Psychiatry, Psychology and Law, 2024
Vol. 31, No. 5, 932–962, https://dx.doi.org/10.1080/13218719.2023.2242435

The effect of surgical masks on identification decisions from masked and unmasked lineups

Annegrete Palu[a], Aire Raidvee[a], Valeri Murnikov[b] and Kristjan Kask[b]

[a]*Institute of Psychology, University of Tartu, Tartu, Estonia;* [b]*School of Natural Sciences and Health, Tallinn University, Tallinn, Estonia*

While research has shown that wearing a disguise hinders lineup identifications, less is known about how to conduct lineups in cases of disguised perpetrators. We examined the influence of surgical masks, worn during a crime event (encoding) and within lineups (retrieval), on eyewitness identification accuracy. In our experiment, 452 participants watched a mock-crime video and identified the perpetrator from either a target-present or a target-absent simultaneous lineup. Contrary to expectations based on the encoding specificity principle, we did not find that matching the presence of masks in the lineup to the encoding condition increased identification accuracy. Instead, compared to the condition with no masks at encoding and retrieval, the presence of masks at either stage negatively affected discriminability and undermined the predictive utility of confidence and decision time. Our findings indicate that when a witness has encountered a masked perpetrator, presenting them with a masked lineup may not be necessary.

Keywords: disguise; encoding specificity principle; eyewitness; face recognition; identification accuracy; lineups; surgical masks.

Article History: Received 9 January 2023; Accepted 29 June 2023

In recent years, there has been a rise in wearing face coverings such as surgical masks in public places, as people have been instructed to do so in many countries since the beginning of the COVID-19 pandemic (World Health Organization, 2021). The recommendation to cover the nose and mouth provides an opportunity to conceal one's identity (Babwin & Dazio, 2020; Southall & Syckle, 2020), which creates a challenge for law enforcement. The perpetrator's decision to wear a face covering as a disguise during the crime is not under the control of the criminal justice system, making it an estimator variable (Wells, 1978). Thus, in these cases, the police must estimate the potential effect of wearing face coverings on subsequent identification decisions. Research focusing specifically on the nature and scope of the effect of partially occluding the face on subsequent identification allows law enforcement to more accurately predict the reliability of witness identifications. In addition to estimating the effect of disguise on eyewitness decisions, the police also need to decide how to conduct and administer a lineup. As the presence of a disguise can be manipulated during identification, wearing a disguise should also be considered a system variable (as has already been done by Manley et al., 2022). This study aims to determine how surgical masks, both as a disguise during a crime event (an estimator variable) and in a lineup (a system variable), affect eyewitness identification accuracy.

Correspondence: Annegrete Palu, Institute of Psychology, University of Tartu, Näituse 2, Tartu 50409, Estonia. Email: annegrete.palu@ut.ee

Different types of disguises that occlude the face entirely or partially or distort facial features can have varying effects on recognition (Mansour et al., 2020). Research has shown that several concealing disguises (Shapiro & Penrod, 1986), such as eyeglasses and sunglasses (Douma et al., 2012; Graham & Ritchie, 2019; Kramer & Ritchie, 2016; Mansour et al., 2020; Noyes et al., 2021; Patterson & Baddeley, 1977; Righi et al., 2012; Terry, 1994), wigs (Douma et al., 2012; Patterson & Baddeley, 1977; Righi et al., 2012), hats (Cutler et al., 1987a, 1987b), beards (Patterson & Baddeley, 1977; Terry, 1994), nylon stockings (Davies & Flin, 1984; Mansour et al., 2020) and ski masks (Manley et al., 2019, 2022), as well as other disguising techniques (Noyes & Jenkins, 2019), impair subsequent person identification. Recent recognition (Freud et al., 2020, 2021; Garcia-Marques et al., 2022; Guerra et al., 2022; Marini et al., 2021) and face-matching studies (Carragher & Hancock, 2020; Estudillo et al., 2021; Noyes et al., 2021) have demonstrated that wearing surgical masks, which cover the lower part of the face during encoding, also negatively affects face recognition. Together, these findings provide strong evidence that any form of face covering or disguise hinders the recognition of unfamiliar and even familiar faces.

Several mechanisms operating at the encoding and/or retrieval stages, distinctly or intertwined with each other, provide potential explanations for how and why disguises impact eyewitness decisions (Mansour et al., 2020). These explanations in relation to surgical masks will be examined in turn.

Difficulties in remembering faces can be partially attributed to difficulties in encoding those faces during the event they are seen in (Megreya & Burton, 2006). It has been established that witnesses who experience poor encoding conditions are more likely to form a low-quality memory representation of the offender, leading to fewer correct identifications and more identifications of innocent persons (Brewer et al., 2007; A. M. Smith et al., 2019). Although not all disguises, including surgical masks, negatively affect the clarity of encoded information, they do reduce the number of facial features available for encoding. This results in an insufficient memory of the perpetrator and less information available during recognition (Brewer et al., 2005; Mansour et al., 2020). Mansour et al. (2020) investigated how different types of disguises and degrees of coverage influence identification decisions. They found that the combined use of a toque and sunglasses led to fewer correct identifications than when either was worn alone. In addition to the number of features occluded, the size of the disguised area is also important (Noyes et al., 2021; Terry, 1994). Although Mansour et al. (2020) demonstrated that identification accuracy generally decreased as disguises became more extensive, they also found that covering two thirds of the face with a stocking was as effective as covering the entire face. This suggests that identification errors depend not only on the extent of the disguise but also on which specific parts and how many of them are disguised during encoding (Mansour et al., 2020).

Individual facial features are not equally important for successful identification of a face. Several studies have shown that disguising the upper half of the face (e.g. the eyes) is more likely to have a larger negative influence on face recognition accuracy than disguising lower facial features (e.g. the mouth; Davies et al., 1977; Mansour et al., 2020; McKelvie, 1976; Nguyen & Pezdek, 2017; Or et al., 2023). For familiar faces, more accurate face identification is associated with greater reliance on eyes (Royer et al., 2018; Stephan & Caine, 2007) and eyebrows (Sadr et al., 2003). However, for identifying unfamiliar faces, external features of a face (e.g. jaw-line, hair, head-shape, hairline) are just as important (Bruce et al., 1999; Ellis et al., 1979; Johnston & Edmonds, 2009; Logan et al., 2017), at least for people in Western societies (Megreya et al., 2012). For example, it has been shown that compared to changes in eyes, changes in hairstyles are better detected in unfamiliar faces (O'Donnell & Bruce, 2001), and using the information about hairline or forehead leads to higher recognition accuracy of unfamiliar faces (Davies et al., 1977). The enhanced ability to recognize unfamiliar faces based on external features might be due to the relatively large size (Logan et al., 2017), which makes them easily discriminable even in poor conditions (Jarudi & Sinha, 2003). Recently, Abudarham and Yovel (2019) showed that both internal and external features are important in recognizing unfamiliar faces. Although surgical masks do not cover all the facial features that are important for unfamiliar face recognition (i.e. eyes, hair, forehead), they may hinder identification of faces by covering a large portion of the face.

Regardless of features that are covered, disguises may have a detrimental effect on face recognition by disrupting holistic processing of faces (Freud et al., 2020). Faces are processed holistically, meaning that the features of a face are processed as a unified whole instead of as separate components (Farah et al., 1998; Tanaka & Farah, 1993; Tanaka & Simonyi, 2016). Disguised and masked faces, in contrast, are processed in a qualitatively different way (Freud et al., 2020, 2021). By leaving only some features visible, masks disrupt holistic processing and promote feature-based processing (Freud et al., 2020; Stajduhar et al., 2022). Given that holistic processing is associated with face recognition (DeGutis et al., 2013; Richler et al., 2011; Wang et al., 2012), surgical masks may impair holistic processing of faces during encoding and, thus, subsequent face recognition.

The presence of a disguise either at the encoding (Shapiro & Penrod, 1986) or at the retrieval (Garcia-Marques et al., 2022; Guerra et al., 2022; Hockley et al., 1999) stage negatively affects face recognition. Thus, it can be inferred that the difficulty of the identification task increases, which can affect witnesses' choosing behaviour. One theory, supported by the results of Guerra et al. (2022), posits that witnesses perceive the identification of a disguised perpetrator as more challenging, leading them to take fewer risks, and, thus, reduces their likelihood of making an identification (Brewer et al., 2005; Mansour et al., 2020). However, Mansour et al. (2020) found that as the degree of disguise increased, lineup selections decreased more slowly than accuracy. This indicates that even if the participants perceived the task as more difficult, they believed that they should make a decision, and, thus, the likelihood of making an identification increased. Research has shown that when either encoding or retrieval conditions get worse and thereby create a weak sense of recognition, witnesses lower their criterion for making an affirmative identification (A. M. Smith, 2020; A. M. Smith et al., 2019). This trend has also been observed when witnesses infer that their memory trace is weak, independent of the quality of encoding and retrieval conditions (Brewer et al., 2022). Lowering one's decision criteria leads to both a reduction in correct identifications and an increase in mistaken identifications under weak memory conditions (strength-based mirror effect; Glanzer & Adams, 1985; Stretch & Wixted, 1998). This pattern of results has also been observed in face recognition studies with surgical masks (Or et al., 2023). Thus, witnessing a perpetrator wearing a surgical mask can impact not only the memory accuracy but the behaviour of the witness as well.

As can be seen, identification of a disguised perpetrator depends not only on the encoding condition but on the retrieval condition and procedures as well. Therefore, to determine the effect of surgical masks on eyewitness identification accuracy, it is essential to consider the interaction of encoding and retrieval conditions.

When a perpetrator wears a disguise during a crime, but an undisguised lineup is presented to the witness, the negative effect of a disguise can be attributed to a mismatch between encoding and retrieval conditions. The idea that the match/mismatch between the encoding and retrieval conditions affects memory performance is central to the encoding specificity principle (Tulving & Thomson, 1973). In a similar notion, the transfer-appropriate processing principle (Morris et al., 1977) states that for an accurate identification to occur, cognitive processes activated at retrieval should match those at encoding. Since disguised and undisguised faces are processed differently (holistic vs. feature-based processing, respectively), presenting an undisguised lineup after a crime involving a disguised perpetrator could create a discrepancy in the cognitive strategies used, potentially impairing identification accuracy. Collectively these principles suggest that if a perpetrator was disguised during the crime, the lineup should also consist of disguised individuals to increase the chances of an accurate decision.

Research has shown that altering one's appearance between the time of encoding and an identification task negatively affects identification accuracy (Carlson et al., 2021; Patterson & Baddeley, 1977; Shapiro & Penrod, 1986), with a decrease in performance as the number of changes increase (Douma et al., 2012; Righi et al., 2012). More specifically, it seems that removing a disguise is more detrimental to identification accuracy than adding one, as has been shown with multiple types of disguises (Davies & Flin, 1984; Douma et al., 2012; Guerra et al., 2022; Righi et al., 2012; Terry, 1994). In addition to the negative effect of change, and consistent with the encoding specificity principle, several face recognition (Or et al., 2023; Toseeb et al., 2014; Toseeb et al., 2012; Ventura et al., 2023) and face matching (Graham & Ritchie, 2019; Kramer & Ritchie, 2016) studies using disguises or face masks have shown that accuracy (as measured by discriminability) in congruent conditions is greater than that in incongruent conditions (only one face is disguised).

However, some studies have only found partial support for the encoding specificity principle. For instance, both Guerra et al. (2022) and Garcia-Marques et al. (2022) found that the interaction between encoding and retrieval conditions did not fully cross: discriminability was highest for faces without a mask in both phases, but when faces were studied with a mask, discriminability remained the same, regardless of whether the faces were presented with or without a mask during retrieval. Notably, Manley et al. (2019) found an opposite pattern: although discriminability in matching conditions was numerically higher than in incongruent conditions, it did not differ for masked and unmasked lineups after an unmasked perpetrator was encoded. Davies and Flin (1984) found that correct identifications for faces disguised both at encoding and recognition were equivalent to hit rate for faces that were undisguised at encoding but disguised at recognition. Furthermore, it seems that the encoding specificity may not apply to all decisions: recent studies by Manley et al. (2019, 2022) revealed matching effects only for correct identifications from target-present lineups but not for correct rejections from target-absent lineups. All these inconsistencies imply that encoding specificity alone may not be able to explain the effect of disguises on identification accuracy.

As can be seen, there are several mechanisms that can influence identification decisions when disguises are present in the encoding and/or retrieval phases. Thus, more research using variety of paradigms and stimuli is needed to understand how different disguises influence identification decisions when a witness encounters a disguised perpetrator and/or views a disguised lineup.

The current study

The COVID-19 pandemic has led to a widespread use of surgical masks. Empirical research that manipulates wearing surgical masks both at encoding and at retrieval can shed light on how law enforcement should conduct and administer a lineup when the crime was committed by a mask-wearing perpetrator. While some studies have already explored the effect of surgical masks on face recognition by using the face recognition paradigm (Freud et al., 2020, 2021; Garcia-Marques et al., 2022; Marini et al., 2021; Stajduhar et al., 2022), no studies, to our knowledge, have examined their effects on eyewitness identification decisions. Thus, the aim of this study is to examine the effect of surgical masks during encoding and/or retrieval phases on lineup decisions by using an eyewitness identification paradigm.

While the combination of various mechanisms likely impacts lineup decisions (Mansour et al., 2020) when the perpetrator and/or lineup members wear surgical masks, we focus on the encoding specificity principle and the transfer-appropriate processing framework, which state that performance on a memory test is best when the conditions and processes activated at retrieval match those at encoding (Morris et al., 1977; Tulving & Thomson, 1973). To our knowledge, the study by Manley et al. (2022) is the only eyewitness identification study that has investigated how matching the lineup condition to the encoding condition in cases of a disguised (a ski mask, in this instance) perpetrator affects eyewitness identification accuracy. Although they concluded that performance is superior when the appearance of lineup members matched the encoded target, caution should be exercised when recommending the use of this principle in the real world, especially with surgical masks. Surgical masks differ from ski masks by covering fewer regions of the face (and head), leaving more details visible for witnesses to encode, which should enhance the memory representation of the perpetrator. The type of disguise can also influence where witnesses allocate their attention (Mansour et al., 2020). Unlike ski masks, which may increase the perception of threat in the witness or draw attention to themselves, surgical masks have become commonplace (Schönweitz et al., 2022). This leaves resources available for encoding, as witnesses are able to allocate their attention to the visible features of the perpetrator. While this reduced coverage and increased attention to visible features might aid with later identification, dividing attention among several features during encoding affects memory as well (Chun & Turk-Browne, 2007; Craik et al., 1996). Therefore, the masked-lineup superiority effect after witnessing a perpetrator wearing a surgical mask may be smaller than what Manley et al. (2019, 2022) found using a ski mask, where only the eyes were visible.

Taken together, drawing from the results of Manley et al. (2019, 2022), we expected that when the encoding and retrieval conditions matched (i.e. perpetrator wearing a surgical mask during encoding and later in lineup, or not wearing a surgical mask during encoding and later in lineup) the proportion of correct target identifications in target-present (TP) lineups would be higher than when the encoding and retrieval conditions did not match (i.e. perpetrator wearing a surgical mask during encoding but not later in lineup, or not wearing a surgical mask during encoding and wearing a mask later in lineup; Hypothesis 1). We also examined the presence of matching effects in target-absent (TA) lineups, but we did not make any explicit predictions due to mixed findings regarding the effect of disguises on target-absent lineup decisions (Manley et al., 2019, 2022; Mansour et al., 2020). Based on previous face recognition research with surgical masks (e.g. Or et al., 2023; Ventura et al., 2023), as well as the combined results of Manley et al. (2019, 2022), we assumed that matching the encoding and retrieval conditions would yield higher discriminability (Hypothesis 2). Furthermore, we explored how the presence of surgical masks influenced criterion setting. Given that surgical masks might introduce challenges to the identification task, they could lead participants to alter their decision-making criteria (Brewer et al., 2022; A. M. Smith et al., 2019).

We considered it important to examine confidence and its relationship with decision accuracy, as eyewitnesses' initial confidence after making an identification has been shown to be a reliable indicator of accuracy of those identifications when fair lineup procedures have been used (Wixted & Wells, 2017). Previous research has suggested that when testing conditions are pristine but encoding conditions are poor, high confidence still indicates accurate target identification, as witnesses are capable of properly adjusting their confidence (Semmler et al., 2018; Wixted & Wells, 2017). This was corroborated by Manley et al. (2022) who showed that after seeing a masked perpetrator, higher confidence was indicative of greater suspect identification accuracy in both unmasked and masked lineups. However, the optimality hypothesis (Bothwell et al., 1987; Deffenbacher, 1980) states that confidence judgments are influenced by witnessing conditions, and, thus, confidence is likely to be a good predictor of accuracy when witnessing conditions are good but not when they are difficult. Consistent with the optimality hypothesis, Giacona et al. (2021) found that poor viewing conditions reduce the accuracy of suspect identifications that are made with high confidence. Furthermore, Mansour et al. (2020) found that for correct identifications, confidence decreased with degree of disguise. These results imply that both confidence and the relationship between confidence and accuracy may be sensitive to viewing conditions. We expected confidence in general to be a reliable indicator of identification accuracy (Hypothesis 3) and investigated how the presence of masks in the encoding and testing condition influences confidence in lineup decisions.

Finally, we examined the effect of surgical masks on decision times for different lineup decisions. Apart from Manley et al. (2019), who found that lineups matching the encoding condition reduced the response time associated with target identifications, there is a lack of research registering decision times regarding disguises. Previous studies so far have demonstrated a negative relationship between identification accuracy and response time for choosers, with accurate identifications being made faster than inaccurate identifications (Brewer et al., 2006; Sauerland & Sporer, 2007, 2009; S. M. Smith et al., 2000; Sporer, 1992, 1994). Thus, we also explored whether decision time could serve as a reliable predictor of identification accuracy in cases involving a masked perpetrator and/or a masked lineup.

Method

Design

In the experiment, a 2 (stimulus video: perpetrator not wearing a mask vs. wearing a mask) × 2 (lineup: lineup members not wearing masks vs. wearing masks) × 2 (target-present vs. target-absent simultaneous lineup) between-subjects design was used. Participants were randomly assigned to one of eight experimental conditions. Although some conditions created by a fully balanced experimental design are unlikely to occur in the real world (e.g. a witness is presented with a masked lineup after witnessing a perpetrator without a mask), it was essential to use the full design to properly test the extent to which the interaction of wearing surgical masks during encoding and retrieval affects lineup performance in an eyewitness identification paradigm. We also balanced target (1, 2) and target position in the lineup (1–6) between subjects and randomly assigned participants to one of these conditions, but treated these variables as covariates to control for possible confounding effects. To increase ecological validity of the study, we used a between-subjects design. This allowed us to avoid the potential effect that knowledge of the task may have had on memory performance (Quigley-McBride & Wells, 2021).

Participants

Sample size was determined by simulating data based on the results from Manley et al. (2019). As we were primarily interested in observing a crossover interaction between encoding and retrieval conditions for target identifications, we used hit rates from Manley et al. (2019, Experiment 4, TP condition) to determine the sample size needed for 80% power. Based on the response rates reported by Manley et al. (2019), we first calculated odds ratios (*OR*s) and then constructed the logistic regression equation and simulated (with a wide range of sample sizes, 10,000 runs each) experimental responses from a binomial distribution with the parameter value determined by the regression outcome. We determined the sample size that yielded at least 80% of interaction terms significant at $\alpha = .05$. This analysis suggested a sample size of 236 with 80% power to detect a statistically significant interaction effect on target identifications. Thus, we aimed to collect at least 59 participants for each experimental condition (we found it appropriate for all conditions, as observing main effects in both TA and TP lineups requires a smaller sample size than observing interactions). To compensate for possible exclusion of some participants, we terminated data collection when each condition reached 61–62 participants.

There were 634 people who started the experiment; 492 of those finished the study. We excluded 25 participants who recognized someone from the video and/or the lineup, two who did not answer all the questions, and four who reported that they did not concentrate on the task. We further excluded nine participants whose lineup decision (*z*-score) time was larger or smaller than the mean ±3 standard deviations (as suggested by Berger & Kiefer, 2021, for excluding outliers). Thus, we ended up with sample sizes ranging from 53 to 61 for each condition. The final sample of this study was 452 adult participants (370 female, 79 male, three other; M_{age} = 36.97 years, $SD = 12.27$), who were recruited through universities and social media advertisements.

Ethics

The experiment was conducted in accordance with the Declaration of Helsinki and with American Psychological Association (APA) ethical standards in human subjects research. Written consent was obtained from healthy adult participants. As the study involved healthy adults, and the data were collected anonymously, acquiring an approval from the institutional review board (IRB) was not required. Participants did not receive any monetary compensation; however, psychology students had the opportunity to request course credit for participating in the experiment. Women who acted as perpetrators or stood as lineup members had also provided written informed consent for the materials to be used in the study.

Materials

Stimulus video

We filmed four stimulus videos of the same theft episode using a Nikon d5100 camera. In the videos, a female target was stealing a wallet from a man's coat pocket. The videos were filmed from an eyewitness's perspective and lasted about 34–35 seconds.

We used two separate targets/perpetrators to increase the generalizability of the results and to diminish the possibility that our findings regarding participants' memory could be attributed only to the identification accuracy of a specific individual (Quigley-McBride & Wells, 2021). Both targets were Caucasian women in their 20s and had no distinctive features. Both targets filmed one video where they were wearing a mask and one video without wearing a mask. Targets were visible in their videos for an equal amount of time, approximately 16 seconds. There were no differences in how many times the targets occurred in different experimental conditions, $\chi^2(7, N = 452) = 1.07$, $p = .99$. Furthermore, target identity did not affect the number of correct lineup decisions, $\chi^2(1, N = 452) = 0.06$, $p = .81$; thus, to increase generalizability, we collapsed across targets when reporting our data.

Lineups

The targets were photographed with and without a mask using a Nikon d5100 camera. The photos of fillers were taken from a database ($n = 150$) that was created for another study. All photos displayed a woman (targets or fillers) looking at the camera with a neutral expression, from the shoulders up, set against a light background. The background tone and brightness of all the photos were standardized. Fillers and targets in the photos wore black shirts and had their hair tied up. To create a masked version of each filler, an image of a surgical mask was superimposed onto each face. Each photograph was 376 × 467 pixels.

The lineups consisted of six photographs, three on the top row and three on the bottom row. Each photo was accompanied by a number under the photo. For participants to reject the lineup, there was a text box saying 'the woman is not in the photos' under the photos. The background of the lineup was light grey.

Lineup validation and construction

To validate the lineups, we had two persons (a male and a female unrelated to the study) observe the target photos separately for about 10 seconds per photo and then describe the two female targets. The first target was described as 'female in her 20s, oval face, long blonde hair, sharp thin nose, middle-size lips, light colour eyes, light eyebrows' and the second as 'female in her 20s, round face, long dark blonde hair, thin nose, thin lips, blue eyes, dark eyebrows'. For both targets, a research assistant chose 10 fillers from the database that best fitted that target's general description. Next, mock witnesses ($N = 20$) read the description of the first target and chose which photo corresponded most to the description of the person using the Doob and Kirshenbaum (1973) procedure. The same was repeated for the second target using the same mock witnesses. Effective lineup sizes, determined as Tredoux's *E* values, were 4.44 and 4.55 (range = 3.23–7.64, Tredoux, 1998).

We constructed TP and TA simultaneous lineups for both targets. The photographs of fillers that were chosen most by the mock witnesses were selected for the lineups. The TP lineups included five filler faces in addition to the perpetrator, and in TA lineups, we replaced the perpetrator with a randomly picked additional filler. The position of targets in TP lineups was balanced between subjects, and the position of each filler face was randomized. The number of times the target or the filler replacing the target appeared in each position did not differ across experimental conditions, $\chi^2(35, N = 452) = 8.55, p = 1$. The number of correct lineup decisions was not dependent on the position of the target, $\chi^2(5, N = 221) = 8.33, p = .14$.

We also measured resultant lineup fairness to see the distribution of filler identifications across lineup members by using responses from the target-absent lineups (Quigley-McBride & Wells, 2021). We computed Tredoux's *E* as a resultant lineup fairness measure using the *r4lineups* package (Tredoux & Naylor, 2018). For one of the targets (B1), $E' = 2.8$, 95% confidence interval, CI [2.2, 5.44], and for the second target (B2), $E' = 4.71$, 95% CI [3.61, 5.79]. The distribution of identifications across fillers for both targets (B1 and B2) is reported in Table S1 in the online Supplemental Materials.

Procedure

Participants were invited to participate in a web-based decision-making and thinking style experiment. Participants were tested individually using a web platform Labvanced (Finger et al., 2017). The experiment was allowed to launch in PCs and tablets and in all browsers. First, the participants filled out a consent form for participating in the experiment and declared that they were adults. Then the participants saw a fixation cross at the centre of the screen for one second, followed by a video clip of the theft. Participants watched the video without knowing they would later have to identify the person in the video. The perpetrator in the video was either wearing or not wearing a mask and was either one (B1) or the other (B2) target. After the video, a distractor task (that took on average 10.77 minutes, $SD = 9.81$) was administered. The distractor task was the Word Meaning Structure Test, which measures the development of conceptual thinking (Toomela, 2003; for a detailed description, see Kask et al., 2019). The time interval between the stimulus video and the lineup did not differ across the experimental conditions, $H(7) = 3.93$, $p = .79$.

Next, the identification phase took place. Participants were informed that they were about to be presented with a lineup of photos, and they would have to decide whether the woman they saw in the videoclip was present in the lineup. If they responded 'yes', they would have to indicate her position in the lineup. Participants were also told that the perpetrator may or may not be present in the lineups. Before moving forward to the lineup, as an attention check, participants were instructed to move the cursor to the left bottom corner and click on a square. Participants saw a fixation cross for one second, followed by a TA or TP lineup. All the lineup members were either wearing or not wearing a mask. Participants had to register their decision by clicking on a photo in the lineup or clicking on the text box saying 'the perpetrator is not in any of the photos'. Following the lineup decision, participants had to rate the degree of their confidence in their decision using a slide ranging from 0 to 100 in increments of 1 (0 = 'not certain at all' to 100 = 'certain', see Sauer & Brewer, 2015). To control for familiarity effects, we asked participants whether they knew someone in the video or lineup personally. We also asked them to indicate on a 10-point scale (from 0 = 'not at all' to 10 = 'very hard') how much they concentrated on the tasks. Finally, participants were presented with demographic questions about their gender, age, mother tongue and education, and with a self-rated question about their visual acuity. At the end of the experiment, the participants were thanked for taking part in the experiment, asked to keep the content of the experiment a secret and presented with instructions for when they would like to receive feedback about the study. The experiment took on average 23.31 minutes ($SD = 62.19$) to complete.

Results

Data were analysed using R, Versions 3.6.3 and 4.2.1 (R Core Team, 2022). We also used parts of the analysis codes by A. M. Smith et al. (2020), Baldassari et al. (2020), and Mansour (2019) that are available on their OSF project pages.

Each participant made one lineup decision, meaning that 452 lineup decisions were made. Of those, 189 (41.81%) were correct, and 263 (58.19%) were not. The proportions of different responses, the average confidence ratings and decision times for different decision types and experimental conditions are presented in Table 1.

Eyewitness identification accuracy

The effect of surgical masks on TP and TA lineup decisions

To examine the effect of surgical masks on eyewitness identification decisions, we first analysed the data from TP and TA lineups separately. We conducted separate binary logistic regression analyses, with perpetrator wearing a mask at encoding (no mask vs. mask), lineup members wearing masks at retrieval (no masks vs. masks) and their interaction entered as predictors. To analyse main effects, we used effect coding for categorical predictors: we specified a contrast centred at 0 such that the first level was coded as −.50 (no mask) and the second level as .50 (mask; Barr, 2019).

For the TP condition, the dependent measures were target identifications (coded as 1, all other decisions as 0), filler identifications (coded as 1, all other decisions as 0) and incorrect rejections (coded as 1, all other decisions as 0). As this set of analyses involved three

Table 1. Lineup decision proportions, confidence ratings and decision times for lineup decisions across all conditions.

	Unmasked perpetrator		Masked perpetrator	
	Unmasked lineup	Masked lineup	Unmasked lineup	Masked lineup
Target-present lineups	($n=53$)	($n=53$)	($n=56$)	($n=59$)
Response proportions				
Target IDs	.59	.36	.21	.22
Filler IDs	.15	.40	.30	.41
Incorrect rejections	.26	.25	.48	.37
Confidence ratings				
Target IDs	82.68	63.68	55.08	61.15
Filler IDs	55.12	51.48	55.71	59
Incorrect rejections	67.43	62.54	55	48.27
Decision time				
Target IDs	16.25	20.28	18.98	21.8
Filler IDs	30.31	24.12	21.62	24.2
Incorrect rejections	21.95	22.13	21.3	23.97
Target-absent lineups	($n=56$)	($n=61$)	($n=58$)	($n=56$)
Response proportions				
Correct rejections	.63	.46	.50	.39
Filler IDs	.38	.54	.50	.61
Confidence ratings				
Correct rejections	66.4	53.68	68.9	52.45
Filler IDs	63.57	51.03	54.52	53.47
Decision time				
Correct rejections	28.00	29.36	19.24	28.75
Filler IDs	27.92	25.67	23.7	24.98

Note: Decision time is presented in seconds and confidence ratings on a scale of 0–100. Proportions may not add up to 1.0 due to rounding error.

regressions, we adopted a per-test Bonferroni-corrected alpha level of .0167 (.05/3). For the TA condition, the dependent measure was correct rejections (coded as 1, filler identifications coded as 0). As participants could only select a filler or reject the lineup, filler identifications were complementary to correct rejections, and we did not analyse them separately. Here we report the results of logistic regressions predicting target identifications and correct rejections as these are most relevant to our hypothesis; the results of logistic regressions predicting TP filler identifications and incorrect rejections can be found in the online Supplemental Materials.

For target identifications, a binary logistic regression indicated that there was no interaction between the perpetrator wearing a mask at encoding and lineup members wearing masks in lineups, $B=0.96$, $SE=0.60$, $z=1.59$, $p = .11$, $OR=2.61$, 95% CI [0.80, 8.62]. However, there was a main effect of the encoding condition, $B=-1.16$, $SE=0.30$, $z=-3.85$, $p < .001$ – namely, when the perpetrator was wearing a mask at encoding, the odds of making a target identification from the lineup decreased ($OR=0.31$, 95% CI [0.17, 0.56]) compared to when the perpetrator was not wearing a mask. We did not find a main effect of lineup members wearing masks at retrieval, $B=-0.44$, $SE=0.30$, $z=-1.47$, $p = .14$, $OR=0.64$, 95% CI [0.35, 1.16].

We did not find an interaction effect for correct rejections, $B=0.24$, $SE=0.53$, $z=0.45$, $p = .65$, $OR=1.27$, 95% CI [0.45, 3.64], nor the main effect of perpetrator wearing a mask at encoding, $B = -\ 0.39$, $SE = 0.27$, $z = -1.46$, $p\ =\ .14$, $OR = 0.68$, 95% CI [0.4, 1.14]. However, we found that retrieval condition had an effect on correct rejections, $B = -0.56$, $SE = 0.27$, $z = -2.08$, $p\ =\ .04$ – namely, when lineup members wore masks in TA lineups, the odds of making a correct rejection decreased ($OR = 0.57$, 95% CI [0.34, 0.97]; and thus, the odds of making a filler identification increased) compared to when lineup members were not wearing masks.

The effect of surgical masks on discriminability

Probit regression analysis. To examine the effect of mask-wearing at encoding and retrieval, and their interaction on discriminability, we used a probit regression analysis, which is one way of conducting a discriminability analysis (A. M. Smith & Neal, 2021; Wright & London, 2009). We included perpetrator wearing a mask at encoding (no mask vs. mask), lineup members wearing masks at retrieval (no masks vs. masks), perpetrator presence (TA vs. TP), all the two-way interactions, and the three-way interaction term as predictors. We used effect coding for all the categorical predictors (–0.5; 0.5).

As fillers are known-innocent persons, and investigators who are mainly interested in suspect identifications do not perpetuate filler identification errors, we categorized only innocent suspect identifications (and not all TA filler identifications), in addition to target identifications, as suspect identifications. We estimated the number of innocent suspect identifications by dividing the number of all TA filler identifications in each experimental condition by the lineup's resultant effective size (Tredoux's *E* 2.8 for target B1 and 4.71 for target B2) and then replaced that many filler identification outcomes with suspect identification outcomes (following A. M. Smith et al., 2022). Thus, we created a dependent variable in which perpetrator identifications from TP lineups and innocent suspect identifications from TA lineups were coded as 1s, while all other lineup decisions were coded as 0s (i.e. TP filler identifications, TP rejections, TA filler identifications, TA rejections).

The probit regression model indicated that identifications of the suspect were more likely to occur when the suspect was guilty as opposed to innocent, $B = 0.65$, $SE = 0.14$, $z = 4.74$, $p < .001$. In addition, we found that regardless of accuracy, suspect identifications were more likely to occur when the perpetrator did not wear a mask, $B = -0.32$, $SE = 0.14$, $z = -2.33$, $p\ =\ .02$, than when they did. We did not find a main effect of lineup members wearing masks at retrieval, $B = -0.05$, $SE = 0.14$, $z = -0.36$, $p = .72$.

The three-way interaction was not significant, $B = 0.76$, $SE = 0.55$, $z = 1.39$, $p\ =\ .16$, which indicates that discriminating guilty from innocent suspects was not dependent on the interaction of mask-wearing at encoding and retrieval. However, we found a significant interaction between target presence and the perpetrator wearing a mask at encoding, $B = -0.78$, $SE = 0.27$, $z = -2.84$, $p\ =\ .005$. We followed up on the interaction by looking at the effect of perpetrator presence on suspect identifications at the two levels (perpetrator not wearing a mask, perpetrator wearing a mask) of the encoding condition (simple effects). When the perpetrator did not wear a mask, perpetrator identifications were significantly more likely to occur than were innocent suspect identifications, $B = 1.04$, $SE = 0.19$, $z = 5.38$, $p < .001$. When the perpetrator wore a mask, the likelihood of a suspect identification did not depend on whether the suspect was guilty or innocent, $B = 0.26$, $SE = 0.19$, $z = 1.33$, $p\ =\ .18$. Finally, we found that neither the interaction between the perpetrator wearing a mask at encoding and lineup members wearing masks at retrieval, $B = 0.22$, $SE = 0.27$, $z = 0.79$, $p\ =\ .43$, nor the interaction between target presence and lineup members wearing masks at retrieval, $B = -0.46$, $SE = 0.27$, $z = -1.68$, $p\ =\ .09$, was significant. To sum up, these results suggest that discriminating between the perpetrator and the innocent suspect was influenced by the presence of a surgical mask at encoding but not at retrieval, nor was it influenced by the interaction of the two conditions.

Compound signal detection model. We also fitted a signal detection model of compound decision task to our data (SDT-CD; Duncan, 2006; see also Palmer et al., 2010). The SDT-CD model is designed to estimate discriminability and response bias in tasks, such as eyewitness decisions in lineups, which involve compound decisions. Compound decisions consist of detection (i.e. whether the perpetrator is present in the lineup) and identification components (i.e. who the perpetrator is in the lineup). The detection component is measured by comparing the proportion of identifications in TP lineups to the proportion of identifications in TA lineups. The identification component is measured by the proportion of target identifications in TP lineups. We applied the independent observation decision rule, as recommended by Wixted et al. (2018), which assumes that eyewitnesses evaluate each lineup member individually and make an identification if at least one of the members exceeds the decision criterion (Duncan, 2006).

We fitted the model using the *sdtlu* package in R (Cohen, 2020; Cohen et al., 2021). We used the unequal variance signal detection model, as it is less restrictive than the equal variance model while allowing for equal variance as well. We combined confidence ratings into three groups to represent different decision criteria: 0–59 (*c1*), 60–79 (*c2*) and 80–100 (*c3*). Consistent with the SDT-CD model, we assumed that a single parameter d' – discriminability of a guilty target from innocent foils – to be a composite descriptor of both the detection and identification components (Duncan, 2006). A d' value of zero indicates chance level performance: the higher the d' value, the better the discriminability in the dataset it was calculated from. Decision criterion (*c*) refers to the amount of evidence required to make an identification: higher values indicate a more conservative criterion and, thus, bias towards rejecting the lineup, while lower values indicate a more liberal criterion, which leads to a bias towards identification. To compare parameter values across experimental conditions, we calculated 95% inferential confidence intervals (ICIs; Tryon, 2001) around each parameter estimate using estimated standard errors that were obtained by calculating a standard deviation of the 10,000 bootstrapped parameter values (Efron & Tibshirani, 1994). In calculating ICIs, we used an average *E* value, which enables comparisons between multiple pairs of conditions (Tryon, 2001). Non-overlapping confidence intervals indicate a significant difference between conditions at the $p < .05$ level. Results from the SDT-CD model fits along with 95% ICIs are presented in Table 2.

As can be seen from Table 2, when participants encoded the perpetrator without a mask,

Table 2. Estimates of discriminability and decision criteria from the SDT-CD model.

		d'[95% ICI]	*c1*[95% ICI]	*c2*[95% ICI]	*c3*[95% ICI]
Unmasked perpetrator	Unmasked lineup	1.89 [1.47, 2.31]	1.45 [1.31, 1.59]	1.76 [1.62, 1.91]	2.23 [1.98, 2.48]
	Masked lineup	1.03 [0.69, 1.38]	1.13 [1.01, 1.24]	1.71 [1.57, 1.84]	2.29 [2.06, 2.51]
Masked perpetrator	Unmasked lineup	0.74 [0.29, 1.19]	1.30 [1.18, 1.41]	1.78 [1.64, 1.92]	2.13 [1.94, 2.31]
	Masked lineup	0.54 [0.12, 0.95]	1.11 [0.99, 1.22]	1.60 [1.48, 1.73]	2.18 [1.98, 2.37]

Note: SDT-CD = signal detection model of compound decision; ICI = inferential confidence interval; d' = discriminability; c = decision criteria. These estimates are based on the SDT-CD independent observations model (also known as the unequal variance model), which fit the data well in each condition (all $G^2 < 5$, all respective $p > .6$).

discriminability was significantly higher when participants saw a lineup without masks ($d' = 1.89$) than a lineup with masks ($d' = 1.03$). When participants witnessed the perpetrator with a mask, discriminability was again higher when participants saw a lineup without masks ($d' = 0.74$) than a lineup with masks ($d' = 0.54$), but this difference did not reach significance. Furthermore, when participants were shown a lineup without masks, discriminability was significantly higher when participants had witnessed an unmasked perpetrator ($d' = 1.89$) than a masked perpetrator ($d' = 0.74$). Decision criterion values indicated conservative responding – that is, overall, participants tended to need more information to make an identification and, thus, were biased to reject the lineup. Furthermore, as can be seen, decision criteria increased with confidence: high-confidence identifications required more evidence. As to the effect of mask-wearing on decision criteria, at the lowest level of confidence, participants were more conservative when an unmasked lineup was seen after witnessing an unmasked perpetrator ($c1 = 1.45$) than when a masked lineup was seen after witnessing either an unmasked perpetrator ($c1 = 1.13$) or a masked perpetrator ($c1 = 1.11$). Decision criteria did not vary significantly across conditions at the highest level of confidence nor at the intermediate level of confidence.

Confidence

Confidence ratings for target identifications and correct rejections

For analyses of confidence, we first used separate analyses of variance (ANOVAs) to explore the effect of perpetrator wearing a mask at encoding and lineup members wearing masks at retrieval on confidence for TP and TA lineup decisions. We adopted a per-test Bonferroni-corrected alpha level of .0167 (.05/3) for confidence analyses for TP lineup decisions and an alpha level of .025 (.05/2) for confidence analyses for TA lineup decisions. Here, we report the results of ANOVAs examining confidence ratings for target identifications and correct rejections; the results of ANOVAs investigating confidence ratings for remaining lineup decisions can be found in the online Supplemental Materials.

A factorial ANOVA examining the effect of mask-wearing at encoding and retrieval on confidence ratings for target identifications revealed a significant interaction, $F(1, 71) = 6.43, p = .01, \eta^2_G = .08$. Pairwise comparisons using Holm correction revealed that when participants saw the perpetrator at encoding without a mask, they were more confident in their target identifications from unmasked lineups ($M = 82.68$, $SD = 14.22$) than from masked lineups ($M = 63.68$, $SD = 18.97$), $p = .008$. However, when participants saw the perpetrator wearing a mask at encoding, they were not significantly more confident in their target identifications from masked lineups ($M = 61.15$, $SD = 22.87$) than from unmasked lineups ($M = 55.08$, $SD = 29.31$), $p = .90$. In addition, we found a main effect of encoding condition, $F(1, 71) = 9.30, p = .003, \eta^2_G = .12$, such that participants made more confident target identifications when they had seen the perpetrator in the video without a mask ($M = 75.46$, $SD = 18.51$) than when they had seen a masked perpetrator ($M = 58.24$, $SD = 25.79$). The main effect of retrieval was not statistically significant, $F(1, 71) = 1.71, p = .20, \eta^2_G = .02$.

For confidence in correct rejections, there was a main effect of retrieval condition, $F(1, 110) = 8.94, p = .003, \eta^2_G = .08$, such that participants were more confident in correctly rejecting the lineup when lineup members did not wear masks ($M = 67.53$, $SD = 23.53$) than when they wore masks ($M = 53.14$, $SD = 27.78$). The main effect of perpetrator wearing a mask at encoding, $F(1, 110) = 0.02$, $p = .90, \eta^2_G < .001$, and the interaction effect on confidence in correct rejections were not statistically significant, $F(1, 110) = 0.15, p = .70, \eta^2_G = .001$.

Confidence for perpetrator and mistaken identifications

We used ANOVA to investigate how mask-wearing at encoding and in lineups at retrieval affects confidence for accurate and inaccurate identifications. We added confidence as a dependent variable and perpetrator wearing a mask at encoding, lineup members wearing masks at retrieval, target presence and decision type (identifications vs. other) as independent variables. Here, identifications included target identifications from TP lineups and all TA filler identifications. Because we measured confidence on a scale of 0 to 100 (in increments of 1), we were not able to estimate innocent suspect identifications as it would have required us to bin confidence and estimate the number of innocent suspects in each confidence group for each experimental group, which would have potentially led to loss in informational value in the data and, thus, bias in these results. We present the key findings here, while the complete results of the ANOVA can be found in Table S2 in online Supplemental Materials.

We found a significant interaction between target presence and decision type, $F(1, 436) = 7.32$, $p = .007$, $\eta^2_G = .02$, indicating that confidence differed for perpetrator and mistaken identifications. Pairwise comparisons with Holm correction revealed that perpetrator identifications were made with higher confidence ($M = 69.72$, $SD = 22.57$) than mistaken identifications ($M = 54.85$, $SD = 24.33$), $p < .001$, as well as filler identifications and incorrect rejections from TP lineups ($M = 56.09$, $SD = 25.91$), $p < .001$. Most importantly, we found that the four-way interaction was not significant, $F(1, 436) = 0.04$, $p = .84$, $\eta^2_G < .0001$, suggesting that confidence for perpetrator and mistaken identifications did not differ as a function of the presence of masks at encoding and retrieval. While none of the other interactions were significant, there was a significant main effect of perpetrator wearing a mask at encoding, $F(1, 436) = 5.15$, $p = .02$, $\eta^2_G = .01$, and lineup members wearing masks at retrieval, $F(1, 436) = 10.53$, $p = .001$, $\eta^2_G = .02$. Confidence was higher when an unmasked perpetrator ($M = 62.32$, $SD = 24.84$) was witnessed than when a masked perpetrator ($M = 56.51$, $SD = 25.96$) was witnessed, and when an unmasked lineup ($M = 64.03$, $SD = 25.4$) was seen than when a masked lineup ($M = 54.74$, $SD = 24.92$) was seen.

Confidence–accuracy relationship

To investigate the relationship between confidence and accuracy, we conducted confidence accuracy characteristic analysis (CAC; Mickes, 2015) using the data from choosers (Sporer et al., 1995). CAC analysis can be used to evaluate the effect of estimator variables on the relationship between eyewitnesses' confidence and their accuracy, which is important to judges and juries as it helps them to assess the reliability of eyewitnesses (Mickes, 2015). As CAC analysis includes only suspect identifications, we estimated the number of innocent suspect identifications for each confidence bin in all the experimental conditions by dividing the number of TA filler identifications by the lineup's resultant effective size (as suggested by A. M. Smith et al., 2021). We conducted a CAC analysis by plotting suspect identification accuracy (TP target identifications/sum of TP target identifications and estimated TA innocent suspect identifications) for each experimental condition across varying levels of confidence. We aimed to bin continuous confidence ratings into categories following Mickes (2015); however, due to low number of data points in the highest confidence bin (90–100%), we decided to use confidence groups of 0–59% (low), 60–79% (medium) and 80–100% (high) instead. The CAC curves, with standard error bars calculated using a bootstrap procedure described in Seale-Carlisle and Mickes (2016), are illustrated in Figure 1 (see Table S3 in online Supplemental Materials for frequency of suspect identification decisions at all levels of confidence).

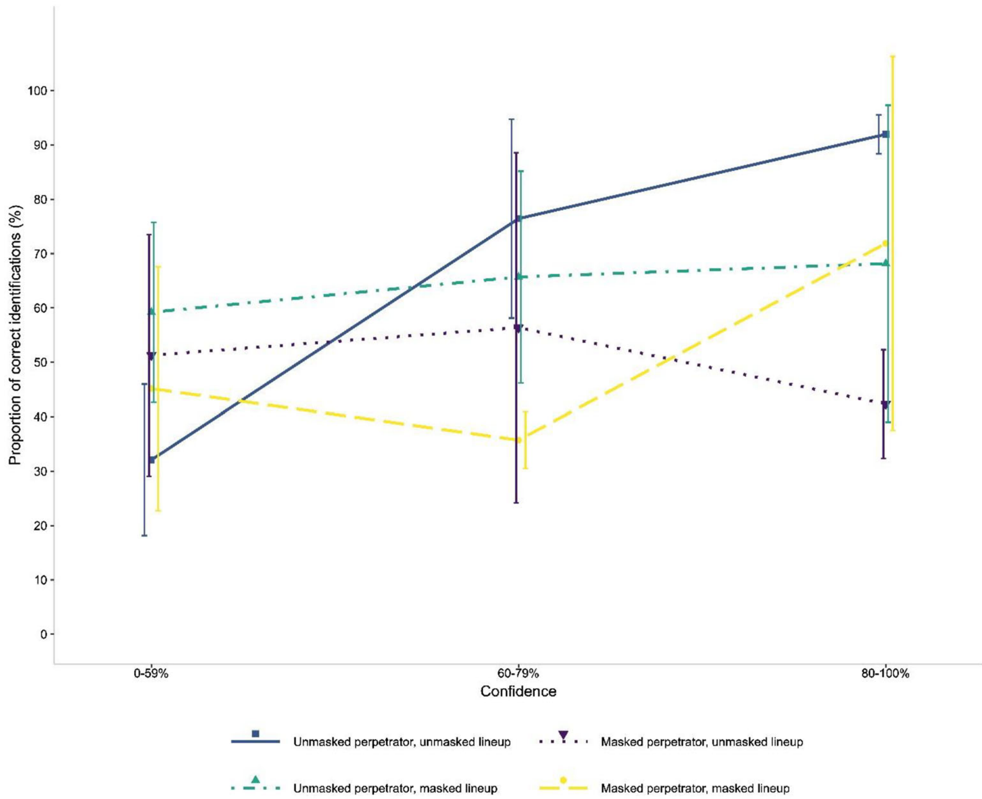

Figure 1. Confidence–accuracy characteristic curves for masked and unmasked lineups after encoding masked and unmasked perpetrators. Error bars reflect ±*SE*.

As can be seen from Figure 1, the level of accuracy varies substantially at each level of confidence across conditions. CAC curves that have a positive slope are indicative of a strong confidence–accuracy relationship as they show that confidence and accuracy increase together. However, only the curves for conditions where an unmasked perpetrator was encoded display a positive slope. Among the two, the slope is steeper when an unmasked lineup was observed after witnessing an unmasked perpetrator, suggesting that confidence was a reliable indicator of accuracy in cases when masks were not present at either encoding or retrieval. The curves for conditions where a masked perpetrator was encoded do not exhibit a clear positive slope, showing that confidence was not predictive of accuracy in these cases.

CAC curves are mainly meant to assess whether identifications made by highly confident witnesses can be trusted. As can be seen from Figure 1, high-confidence suspect identifications were not associated with high levels of accuracy in all the conditions. When the participants observed an unmasked lineup after encoding an unmasked perpetrator, the proportion of correct identifications was 92%. Despite high-confidence suspect identifications being more accurate than identifications made at lower levels of confidence in conditions where a masked lineup was observed after encoding either a masked perpetrator or an unmasked perpetrator, they were still correct only 71.9% and 68.2% of the time, respectively. Thus, the level of accuracy associated with the highest level of confidence varies substantially across conditions where participants saw either a masked perpetrator or a masked lineup.

We also calculated calibration indices – calibration (C); over/underconfidence (O/U); discrimination (adjusted normalized discrimination index, ANDI) – using the CAC data. Calibration indices provide further information about the confidence–accuracy relationship (Brewer & Wells, 2006). The calibration statistic (C) indicates how well calibrated the participants' confidence is overall, with 0 corresponding to perfect calibration and 1 to no calibration. The over/underconfidence (O/U) statistic, which varies from −1 to +1, indicates the extent to which confidence reports tend to over- or underestimate accuracy. Negative scores of O/U reflect underconfidence (less confident than accurate), and positive scores show overconfidence (more confident than accurate; Brewer & Wells, 2006). The adjusted normalized discrimination index (ANDI; Vredeveldt & Sauer, 2015; Yaniv et al., 1991), which varies from 0 (no discrimination) to 1 (perfect discrimination), reflects how effectively confidence discriminates between accurate and inaccurate identifications. We conducted inferential comparisons of the calibration indices by using an average *E* value to calculate 95% ICIs around each statistic (Tryon, 2001). For each statistic, the standard deviation of the 10,000 bootstrapped statistic values provided the estimated standard error. Non-overlapping confidence intervals indicate a significant difference at the $p < .05$ level between conditions.

The calibration indices with ICIs are depicted in Table 3. Confidence and accuracy were fairly well calibrated in all the conditions, with the unmasked lineup after unmasked perpetrator condition showing the best calibration. None of the conditions differed in calibration as indicated by overlapping confidence intervals. While most conditions produced almost no over- or underconfidence, participants who saw a masked lineup after encoding a masked perpetrator were slightly overconfident. However, this did not significantly differ from other conditions. Low ANDI values indicate that confidence was unable to discriminate between correct and incorrect suspect identifications. The condition where participants saw an unmasked lineup after encoding an unmasked perpetrator showed the best discrimination; however, none of the conditions differed significantly from one another.

The effect of confidence on perpetrator identifications and mistaken identifications

When a witness identifies a suspect with some level of confidence, the task of the police officer is to postdict the probability that the suspect is the culprit using that decision and the corresponding confidence rating. Thus, we conditioned suspect guilt on confidence, which provides information that is relevant to the criminal justice system (A. M. Smith et al.,

Table 3. Calibration, over/underconfidence, and adjusted normalized discrimination index for all conditions.

	C [95% ICI]	O/U [95% ICI]	ANDI [95% ICI]
Unmasked perpetrator			
Unmasked lineup	.007 [.004, .01]	.02 [–.06, .10]	.22 [–.02, .45]
Masked lineup	.05 [.008, .09]	−.05 [–.25, .15]	−.06 [–.10, −.03]
Masked perpetrator			
Unmasked lineup	.09 [–.05, .22]	.04 [–.25, .34]	−.07 [–.22, .07]
Masked lineup	.06 [.01, .10]	.11 [–.09, .31]	.02 [–.12, .17]

Note: C = calibration; O/U = over/underconfidence; ANDI = adjusted normalized discrimination index; ICI = inferential confidence interval.

2021). To determine whether confidence could discriminate between perpetrator and mistaken identifications, and whether this effect was dependent on the presence of the culprit or the presence of masks during both encoding and retrieval stages, we conducted a probit regression analysis where we added confidence as a predictor. Confidence was centred around the mean, and all categorical predictors were effect coded (–0.5; 0.5). As we were not able to estimate the number of innocent suspects for each confidence rating in each experimental condition, we created a dependent variable in which target identifications from TP lineups and all filler identifications from TA lineups were coded as 1s, while all other responses were coded as 0s. We present the findings relating to the relationship between confidence and accuracy here; the complete results of the model can be found in in the online Supplemental Materials (Table S4).

First, we found that mistaken identifications were more likely to occur than perpetrator identifications, $B = -0.58$, $SE = 0.13$, $z = -4.39$, $p < .001$, which was predictable, as mistaken identifications included all filler identifications from TA lineups and not just innocent suspect identifications. We found a significant interaction between target presence and confidence, $B = 0.02$, $SE = 0.006$, $z = 3.45$, $p < .001$, indicating that accuracy (the likelihood of a perpetrator identification) increased as confidence increased. Although we found that confidence was able to discriminate between perpetrator and mistaken identifications, the predictive utility of confidence was not dependent on the interaction of masks at encoding and retrieval, indicated by the non-significant four-way interaction, $B = 0.02$, $SE = 0.02$, $z = 0.69$, $p = .49$. Regardless, we examined the simple slopes of confidence at each different level and found that confidence was predictive of identification accuracy when the perpetrator and the lineup were unmasked, $B = 0.04$, $SE = 0.01$, $z = 2.97$, $p = .003$, but not when masks were present at either encoding, retrieval, or both, $ps > .09$.

Decision time

Decision time for target identifications and correct rejections

For analyses of decision time, we followed the same procedure as with analyses of confidence. First, we used separate ANOVAs (with per-test Bonferroni corrections) to investigate the effect of mask-wearing on decision time for TP and TA lineup decisions. Here, we report the results for target identifications and correct rejections; the results for remaining lineup decisions can be found in the online Supplemental Materials.

We found that target identification decision times were not influenced by perpetrator wearing a mask at encoding, $F(1, 71) = 1.3$, $p = .26$, $\eta^2_G = .02$, lineup members wearing masks at retrieval, $F(1, 71) = 3.38$, $p = .07$, $\eta^2_G = .05$, or the interaction of conditions, $F(1, 71) = 0.10$, $p = .75$, $\eta^2_G = .001$. Decision times for TA correct rejections were also not affected by perpetrator wearing a mask at encoding, $F(1, 110) = 2.11$, $p = .15$, $\eta^2_G = .02$, lineup members wearing masks at retrieval, $F(1, 110) = 2.85$, $p = .09$, $\eta^2_G = .03$, or their interaction, $F(1, 110) = 1.6$, $p = .21$, $\eta^2_G = .01$.

Decision time for perpetrator and mistaken identifications

Similarly to analyses regarding confidence, we used a factorial ANOVA to investigate how mask-wearing at encoding and in lineups at retrieval affects decision time for accurate and inaccurate identifications. The complete results of the ANOVA are presented in Table S5 in online Supplemental Materials. Although we did not find a significant interaction between decision type and target presence, $F(1, 436) = 1.12$, $p = .29$, $\eta^2_G = .003$, pairwise comparisons using Holm correction revealed that perpetrator identifications were made faster ($M = 18.67$, $SD = 7.7$) than mistaken identifications ($M = 25.38$, $SD = 18.47$), $p = .02$. However, perpetrator and mistaken identification decision times did not differ as a function of mask-wearing at encoding and retrieval, $F(1, 436) = 0.005$, $p = .94$, $\eta^2_G < .001$. We also found a main effect of target presence, $F(1, 436) = 8.21$, $p = .004$, $\eta^2_G = .02$, indicating that all decisions from TP lineups ($M = 21.7$, $SD = 13.32$) were made quicker than all decisions from TA lineups ($M = 25.81$, $SD = 17.84$).

The effect of decision time on perpetrator identifications and mistaken identifications

To determine whether decision time could discriminate between perpetrator and mistaken identifications, and whether this effect was dependent on the presence of the culprit or the presence of masks during both encoding and retrieval stages, we conducted a probit regression analysis where we added decision time as a predictor. Decision time was centred around the mean, and all categorical predictors were effect coded (–0.5; 0.5). As with our earlier analysis involving confidence, we created a dependent variable in which target identifications from TP lineups and all filler identifications from TA lineups were coded as 1s, while all other responses were coded as 0s. The most relevant findings about the decision time–accuracy relationship are presented here; the rest of the results can be found in the online Supplemental Materials (Table S6).

First, we found a significant interaction between target presence and decision time, $B = -0.02$, $SE = 0.01$, $z = -2.40$, $p = .02$, indicating that the likelihood of a perpetrator identification decreased as decision time increased. Although the predictive utility of decision time was not dependent on the presence of masks at encoding and retrieval both, $B = -0.04$, $SE = 0.04$, $z = -1.03$, $p = .30$, we found that it was dependent on lineup members wearing masks at retrieval, as indicated by a significant three-way interaction, $B = 0.04$, $SE = 0.02$, $z = 2.12$, $p = .03$. We followed this up by looking at the interaction between decision time and target presence at different levels of mask-wearing at retrieval. When the lineup members were not wearing masks, the likelihood of a perpetrator identification decreased as decision time increased, $B = -0.05$, $SE = 0.02$, $z = -2.85$, $p = .004$. When the lineup members wore masks, decision time was not predictive of an accurate identification, $B = -0.003$, $SE = 0.01$, $z = -0.23$, $p = .82$. Thus, quicker identifications from unmasked lineups were likely to be accurate.

Discussion

Considering that disguises and face coverings have a negative impact on eyewitness identification accuracy (e.g. Shapiro & Penrod, 1986), our aim was to examine how the presence of surgical masks at both encoding and retrieval affects eyewitness identification decisions and accuracy. In line with the encoding specificity (Tulving & Thomson, 1973) and the transfer-appropriate processing principles (Morris et al., 1977), we sought to investigate whether presenting a masked lineup would help witnesses to identify a perpetrator who concealed themselves by wearing a surgical mask.

The effect of matching surgical masks at encoding and retrieval on lineup decisions

In contrast to our prediction based on the encoding specificity principle and the transfer-appropriate processing theory, and to the findings of Manley et al. (2019, 2022), we found that matching the presence of surgical masks at the encoding and retrieval conditions did not lead to a higher proportion of target identifications in TP lineups (Hypothesis 1). Rather, we observed more target identifications (consistent with Or et al. 2023) and fewer incorrect rejections when participants had witnessed an unmasked perpetrator, irrespective of the presence of masks in the lineup. These results imply the importance of encoding conditions, particularly the absence of disguises, in facilitating accurate lineup identifications (Mansour et al., 2020; Shapiro & Penrod, 1986). Surgical masks decrease facial information available for the witness to encode by covering a large area of the face and many facial features, potentially leading to a less detailed memory representation of the perpetrator. Additionally, encountering a masked perpetrator interferes with holistic processing of the face (Freud et al., 2020), which might lead the witness to split their attention between the visible features of the perpetrator and the unfolding events. Division of attention can diminish available resources for processing the face, thereby increasing perceptual load (Lavie, 1995). Increased perceptual load at encoding impairs the ability to encode visible stimuli (Greene et al., 2017), leading to less accurate memory (Murphy & Greene, 2016; Shapiro & Penrod, 1986). As a result, even when the witness is presented with a matching lineup featuring masked members, their memory representation of the masked perpetrator might not contain enough information to facilitate a successful identification, thus leading to less accurate identifications (Brewer et al., 2005). To sum up, this finding indicates that witnessing a masked perpetrator probably results in a less detailed memory representation, which is necessary for accurate identifications.

Similarly to TP lineups and in accordance with Manley et al. (2019, 2022), we found that presenting participants with a masked lineup after witnessing a masked perpetrator did not improve their ability to make correct rejections from TA lineups. However, we found a main effect of lineup members wearing surgical masks. Unlike Manley et al. (2019, Experiment 3), who found that masked lineups increased correct rejections, we observed the opposite. Namely, the correct rejection rate was lower (and thus, the filler identification rate was higher) when lineup members wore surgical masks than when they did not. This pattern, where disguises at retrieval increase false alarms, is consistent with findings from prior face recognition studies (Hockley et al., 1999; Or et al., 2023).

Past research has shown that a weak recognition experience leads witnesses to lower the amount of evidence required for an identification and, thus, be more willing to select someone from a lineup (Kent et al., 2018; Shapiro & Penrod, 1986; A. M. Smith et al., 2019). A weak recognition experience can potentially explain why false alarms increased from masked TA lineups, compared to unmasked TA lineups (A. M. Smith et al., 2019). According to the culprit present-absent criteria discrepancy hypothesis (A. M. Smith et al., 2018), a TA lineup induces a weaker match-to-memory than a TP lineup, as the perpetrator's absence essentially causes a mismatch between encoding and retrieval conditions (Tulving & Thomson, 1973). This weak recognition experience could be further compromised when lineup members wear masks. First, masks can induce a sense of uncertainty by reducing the information available about lineup members (Kamal & Burkell, 2011), which in turn can increase cognitive load (Coutinho et al., 2015; Mushtaq et al., 2011). Increased cognitive load reduces resources available for the identification task (Coutinho et al., 2015) and thereby affects lineup decisions (Shapiro & Penrod, 1986). Second, witnesses may perceive identifications from a masked lineup more difficult. Although we did not measure perception of task difficulty, previous research suggests that masked conditions lead to higher difficulty ratings (Cash & Pazos, 2023; Maiorana et al., 2022). Furthermore, we also observed that all the lineup decisions, including correct rejections, from masked lineups were made with lower confidence than decisions from unmasked lineups, which suggests that participants perceived identifications from masked lineups to be more difficult. Taken together, our finding about TA lineup decisions indicates that participants found the identification task more challenging when both the lineup was masked and the perpetrator was absent, resulting in an increase in incorrect identifications and a decrease in correct rejections.

The effect of matching surgical masks at encoding and retrieval on discriminability

From the practitioner's perspective, understanding the effects of mask-wearing on guilty and suspect innocent identifications is of paramount importance. We did not find support for our prediction (Hypothesis 2) that matching the presence of surgical masks at retrieval and encoding stage would yield higher discriminability. Rather, we discovered that discriminability was influenced by the presence of surgical masks at encoding alone. More specifically, suspect identifications were more likely to be accurate after participants had encoded an unmasked perpetrator, but not after encoding a masked perpetrator. We also examined the matching effects separately for when an unmasked or a masked perpetrator was encoded, as this provides crucial information to the investigators for constructing lineups. As was expected, a matching effect was present when participants encoded an unmasked perpetrator; unmasked lineups yielded higher discriminability than masked lineups. However, discriminability did not differ between masked and unmasked lineups when the perpetrator was seen wearing a surgical mask (Garcia-Marques et al., 2022; Guerra et al., 2022). These findings further support the suggestion that the presence of surgical masks at the encoding stage results in a less-detailed memory representation of the perpetrator, to the extent that presenting a matching lineup to a witness may not increase discriminability. Thus, it may be that under certain circumstances (see discussion below), memory processes or attentional factors may compensate for deficits in holistic processing that appear in incongruent conditions (Garcia-Marques et al., 2022). Moreover, the fact that discriminability remained unaffected when an unmasked lineup was presented (rather than a masked lineup) following the encoding of a masked perpetrator suggests that the additional features visible in the lineup did not interfere with participants' recognition of faces, as they were non-diagnostic of guilt (Carlson et al., 2021). Finally, while several previous studies have found that removal of a disguise is more detrimental to identification accuracy than adding one during retrieval (Davies & Flin, 1984; Douma et al., 2012; Guerra et al., 2022; Righi et al., 2012; Terry, 1994), our findings showed no differences in discriminability between the two in the context of surgical masks. In conclusion, our results indicate that the presence of surgical masks at either encoding or retrieval, or both, impairs identification accuracy. Thus, when a witness encounters a perpetrator wearing a surgical mask, presenting an unmasked lineup as opposed to a masked lineup does not harm identification accuracy.

There are several potential explanations for why we did not observe matching effects for masked faces found in previous face-recognition studies (Manley et al., 2019; Or et al., 2023; Ventura et al., 2023). First, compared to eyewitness identification paradigms, face-recognition studies typically use identical photographs for both the study and test phases, allowing for the matching of low-level visual patterns as opposed to high-level face identity information (Or et al., 2023). Additionally, participants in our study had to observe a mock-crime and make identifications from a lineup, contrasting with face-recognition studies where participants typically view photos one by one, without any additional distractors. This allows them to focus solely on processing and memorizing faces. As a result, it is possible that congruent conditions might enhance recognition in straightforward tasks. However, as conditions and tasks at encoding and retrieval become more complex, as is the case with eyewitness identification studies, memory processes and attentional factors may assume a more crucial role in face identification than matching the processing of faces between conditions.

Second, we speculate that conflicting results with Manley et al. (2022), who also implemented an eyewitness identification paradigm, might arise from differences in the types of masks used. We used surgical masks in our study, whereas Manley et al. (2022) used ski masks that only exposed the eyes. Surgical masks cover the lower part of the face while leaving the upper part of the face and the external features of the face exposed, which are important for identifying unfamiliar faces (Logan et al., 2017; Megreya et al., 2012). Thus, compared to ski masks, surgical masks cover fewer features of the head and face, which could account for the absence of matching effects in our study. The increased visibility of a larger proportion of the face might have assisted participants in making identification decisions after witnessing a masked perpetrator. However, some face recognition studies have found that compared to face masks, covering the eyes with sunglasses have led to larger reductions in recognition accuracy (Carlaw et al., 2022; Nguyen & Pezdek, 2017; Or et al., 2023). This suggests that the eye region might be more important for holistic face-identity processing than the lower region of the face (Or et al., 2023), even though the lower region contains multiple features. Consequently, to increase face recognition, congruence between encoding and retrieval conditions may be more important for disguises that either cover a large area of a face or obscure the eyes.

However, Manley et al. (2019, 2022) found higher discriminability in congruent conditions, despite using ski masks, which left the eyes exposed, and an eyewitness identification paradigm in one of their studies (Manley et al., 2022). Surprisingly, in contrast to our study, their results showed highest discriminability when a masked perpetrator was identified from a masked lineup. We speculate that attentional factors combined with the type of disguise used may account for these discrepancies. While we instructed participants to watch the video (as real-life witnesses observe the events of the crime as well as people involved in it), Manley et al. instructed participants to focus on the perpetrator (Manley et al., 2022) or on remembering the eyes of the person (Manley et al., 2019). Although it has been recently shown that pre-event instructions asking participants to attend to the crime and informing them of a subsequent identification task does not influence identification accuracy (Baldassari et al., 2023; but see Kerstholt et al., 1992), instructing participants to intentionally encode one type of target object has been shown to increase memory for that target, regardless of expectation of a subsequent memory test (Williams, 2010). As the eyes were the only element visible in the masked condition in the Manley et al. (2019, 2022) studies, this focused attention on the perpetrator could have simplified the task for participants, thereby reducing perceptual load at encoding. This in turn may have facilitated a stronger memory representation of those eyes (Murphy & Greene, 2016; Shapiro & Penrod, 1986), which could have made the identification from a masked lineup easier, thereby explaining their finding that discriminability was highest in the condition where masks were present both at encoding and at retrieval. In contrast, participants in our study, without specific instructions, likely divided their attention between the events and various features of the perpetrator during the mock-crime in both masked and unmasked conditions. This divided attention (Chun & Turk-Browne, 2007; Craik et al., 1996) could have resulted in a weaker memory representation of the perpetrator, especially in the masked condition where fewer facial features were visible. To sum up, it seems that the efficacy of congruent masked lineups may also depend on both the type of disguise and the attentional factors during the encoding phase.

Decision criterion

Research indicates that in addition to memory, disguises, including surgical masks, affect decision-making as well (Garcia-Marques et al., 2022; Guerra et al., 2022; Hockley et al., 1999; Manley et al., 2019, 2022; Or et al., 2023). First, we found that participants tended to require more evidence for a high- or a medium-confidence identification (i.e. their response criterion became stricter) than for making a low-confidence identification, regardless of the presence of surgical masks. This indicates that participants' metacognitions appeared to influence their lineup decisions: when a memory trace was perceived as weak, they tended to apply a more lenient response criterion (Brewer et al., 2022). Second, we also found that compared to when unmasked faces were presented at encoding and retrieval, decision criteria were lower when surgical masks were involved – either at encoding or retrieval, or both – at the lowest level of confidence. Prior research (A. M. Smith, 2020; A. M. Smith et al., 2019) has shown that poor encoding or retrieval conditions can induce a weak sense of recognition in witnesses, which leads them to lower their criterion for making an affirmative identification. Hence, regardless of why surgical masks might have created a weak sense of recognition – whether due to limited visual cues at encoding or retrieval or the perceived difficulty of the identification task – their presence made participants who were unsure in their memory lower their decision criteria and, thus, be more willing to choose the suspect. Our results suggest that witnesses who encounter either a masked perpetrator or a masked lineup, and who are unsure in their memory strength, might have a tendency to use a lower criterion for identification than witnesses who are more confident in their memory or who encounter an unmasked lineup after an unmasked perpetrator.

Confidence and decision time as predictors of accuracy

We were also interested in how confidence in different lineup decisions varied with the presence of surgical masks. Overall, we found that confidence was lower when the perpetrator wore a mask during the crime, or when lineup members wore surgical masks, than in conditions with no masks. We did not find fully crossed matching effects for confidence in any of the decisions. However, we did discover that confidence in correct identifications was higher after witnessing an unmasked perpetrator than after witnessing a masked one, and that confidence was higher for correct rejections made from unmasked lineups than for those made from masked lineups. These results suggest that factors that influence memory strength (Bothwell et al., 1987; Deffenbacher, 1980), such as disguises worn by the perpetrator, and factors that affect perceptions of task difficulty, such as masked lineups, could both influence confidence ratings. Yet, despite the fact that perpetrator identifications were made with higher confidence than mistaken identifications, confidence in them was surprisingly not influenced by mask-wearing at either the encoding or the retrieval stage. In summary, these results suggest that participants displayed sensitivity to both encoding and retrieval conditions by reducing their confidence when either the perpetrator or lineup members were masked (Mansour et al., 2020). However, they appeared to be unable to acknowledge the effect that surgical masks at either encoding or retrieval had on their memory and, thus, failed to adjust their confidence accordingly.

When a witness identifies a suspect, confidence is generally a reliable indicator of accuracy of that identification (Wixted & Wells, 2017). Confirming this notion (Hypothesis 3), our results showed that as confidence increased, so did the accuracy of suspect identifications. However, we found that this predictive utility held true only for identifications made from unmasked lineups after encoding an unmasked perpetrator, aligning with the findings of O'Rourke et al. (1989). When masks were worn either at encoding or retrieval, or at both stages, confidence no longer effectively discriminated between correct and mistaken identifications. Therefore, consistent with the optimality hypothesis (Bothwell et al., 1987; Deffenbacher, 1980), surgical masks at encoding negatively influence subsequent identification accuracy, which also has a detrimental effect on the confidence–accuracy relationship. This supports the notion that as memory becomes weaker, witnesses are not always able to adjust their confidence accordingly (Giacona et al., 2021). Moreover, our findings suggest that the confidence–accuracy relationship could be undermined not just by encoding conditions, but by retrieval conditions as well. Prior research has shown that increases in task difficulty, for instance due to poor retrieval conditions, can have a negative effect on the confidence–accuracy relationship (A. M. Smith et al., 2019; Weber & Brewer, 2004). Therefore, our results indicate that confidence might not serve as a reliable indicator of suspect identification accuracy when either the perpetrator or lineup members are masked.

A key question for judges and juries is whether high-confidence suspect identifications are reliable. Current literature generally supports the idea that suspect identifications made with high confidence are unaffected by estimator variables (Semmler et al., 2018; Wixted & Wells, 2017). However, in our study, high-confidence identification accuracy in conditions involving surgical masks only reached up to 72%, a rate that cannot be considered as highly accurate (A. M. Smith et al., 2021). Nevertheless, we advise caution in interpreting these findings due to low cell sizes in the CAC analysis. Consequently, more research is needed to fully understand how various disguises, including surgical masks, affect the relationship between confidence and accuracy. Yet, it seems that surgical masks affect accuracy rates even at the highest levels of confidence (Sauer et al., 2019).

In terms of decision time, we found evidence to support the commonly observed pattern that perpetrator identifications are quicker than inaccurate identifications (Brewer et al., 2006; Sauerland & Sporer, 2007, 2009; S. M. Smith et al., 2000; Sporer, 1992, 1994). We also discovered that all decisions were made quicker from TP lineups than from TA lineups, suggesting that in the absence of the perpetrator and, thus, a strong match to memory, lineup decisions take more time. As to the effect of surgical masks on decision times for different lineup decisions, we did not find any matching effects. However, we found that whether the perpetrator was masked or not during encoding did impact decision times. Identifications, both accurate and mistaken, were made quicker when an unmasked perpetrator was encoded, implying that an unmasked perpetrator provides a more detailed memory representation, which led participants to choose faster. But can decision time be considered a reliable predictor for accuracy when an eyewitness has identified someone? Our results echo previous findings (Seale-Carlisle et al., 2019) suggesting that it can: identifications were more likely to be accurate as decision time decreased. However, this relationship held true only when the lineup members did not wear masks. Furthermore, although encoding a masked perpetrator increased the likelihood of a mistaken identification, the predictive utility of decision time was not dependent on whether the perpetrator was masked or not. In conclusion, our results suggest that irrespective of the presence of surgical masks at encoding, quicker identifications seem to indicate accuracy only when made from lineups without surgical masks.

Limitations

In addition to the aspects that increased the ecological validity of our study and the generalizability of our results, there are two limitations to consider. First, to create masked lineups, we superimposed surgical masks on fillers, which could have affected our results (Estudillo et al., 2021; Noyes et al., 2021). We photographed targets both with and without the masks, whereas for fillers we superimposed surgical masks that the targets wore to the existing images. Although surgical masks cover the lower part of the face, they fit quite tightly to the face, and, thus, the face shape information (e.g. jawline) could still be available. However, participants in our study could not use this feature in making a lineup decision as all the surgical masks of the lineup members (and face shape information) were identical to the mask of the target. Thus, superimposing surgical masks on lineup members could have hindered participants in making accurate lineup decisions. Second, while most research in this area is conducted with male targets, we used female targets in our experiment.

Implications and conclusions

The current study is the first to explore the effect of surgical masks worn by a perpetrator during encoding and lineup members during identification on eyewitness decisions. These findings have important implications, both theoretically and practically. First, our findings question the applicability of the encoding specificity principle when a perpetrator wears a surgical mask. While the principle predicts improved identification performance when disguises match at encoding and retrieval, our study suggests this is not always the case. It appears that a mix of factors at the encoding and retrieval stages can help explain the effect that disguises have on eyewitness identification decisions. However, one thing is clear from our results: surgical masks – whether worn at encoding, or retrieval – hurt identification accuracy. Our findings indicate that witnessing a perpetrator wearing a surgical mask and viewing a lineup with masks might lead to fewer perpetrator identifications and more mistaken identifications, respectively, as well as make it more challenging to discriminate between guilty and innocent suspects when compared to scenarios where no masks are present. Additionally, it appears that neither confidence nor decision time reliably predicts suspect identification accuracy in these cases.

From a practical perspective, these findings suggest that if a witness comes across a perpetrator wearing a surgical mask, presenting that witness with an unmasked lineup could be as effective as presenting a masked lineup. However, we emphasize that before making procedural suggestions, more research is needed to discover whether and in what conditions (different disguises, conditions of encoding and retrieval stages, attentional factors, etc.) matching the lineup to the encoding condition improves lineup decision accuracy and discriminability, as many relevant variables and interactions remain unexplored regarding the effect of disguises on lineup accuracy.

Acknowledgements

We wish to thank Erlend Art Arras for help with creating the photo and video stimuli, Maris Soonsein and Marjaliisa Palu for editing the photo stimuli, Ats Veske and Christofer Robert Mäeumbaed for help with data collection and Kirke Kaur for help with the theoretical overview.

Ethical standards

Declaration of conflicts of interest

Annegrete Palu has declared no conflicts of interest.
Aire Raidvee has declared no conflicts of interest.
Valeri Murnikov has declared no conflicts of interest.
Kristjan Kask has declared no conflicts of interest.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the Research Ethics Committee of the University of Tartu and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent

Informed consent was obtained from all individual participants included in the study

Funding

This research was partially supported by the Estonian Research Council [grant number PRG1151].

Supplemental data

Supplemental material is available via the 'Supplementary' tab on the article's online page (https://dx.doi.org/10.1080/13218719.2023.2242435).

ORCID

Annegrete Palu http://orcid.org/0000-0003-4738-3558
Aire Raidvee http://orcid.org/0000-0003-2385-9099
Valeri Murnikov http://orcid.org/0000-0002-0129-2362
Kristjan Kask http://orcid.org/0000-0002-9396-6941

Data availability statement

The data that support the findings of this study are available from the corresponding author upon request.

References

Abudarham, N., & Yovel, G. (2019). Same critical features are used for identification of familiarized and unfamiliar faces. *Vision Research*, *157*, 105–111. https://doi.org/10.1016/j.visres.2018.01.002

Babwin, D., Dazio, S. (2020, May 16). Coronavirus masks a boon for crooks who hide their faces. *AP NEWS*. https://apnews.com/article/nyc-wire-virus-outbreak-us-news-ap-top-news-ca-state-wire-f97b4914b4159dec0c98359fac123d52.

Baldassari, M. J., Moore, K., Lindsay, D. S., Hyman, I., Hope, L., Mah, E. Y., Horry, R., Clark, A., Mileva, M., Karanian, J., Cullen, H., van Golde, C., Palma, T. A., Jeye, B., Van der Cruyssen, I., Wiechert, S., Saraiva, R. B., Mansour, J. K., Mestry, N., … Cahill, B. S. (2020, March 13). Investigating Crime Video Instructions in Eyewitness Research. *OSF*. https://osf.io/zb85d/

Baldassari, M. J., Moore, K. N., Hyman, I. E., Hope, L., Mah, E. Y., Lindsay, D. S., Mansour, J., Saraiva, R., Horry, R., Rath, H., Kelly, L., Jones, R., Vale, S., Lawson, B., Pedretti, J., Palma, T. A., Cruz, F., Quarenta, J., Van der Cruyssen, I., … Wiechert, S. (2023). The effect of pre-event instructions on eyewitness identification. *Cognitive Research*, *8*(1), 16. https://doi.org/10.1186/s41235-023-00471-4

Barr, D. J. (2019, April 2). Coding categorical predictor variables in factorial designs. https://talklab.psy.gla.ac.uk/tvw/catpred/

Berger, A., & Kiefer, M. (2021). Comparison of different response time outlier exclusion methods: A simulation study. *Frontiers in Psychology*, *12*, 675558. https://doi.org/10.3389/fpsyg.2021.675558

Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology*, *72*(4), 691–695. https://doi.org/10.1037/0021-9010.72.4.691

Brewer, N., Caon, A., Todd, C., & Weber, N. (2006). Eyewitness identification accuracy and response latency. *Law and Human Behavior*, *30*(1), 31–50. https://doi.org/10.1007/s10979-006-9002-7

Brewer, N., Weber, N. & Semmler, C. (2005). Eyewitness identification. In N. Brewer & K. D. Williams (Eds.), *Psychology and law: An empirical perspective* (pp. 177–221). The Guilford Press.

Brewer, N., Weber, N., & Semmler, C. (2007). A role for theory in eyewitness identification research. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol II: Memory for people.* (pp. 201–218)Lawrence Erlbaum Associates Publishers. https://doi.org/10.4324/9780203936368

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology. Applied*, *12*(1), 11–30. https://doi.org/10.1037/1076-898X.12.1.11

Brewer, N., Zweck, T., Lucas, C. A., & Guidolin, M. (2022). Eyewitnesses' pre-lineup memory strength inferences can influence identification decisions. *Journal of Police and Criminal Psychology*, *37*(2), 339–350. https://doi.org/10.1007/s11896-021-09462-x

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology*, *5*(4), 339–360. https://doi.org/10.1037/1076-898X.5.4.339

Carlaw, B. N., Huebert, A. M., McNeely-White, K. L., Rhodes, M. G., & Cleary, A. M. (2022). Detecting a familiar person behind the surgical mask: Recognition without identification among masked versus sunglasses-covered faces. *Cognitive Research*, *7*(1), 90. https://doi.org/10.1186/s41235-022-00440-3

Carlson, C. A., Hemby, J. A., Wooten, A. R., Jones, A. R., Lockamyeir, R. F., Carlson, M. A., Dias, J. L., & Whittington, J. E. (2021). Testing encoding specificity and the diagnostic feature-detection theory of eyewitness identification, with implications for showups, lineups, and partially disguised perpetrators. *Cognitive Research*, *6*(1), 14. https://doi.org/10.1186/s41235-021-00276-3

Carragher, D. J., & Hancock, P. J. B. (2020). Surgical face masks impair human face matching performance for familiar and unfamiliar faces. *Cognitive Research*, *5*(1), 59. https://doi.org/10.1186/s41235-020-00258-x

Cash, D. K., & Pazos, L. A. (2023). Masking the truth: The impact of face masks on deception detection. *The Journal of Social Psychology*, *2023*, 1–14. https://doi.org/10.1080/00224545.2023.2195092

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology*, *17*(2), 177–184. https://doi.org/10.1016/j.conb.2007.03.005

Cohen, A. L. (2020). *sdtlu: Signal Detection Models for Lineups* (R package version 0.0.0.90100).

Cohen, A. L., Starns, J. J., & Rotello, C. M. (2021). sdtlu: An R package for the signal detection analysis of eyewitness lineup data. *Behavior Research Methods*, *53*(1), 278–300. https://doi.org/10.3758/s13428-020-01402-7

Coutinho, M. V. C., Redford, J. S., Church, B. A., Zakrzewski, A. C., Couchman, J. J., & Smith, J. D. (2015). The interplay between uncertainty monitoring and working memory: Can metacognition become automatic? *Memory & Cognition*, *43*(7), 990–1006. https://doi.org/10.3758/s13421-015-0527-1

Craik, F. I., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology*, *125*(2), 159–180. https://doi.org/10.1037/0096-3445.125.2.159

Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987a). 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

Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987b). The reliability of eyewitness identification: The role of system and estimator variables. *Law and Human Behavior*, *11*(3), 233–258. https://doi.org/10.1007/BF01044644

Davies, G., Ellis, H., & Shepherd, J. (1977). Cue saliency in faces as assessed by the 'Photofit' Technique. *Perception*, *6*(3), 263–269. https://doi.org/10.1068/p060263

Davies, G., & Flin, R. (1984). The man behind the mask – Disguise and face recognition. *Journal of Practical Research & Applications*, *3*(2), 83–95.

Deffenbacher, K. A. (1980). Eyewitness accuracy and confidence: Can we infer anything about their relationship? *Law and Human Behavior*, *4*(4), 243–260. https://doi.org/10.1007/BF01040617

DeGutis, J., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition*, *126*(1), 87–100. https://doi.org/10.1016/j.cognition.2012.09.004

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

Douma, A., Moniz, E., Tarr, M., & Peissig, J. (2012). Familiarity and the recognition of disguised faces. *Journal of Vision*, *12*(9), 980–980. https://doi.org/10.1167/12.9.980

Duncan, M. (2006). *A signal detection model of compound decision tasks* (Technical Report TR 2006-256). Defence Research and Development Canada. https://apps.dtic.mil/sti/citations/ADA473015.

Efron, B., & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429246593

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception*, *8*(4), 431–439. https://doi.org/10.1068/p080431

Estudillo, A. J., Hills, P., & Wong, H. K. (2021). The effect of face masks on forensic face matching: An individual differences study. *Journal of Applied Research in Memory and Cognition*, *10*(4), 554–563. https://doi.org/10.1016/j.jarmac.2021.07.002

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is 'special' about face perception? *Psychological Review*, *105*(3), 482–498. https://doi.org/10.1037/0033-295X.105.3.482

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., König, P. (2017). LabVanced: A unified JavaScript framework for online studies. *International Conference on Computational Social Science*.

Freud, E., Stajduhar, A., Rosenbaum, R. S., Avidan, G., & Ganel, T. (2020). The COVID-19 pandemic masks the way people perceive faces. *Scientific Reports*, *10*(1), 1–8. https://doi.org/10.1038/s41598-020-78986-9

Freud, E., Stajduhar, A., Rosenbaum, R. S., Avidan, G., & Ganel, T. (2021). Recognition of masked faces in the era of the pandemic: No improvement, despite extensive, natural exposure. *PsyArXiv*. https://doi.org/10.31234/osf.io/x3gzq

Garcia-Marques, T., Oliveira, M., & Nunes, L. (2022). That person is now with or without a mask: How encoding context modulates identity recognition. *Cognitive Research*, *7*(1), 29. https://doi.org/10.1186/s41235-022-00379-5

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

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*(1), 8–20. https://doi.org/10.3758/BF03198438

Graham, D. L., & Ritchie, K. L. (2019). Making a spectacle of yourself: The effect of glasses and sunglasses on face perception. *Perception*, *48*(6), 461–470. https://doi.org/10.1177/0301006619844680

Greene, C. M., Murphy, G., & Januszewski, J. (2017). Under high perceptual load, observers look but do not see. *Applied Cognitive Psychology*, *31*(4), 431–437. https://doi.org/10.1002/acp.3335

Guerra, N., Pinto, R., Mendes, P. S., Rodrigues, P. F. S., & Albuquerque, P. B. (2022). The impact of COVID-19 on memory: Recognition for masked and unmasked faces. *Frontiers in Psychology*, *13,* 960941. https://doi.org/10.3389/fpsyg.2022.960941

Hockley, W. E., Hemsworth, D. H., & Consoli, A. (1999). Shades of the mirror effect: Recognition of faces with and without sunglasses. *Memory & Cognition*, *27*(1), 128–138. https://doi.org/10.3758/BF03201219

Jarudi, I. N., Sinha, P. (2003). Relative contributions of internal and external features to face recognition. https://dspace.mit.edu/handle/1721.1/7274.

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, *17*(5), 577–596. https://doi.org/10.1080/09658210902976969

Kamal, A. M., & Burkell, J. (2011). Addressing Uncertainty: When Information is Not Enough/Faire face à l'incertitude : quand l'information ne suffit pas. *Canadian Journal of Information and Library Science*, *35*(4), 384–396. https://doi.org/10.1353/ils.2011.0030

Kask, K., Ventsel, R., & Toomela, A. (2019). The relationship between the development of conceptual thinking and children's responses to different question types. *Nordic Psychology*, *71*(4), 235–248. https://doi.org/10.1080/19012276.2019.1586572

Kent, C., Lamberts, K., & Patton, R. (2018). Cue quality and criterion setting in recognition memory. *Memory & Cognition*, *46*(5), 757–769. https://doi.org/10.3758/s13421-018-0796-6

Kerstholt, J. H., Raaijmakers, J. G., & Valeton, J. M. (1992). The effect of expectation on the identification of known and unknown persons. *Applied Cognitive Psychology*, *6*(2), 173–180. https://doi.org/10.1002/acp.2350060207

Kramer, R. S. S., & Ritchie, K. L. (2016). Disguising superman: How glasses affect unfamiliar face matching. *Applied Cognitive Psychology*, *30*(6), 841–845. https://doi.org/10.1002/acp.3261

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology*, *21*(3), 451–468. https://doi.org/10.1037/0096-1523.21.3.451

Logan, A. J., Gordon, G. E., & Loffler, G. (2017). Contributions of individual face features to face discrimination. *Vision Research*, *137*, 29–39. https://doi.org/10.1016/j.visres.2017.05.011

Maiorana, N., Dini, M., Poletti, B., Tagini, S., Rita Reitano, M., Pravettoni, G., Priori, A., & Ferrucci, R. (2022). The effect of surgical masks on the featural and configural processing of emotions. *International Journal of Environmental Research and Public Health*, *19*(4), 2420. https://doi.org/10.3390/ijerph19042420

Manley, K. D., Chan, J. C. K., & Wells, G. L. (2019). Do masked-face lineups facilitate eyewitness identification of a masked individual? *Journal of Experimental Psychology*, *25*(3), 396–409. https://doi.org/10.1037/xap0000195

Manley, K. D., Chan, J. C. K., & Wells, G. L. (2022). Improving face identification of mask-wearing individuals. *Cognitive Research*, *7*(1), 27. https://doi.org/10.1186/s41235-022-00369-7

Mansour, J. K. (2019, November 24). Data and analysis scripts. *OSF*. https://osf.io/qydb4/

Mansour, J. K., Beaudry, J. L., Bertrand, M. I., Kalmet, N., Melsom, E. I., & Lindsay, R. C. L. (2020). Impact of disguise on identification decisions and confidence with simultaneous and sequential lineups. *Law and Human Behavior*, *44*(6), 502–515. https://doi.org/10.1037/lhb0000427

Marini, M., Ansani, A., Paglieri, F., Caruana, F., & Viola, M. (2021). The impact of facemasks on emotion recognition, trust attribution and re-identification. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-84806-5

McKelvie, S. J. (1976). The role of eyes and mouth in the memory of a face. *The American Journal of Psychology*, *89*(2), 311–323. https://doi.org/10.2307/1421414

Megreya, A. M., & Burton, A. M. (2006). Recognising faces seen alone or with others: When two heads are worse than one. *Applied Cognitive Psychology*, *20*(7), 957–972. https://doi.org/10.1002/acp.1243

Megreya, A. M., Memon, A., & Havard, C. (2012). The headscarf effect: Direct evidence from the eyewitness identification paradigm. *Applied Cognitive Psychology*, *26*(2), 308–315. https://doi.org/10.1002/acp.1826

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

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

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

Mushtaq, F., Bland, A., & Schaefer, A. (2011). Uncertainty and cognitive control. *Frontiers in Psychology*, *2*, 249. https://doi.org/10.3389/fpsyg.2011.00249

Nguyen, T. B., & Pezdek, K. (2017). Memory for disguised same and cross-race faces: The eyes have it. *Visual Cognition*, *25*(7–8), 762–769. https://doi.org/10.1080/13506285.2017.1329762

Noyes, E., Davis, J. P., Petrov, N., Gray, K. L. H., & Ritchie, K. L. (2021). The effect of face masks and sunglasses on identity and expression recognition with super-recognizers and typical observers. *Royal Society Open Science*, *8*(3), 201169. https://doi.org/10.1098/rsos.201169

Noyes, E., & Jenkins, R. (2019). Deliberate disguise in face identification. *Journal of Experimental Psychology: Applied*, *25*(2), 280–290. https://doi.org/10.1037/xap0000213

O'Donnell, C., & Bruce, V. (2001). Familiarisation with faces selectively enhances sensitivity to changes made to the eyes. *Perception*, *30*(6), 755–764. https://doi.org/10.1068/p3027

Or, C. C.-F., Ng, K. Y. J., Chia, Y., Koh, J. H., Lim, D. Y., & Lee, A. L. F. (2023). Face masks are less effective than sunglasses in masking face identity. *Scientific Reports*, *13*(1), 4284. https://doi.org/10.1038/s41598-023-31321-4

O'Rourke, T. E., Penrod, S. D., Cutler, B. L., & Stuve, T. E. (1989). The external validity of eyewitness identification research: Generalizing across subject populations. *Law and Human Behavior*, *13*(4), 385–395. https://doi.org/10.1007/BF01056410

Palmer, M. A., Brewer, N., & Weber, N. (2010). Postidentification feedback affects subsequent eyewitness identification performance. *Journal of Experimental Psychology. Applied*, *16*(4), 387–398. https://doi.org/10.1037/a0021034

Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology*, *3*(4), 406–417. https://doi.org/10.1037/0278-7393.3.4.406

Quigley-McBride, A., & Wells, G. L. (2021). Methodological considerations in eyewitness identification experiments. In A. M. Smith, M. P. Toglia, & J. M. Lampinen (Eds.), *Methods, Measures, and Theories in Eyewitness Identification Tasks*. Routledge.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic processing predicts face recognition. *Psychological Science*, *22*(4), 464–471. https://doi.org/10.1177/0956797611401753

Righi, G., Peissig, J. J., & Tarr, M. J. (2012). Recognizing disguised faces. *Visual Cognition*, *20*(2), 143–169. https://doi.org/10.1080/13506285.2012.654624

Royer, J., Blais, C., Charbonneau, I., Déry, K., Tardif, J., Duchaine, B., Gosselin, F., & Fiset, D. (2018). Greater reliance on the eye region predicts better face recognition ability. *Cognition*, *181*, 12–20. https://doi.org/10.1016/j.cognition.2018.08.004

Sadr, J., Jarudi, I., & Sinha, P. (2003). The role of eyebrows in face recognition. *Perception*, *32*(3), 285–293. https://doi.org/10.1068/p5027

Sauer, J. D., & Brewer, N. (2015). Confidence and accuracy of eyewitness identification. In T. Valentine & J. P. Davis (Eds.), *Forensic facial identification: Theory and practice of identification from eyewitnesses, composites and CCTV* (pp. 185–208). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118469538.ch8

Sauer, J. D., Palmer, M. A., & Brewer, N. (2019). Pitfalls in using eyewitness confidence to diagnose the accuracy of an individual identification decision. *Psychology, Public Policy, and Law*, *25*(3), 147–165. https://doi.org/10.1037/law0000203

Sauerland, M., & Sporer, S. L. (2007). Post-decision confidence, decision time, and self-reported decision processes as postdictors of identification accuracy. *Psychology, Crime & Law*, *13*(6), 611–625. https://doi.org/10.1080/10683160701264561

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

Schönweitz, F., Eichinger, J., Kuiper, J., Ongolly, F., Spahl, W., Prainsack, B., & Zimmermann, B. (2022). The social meanings of artefacts: Face masks in the COVID-19 pandemic. *Frontiers in Public Health*, *10*, 829904. https://doi.org/10.3389/fpubh.2022.829904

Seale-Carlisle, T. M., Colloff, M. F., Flowe, H. D., Wells, W., Wixted, J. T., & Mickes, L. (2019). Confidence and response time as indicators of eyewitness identification accuracy in the lab and in the real world. *Journal of Applied Research in Memory and Cognition*, *8*(4), 420–428. https://doi.org/10.1016/j.jarmac.2019.09.003

Seale-Carlisle, T. M., & Mickes, L. (2016). US line-ups outperform UK line-ups. *Royal Society Open Science*, *3*(9), 160300. https://doi.org/10.1098/rsos.160300

Semmler, C., Dunn, J., 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. (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. (2020). Why do mistaken identification rates increase when either witnessing or testing conditions get worse? *Journal of Applied Research in Memory and Cognition*, *9*(4), 495–507. https://doi.org/10.1016/j.jarmac.2020.08.002

Smith, A. M., & Neal, T. M. S. (2021). The distinction between discriminability and reliability in forensic science. *Science & Justice*, *61*(4), 319–331. https://doi.org/10.1016/j.scijus.2021.04.002

Smith, A. M., Smalarz, L., Ditchfield, R., & Ayala, N. T. (2021). Evaluating the claim that high confidence implies high accuracy in eyewitness identification. *Psychology, Public Policy, and Law*, *27*(4), 479–491. https://doi.org/10.1037/law0000324

Smith, A. M., Smalarz, L., Wells, G. L., Lampinen, J. M., & Mackovichova, S. (2022). Fair lineups improve outside observers' discriminability, not eyewitnesses' discriminability: Evidence for differential filler-siphoning using empirical data and the WITNESS computer-simulation architecture. *Journal of Applied Research in Memory and Cognition*, *11*(4), 534–544. https://doi.org/10.1037/mac0000021

Smith, A. M., Wells, G. L., Lindsay, R. C. L., & Myerson, T. (2018). Eyewitness identification performance on showups improves with an additional-opportunities instruction: Evidence for present-absent criteria discrepancy. *Law and Human Behavior*, *42*(3), 215–226. https://doi.org/10.1037/lhb0000284

Smith, A. M., Wilford, M. M., Quigley-McBride, A., & Wells, G. L. (2019).

Mistaken eyewitness identification rates increase when either witnessing or testing conditions get worse. *Law and Human Behavior*, *43*(4), 358–368. https://doi.org/10.1037/lhb0000334

Smith, A., Wells, G. L., Smalarz, L., Lampinen, J., Mackovichova, S. (2020, July 3). Filler similarity × quality of view × culprit presence. *OSF*. https://osf.io/djpqs/?view_only=b43770f40d6142ecad176d9df69399cf.

Smith, S. M., Lindsay, R. C. L., & Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed? *The Journal of Applied Psychology*, *85*(4), 542–550. https://doi.org/10.1037/0021-9010.85.4.542

Southall, A., Syckle, K. V. (2020, March 8). Coronavirus bandits? 2 Armed men in surgical masks rob racetrack. *The New York Times*. https://www.nytimes.com/2020/03/08/nyregion/aqueduct-racetrack-robbery.html.

Sporer, S. L. (1992). Post-dicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and non-choosers. *European Journal of Social Psychology*, *22*(2), 157–180. https://doi.org/10.1002/ejsp.2420220205

Sporer, S. L. (1994). Decision times and eyewitness identification accuracy in simultaneous and sequential lineups. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 300–327). Cambridge University Press. https://doi.org/10.1017/CBO9780511759192.015

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

Stajduhar, A., Ganel, T., Avidan, G., Rosenbaum, R. S., & Freud, E. (2022). Face masks disrupt holistic processing and face perception in school-age children. *Cognitive Research*, *7*(1), 9. https://doi.org/10.1186/s41235-022-00360-2

Stephan, B. C. M., & Caine, D. (2007). What is in a view? The role of featural information in the recognition of unfamiliar faces across viewpoint transformation. *Perception*, *36*(2), 189–198. https://doi.org/10.1068/p5627

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology*, *24*(6), 1379–1396. https://doi.org/10.1037/0278-7393.24.6.1379

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology*, *46*(2), 225–245. https://doi.org/10.1080/14640749308401045

Tanaka, J. W., & Simonyi, D. (2016). The 'parts and wholes' of face recognition: A review of the literature. *Quarterly Journal of Experimental Psychology*, *69*(10), 1876–1889. https://doi.org/10.1080/17470218.2016.1146780

Terry, R. L. (1994). Effects of facial transformations on accuracy of recognition. *The Journal of Social Psychology*, *134*(4), 483–492. https://doi.org/10.1080/00224545.1994.9712199

Toomela, A. (2003). Relationships between personality structure, structure of word meaning, and cognitive ability: A study of cultural mechanisms of personality. *Journal of Personality and Social Psychology*, *85*(4), 723–735. https://doi.org/10.1037/0022-3514.85.4.723

Toseeb, U., Bryant, E. J., & Keeble, D. R. T. (2014). The Muslim headscarf and face perception: "They all look the same, don't they?" *PLoS One*, *9*(2), e84754. https://doi.org/10.1371/journal.pone.0084754

Toseeb, U., Keeble, D. R. T., & Bryant, E. J. (2012). The significance of hair for face recognition. *PLoS One*, *7*(3), e34144. https://doi.org/10.1371/journal.pone.0034144

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

Tredoux, C., Naylor, T. (2018). *r4lineups: Statistical Inference on Lineup Fairness* (0.1.1). https://CRAN.R-project.org/package=r4lineups.

Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical tests. *Psychological Methods*, *6*(4), 371–386. https://doi.org/10.1037/1082-989X.6.4.371

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

Ventura, M., Palmisano, A., Innamorato, F., Tedesco, G., Manippa, V., Caffò, A. O., & Rivolta, D. (2023). Face memory and facial expression recognition are both affected by wearing disposable surgical face masks.

Cognitive Processing, *24*(1), 43–57. https://doi.org/10.1007/s10339-022-01112-2

Vredeveldt, A., & Sauer, J. D. (2015). Effects of eye-closure on confidence-accuracy relations in eyewitness testimony. *Journal of Applied Research in Memory and Cognition*, *4*(1), 51–58. https://doi.org/10.1016/j.jarmac.2014.12.006

Wang, R., Li, J., Fang, H., Tian, M., & Liu, J. (2012). Individual differences in holistic processing predict face recognition ability. *Psychological Science*, *23*(2), 169–177. https://doi.org/10.1177/0956797611420575

Weber, N., & Brewer, N. (2004). Confidence-accuracy calibration in absolute and relative face recognition judgments. *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

Williams, C. C. (2010). Incidental and intentional visual memory: What memories are and are not affected by encoding tasks? *Visual Cognition*, *18*(9), 1348–1367. https://doi.org/10.1080/13506285.2010.486280

Wixted, J. T., Vul, E., Mickes, L., & Wilson, B. M. (2018). Models of lineup memory. *Cognitive Psychology*, *105*, 81–114. https://doi.org/10.1016/j.cogpsych.2018.06.001

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

World Health Organization. (2021, December). *When and how to use masks*. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/when-and-how-to-use-masks.

Wright, D. B., & London, K. (2009). Multilevel modelling: Beyond the basic applications. *The British Journal of Mathematical and Statistical Psychology*, *62*(Pt 2), 439–456. https://doi.org/10.1348/000711008X327632

Yaniv, I., Yates, J. F., & Smith, J. K. (1991). Measures of discrimination skill in probabilistic judgment. *Psychological Bulletin*, *110*(3), 611–617. https://doi.org/10.1037/0033-2909.110.3.611