the photographic lineups, and that, if included, target appearance and/or clothing may have changed.

Participants were randomly allocated to one of two lineup conditions. Participants in the simultaneous lineup condition viewed all eight lineup members at the same time

(lineup members were presented in a single row of eight). Conversely, participants in the sequential lineup condition viewed lineup members one at a time (lineup members were presented individually). For sequential lineups, the Lindsay and Wells (1985) variant was adopted, which allows participants a single view of each lineup member and immediately ends when participants make a positive identification (i.e. when a member of the lineup is selected), or in the absence of a positive identification, when participants viewed all eight lineup members. Participants within each lineup condition were then randomly allocated to one of eight possible sub-conditions. Within each sub-condition, appearance, position, and target were varied across the four lineups. Regarding position, targets and replacements presented third from the left (simultaneous) or third (sequential) in each eight-person lineup constituted early position, and targets and replacements presented sixth from the left (simultaneous) or sixth (sequential), constituted late position. The order in which the four experimental lineups were presented to participants was randomised.

During each experimental lineup trial, participants viewed a target person for three seconds, engaged in a brief distractor task (i.e. watched a two-to-three-minute video and answered a related question), and made an identification decision and an associated post-decisional confidence judgement from a photographic lineup. To record a positive identification decision, participants selected the face of the lineup member they believed to be the target. To make a non-identification (i.e. when no member of the lineup is selected), participants in the simultaneous condition selected the 'not present' option for a given lineup, and participants in the sequential condition selected the 'not present' option for all eight lineup members. Directly following a decision (positive identification or non-identification), participants were asked to rate their confidence in the accuracy of their decision on a 6-point scale, ranging from 1 'uncertain' to 6 'absolutely certain'.

Once participants had completed the four lineup trials, they were asked to provide some basic demographic information (age, gender, and ethnicity). Finally, participants were thanked for completing the study and provided with a debrief statement. Participation took approximately 20–25 min.

## Results

The results present three sets of analyses. The first set examines the impact of appearance, procedure and position on the frequency of identification decisions and participants' mean post-decisional confidence ratings in target present and absent lineups. The second set examines the impact of appearance, procedure and position on measures of discrimination accuracy. The third set examines the impact of appearance, procedure and position on confidence-specific accuracy (i.e. the relationship between confidence and accuracy in 'suspect' identifications, herein referred to as target/replacement identifications).

Given the mock witness task revealed variation in defendant bias across lineups, preliminary analyses were conducted to examine if identification decisions made from unbiased (i.e. lineups 1 and 2) and biased (i.e. lineups 3 and 4) lineups varied as a function of appearance, procedure and position. These analyses revealed that the pattern of findings was consistent across both sets of lineups (see supplemental materials). Therefore, consistent with previous research, each lineup trial, as opposed to each participant,

was treated as an independent case for the purpose of analysis (Carlson et al., 2019b; Lucas & Brewer, 2022).

### **Identification decisions and post-decisional confidence ratings**

Participants' identification decisions were divided into six categories: correct target identifications (where the target is identified), incorrect foil identifications (where a foil is identified) and incorrect non-identifications (where no one is identified) in target present lineups; and correct non-identifications (where no one is identified), incorrect foil identifications (where a foil is identified), and incorrect replacement identifications (where the replacement is identified) in target absent lineups. As the nature of a correct identification varies as a function of target presence (i.e. target identification or non-identification) data from target present and target absent lineups were analysed separately (Carlson et al., 2016; Manley et al., 2021). Table 1 shows the proportions and frequencies of participants' target present and absent identification decisions, and Table 2 shows the means and standard deviations of confidence ratings for all target present and absent identification decisions, as a function of appearance, procedure, and position.

To examine the impact of appearance, procedure and position on the frequency of identification decisions, two hierarchical loglinear (HILOG) analyses were conducted. HILOG analysis is an appropriate method for examining eyewitness identification data, as all decision types can be included without being recategorised or recoded (see Colloff et al., 2017; Lucas & Brewer, 2022; Wells et al., 2011). Note, that in HILOG analysis no single variable is classified as a dependent measure, therefore, a two-way interaction is akin to a main effect (e.g. appearance and identification decisions [ID]). Consistent with previous research (e.g. Colloff et al., 2017), z-scores, which indicate the difference between observed and expected frequencies (Field, 2018), were used as a measure of statistical significance. As HILOG analyses can sometimes exclude significant associations with small effect sizes (Tabachnick & Fidell, 2013), all higher order interactions of marginal significance (at  $\alpha < .08$ ) were investigated using likelihood ratio chi-square tests. To examine the impact of appearance, procedure and position on participants' post-decisional confidence ratings, factorial between-participant ANOVAs were conducted.

**Table 1.** Proportions and frequencies (n) of target present and absent identification decisions as a function of appearance, procedure, and position.

			Present % (n)			Absent % (n)		
			Target ID	Foil ID	No ID	TR ID	Foil ID	No ID
No change	Simultaneous	Early	0.71 (60)	0.11 (9)	0.18 (15)	0.07 (6)	0.20 (17)	0.73 (61)
		Late	0.66 (58)	0.13 (11)	0.22 (19)	0.06 (5)	0.30 (26)	0.65 (57)
	Sequential	Early	0.71 (62)	0.07 (6)	0.23 (20)	0.08 (7)	0.19 (17)	0.73 (65)
		Late	0.56 (50)	0.22 (20)	0.22 (20)	0.07 (6)	0.27 (24)	0.66 (59)
Change	Simultaneous	Early	0.33 (29)	0.17 (15)	0.50 (44)	0.05 (4)	0.26 (23)	0.69 (61)
		Late	0.34 (29)	0.18 (15)	0.48 (40)	0.06 (5)	0.21 (18)	0.73 (61)
	Sequential	Early	0.37 (33)	0.17 (15)	0.47 (42)	0.06 (5)	0.26 (23)	0.69 (62)
		Late	0.22 (19)	0.18 (16)	0.60 (53)	0.06 (5)	0.21 (18)	0.74 (65)

Note: All proportions are rounded to two decimal places. TR = Target replacement.

**Table 2.** Mean (SD) post-decisional confidence in target present and absent identification decisions as a function of appearance, procedure, and position.

		Present Mean (SD)			Absent Mean (SD)		
		Target ID	Foil ID	No ID	TR ID	Foil ID	No ID
No change	Simultaneous	4.57 (1.29)	3.22 (1.30)	3.93 (1.58)	4.34 (1.38)	3.29 (1.49)	2.33 (1.37)
	Late	4.55 (1.35)	3.64 (1.21)	3.68 (1.42)	4.32 (1.34)	2.85 (1.16)	2.80 (1.30)
Change	Sequential	4.76 (1.14)	3.50 (0.84)	3.05 (1.70)	3.57 (1.61)	3.18 (0.95)	4.57 (0.54)
	Simultaneous	4.96 (1.07)	3.35 (1.18)	3.45 (1.96)	3.81 (1.60)	3.54 (1.06)	3.67 (1.37)
Change	Sequential	4.28 (1.28)	2.73 (0.96)	4.36 (1.46)	4.48 (1.26)	2.78 (1.70)	2.75 (0.96)
	Late	3.79 (1.40)	3.13 (1.60)	4.28 (1.47)	4.64 (1.35)	2.78 (1.74)	3.80 (1.48)
Change	Sequential	4.30 (1.13)	3.53 (1.18)	3.93 (1.79)	4.10 (1.54)	3.26 (1.45)	4.20 (1.10)
	Late	4.53 (.91)	3.81 (1.17)	3.49 (1.73)	4.43 (1.60)	3.94 (1.00)	3.20 (0.84)

Note: The scale ranged from 1 (uncertain) to 6 (absolutely certain). TR = Target replacement.

### **Target present lineups**

A  $2 \times 2 \times 2 \times 3$  backward elimination HILOG was conducted to examine the impact of appearance, procedure and position on the frequency of target present identifications. Examination of the global test of order terms indicated that the highest order effect that reached significance (at  $\alpha = .05$ ) was a two-way interaction,  $\chi^2(18) = 107.81$ ,  $p < .001$ . Examination of the partial associations pointed to the presence of two significant two-way interactions for appearance and ID,  $\chi^2(2) = 91.73$ ,  $p < .001$ , and for position and ID,  $\chi^2(2) = 6.68$ ,  $p = .035$ . Regarding the appearance and ID interaction, participants made significantly more correct identifications in no change (65.7%,  $n = 230$ ) than in change conditions (31.4%,  $n = 110$ ;  $z = 8.05$ ,  $p < .001$ ), and significantly more incorrect non-identifications in change (51.1%,  $n = 179$ ) than in no change conditions (21.1%,  $n = 74$ ;  $z = 5.88$ ,  $p < .001$ ). Incorrect foil identifications were not influenced by appearance.

Regarding the position and ID interaction, participants made significantly more correct identifications when the target was presented early (54.1%,  $n = 184$ ) than late (45.9%,  $n = 156$ ;  $z = 2.48$ ,  $p = .013$ ). Examination of the response frequencies and parameter estimates, however, revealed that the early position advantage was far more pronounced within sequential than simultaneous lineups; and that a three-way interaction for procedure, position and correct identifications was also marginally significant ( $z = 1.89$ ,  $p = .058$ ). To explore the potential procedure, position and ID interaction further, four likelihood ratio chi-square tests were conducted. Of these, one was found to be significant (at Bonferroni corrected  $\alpha = .013$ ). This test confirmed that within sequential lineups, participants made significantly more correct identifications when the target was positioned early (53.4%,  $n = 95$ ) than late (38.8%,  $n = 69$ ), and significantly more incorrect foil identifications when the target was positioned late (20.2%,  $n = 36$ ) than early (11.8%,  $n = 21$ ),  $\chi^2(2) = 8.97$ ,  $p = .011$ ,  $w = 0.16$ . Incorrect non-identifications were not influenced by procedure or position.

Three  $2 \times 2 \times 2$  factorial between-participant ANOVAs were also conducted to examine the impact of appearance, procedure and position on mean confidence ratings for each target present identification response (i.e. correct identifications, incorrect foil identifications, and incorrect non-identifications). The analyses revealed that appearance significantly influenced confidence in correct identifications (at Bonferroni corrected  $\alpha = 0.17$ ), with higher mean confidence ratings in no change ( $M = 4.70$ ,  $SD = 1.23$ ) than change conditions ( $M = 4.20$ ,  $SD = 1.23$ ),  $F(1, 332) = 11.36$ ,  $p < .001$ ,  $\eta p^2 = .033$ . Furthermore, procedure significantly influenced confidence in incorrect non-identifications, with higher mean confidence ratings in simultaneous lineups ( $M = 4.17$ ,  $SD = 1.48$ ) than sequential lineups ( $M = 3.56$ ,  $SD = 1.78$ ),  $F(1, 253) = 6.54$ ,  $p = .011$ ,  $\eta p^2 = .026$ . No other main or interaction effects were significant.

### **Target absent lineups**

A  $2 \times 2 \times 2 \times 3$  backward elimination HILOG was conducted to examine the impact of appearance, procedure and position on the frequency of target-absent identifications. Examination of the global test of order terms indicated that none of the higher order effects reached significance. Examination of the partial associations and parameter estimates confirmed that no interactions or effects reached significance.



Three  $2 \times 2 \times 2$  ANOVAs were then conducted to examine the impact of appearance, procedure and position on mean confidence ratings obtained for each target absent identification response (i.e. correct non-identifications, incorrect foil identifications, and incorrect replacement identifications). The analyses revealed that appearance and procedure significantly influenced confidence in correct non-identifications. Mean confidence ratings were higher in change ( $M = 4.41$ ,  $SD = 1.45$ ) than no change conditions ( $M = 4.00$ ,  $SD = 1.51$ ),  $F(1, 483) = 9.07$ ,  $p = .003$ ,  $\eta^2 = .018$ , and in simultaneous lineups ( $M = 4.45$ ,  $SD = 1.33$ ) than sequential lineups ( $M = 3.98$ ,  $SD = 1.61$ ),  $F(1, 483) = 12.32$ ,  $p < .001$ ,  $\eta^2 = .025$ . Procedure also significantly influenced confidence in incorrect foil identifications, with higher mean confidence ratings in sequential lineups ( $M = 3.48$ ,  $SD = 1.17$ ) than simultaneous lineups ( $M = 2.90$ ,  $SD = 1.50$ ),  $F(1, 158) = 6.85$ ,  $p = .010$ ,  $\eta^2 = .042$ . No other main or interaction effects were significant.

### Discrimination accuracy

To evaluate the impact of appearance, procedure and position on the accuracy of participants' identification decisions, multiple measures of empirical discriminability were computed using the multi- $d'$  model (see Lee & Penrod, 2019).<sup>2</sup> The multi- $d'$  model provides information regarding participants' capacity to discriminate between: targets and replacements,  $d'(TR)$ ; targets and foils in target-present (TP) lineups,  $d'(TFp)$ ; replacements and foils in target-absent (TA) lineups,  $d'(RFa)$ ; and the differential appeal of foils in TA lineups and foils in TP lineups,  $d'(FaFp)$ .

Group-level multi- $d'$  scores were computed for each experimental condition. Consistent with recommendations in the literature (see Mickes et al., 2014) and recent research (e.g. Lee & Penrod, 2022; Rubínová et al., 2021), Gourevitch and Galanter's (1967)  $G$  test, was used to make statistical comparisons between  $d'$  scores. Note, that statistical comparisons are limited to traditional measures of  $d'$  only (i.e.  $d'(TR)$ ). Descriptive comparisons of the other three measures of discrimination (i.e.  $d'(TFp)$ ,  $d'(RFa)$ ,  $d'(FaFp)$ ) are included to provide additional insight into the differences observed. Table 3 reports multi- $d'$  estimates for each experimental condition.

As each  $G$  test only allows for the comparison of two  $d'$  scores, initial comparisons focused on the main effects of appearance, procedure and position. Of the three  $G$  tests conducted, one was significant (at Bonferroni corrected  $\alpha = 0.17$ ),  $G = 4.29$ ,  $p < .001$ . This test showed that participants' capacity to correctly discriminate between

**Table 3.** Multi- $d'$  as a function of appearance, procedure, and position.

			$d'(TF_p)$	$d'(RF_a)$	$d'(F_a F_p)$	$d'(TR)$
No change	Simultaneous	Early	1.81	-0.63	0.41	2.03
		Late	1.56	-1.04	0.61	1.99
	Sequential	Early	2.03	-0.54	0.62	1.95
		Late	0.90	-0.88	0.15	1.64
Change	Simultaneous	Early	0.51	-1.05	0.31	1.25
		Late	0.52	-0.77	0.13	1.16
	Sequential	Early	0.63	-0.94	0.31	1.25
		Late	0.12	-0.76	0.08	0.80

Note:  $d'(TFp)$  = capacity to discriminate between targets and foils in TP lineups;  $d'(RFa)$  = capacity to discriminate between replacements and foils in TA lineups;  $d'(FaFp)$  = differential appeal of foils in TA lineups from foils in TP lineups; and  $d'(TR)$  = capacity to discriminate between targets and replacements.

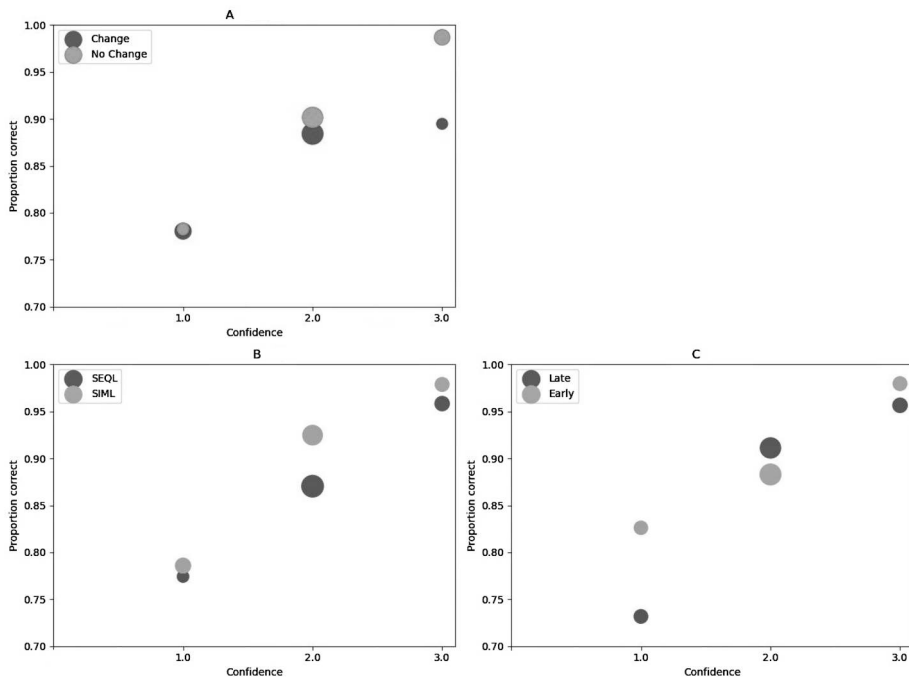
targets (target present) and replacements (target absent) was significantly higher in no change than change conditions (1.89 vs. 1.12). Examination of the multi- $d'$  measures, suggests that the advantage in discrimination accuracy observed in no change conditions, was primarily driven by differences in the underlying memory strength distributions within target present lineups. Indeed, as shown in Table 3, participants' capacity to discriminate between targets and foils in target-present lineups (i.e.  $d'[TFp]$ ), was far higher in no change than change conditions (1.52 vs. 0.45). Comparatively, participants' capacity to discriminate between replacements and foils in target-absent lineups (i.e.  $d'[RFa]$ ) did not appear to vary across no change and change conditions ( $-0.78$  vs.  $-0.88$ ). Further, the differential appeal of foils in target absent and target present lineups (i.e.  $d'[FaFp]$ ) showed limited variability across appearance conditions (0.41 vs. 0.21).

As the main effect of appearance was well established, 12 additional G tests were conducted within both levels of the appearance manipulation to test for higher order interactions between  $d'$  scores that varied as a function of procedure and position. Of these tests, none reached statistical significance (at either Bonferroni corrected  $\alpha = .004$ , or conventional  $\alpha = .05$ ).

### Confidence-specific accuracy

To evaluate the impact of appearance, procedure and position on confidence-specific accuracy, confidence-accuracy characteristic (CAC) curves were computed (see Mickes, 2015). CAC curves provide a visual indication as to whether high post-decisional confidence is associated with high decision accuracy, by plotting the probability that a target/replacement identification is accurate, at each specified level of confidence. Consistent with previous research (e.g. Arndorfer & Charman, 2022; Manley et al., 2021), confidence ratings were collapsed (i.e. binned) into three groups, to reflect low (1-3), medium (4-5) and high (6) confidence. Incorrect replacement identifications were made infrequently (i.e. in six of the eight experimental conditions, zero incorrect replacement identifications were recorded at one or more levels of confidence), therefore, CAC curves were computed to assess the reliability of target/replacement identifications made at low, medium and high confidence, as a function of each main effect only.

Figure 1(A–C) presents CAC curves that plot the relationship between confidence and accuracy, as a function of appearance, procedure and position respectively. Note, points on the CAC plots vary in size, reflecting differences in the number of participant decisions obtained at each level of confidence (i.e. larger points reflect a greater number of participant decisions). As shown in the figures, all CAC curves have a positive trajectory, indicating that identification accuracy increased with confidence, regardless of the experimental manipulation. Figure 1(B and C) shows that the diagnostic value of confidence remains similar across procedure type and position placement. That is, target/replacement identifications made from simultaneous and sequential lineups, and early and late suspect positions, were likely to be accurate when associated with the highest level of confidence. While a similar pattern was observed regarding appearance, at the highest level of confidence, participants' identification accuracy sharply decreased (though remained measurably high) in change (89%) relative to no change (98%) conditions.



**Figure 1.** Confidence accuracy characteristic curves (generated using the Python toolkit, Pywitness; Mickes et al., 2022) for appearance (A), procedure (B) and position (C). Note the CAC curves do not have error bars because too few incorrect replacement IDs were made at all levels of confidence to compute stable bootstrap estimates. On the x-axis, numbers 1.0, 2.0, and 3.0 denote, low, medium and high confidence, respectively.

## Discussion

The present study examined the influence of appearance, procedure, and position on identification decisions, post-decisional confidence ratings, and estimates of discrimination and confidence-specific accuracy. Regarding appearance, consistent with the proposed hypothesis, correct identifications and mean post-decisional confidence in correct identifications both decreased in target-present lineups when target appearance changed. These findings support the assumptions of signal detection theory and demonstrate that when removal-based naturalistic appearance changes were made, match-to-memory signals less frequently exceeded participants' decision criteria; and when they did, the distance with which match-to-memory exceeded participants' decision criteria decreased. The findings are also consistent with previous research which has found that when target appearance is disguised, targets are much less likely to be correctly identified (e.g. Cutler et al., 1987; Mansour et al., 2012; O'Rourke et al., 1989; Pozzulo & Balfour, 2006; Pozzulo & Marciniak, 2006). The present findings add to this body of knowledge by demonstrating that removal-based day-to-day naturalistic appearance changes contribute to a substantial decline in identification decision accuracy. In this regard, the findings are consistent with the one prior study to have examined additive day-to-day naturalistic appearance changes in an identification context (Gronlund et al., 2009). Contrary to some findings in the face recognition literature (e.g. Righi et al., 2012; Terry, 1994),

the current findings in combination with Gronlund et al. suggest that both removal-based and additive day-to-day naturalistic appearance changes have a comparable effect, and that these may be as disruptive to eyewitness decision-making as distinct-disguise related changes. It is important to acknowledge, however, that additional research is needed that directly compares different appearance change manipulations (e.g. additive and removal-based; and naturalistic and disguise-related) to better understand how appearance impacts the internal cognitive processes underpinning recognition.

In target-absent lineups, correct non-identifications were not affected by appearance but mean post-decisional confidence in correct non-identifications increased in change conditions compared to no change conditions. This finding is consistent with the assumptions of decisional locus theories of confidence processing, such as signal detection theory (Sauer et al., 2008), and demonstrates that when lineup members appeared less like the target in change conditions, the distance with which match-to-memory fell short of participants' decision criteria increased (Horry & Brewer, 2016). Lineup members were selected to match the appearance of the unchanged target, therefore, participants may have felt more confident making non-identifications in change conditions, because there were fewer latent similarities between lineup members and targets that could resonate in memory, generating feelings of match. While this explanation is somewhat intuitive, it is worth noting that the one known study to have reported how appearance impacts on confidence in correct non-identifications reported contrary findings (Mansour et al., 2012). Mansour et al. observed that the addition of a hat or sunglasses at encoding had no effect, and reduced confidence in correct non-identifications, respectively. Although difficult to draw clear conclusions, the divergent findings of Mansour et al. and the present research might indicate that the influence of appearance on post-decisional confidence is contingent upon the types of changes made. To avoid speculation, additional research is needed to directly compare the impacts of varied appearance manipulations on post-decisional confidence across decision types, with specific consideration for how the manipulations obscure or alter featural and configurational facial properties.

Findings for group-level estimates of decision accuracy were also consistent with the proposed hypothesis, as both discrimination and confidence-specific accuracy decreased when target appearance changed. Regarding discrimination, the current findings revealed that when appearance changed, the decrease in discrimination of the target from the replacement (i.e.  $d'(TR)$ ), was driven primarily by the decrease in discrimination of the target from foils in target-present lineups (i.e.  $d'(TP)$ ). That is, participants were less capable of accurately distinguishing changed targets from foils but were not more likely to erroneously distinguish replacements from foils in change conditions. That the effect of appearance varied as a function of target presence is not surprising, because in target-absent conditions participants only view a single representation of the target at encoding, and therefore no actual 'change' to appearance occurs. This is not to say, however, that the current findings suggest that estimates of decision accuracy as they relate to replacement identifications, are completely immune to the effects of appearance. In fact, the relationship between high-confidence target/replacement identifications (i.e. suspect identifications) was notably impacted by appearance, such that the odds that an identification made at the highest level of confidence was the replacement rather than

the target increased to 1 in 10 in change conditions, compared to just 1 in 50 in no change conditions.

Prior research on appearance in lineup-based recognition contexts has not typically reported traditional estimates of either discrimination or confidence-specific accuracy measures. Nevertheless, the current findings regarding discrimination are broadly consistent with research that has found appearance has a differential influence across target present and absent lineups (Mansour et al., 2012; Pozzulo & Marciniak, 2006), and aligns with recent research which has shown that estimator variables impacting memory (e.g. distance to the target, exposure duration, retention interval, availability of external features) have a considerable deleterious effect on traditional estimates of discrimination accuracy (Giacona et al., 2021; Manley et al., 2021; Semmler et al., 2018). Further, the confidence-specific accuracy findings support the one recently published study to have examined the association between confidence and accuracy with reference to appearance, which found that confidence-specific accuracy declined sharply as target similarity between encoding and retrieval weakened (Charman et al., 2022).

Regarding procedure and position, both had a limited impact on participants' lineup decisions and did not affect group-level estimates of decision accuracy. In fact, there were no significant main effects of procedure on target present or absent identifications or estimates of decision accuracy. There was, however, a non-significant trend in the data favouring a simultaneous lineup advantage, which is consistent with recently published research (e.g. Seale-Carlisle et al., 2019). Position, by comparison, did influence participants' target present identifications, though its effect was mostly conditional on procedure. Within sequential target-present lineups, correct identifications were made more frequently when targets were positioned early rather than late, and incorrect foil identifications were made more frequently when targets were positioned late rather than early. The findings are consistent with several studies, which have observed that participants 'spend' their identifications on foils preceding the target when targets are positioned late in sequential lineups, (Carlson et al., 2016; Clark & Davey, 2005). However, the findings do not support the assumptions of the diagnostic-feature detection hypothesis, and directly contrast with the emerging body of research which has found evidence of a late position, sequential advantage (Carlson et al., 2008; Gronlund et al., 2009; Meisters et al., 2018). Although not clear cut, it is possible that methodological differences, particularly those associated with lineup composition, have contributed to the discrepant findings between studies.

Although the effect of procedure was relatively limited, it did impact on post-decisional confidence ratings, with confidence higher for incorrect and correct non-identifications, and lower for incorrect foil identifications, in simultaneous than sequential lineups. This finding is mostly inconsistent with prior research, which has reported a range of variable outcomes (e.g. Mansour et al., 2012; Lindsay & Wells, 1985). In fact, to the best of the authors' knowledge, only one published study has found that participants display higher mean post-decisional confidence in non-identifications made from simultaneous than sequential lineups (Gronlund et al., 2009). Additional research is, therefore, needed to better understand the variable influence of procedure on post-decisional confidence, and to test Dobolyi and Dodson's (2013) suggestion that methodological differences likely contributed to these inconsistent findings.

## ***Practical implications***

This research has several important implications for the criminal justice system. First, the findings regarding appearance provide some evidence to suggest that eyewitness researchers and policy makers may have overestimated the accuracy with which eyewitnesses are able to make decisions from lineups. The rather sobering assessment presented in this paper, was that participants may rely almost exclusively on person-specific cues available at encoding and retrieval to make inferences regarding identity, and yet even naturalistic changes that occur in a matter of days are enough to consistently derail match-to-memory, and subsequent decision accuracy. Indeed, participants seemed unable to reliably look beyond appearance to establish identity, with many erroneously adjudging that a changed target was a different person to the target seen at encoding, as opposed to the same person with different features. That there do not appear to be any feasible and practicable solutions in real-world contexts, particularly to the issue of day-to-day appearance changes, is cause for some concern. Charman et al. (2022) proposed police agencies might be able to mitigate the effects of appearance by ensuring ‘up-to-date’ suspect photographs (i.e. obtained at the time of arrest) be used in lineups. This proposition while sensible, would still have limited utility in cases where a suspect is not apprehended at the scene of a crime. As this study and prior research show, changes which occur naturally over a span of days can be as distortive as deliberate disguises. For this reason, policy makers and eyewitness researchers need to be cognisant that in real-world cases, naturalistic changes to suspect appearance can occur easily and often, and will have a notable detrimental influence on eyewitness decision-making.

Second, the findings regarding confidence-specific accuracy suggest recent claims that highly confident eyewitnesses are likely to be accurate in real-world contexts need to be interpreted with caution. It is important to clarify that the present findings do not challenge the view that confidence and identification decision accuracy (for positive identifications) are, for the most part, well-calibrated and robustly associated. In fact, the present findings clearly show that as confidence in positive identifications increases, generally so too does decision accuracy. The present findings do suggest, however, that the generalisability of this relationship to real-world contexts may be limited by the presence of estimator variables like appearance. In this paper, the accuracy of identifications made at the highest level of confidence, experienced a decline when target appearance changed. This meant, that relative to no change conditions, highly confident witnesses in change conditions were much more likely to mistakenly identify a replacement (i.e. innocent suspect). This finding directly contributes to a worrying trend observed in the recently published literature, which has shown that the presence of estimator variables can undermine the relationship between confidence and accuracy (e.g. Charman et al., 2022; Giacona et al., 2021), reducing both the perceived and probative value of confidence in investigations and subsequent trials. Given this trend, and the propensity for naturalistic appearance changes to occur in the real world, it remains unclear if confidence should be relied on in actual cases to measure or predict the decision accuracy of eyewitnesses. It is important to acknowledge, however, that the present study lacked sufficient data points to compute inferential confidence intervals around specific CAC points. Therefore, additional research that varies the base rates of target present and absent lineups (e.g. Giacona et al., 2021), and/or obtains a much larger sample (e.g. Colloff et al., 2016) is needed to determine the

replicability of the current findings, and further explore the impact of appearance on confidence-specific accuracy. In the meantime, decision-makers and triers of fact should exhibit caution when evaluating the testimony of highly confident eyewitnesses and consider the possibility and probability that estimator variables found to influence the relationship between confidence and accuracy were present or occurred.

Third, the findings regarding procedure and position demonstrate that neither system variable was able to mitigate or minimise the deleterious effects of appearance. These findings are especially concerning given the simple nature of the experimental lineup task (e.g. short retention, front-on image only, no stress) and the subtlety of the appearance manipulations adopted in this study. In real-life settings, where retention is likely to be longer (i.e. further reducing memory strength) and there are likely to be other event and witness factors present that can distort memory, decision accuracy may be even poorer, regardless of how key system variables are manipulated. Unfortunately, there do not appear to be any immediate steps the criminal justice system can take to safeguard identification decisions from the damaging influence of estimator variables like appearance. Therefore, moving forward, researchers must be encouraged to develop and explore the utility of radical and novel alternative system factors (Wells et al., 2006). In exploring alternatives, researchers should consider approaches that currently exist outside the realm of what is considered reasonable or practicable, as well as those that eschew the constraints of the enduring but flawed identification paradigm. For example, researchers could examine if promising alternative approaches to identification, that involve collecting multiple non-categorical confidence or similarity judgments in place of (e.g. Brewer & Doyle, 2021; Jordan, 2021; Sauer et al., 2008; Zwart, 2016), prior to (Carlson et al., 2019a), or following (e.g. Huang & Fitzgerald, 2023; Smith et al., 2023) a single categorical decision, are as susceptible to the influences of appearance and other estimator variables.

### ***Limitations and future research***

Like all experimental research, the present study was subject to several limitations. First, still photographs of the target, taken from a single angle were used as encoding stimuli. Still, photographs are not representative of a criminal event in the real world, where eyewitnesses often encode a moving perpetrator, from multiple angles, while being exposed to several additional contextual factors. The use of still photographs, therefore, has limited ecological validity, which may affect the generalisability of the findings. While presenting participants with a mock crime video at encoding may better capture the 'mundane realism' associated with a criminal event, to the best of the authors' knowledge there were no openly available mock crime materials which captured subtle, day-to-day naturalistic changes to target appearance. As the present findings suggest that subtle changes to appearance can have a substantial impact on participants' recognition capabilities, future research should develop and utilise more ecologically valid encoding stimuli to better understand how these appearance changes impact witnesses in an applied context.

Second, participants in this study completed four lineup trials. In actual criminal investigations, eyewitnesses are rarely required to respond to multiple lineups with different targets. Within an experimental context, participants presented with multiple lineups

might experience practice or learning effects and amend how they respond to each lineup in the sequence (Van Lehn, 1996). However, recent research suggests that participants can complete many lineup trials in a sequence (e.g. 24), with no evidence of practice or learning effects (Mansour et al., 2017).

Third, the appearance manipulation adopted in the present study altered three person-specific characteristics (i.e. a reduction in hair length, the removal of stubble, and a change of clothing). While these changes could all occur on a day-to-day basis, research on context effects in person identification suggests that the changes likely had a cumulative effect (Righi et al., 2012; Terry, 1994; Thomson et al., 1982), exacerbating the size of the observed difference in identifications across no change and change conditions. Additional research, therefore, should focus on the development of research materials which systematically vary the appearance of selected target persons, allowing researchers to better quantify the contribution that individual changes to target appearance have on lineup-based decision-making. By doing this, researchers will be able to further explore if the influence of appearance is a consequence of specific featural, or general configurational changes.

## Conclusion

The inferences that eyewitnesses make regarding the *likely identity* of an unknown perpetrator, appear to be based primarily on the person-specific cues, or visual patterns of information available during the criminal event and during a lineup presentation. The present findings highlight one of the pitfalls of relying primarily on person-specific information to form identity judgements, because appearances can be deceiving. As discussed, making naturalistic changes to target appearance contributed to a large decline in correct identifications, confidence in correct identifications, and overall estimates of decision accuracy. Further, two important system-controlled variables were unable to provide any improvements regarding the accuracy or reliability of identification decisions, when appearance had changed. Collectively, these findings highlight the risks of presenting what appear to be subjective assessments of match between a perceptual representation of a person and a memorial representation of a person, in terms of sameness (i.e. identity). For when the appearance of a perpetrator presented as part of a lineup does not provide an exact match to the perpetrator seen during a criminal event, witnesses tend to mistake these representations as different people rather than the same person with different features. Therefore, in real cases involving unknown suspects, it is important that triers of fact and the criminal justice system continue to treat identifications and the confidence associated with them, with the utmost caution.

## Notes

1. Research has shown that people process and recognise familiar (i.e. known) and unfamiliar (i.e. unknown) faces in qualitatively different ways (e.g. Hancock et al., 2000). Therefore, it is important to acknowledge that this paper focuses on the identification of unknown targets (i.e. perpetrators) only.
2. To avoid confusion, acronyms from the original multi-d' model have been retitled to reflect the terminology used in this paper. For example, as the terms target (T) and replacement (R)



were used in place of guilty (G) and innocent suspect (I) respectively, the acronym d'(TR) is used in place of d'(GI).

## Acknowledgements

The authors would like to thank Pamela Henry for her support throughout the duration of the research, and Aimee Wrightson-Hester for her helpful comments and feedback on draft versions of the paper. The first author would like to acknowledge the support of the Australian Government through an Australian Government Research Training Program Scholarship.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Ethics statement

Ethics approval for the research presented in this study was granted by the Edith Cowan University Human Research Ethics Committee (ref. 17470), and the Goldsmiths Research Ethics and Integrity Sub-Committee (ref. 1517). The guidelines of these committees were followed.

## Data availability statement

The data that support the findings of this study are openly available in The Open Science Framework (OSF) at <https://doi.org/10.17605/OSF.IO/4XR6Z>.

## ORCID

Dominic T. Jordan  <http://orcid.org/0000-0002-6537-7330>

Adrian J. Scott  <http://orcid.org/0000-0002-4171-6318>

## References

- Arndorfer, A., & Charman, S. D. (2022). Assessing the effect of eyewitness identification confidence assessment method on the confidence-accuracy relationship. *Psychology, Public Policy, and Law*, 28(3), 414–432. <https://doi.org/10.1037/law0000348>
- Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 929–945. <https://doi.org/10.1037/0096-1523.24.3.929>
- Bergold, A. N., & Heaton, P. (2018). Does filler database size influence identification accuracy? *Law and Human Behavior*, 42(3), 227–243. <https://doi.org/10.1037/lhb0000289>
- Brewer, N. (2006). Uses and abuses of eyewitness identification confidence. *Legal and Criminological Psychology*, 11(1), 3–23. <https://doi.org/10.1348/135532505X79672>
- Brewer, N., & Doyle, J. (2021). Changing the face of police lineups: Delivering more information from witnesses. *Journal of Applied Research in Memory and Cognition*, 10(2), 180–195. <https://doi.org/10.1016/j.jarmac.2020.12.004>
- Brewer, N., & Wells, G. L. (2011). Eyewitness identification. *Current Directions in Psychological Science*, 20(1), 24–27. <https://doi.org/10.1177/0963721410389169>

- Carlson, C. A., & Carlson, M. A. (2014). An evaluation of lineup presentation, weapon presence, and a distinctive feature using ROC analysis. *Journal of Applied Research in Memory and Cognition*, 3(2), 45–53. <https://doi.org/10.1016/j.jarmac.2014.03.004>
- Carlson, C. A., Carlson, M. A., Weatherford, D. R., Tucker, A., & Bednarz, J. (2016). The effect of back-loading instructions on eyewitness identification from simultaneous and sequential lineups. *Applied Cognitive Psychology*, 30(6), 1005–1013. <https://doi.org/10.1002/acp.3292>
- Carlson, C. A., Gronlund, S. D., & Clark, S. E. (2008). Lineup composition, suspect position, and the sequential lineup advantage. *Journal of Experimental Psychology: Applied*, 14(2), 118–128. <https://doi.org/10.1037/1076-898X.14.2.118>
- Carlson, C. A., Jones, A. R., Goodsell, C. A., Carlson, M. A., Weatherford, D. R., Whittington, J. E., & Lockamy, R. F. (2019a). A method for increasing empirical discriminability and eliminating top-row preference in photo arrays. *Applied Cognitive Psychology*, 33(6), 1091–1102. <https://doi.org/10.1002/acp.3551>
- Carlson, C. A., Jones, A. R., Whittington, J. E., Lockamy, R. F., Carlson, M. A., & Wooten, A. R. (2019b). Lineup fairness: Propitious heterogeneity and the diagnostic feature-detection hypothesis. *Cognitive Research: Principles and Implications*, 4(20), 1–16. <https://doi.org/10.1186/s41235-019-0185-0>
- Charman, S. D., Shambaugh, L. J., Cahill, B. S., & Molinaro, P. F. (2022). The ability to infer witness accuracy from high-confidence lineup identifications is undermined by the appearance-change instruction and target appearance change. *Psychology, Public Policy, and Law*, 28(4), 491–504. <https://doi.org/10.1037/law0000368>
- Charman, S. D., & Wells, G. L. (2007). Eyewitness lineups: Is the appearance-change instruction a good idea? *Law and Human Behavior*, 31(1), 3–22. <https://doi.org/10.1007/s10979-006-9006-3>
- Clark, S. E. (2011). Blackstone and the balance of eyewitness identification evidence. *Albany Law Review*, 74(3), 1105–1156. <http://www.heinonline.org>
- Clark, S. E., & Davey, S. L. (2005). The target to foil shift in simultaneous and sequential lineups. *Law and Human Behavior*, 29(2), 151–172. <https://doi.org/10.1007/s10979-005-2418-7>
- Clark, S. E., Howell, R. T., & Davey, S. L. (2008). Regularities in eyewitness identification. *Law and Human Behavior*, 32(3), 187–218. <https://doi.org/10.1007/s10979-006-9082-4>
- Colloff, M. F., & Wixted, J. T. (2020). Why are lineups better than showups? A test of the filler siphoning and enhanced discriminability accounts. *Journal of Experimental Psychology: Applied*, 26(1), 124–143. <https://doi.org/10.1037/xap0000218>
- Colloff, M. K., Wade, K. A., & Strange, D. (2016). Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychological Science*, 27(9), 1227–1239. <https://doi.org/10.1177/0956797616655789>
- Colloff, M. K., Wade, K. A., Wixted, J. T., & Maylor, E. A. (2017). A signal-detection analysis of eyewitness identification across the adult lifespan. *Psychology and Aging*, 32(3), 243–258. <https://doi.org/10.1037/pag0000168>
- Cutler, B. L., & Penrod, S. D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. *Journal of Applied Psychology*, 73(2), 281–290. <https://doi.org/10.1037/0021-9010.73.2.281>
- Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). Improving the reliability of eyewitness identification: Putting context into context. *Journal of Applied Psychology*, 72(4), 629–637. <https://doi.org/10.1037/0021-9010.72.4.629>
- Davies, G., & Flin, R. (1984). The man behind the mask – disguise and face recognition. *Human Learning: Journal of Practical Research & Applications*, 3(2), 83–95.
- Dobolyi, D. G., & Dodson, C. S. (2013). Eyewitness confidence in simultaneous and sequential lineups: A criterion shift account for mistaken identification overconfidence. *Journal of Experimental Psychology*, 19(4), 345–357. <https://doi.org/10.1037/a0034596>
- Erickson, W. B., Lampinen, J. M., Frowd, C. D., & Mahoney, G. (2017). When age-progressed images are unreliable: The roles of external features and age range. *Science & Justice*, 57(2), 136–143. <https://doi.org/10.1016/j.scijus.2016.11.006>
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed). Sage.

- Giacona, A. M., Lampinen, J. M., & Anastasi, J. S. (2021). Estimator variables can matter even for high-confidence lineup identifications made under pristine conditions. *Law and Human Behavior, 45*(3), 256–270. <https://doi.org/10.1037/lhb0000381>
- Gourevitch, V., & Galanter, E. (1967). A significance test for one parameter isosensitivity functions. *Psychometrika, 32*(1), 25–33. <https://doi.org/10.1007/BF02289402>
- Gronlund, S. D., Carlson, C. A., Dailey, S. B., & Goodsell, C. A. (2009). Robustness of the sequential lineup advantage. *Journal of Experimental Psychology, 15*(2), 140–152. <https://doi.org/10.1037/a0015082>.
- Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating eyewitness identification procedures using ROC analysis. *Current Directions in Psychological Science, 23*(1), 3–10. <https://doi.org/10.1177/0963721413498891>
- Gross, R., Matthews, I., Cohn, J., Kanade, T., & Baker, S. (2010). Multi-PIE. Proceedings of the international conference on automatic face and gesture recognition. *IEEE International Conference on Automatic Face & Gesture Recognition, 28*(5), 807–813. <https://doi.org/10.1016/j.imavis.2009.08.002>.
- Hancock, P. J., Bruce, V. V., & Burton, M. A. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences, 4*(9), 330–337. [https://doi.org/10.1016/S1364-6613\(00\)01519-9](https://doi.org/10.1016/S1364-6613(00)01519-9)
- Hope, L., & Sauer, J. D. (2014). Eyewitness memory and mistaken identifications. In M. Yves (Ed.), *Investigative interviewing: The essentials* (pp. 97–124). Carswell.
- Horry, R., & Brewer, N. (2016). How target–lure similarity shapes confidence judgments in multiple-alternative decision tasks. *Journal of Experimental Psychology: General, 145*(12), 1615–1634. <https://doi.org/10.1037/xge0000227>
- Horry, R., Halford, P., Brewer, N., Milne, R., & Bull, R. (2014). Archival analysis of eyewitness identification test outcomes: What can they tell us about eyewitness memory? *Law and Human Behavior, 38*(1), 94–108. <https://doi.org/10.1037/lhb0000060>
- Horry, R., Palmer, M. A., & Brewer, N. (2012). Backloading in the sequential lineup prevents within-lineup criterion shifts that undermine eyewitness identification performance. *Journal of Experimental Psychology: Applied, 18*(4), 346–360. <https://doi.org/10.1037/a0029779>
- Huang, C., & Fitzgerald, R. (2023, March 16–18). *An alternative approach to the rule-out lineup procedure* [Conference Session]. American Psychology-Law Society Annual Conference, Philadelphia, PA, United States. <https://ap-ls.org/conferences/schedule.html>
- Jordan, D. T. (2021). *Identity: A crisis of confidence? Or is it resemblance? An exploration of the different approaches by which eyewitness evidence can be obtained from lineups* [Doctoral thesis, Edith Cowan University]. Research Online Institutional Repository <https://ro.ecu.edu.au/theses/2449>
- Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence accuracy correlation. *Journal of Experimental Psychology, 22*(5), 1304–1316. <https://doi.org/10.1027-7393/96/53.00>.
- Lee, J., & Penrod, S. D. (2019). New signal detection theory-based framework for eyewitness performance in lineups. *Law and Human Behavior, 43*(5), 436–454. <https://doi.org/10.1037/lhb0000343>
- Lee, J., & Penrod, S. D. (2022). Three-level meta-analysis of the other-race bias in facial identification. *Applied Cognitive Psychology, 36*(5), 1106–1130. <https://doi.org/10.1002/acp.3997>
- Leippe, M. R., Eisenstadt, D., & Rauch, S. M. (2009). Cueing confidence in eyewitness identifications: Influence of biased lineup instructions and pre-identification memory feedback under varying lineup conditions. *Law and Human Behavior, 33*(3), 194–212. <https://doi.org/10.1007/s10979-008-9135-y>
- Levi, A. M., & Jungman, N. (1995). The police lineup: Basic weaknesses and radical solutions. *Criminal Justice and Behaviour, 22*(4), 347–372. <https://doi.org/10.1177/0093854895022004001>
- Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identification from line-ups: Simultaneous versus sequential line-up presentation. *Journal of Applied Psychology, 70*(3), 556–564. <https://doi.org/10.1037/0021-9010.70.3.556>
- Lucas, C. A., & Brewer, N. (2022). Could precise and replicable manipulations of suspect-filler similarity optimize eyewitness identification performance? *Psychology, Public Policy, and Law, 28*(1), 108–122. <https://doi.org/10.1037/law0000329>

- Macmillan, N. A., & Creelman, D. C. (1991). *Detection theory: A user's guide*. Cambridge University Press.
- Malpass, R. (1981). Effective size and defendant bias in eyewitness identification lineups. *Law and Human Behaviour*, 5(4), 299–309. <https://doi.org/10.1007/BF01044945>
- Manley, K. D., Chan, J. C. K., & Wells, G. L. (2021). Improving face identification accuracy of mask wearing individuals. *Cognitive Research: Principles and Implications*, 7(27), 1–17. <https://doi.org/10.1186/s41235-022-00369-7>.
- Mansour, J. K., Beaudry, J. L., Bertrand, M. I., Kalmes, N., Melsom, E. I., & Lindsay, R. C. L. (2012). Impact of disguise on identification decision and confidence with simultaneous and sequential lineups. *Law and Human Behavior*, 36(6), 513–526. <https://doi.org/10.1037/h0093937>
- Mansour, J. K., Beaudry, J. L., & Lindsay, R. C. L. (2017). Are multiple lineup trial experiments appropriate for eyewitness identification studies? Accuracy, choosing and confidence across trials. *Behaviour Research Methods*, 49(6), 2235–2254. <https://doi.org/10.3758/s13428-017-0855-0>
- Meisters, J., Deidendorf, B., & Musch, J. (2018). Eyewitness identification in simultaneous and sequential lineups: An investigation of position effects using receiver operating characteristics. *Memory (Hove, England)*, 26(9), 1297–1309. <https://doi.org/10.1080/09658211.2018.1464581>
- Memon, A., & Gabbert, F. (2003). Unravelling the effects of sequential presentation in culprit-present lineups. *Applied Cognitive Psychology*, 17(6), 703–714. <https://doi.org/10.1002/acp.909>
- Mickes, L. (2015). Receiver operating characteristic analysis and confidence-accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition*, 4(2), 93–102. <https://doi.org/10.1016/j.jarmac.2015.01.003>
- Mickes, L., Moreland, M. B., Clark, S. E., & Wixted, J. T. (2014). Missing the information needed to perform ROC analysis? Then compute  $d'$ , not the diagnosticity ratio. *Journal of Applied Research in Memory and Cognition*, 3(2), 58–62. <https://doi.org/10.1016/j.jarmac.2014.04.007>
- Mickes, L., Seale-Carlisle, T. M., Chen, X., & Boogert, S. (2022). *Pywitness 1.0: A python eyewitness identification analysis toolkit*. *PsyArXiv*. <https://doi.org/10.31234/osf.io/5ruks>
- Molinaro, P. F., Arndorfer, A., & Charman, S. D. (2013). Appearance-change instruction effects on eyewitness lineup identification accuracy are not moderated by amount of appearance change. *Law and Human Behavior*, 37(6), 432–440. <http://doi.org/10.1037/lhb0000049>
- O'Rourke, T. E., Penrod, S. D., Cutler, B. L., & Stuve, T. E. (1989). The external validity of eyewitness identification research: Generalising across subject populations. *Law and Human Behavior*, 13(4), 385–395. <https://doi.org/10.1007/BF01056410>
- Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less biased criterion setting but does not improve discriminability. *Law and Human Behavior*, 36(3), 247–255. <https://doi.org/10.1037/h0093923>
- Pozzulo, J. D., & Balfour, J. (2006). Children's and adults' eyewitness identification accuracy when a culprit changes his appearance: Comparing simultaneous and elimination lineup procedures. *Legal and Criminological Psychology*, 11(1), 25–34. <https://doi.org/10.1348/135532505X52626>
- Pozzulo, J. D., & Marciniak, S. (2006). Comparing identification procedures when the perpetrator has changed appearance. *Psychology, Crime & Law*, 12(4), 429–438. <https://doi.org/10.1080/10683160500050690>
- Righi, G., Peissig, J. J., & Tarr, M. J. (2012). Recognising disguised faces. *Visual Cognition*, 20(2), 143–169. <https://doi.org/10.1080/13506285.2012.654624>
- Rubínová, E., Fitzgerald, R. J., Juncu, S., Ribbers, E., Hope, L., & Sauer, J. D. (2021). Live presentation for eyewitness identification is not superior to photo or video presentation. *Journal of Applied Research in Memory and Cognition*, 10(1), 167–176. <https://doi.org/10.1016/j.jarmac.2020.08.009>
- Sauer, J. D., & Brewer, N. (2015). Confidence and accuracy of eyewitness identification. In T. Valentine, & J. Davis (Eds.), *Forensic facial identification: Theory and practice of identification from eyewitnesses, composites and CCTV* (pp. 185–208). Wiley Blackwell.
- Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General*, 137(3), 528–547. <https://doi.org/10.1037/a0012712>

- Seale-Carlisle, T. M., & Mickes, L. (2016). US line-ups outperform UK line-ups. *Royal Society Open Science*, 3(9), 160–300. <https://doi.org/10.1098/rsos.160300>.
- Seale-Carlisle, T. M., Wetmore, S. A., Flowe, H. D., & Mickes, L. (2019). Designing police lineups to maximize memory performance. *Journal of Experimental Psychology: Applied*, 25(3), 410–430. <https://doi.org/10.1037/xap0000222>
- Semmler, C. A., 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>
- Shapiro, P. N., & Penrod, S. D. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, 100(2), 139–156. <https://doi.org/10.1037/0033-2909.100.2.139>
- Smith, A. M., Ayala, N. T., & Ying, R. C. (2023). The rule out procedure: A signal-detection-informed approach to the collection of eyewitness identification evidence. *Psychology, Public Policy, and Law*, 29(1), 19–31. <https://doi.org/10.1037/law0000373>
- Smith, A. M., Wells, G. L., Lindsay, R. C. L., & Penrod, S. D. (2017). Fair lineups are better than biased lineups and showups, but not because they increase underlying discriminability. *Law and Human Behavior*, 41(2), 127–145. <https://doi.org/10.1037/lhb0000219>
- Stebly, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of the sequential line-up superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy and Law*, 17(1), 99–139. <https://doi.org/10.1037/a0021650>
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multi-variate statistics* (6th ed.). Pearson Education.
- Terry, R. L. (1994). Effect of facial transformations on the accuracy of recognition. *The Journal of Social Psychology*, 134(4), 483–492. <https://doi.org/10.1080/00224545.1994.9712199>
- Thomson, D. M. (1982). The realities of eyewitness identification. *Australian Journal of Forensic Sciences*, 14(4), 150–157. <https://doi.org/10.1080/00450618209411171>
- Thomson, D. M. (1986). Face recognition: More than a feeling of familiarity? In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. Young (Eds.), *Aspects of face processing* (pp. 391–399). Elsevier.
- Thomson, D. M., Robertson, S. L., & Vogt, R. (1982). Person recognition: The effect of context. *Human Learning*, 1, 497–511.
- Tredoux, C. G. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behaviour*, 22(2), 217–237. <https://doi.org/10.1023/A:1025746220886>
- Tredoux, C. G. (1999). Statistical considerations when determining measures of lineup size and lineup bias. *Applied Cognitive Psychology*, 13(5), S9–S26. [https://doi.org/10.1002/\(SICI\)1099-0720\(199911\)13:1+<S9::AID-ACP634>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1099-0720(199911)13:1+<S9::AID-ACP634>3.0.CO;2-1)
- Tulving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning and Verbal Behaviour*, 20(5), 479–496. [https://doi.org/10.1016/S0022-5371\(81\)90129-8](https://doi.org/10.1016/S0022-5371(81)90129-8)
- Tunnicliff, J. L., & Clark, S. E. (2000). Selecting foils for identification lineups: Matching suspects or descriptions? *Law and Human Behavior*, 24(2), 231–258. <https://doi.org/10.1023/A:1005463020252>
- Van Lehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, 47(1), 513–539. <https://doi.org/10.1146/annurev.psych.47.1.513>
- Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(3), 582–600. <https://doi.org/10.1037/0278-7393.26.3.582>
- Weber, N., & Brewer, N. (2004). Confidence-accuracy calibration in absolute and relative face recognition judgements. *Journal of Experimental Psychology: Applied*, 10(3), 156–172. <https://doi.org/10.1037/1076-898X.10.3.156>
- Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology*, 36(12), 1546–1557. <https://doi.org/10.1037/0022-3514.36.12.1546>
- Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, 7(2), 45–75. <https://doi.org/10.1111/j.1529-1006.2006.00027.x>
- Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, 54(1), 277–295. <https://doi.org/10.1146/annurev.psych.54.101601.145028>

- Wells, G. L., Steblay, N. K., & Dysart, J. E. (2011). A test of the simultaneous vs. sequential lineup Methods: An initial report of the AJS national eyewitness identification field studies. *American Judicature Society*. <https://mn.gov/law-library-stat/>
- Wetmore, S. A., McAdoo, R. M., Gronlund, S. D., & Neaushatz, J. S. (2017). The impact of fillers on lineup performance. *Cognitive Research: Principles and Implications*, 2(48), 1–13. <https://doi.org/10.1186/s41235-017-0084-1>
- Wixted, J. T. (2007). Dual process theory and signal detection theory of recognition memory. *Psychological Review*, 114(1), 152–176. <https://doi.org/10.1037/0033-295X.114.1.152>
- Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review*, 121(2), 262–276. <https://doi.org/10.1037/a0035940>
- Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. L. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist*, 70(6), 515–526. <https://doi.org/10.1037/a0039510>
- 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/10.1177/1529100616686966>
- Yarmey, A. D. (2004). Eyewitness recall and photo identification: a field experiment. *Psychology, Crime & Law*, 10(1), 53–68. <http://doi.org/10.1080/1068316021000058379>
- Zwartz, M. (2016). *Identity crisis in identification evidence: Similarity judgments as an alternative to identification decisions* [Doctoral thesis, Deakin University]. Deakin Research Online <http://dro.deakin.edu.au/view/DU:30089085>

## Appendix. Appearance, position, and target variations for the four lineup trials within each randomly allocated viewing condition.

Condition	Lineup 1	Lineup 2	Lineup 3	Lineup 4
1	TP-NC-E	TA-C-L	TP-C-L	TA-NC-E
2	TP-NC-L	TA-C-E	TP-C-E	TA-NC-L
3*	TA-C-E	TP-NC-L	TA-NC-L	TP-C-E
4*	TA-C-L	TP-NC-E	TA-NC-E	TP-C-L
5	TP-C-E	TA-NC-L	TP-NC-L	TA-C-E
6	TP-C-L	TA-NC-E	TP-NC-E	TA-NC-L
7	TA-NC-E	TA-C-L	TA-C-L	TP-NC-E
8	TA-NC-L	TA-C-E	TA-C-E	TP-NC-L

Note: TP/TA = target present/absent, NC/C = no change/change, and E/L = early/late position. \*In the sequential lineup condition, a Qualtrics coding error occurred in viewing conditions 3 and 4, such that participants viewed the late and early variants of lineup 4 respectively, as opposed to the early and late variants (as per the simultaneous lineup condition).