



Check for updates

**AUTHORS:**

Kyra Scott<sup>1</sup>   
Colin Tredoux<sup>1</sup>   
Alicia Nortje<sup>1</sup>

**AFFILIATION:**

<sup>1</sup>Eyewitness and ACSENT Laboratories, Department of Psychology, University of Cape Town, Cape Town, South Africa

**CORRESPONDENCE TO:**

Kyra Scott

**EMAIL:**

kyra.em.scott@gmail.com

**DATES:**

**Received:** 23 Aug. 2021

**Revised:** 19 Nov. 2022

**Accepted:** 05 Dec. 2022

**Published:** 29 Mar. 2023

**HOW TO CITE:**

Scott K, Tredoux C, Nortje A. Evaluating the utility of facial identification information: Accuracy versus precision. *S Afr J Sci.* 2023;119(3/4), Art. #12067. <https://doi.org/10.17159/sajs.2023/12067>

**ARTICLE INCLUDES:**

- Peer review
- Supplementary material

**DATA AVAILABILITY:**

- Open data set
- All data included
- On request from author(s)
- Not available
- Not applicable

**EDITOR:**

Leslie Swartz

**KEYWORDS:**

facial composites, facial descriptions, identification, eyewitness, face recognition

**FUNDING:**

National Research Foundation of South Africa; University of Cape Town



# Evaluating the utility of facial identification information: Accuracy versus precision

Facial identification evidence obtained from eyewitnesses, such as person descriptions and facial composites, plays a fundamental role in criminal investigations and is regularly regarded as valuable evidence for apprehending and prosecuting perpetrators. However, the reliability of such facial identification information is often queried. Person descriptions are frequently reported in the research literature as being vague and generalisable, whilst facial composites often exhibit a poor likeness to an intended target face. This raises questions regarding the accuracy of eyewitness facial identification information and its ability to facilitate efficient searches for unknown perpetrators of crimes. More specifically, it questions whether individuals, blind to the appearance of a perpetrator of a crime (i.e. the public), can correctly identify the intended target face conveyed by facial identification information recalled from eyewitness memory, and which of the two traditional facial identification formats would be better relied upon by law enforcement to enable such searches. To investigate this, in the current study ( $N=167$ ) we employed two metrics – identification accuracy and identification precision – to assess the utility of different formats of eyewitness facial identification information in enabling participants to correctly identify an unknown target face across three different formats: facial descriptions, facial composites and computer-generated description-based synthetic faces. A statistically significant main effect for the format of facial identification information on identification accuracy ( $p<0.001$ ) was found, with a higher target identification accuracy yielded by facial descriptions in comparison to composites and description-based synthetic faces. However, the reverse relationship was established for identification precision, where composites and description-based synthetic faces enabled significantly greater precision in the narrowing down of a suspect pool than did facial descriptions, but did not necessarily result in the retainment of the intended target face ( $p<0.001$ ).

**Significance:**

- This study highlights the relative importance of person descriptions in being as effective as, if not better than, facial composites in allowing for accurate identifications when solely relying upon eyewitness facial identification information to facilitate the search for unknown perpetrators.
- We introduce the metric of identification precision to evaluate the utility of facial identification information obtained by eyewitnesses.
- The study provides a novel approach to directly model facial composites based on a person description using traditional fourth-generation composite systems, thus producing a computer-generated description-based synthetic face that resembles a target face observed by an eyewitness.

## Introduction

Eyewitness identification evidence (e.g. person descriptions and facial composites) refers to any information provided by an eyewitness, from memory, regarding the physical appearance of the perpetrator. Such evidence frequently plays a pivotal role in the criminal justice system<sup>1,2</sup>, often impacting the outcome of criminal cases, especially in the absence of physical incriminating evidence, such as DNA or fingerprints, or still and video imagery, such as CCTV footage, enabling identification of the offender<sup>3</sup>. When there is a lack of incriminating evidence and suitable suspects, police officials regularly depend upon eyewitness accounts and descriptive evidence to facilitate wider searches amongst the public in order to apprehend unidentified perpetrators.<sup>4,5</sup> The general sentiment towards eyewitness identification evidence held by society and law enforcement is that it is a reliable source of evidence to correctly establish the identity of a perpetrator.<sup>1,6-8</sup> However, misidentifications and convictions of innocent people often occur when such convictions are primarily dependent upon this form of evidence.<sup>9,10</sup> Furthermore, researchers have repeatedly raised concerns surrounding the risks of solely relying upon eyewitness identification evidence to aid in the search and accurate identification of offenders, as it has been determined to be easily susceptible to suggestibility, bias and error.<sup>11-14</sup> Moreover, facial identification information recalled by witnesses may be particularly unreliable because of a frequently reported lack of specificity of facial descriptors<sup>15,16</sup> and a poor similarity of facial composites to intended target faces<sup>9</sup>. Nonetheless, evidence obtained from eyewitnesses is essential to criminal investigations, specifically in the absence of other incriminating evidence and in time-sensitive cases, where retrieving camera imagery may take too long.<sup>17</sup>

Law enforcement officers routinely collect eyewitness person descriptions of offenders<sup>5</sup>, and indeed are expected to do so by South African law (see *R v Shekelele*<sup>18</sup>). Typically the first evidence gathered during a criminal investigation, person descriptions are verbal or written recollections of a perpetrator's appearance given by eyewitnesses.<sup>19</sup> Various interview procedures, which attempt to achieve maximum reliability and comprehensive recall of details<sup>20</sup>, are employed by police officers to elicit person descriptions from eyewitness memory<sup>11</sup>. These include the use of person-feature checklists<sup>21</sup>, the Cognitive Interview<sup>22,23</sup> and its variations such as the Holistic-Cognitive Interview<sup>24</sup> for facial-composite construction, the Person Description Interview<sup>25</sup>, Self-Administered Interview<sup>26</sup> and free-recall descriptions<sup>11</sup>. Despite recalled descriptors achieving accuracy rates of over 80%<sup>15,27-29</sup>, person descriptions are still

frequently perceived as vague and incomplete in terms of portraying an offender's appearance<sup>1,8,15,16</sup>. This is because witnesses tend to provide information regarding basic physical attributes (e.g. gender, age, height, build and clothing) instead of more distinct and identifiable attributes that are vital for person identification, such as inner-facial features (e.g. eye shape and colour, and relative size and shape of the nose, lips and mouth).<sup>29</sup> Moreover, eyewitness reporting of inner-facial features is typically less than 40% accurate.<sup>29</sup> Hence, facial composites are predominately relied upon by law enforcement to locate unidentified offenders.<sup>5</sup>

Facial composites are attempted visual depictions of a perpetrator's face, constructed by an eyewitness's recall with the aid of a software operator or sketch artist who guides the process.<sup>30</sup> Composites facilitate the search for suspects by providing an image of the perpetrator that it is hoped members of the public familiar with the perpetrator will recognise, so they can then report the individual in question.<sup>5</sup> Originally, composites were drawn by artists working alongside eyewitnesses to create a sketch of the offender.<sup>6</sup> This later evolved to the use of 'featural' systems (mechanical systems and computerised 'feature' systems), allowing police with less artistic expertise to facilitate the construction process for facial composites.<sup>17</sup> These systems enable witnesses to build a likeness of the perpetrator's face using a selection of set facial features.<sup>6</sup> However, the composites produced by such 'featural' systems tended to bear a poor similarity to the suspect<sup>31</sup>, and so led to the introduction of 'holistic' fourth-generation composite systems, such as EvoFit<sup>32</sup>, EFIT-V<sup>33</sup> and ID<sup>34</sup>, which utilise holistic-configural facial processing<sup>35</sup>. Currently, the most widely used facial composite systems, fourth-generation composite systems, outperform previous sketch and 'featural' systems in enabling accurate identifications.<sup>31</sup> Nonetheless, facial composites still attain low correct identification naming rates, with laboratory studies documenting naming rates of less than 50% under forensically valid conditions.<sup>36</sup> Similarly, police utilising a fourth-generation composite system reported an identification naming rate of approximately 60% calculated on suspect arrests.<sup>37</sup>

The reliance on both person descriptions and face composites to search for unidentified offenders raises the question of which of these methods law enforcement should be emphasising for better identification accuracy. Several studies have sought to directly compare and determine the effectiveness of person descriptions and facial composites, and the consensus is that facial descriptions significantly outperform facial composites, resulting in fewer false identifications of wanted offenders.<sup>38-41</sup>

In the current study we sought to determine the effective utility of facial identification information obtained from eyewitnesses by assessing its effectiveness in facilitating efficient and accurate searches for unknown perpetrators of crimes. More specifically we considered whether individuals, blind to the appearance of a perpetrator of a crime (i.e. the public), could correctly identify the intended target face conveyed by the facial identification information recalled from eyewitness memory. This was attempted by evaluating the performance of three different formats of eyewitness facial identification information in enabling participants blinded to the appearance of simulated offenders to identify accurately and unambiguously the said persons. To do this, two performance metrics, identification accuracy and identification precision, were employed to measure differences in the effectiveness of different formats of eyewitness facial identification information to enable a correct identification. The three different formats of eyewitness facial identification information were (1) eyewitness facial descriptions, (2) traditional facial composites and (3) novel computer-generated description-based synthetic faces produced by fourth-generation facial composite software.

Given that facial descriptions have been shown to facilitate better recognition accuracy than facial composites<sup>38-41</sup>, computer-generated description-based synthetic faces were used in the present study as a visual form of eyewitness facial descriptions. This produced a target face observed by an eyewitness whilst minimising human involvement in facial composite development, in an attempt to control and limit human-induced inaccuracies and potential eyewitness memory issues during composite construction, such as further memory decay, operator bias<sup>35</sup> and pattern-specific interference<sup>21</sup>. Description-based synthetic faces were also generated with the aid of fourth-generation facial composite software to standardise the creation medium for the visual facial identification

information across the two different visual formats. Furthermore, by incorporating these computer-generated description-based synthetic faces, it becomes possible to explore whether facial descriptions do capture important information that facial composites typically overlook. The current study hypothesises that (1) facial descriptions will exhibit significantly higher target identification accuracy than both facial composites and computer-generated description-based synthetic faces, and (2) traditional facial composites will have significantly poorer target identification accuracy than computer-generated description-based synthetic faces.

## Method

### Design and setting

This study was conducted in two stages: (1) facial identification information was obtained from mock witnesses for use as the primary experimental material, and (2) the utility of the respective facial identification information was evaluated in terms of identification accuracy and identification precision. To obtain facial identification information, participants acting as eyewitnesses were exposed to two simulated offenders and asked to (1) describe the face and (2) construct a traditional facial composite based on (a) memory or (b) in-view observation. These descriptions and composites were then evaluated via an online experiment hosted on the survey platform Qualtrics. New participants, independent of stage one, were required to complete two identification tasks to assess the relative identification accuracy and identification precision of eyewitness facial descriptions and face composites.

### Factors

A mixed repeated-measures design, consisting of  $3 \times 2 \times 2$  randomised experimental cells, was implemented. The three factors were: (1) 'Format of Facial Identification Information' (within-subjects factor; description versus computer-generated description-based synthetic face versus facial composite), (2) 'Mode of Recall' (between-subjects factor; from memory versus in-view) and (3) 'Target' (between-subjects factor; Target A versus Target B). For this study, two simulated offenders' faces (target faces) were introduced to control for possible target bias (e.g. through differential facial distinctiveness).

### Dependent variables

The utility of different formats of facial identification information was measured through two dependent variables: 'Identification Accuracy' and 'Identification Precision'. 'Identification Accuracy' was operationalised as correct identification of an intended target face from an array. 'Identification Precision' measured how efficient identification information was in narrowing down a pool of potential suspects from a starting face-matrix set of 24 faces.

### Participants

The initial sample consisted of 169 volunteers recruited electronically with opportunistic sampling. It was established that no participants knew or recognised either of the simulated offenders. Two participants were excluded as they had not resided in southern Africa for at least the past 5 years (all target faces were South African). This criterion was imposed due to the sensitivity of facial identification, which can be negatively impacted by own group bias.<sup>22</sup> Thus, the final sample size was 167, with the majority (59.28%) being female participants. Mean age was 38.65 years (standard deviation (SD) = 15.55). Approximately 68% of the sample self-identified as white, 14% as coloured, 11% as Indian, 5% as black and 1% as Asian.

### Materials

#### Face stimuli

The simulated offenders were two coloured South African men in their early 20s. The two identification tasks were constructed to assess identification accuracy and identification precision. To do so, separate  $6 \times 4$  photographic face matrices were constructed for each simulated offender. These face matrices acted as the suspect pools, each

consisting of the target faces and 23 standardised filler faces that bore a resemblance to the simulated offenders. A modal facial description was produced for each simulated offender based on the most frequently mentioned facial attributes given by 17 individuals in delayed-matching, free-recall facial descriptions. These individuals did not participate in any other part of the study. Subsequently, 23 filler faces were selected from a database of coloured South African men's faces for each simulated offender's suspect pool, based on the modal facial descriptions. Facial photographs were presented in colour, with faces bearing neutral expressions. To ensure unbiased arrangements of photographs in the face matrices, placement of photographs was randomised.

### Facial identification information

Facial identification information was gathered independently of the online evaluation tasks. The process used to gather the facial identification information is depicted in Figure 1. A total of 16 undergraduate psychology students were recruited, via convenience sampling, to act as mock witnesses and produce the required eyewitness facial identification information of the two simulated offenders. Individuals were recruited under the pretense of a 'tarot-card reading' study to mitigate potential priming and subject-expectancy effects.

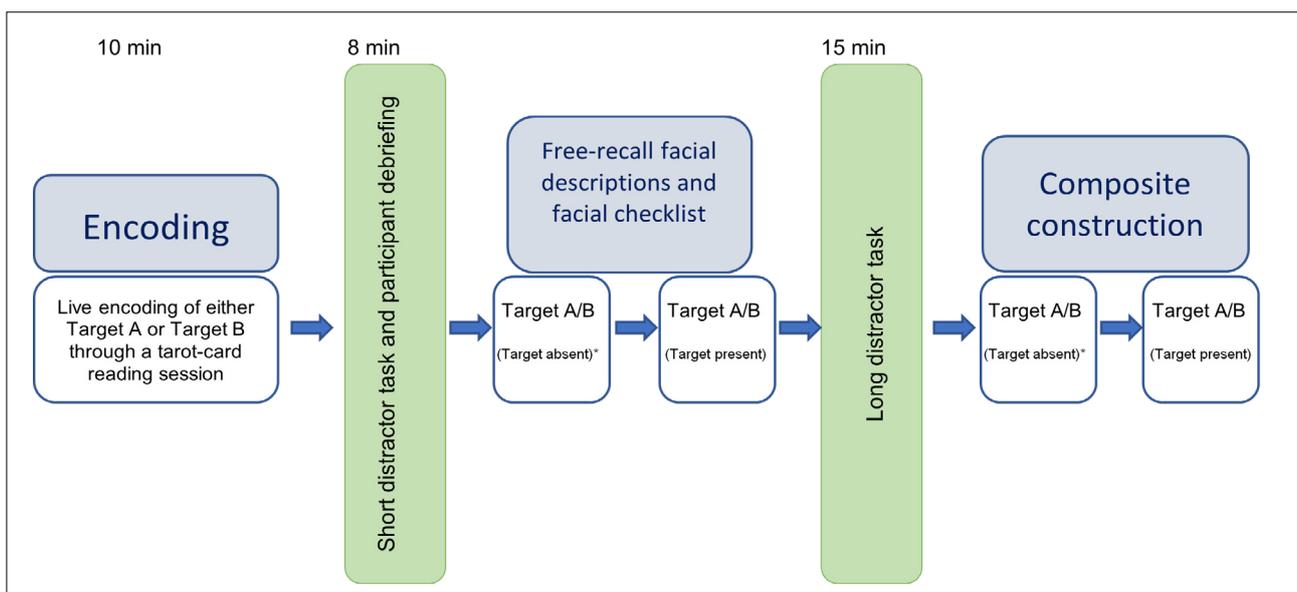
To gather facial identification information from memory, the mock witnesses were randomly exposed to a live, in-person, 10-minute encoding to one of the two simulated offenders, who acted as a tarot-card reader. Exposure to the two simulated offenders during the encoding was counter-balanced between participants. Following an 8-minute distractor task, witnesses were required to give free-recall facial descriptions and complete a facial checklist, which was a modified version of the Aberdeen University Face Rating Schedule (FRS)<sup>21</sup>, for the two simulated offenders. This consisted of 41 ratings of facial attributes on a five-point Likert scale. Subsequently, facial descriptions of the simulated offender (tarot-card reader) were elicited from witness memory, followed by another elicited facial description of the other simulated offender whilst they sat in view. A 15-minute distractor task followed the elicitation of facial descriptions before witnesses constructed facial composites for both simulated offenders. First, witnesses produced the facial composite of the tarot-card reader from memory, and then they produced an in-view facial composite of the other simulated offender. This process produced a total of 32 free-recall facial descriptions (16 from memory and 16 in view) and 32 facial composites (16 from memory and 16 in view). Refer to Figure 2 for examples of the facial identification information collected.

### Composite software

A fourth-generation composite system, ID<sup>34</sup>, was used to produce all visual facial identification information (i.e. facial composites and computer-generated description-based synthetic faces). ID is a contemporary eigenface composite construction software similar to EvoFit<sup>32</sup> and EFIT-V<sup>33</sup>, which are currently utilised by police personnel<sup>17</sup>. The system promotes holistic-configural facial processing by presenting witnesses with an array of 12 synthetic faces and allowing for the repeated selection, morphing and blending of multiple faces together. This iterative process utilises underlying eigenfaces and evolutionary genetic algorithms, such as Population Incremental Learning and M-Choice, to yield new 'generations' of faces bearing a likeness to previous selected faces. Overall variation between generated faces in the arrays continuously lessens upon each iteration as formerly selected faces are combined to create new 'generations' until a synthetic face is produced that resembles the intended target face. Witnesses are also capable of altering individual facial features and featural spacing.

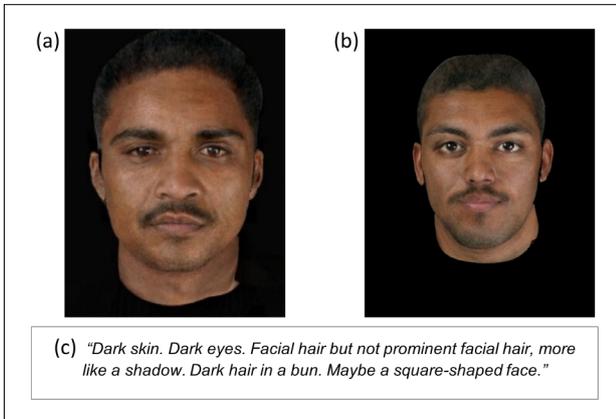
### Synthetic description-to-face generation

As for the traditional facial composites, ID was also used as the primary system to generate the description-based synthetic faces. This enabled standardisation across visual facial identification information formats in the creation medium so that the utility of eyewitness facial identification evidence could be better assessed. Computer-generated description-based synthetic faces were produced by initially training a front-end multivariate regression model to model the underlying eigenface coefficients required by ID to construct facial composites. To create this model, an independent online survey was circulated to a total of 72 individuals that required them to rate ID-generated facial composites using the modified FRS (checklist rating schedule). Sixty composites were rated in total, by between five and seven raters. An average rating was computed for each composite on each of the 41 dimensions in the FRS, thus averaging out differences between raters on all dimensions. The composite faces were created using an appearance model that had 138 eigenfaces (basis dimensions) underlying it (see Tredoux et al.<sup>34</sup> for a description of how ID works). In other words, each of the 60 synthetic or composite faces had a known set of coefficients that created the composite face from the underlying eigenfaces. The perceptual ratings on the FRS were then entered into a multivariate regression model, in which each of the underlying eigenfaces in the ID system was modelled by ratings on the perceived (FRS) face dimensions. In other words, a predictive model was built that allowed us to generate coefficient values from ratings, and the coefficient values could in



**Figure 1:** Procedure for gathering facial identification information. Target encoding was counterbalanced between individuals. Target-absent condition was used for the simulated offender encoded during the tarot-card reading session. The target-present condition was used as an in-view control of strong encoding.

turn be used to generate synthetic faces. FRS facial descriptions obtained for the two simulated offenders were then entered into the multivariate regression model and the relevant description-based synthetic faces were produced in ID, based on the model's output. This created a total of 32 computer-generated description-based synthetic faces (16 from memory and 16 in view). Refer to Figure 2 for an example of a computer-generated description-based synthetic face.



**Figure 2:** Samples of gathered facial identification information: (a) computer-generated synthetic face produced using an in-view description; (b) in-view facial composite constructed in ID; and (c) in-view facial description.

### Procedure

To assess the utility of the three sources of facial identification information, a ranking task and a set-reduction task were conducted online to test identification accuracy and identification precision, respectively. Demographic information was gathered virtually from participants before they commenced the online tasks. This was done to measure the potential impact of own race bias<sup>42</sup> on facial identification results.

Participants were randomly assigned to one of four groups based upon 'Target' and 'Mode of Recall'. Hence, for both tasks they were exposed to only one of the two target faces and facial identification information produced either in view or from memory. However, all participants were presented with at least one instance of each source of eyewitness facial identification information (facial description, facial composite and computer-generated description-based synthetic face). Order of exposure to the different formats of facial identification information was randomised to control for order effects. Additionally, facial identification information was arbitrarily allocated to participants for each respective format.

Jointly presented with the  $6 \times 4$  face matrix and the relevant facial identification information, participants were required first to complete a set-reduction task and then a ranking task for each format of facial identification information. The set-reduction task aimed to evaluate identification precision by asking participants to eliminate all faces from the face matrix that were not deemed to bear a resemblance to the provided facial identification information, resulting in a reduced subset of faces. The ranking task assessed identification accuracy by requiring participants to rank the 24 faces in the face matrix from least to most likely to resemble the given facial identification information.

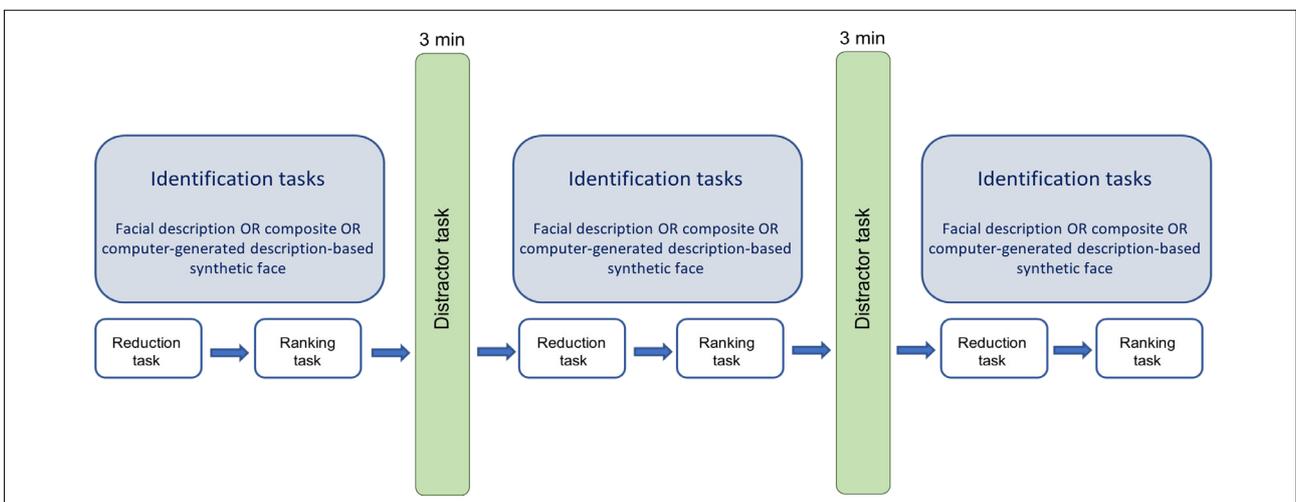
Between the two identification tasks for each format of facial identification information, participants engaged in a short 3-minute distractor task, mitigating potential interference effects. See Figure 3 for an overview of the procedure.

### Data analyses

All statistical analyses were completed in SPSS<sup>43a</sup>, with the alpha value set to 0.05 as per convention<sup>44</sup>. Analyses commenced with descriptive summaries of each variable. Further examination of group differences for 'Identification Accuracy' and 'Identification Precision' across formats of facial identification information and mode of recall (from memory versus in view) were carried out via a series of mixed analyses of variance (ANOVA). Analysis of the data pertaining to 'Identification Accuracy' found strong negative skewing of the standardised residual distributions across most experimental cells. To correct for this, a 'ln(25 - X)' transformation was applied to the data. Further assessment of the transformed data also revealed significant outliers<sup>45</sup> ( $n=16$ ), which were subsequently omitted. Outliers were identified as transformed standardised residuals greater than 2.30 or less than -2.30.<sup>46,47</sup> The Greenhouse-Geisser estimate ( $\epsilon=0.79$ ) was used in the interpretation of the ANOVA results relating to 'Identification Accuracy' to correct degrees of freedom, as Mauchly's test of sphericity was found to be violated ( $\chi^2(2)=46.05, p<0.001$ ). Visual assessment of the standardised residual distributions for 'Identification Precision' also found the data to be strongly skewed and thus an appropriate two-step transformation<sup>47</sup> was applied to correct for this. The Greenhouse-Geisser estimate ( $\epsilon=0.85$ ) was also used in the interpretation of the ANOVA results for 'Identification Precision', as Mauchly's test of sphericity indicated violation of the assumption of sphericity ( $\chi^2(2)=32.25, p<0.001$ ).

### Coding of dependent variables

'Identification Accuracy' was scored as a value between 1 and 24, determined by the positional rank a participant placed on the target face of the simulated offender during the ranking task. Higher scores are indicative of a greater accuracy than lower scores, with a score of 24 reflecting an accuracy of 100%. 'Identification Precision' was coded as the size of the remaining suspect pool after the set-reduction task took



**Figure 3:** Procedure for testing identification accuracy and identification precision of facial identification information. The order of presentation of the type of facial identification information was counterbalanced.

place to narrow down the face matrix. This variable was measured as a value ranging between 0 and 24, where smaller set sizes related to higher levels of identification precision than larger set sizes.

## Results

### Identification accuracy

Descriptive statistics of identification accuracy across the different experimental conditions are reported in Table 1. On average, facial descriptions were able to facilitate the highest identification accuracy (mean = 16.74 (SD=6.87)). Facial composites (mean = 12.38 (SD=6.77)) and computer-generated description-based synthetic faces (mean = 11.05 (SD=5.99)) achieved lower identification accuracy scores. This resulted in facial descriptions performing on average, 18.17% and 23.71% better in accurately identifying a simulated offender than facial composites and computer-generated description-based synthetic faces, respectively. This claim in performance difference is further supported by the statistical analysis that follows.

**Table 1:** Descriptive statistics for identification accuracy ( $N=167$ )

	Target A		Target B		Total
	Memory ( $n=37$ )	In view ( $n=39$ )	Memory ( $n=51$ )	In view ( $n=40$ )	
Facial description	16.89 (5.88)	16.41 (5.69)	16.71 (8.05)	16.98 (7.36)	16.74 (6.87)
Facial composite	12.95 (5.65)	15.00 (6.47)	10.00 (7.21)	12.33 (6.60)	12.38 (6.77)
Computer-generated description-based synthetic face	12.97 (6.37)	12.38 (5.05)	9.08 (5.36)	10.50 (6.57)	11.05 (5.99)

Note: Identification accuracy was scored from 1 to 24, where higher scores indicate better identification accuracy.

Values shown are mean (standard deviation).

A summary of the results from the  $3 \times 2 \times 2$  mixed-designs ANOVA are reported in Table 2. A significant main effect of the type of facial identification information format on identification accuracy was indicated:  $F(1.57, 231.40) = 56.79, p < 0.001, \eta_p^2 = 0.28$ . Post-hoc analysis, adopting a Bonferroni adjustment as a correction for multiple

significance testing, found facial descriptions (mean = 1.95 (standard error (SE)=0.07), 95% confidence interval (CI) [1.81, 2.08]) allow for significantly greater identification of an intended target face than facial composites (mean = 2.50 (SE=0.05), 95% CI [2.42, 2.59]) or computer-generated description-based synthetic faces (mean = 2.66 (SE=0.03), 95% CI [2.60, 2.73]), which tended to result in poorer identification accuracy. However, no significant difference in identification accuracy was revealed between facial composites and computer-generated description-based synthetic faces.

Interpretation of interactions found a significant interaction between 'Mode of Recall' and 'Format of Facial Identification Information', with facial composites produced from memory yielding significantly poorer identification accuracy than facial composites generated in view of a simulated offender:  $F(1, 147)=8.99, p=0.003$ . Facial descriptions and computer-generated description-based synthetic faces performed similarly in relation to identification accuracy, regardless of 'Mode of Recall'. See Figure 4 for an illustration of the results. 'Format of Facial Identification Information' and 'Target' showed a relatively small two-way interaction effect:  $F(1.57, 231.40)=7.06, p=0.003$ . Further analyses of this two-way interaction, via simple main effects and pairwise comparisons, revealed a significantly lower identification accuracy for all facial composites of Target B over all facial composites of Target A: mean = 0.30 (SE = 0.09),  $p=0.001$ , 95% CI [0.12, 0.48];  $F(1, 147) = 11.19, p=0.001$ . A similar result with respect to poorer identification accuracy on Target B over Target A was also found for the computer-generated description-based synthetic faces: mean = 0.22 (SE=0.06),  $p=0.001$ , 95% CI [0.10, 0.35];  $F(1, 147)=12.61, p=0.001$ . No significant difference was indicated between facial descriptions of the two different target faces: mean = 0.19 (SE=0.14),  $p=0.17$ , 95% CI [-0.08, 0.46];  $F(1, 147)=1.90, p=0.170$  (Figure 5).

### Identification precision

Computer-generated description-based synthetic faces enabled the greatest narrowing down of the suspect pool by having the fewest remaining faces in the reduced face-matrix set (mean = 7.32 (SD=6.07)), resulting in an overall better identification precision than the other formats of facial identification information. Facial composites were revealed to perform only slightly worse than computer-generated description-based synthetic faces (mean = 7.40 (SD=6.19)). However, facial descriptions, on average, yielded the most retained faces within the face-matrix set (mean = 9.28 (SD=7.47)), leading to the worst identification precision amongst the different formats of facial identification information. Full descriptive statistics are shown in Table 3.

**Table 2:** Summary of ANOVA tests of between-subject effects and within-subject effects for identification accuracy ( $N=151$ )

Source	df	Sum of squares	Mean square	F	p	$\eta_p^2$
Between subjects	150.00	25.820				
Target	1.00	0.470	0.47	3.30	0.071	0.02
Mode of Recall	1.00	0.350	0.35	2.49	0.117	0.02
Target $\times$ Mode of Recall	1.00	0.000	0.00	0.00	0.958	0.00
Error	147.00	20.770	0.14			
Within subjects	237.68	157.800				
Facial Identification Information Format	1.57	41.980	26.67	56.79	<0.001**	0.28
Facial Identification Information Format $\times$ Target	1.57	5.220	3.32	7.06	0.003*	0.05
Facial Identification Information Format $\times$ Mode of Recall	1.57	1.770	1.13	2.40	0.106	0.02
Facial Identification Information Format $\times$ Target $\times$ Mode of Recall	1.57	0.170	0.11	0.23	0.741	0.00
Error	231.40	108.660	0.47			
Total	387.68	134.480				

\* $p < 0.05$ ; \*\* $p < 0.001$ . All listed p-values are two-tailed; df, degrees of freedom;  $\eta_p^2$ , partial eta squared

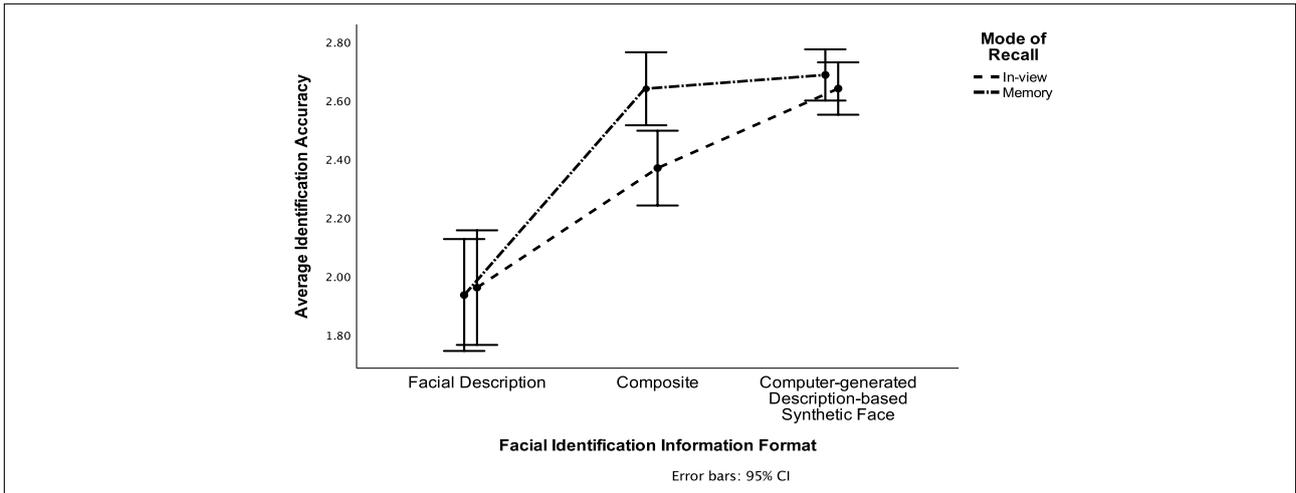


Figure 4: Interaction effect for the type of facial identification information format on mode of recall.

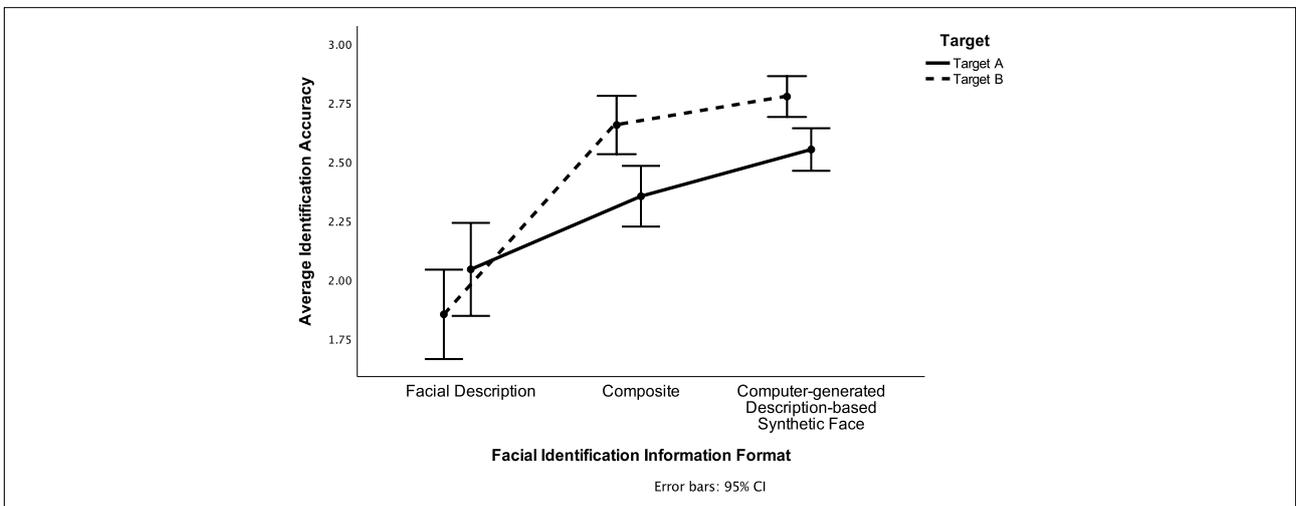


Figure 5: Interaction effect for the type of facial identification information format on target.

Table 3: Descriptive statistics for identification precision (N=167)

	Target A		Target B		Total
	Memory (n=37)	In view (n=39)	Memory (n=51)	In view (n=40)	
Facial description	9.22 (7.01)	9.79 (7.22)	7.71 (7.13)	10.85 (8.37)	9.28 (7.47)
Facial composite	6.46 (6.19)	7.64 (6.54)	7.39 (6.00)	8.02 (6.22)	7.40 (6.19)
Computer-generated description-based synthetic face	7.43 (6.37)	6.23 (5.64)	7.57 (6.01)	7.98 (6.36)	7.32 (6.07)

Note: Identification accuracy was scored from 1 to 24, where higher scores indicate better identification accuracy. Values shown are mean (standard deviation).

As no significant main effects or interactions were indicated in relation to 'Target', we collapsed across this factor. A 3 × 2 mixed-designs ANOVA ('Facial Identification Information Format' × 'Mode of Recall') was run on the identification precision outcome. These results are shown in Table 4. A significant main effect of the format of facial identification information in relation to identification precision was revealed:  $F(1.70, 427.18)=9.48, p<0.001, \eta_p^2=0.05$ .

Table 4: Summary of ANOVA tests of between-subject effects and within-subject effects for identification precision (N=167)

Source	df	Sum of squares	Mean square	F	p	$\eta_p^2$
Between subjects	166.00	12 742.06				
Mode of Recall	1.00	66.09	66.09	0.86	0.355	0.01
Error	165.00	12 675.97	76.82			
Within subjects	283.41	7958.47				
Facial Identification Information Format	1.70	427.18	251.72	9.48	<0.001*	0.05
Facial Identification Information Format × Mode of Recall	1.70	94.83	55.88	2.10	0.132	0.01
Error	280.01	7436.46	26.56			
Total	449.41	20 700.53				

\* $p<0.05$ ; \*\* $p<0.001$ . All listed p-values are two-tailed; df, degrees of freedom;  $\eta_p^2$ , partial eta squared

Further post-hoc analysis of this effect, adopting a Sidak adjustment as all assumptions were met, determined that facial descriptions (mean = 9.45 (SE=0.55), 95% CI [8.36, 10.53]) resulted in a significantly worse identification precision than both facial composites (mean = 7.54 (SE=0.47), 95% CI [6.61, 8.46]) and computer-generated description-based synthetic faces (mean = 7.44 (SE=0.46), 95% CI [6.53, 8.34]). Thus, producing a mean difference of 1.91 (SE=0.57,  $p=0.003$ , 95% CI [0.54, 3.28]) and 2.01 ((SE=0.58,  $p=0.002$ , 95% CI [0.61, 3.41]) between facial descriptions and the other two formats of facial identification information, respectively. No significant difference was established between facial composites and computer-generated description-based synthetic faces in relation to the number of faces eliminated from the face-matrix set. Thus, on average, the use of facial descriptions resulted in the largest groups of suspects after reduction when compared with groups of suspects reduced using composite faces or computer-generated synthetic faces derived from descriptions.

## Discussion

The purpose of the present research was to evaluate the utility of eyewitness facial identification information by examining whether individuals, blind to the appearance of a perpetrator of a crime, could accurately and precisely identify the target face conveyed by eyewitness facial identification evidence. This was achieved by independently gathering facial identification information from mock witnesses and then conducting an online experiment, consisting of a sorting task and a set-reduction task, to assess the eyewitness facial identification information in relation to identification accuracy and identification precision, respectively.

In line with previous studies<sup>38-40</sup>, and the current hypotheses, the results indicate that facial descriptions outperformed facial composites and computer-generated description-based synthetic faces for identification accuracy. However, facial composites and computer-generated description-based synthetic faces were found to achieve similar rates of identification accuracy, not supporting the original hypothesis that computer-generated description-based synthetic faces would outperform facial composites. In relation to identification precision, facial composites and computer-generated description-based synthetic faces enabled significantly higher levels of precision in narrowing down a suspect pool than facial descriptions did. This inverse relationship between identification accuracy and identification precision could be accredited to numerous factors.

Differing description modalities (i.e. visual versus written) could account for some of the observed identification differences. Facial composites and computer-generated description-based synthetic faces convey facial information visually to enable the search and identification of an unknown target face, which is also a visual stimulus. This allows a straight comparison between the ability of the two visual stimuli to facilitate an identification. Written facial descriptions, on the other hand, convey facial information non-visually, making comparisons between this identification information and the other stimuli less direct.<sup>19</sup> This may account for the higher levels of identification precision achieved for facial composites and computer-generated description-based synthetic faces. Human observers, blind to the true appearance of a target, might well accept a visual facial likeness as an absolute depiction of a target face instead of an approximate, vague impression.<sup>21,40</sup> This may encourage the search for an exact match to the facial composite or computer-generated description-based synthetic face in a set of faces. Moreover, this may consequently produce more precise identifications, as individuals invoke stricter judgement and selection criteria of a chosen target face, resulting in more potential suspects being eliminated from a suspect pool, but not necessarily resulting in better identification accuracy. In comparison, facial descriptions convey non-visual facial information, which requires individuals to internally construct a mental image of the reported target face before searching for it. This could produce more subjective interpretations of a target, potentially leading to lower levels of congruency with a chosen target face as a less stringent judgement and selection criterion is invoked. Ultimately, this might produce less precise identifications as individuals eliminate fewer potential suspects from a suspect pool.

Despite finding that facial descriptions were associated with a poorer filtering ability, they did enable the most accurate identifications in comparison with facial composites and computer-generated description-based synthetic faces. This finding is in line with those of previous studies in which facial descriptions have been shown to outperform facial composites in enabling accurate identification of intended targets. Modality-specific interference and high cognitive loads induced by transferring an internalised, visual representation of an encoded target face into an externalised visual facial composite have been suggested as potential reasons for why facial composites yield poorer identification accuracy than facial descriptions.<sup>21,40</sup> However, we propose some alternative reasons that may account for the differences seen in identification accuracy amongst the differing formats of facial identification information.

Facial descriptions enable less facial information to be conveyed in comparison to visual facial identification information formats, as not all facial features and aspects are necessarily described by an eyewitness. Visual forms of facial identification, such as face composite images, require the inclusion of all facial details to illustrate the face of a perpetrator, regardless of whether the eyewitness is confident regarding all aspects of the face. The necessary inclusion of more information in facial composites and computer-generated description-based synthetic faces creates a higher probability of incorporating erroneous details. This enables greater subjective identification precision (set reduction) in third parties, as a greater number of comparisons can be made against a potential target and the visual facial identification information. However, the increased probability of error-prone facial information increases the chances of poorer identification accuracy rates. Facial descriptions implicitly allow witnesses to omit facial features and details regarding the suspect's face that are uncertain to them. This is likely to lower the probability of error within a facial description, allowing for better identification accuracy. Facial descriptions also allow eyewitnesses to place more emphasis upon facial attributes that they are more confident of, whilst less certain details can remain generalisable through vague or ambiguous word choice, or simple omission. Facial composites and computer-generated description-based synthetic faces, on the other hand, do not allow for emphasis on specific facial features. If better composite systems become available which address this difference, then facial composites could potentially outperform facial descriptions as they already outperform in terms of identification precision, which currently enables better narrowing down of a suspect pool than facial descriptions, although does not always lead to the correct identification of a perpetrator.

The significant difference between identification accuracy targets set for facial composites and for computer-generated description-based synthetic faces may reflect limitations in the ability of the composite software (ID) to adequately capture the visual likeness of the one simulated offender's face. This speculation is justified by the lack of a significant difference being established between facial descriptions of the two simulated offenders, which removes the possibility of facial distinctiveness inflating identification accuracy for the one simulated offender for the facial composites and computer-generated description-based synthetic faces.<sup>48</sup>

Facial composites constructed in view of the simulated offender enabled better identification accuracy than facial composites constructed from memory. This finding was in line with previous research that proposes constructing facial composites from memory enables memory decay and potentially other factors that can negatively impact accurate portrayal of a perpetrator as a result of misleading information leaking into the construction of the facial composite.<sup>49</sup>

## Limitations and future research directions

From this study, it is reasonable to assume that facial descriptions will always be lacking in identification precision because of the lower level of facial information conveyed in this format, as it will always be near impossible to verbally describe all aspects of a face. Visual facial identification information, on the other hand, appears to provide a higher level of identification precision because of the amount and depth of



facial information conveyed by a visual likeness of a face, even if much of this information is inaccurate. If facial composites and computer-generated description-based synthetic faces can improve the accuracy of features in the visual medium, they may consistently outperform facial descriptions in identification accuracy. This would further increase the ecological validity of the use of facial composites by the criminal justice system to facilitate the search for unidentified perpetrators.

Given recent advancements in generative models, such as generative adversarial networks<sup>50,51</sup> and diffusion networks<sup>52</sup>, as well as conditionalised text-to-image models<sup>53-57</sup>, it may be feasible to increase the accuracy of computer-generated description-based synthetic faces. These networks could be utilised to fill in features that the eyewitness is not certain of by constructing the feature based on what is statistically most likely. Furthermore, a composite software that blurs facial features based on the confidence of the eyewitness's memory of a perpetrator's face could reduce the probability of erroneous facial information being introduced. This would lead to increased identification accuracy but at the cost of identification precision. This approach may allow composite operators to scale identification precision against identification accuracy. We are currently conducting research on both aspects.

## Conclusion

In this study, the utility of varying formats of eyewitness facial identification information were evaluated in relation to identification accuracy and identification precision. Verbal (or written) facial descriptions from eyewitnesses were found to facilitate the most accurate identification of a conveyed offender from a pool of suspects, by individuals blind to the appearance of the offender. On the other hand, visual formats of eyewitness facial identification information (i.e. facial composites and computer-generated description-based synthetic faces) enabled the most precise narrowing down of a suspect pool, but at the cost of lowering identification accuracy.

## Ethical considerations

Ethical clearance to conduct the study was granted by the University of Cape Town's Ethics Review Committee of the Faculty of Humanities (ethical clearance reference no. PSY2019-051). All aspects of the study were ethically approved. All participation was voluntary with informed consent given before engagement with any aspect of the study. Additionally, all participants were informed of their right to withdraw from the study at any stage without justification or penalty. Confidentiality and anonymity of participants were ensured by organising data according to randomly allocated participant numbers as well as encrypting and password-protecting all electronically stored data.

## Acknowledgements

We thank the National Research Foundation of South Africa and the University of Cape Town for their financial support.

## Competing interests

We have no competing interests to declare.

## Authors' contributions

All authors were responsible for the conceptualisation and methodology of the study. K.S. was the primary researcher for the study, whilst C.T. and A.N. provided student supervision and critical feedback throughout the entire process. K.S. conducted the data collection and wrote the original draft of the manuscript. C.T. reviewed the manuscript draft and aided in subsequent writing revisions.

## References

1. Brown C, Lloyd-Jones TJ, Robinson M. Eliciting person descriptions from eyewitnesses: A survey of police perceptions of eyewitness performance and reported use of interview techniques. *Eur J Cogn Psychol.* 2008;20(3):529–560. <https://doi.org/10.1080/09541440701728474>
2. McQuiston-Surrett D, Topp LD, Malpass RS. Use of facial composite systems in US law enforcement agencies. *Psychol Crime Law.* 2006;12(5):505–517. <http://dx.doi.org/10.1080/10683160500254904>

3. Keval HU, Sasse MA. Can we ID from CCTV? Image quality in digital CCTV and face identification performance. In: Agaian SS, Jassim SA, editors. *Mobile multimedia/image processing, security, and applications.* SPIE; 2008. p. 189–203. <https://doi.org/10.1117/12.774212>
4. Laughery KR, Duval C, Wogalter MS. Dynamics of facial recall. In: *Aspects of face processing.* Dordrecht: Springer; 1986. p. 373–387. [https://doi.org/10.1007/978-94-009-4420-6\\_40](https://doi.org/10.1007/978-94-009-4420-6_40)
5. Kovera MB, Penrod SD, Pappas C, Thill DL. Identification of computer-generated facial composites. *J Appl Psychol.* 1997;82(2):235–246. <http://dx.doi.org/10.1037/0021-9010.82.2.235>
6. Davies G, Valentine T. Facial composites: Forensic utility and psychological research. In: Lindsay RCL, Ross DF, Read JD, Toglia MP, editors. *Handbook of eyewitness psychology.* New York: Routledge; 2007. p. 59–83. <https://doi.org/10.4324/9781315805535>
7. Gabbert F, Brown C. Interviewing for face identification. In: Valentine T, Davis JP, editors. *Forensic facial identification: Theory and practice of identification from eyewitnesses, composites and CCTV.* Chichester, UK: John Wiley & Sons; 2015. p. 15–41. <https://doi.org/10.1002/9781118469538>
8. Kebbell MR, Milne R. Police officers' perceptions of eyewitness performance in forensic investigations. *J Soc Psychol.* 1998;138(3):323–330. <http://dx.doi.org/10.1080/00224549809600384>
9. Innocence Project. Eyewitness identification reform [webpage on the Internet]. c2016 [2022 Nov 10]. Available from: <https://www.innocenceproject.org/eyewitness-identification-reform/>
10. Albright TD. Why eyewitnesses fail. *Proc Natl Acad Sci USA.* 2017;114(30):7758–7764. <http://dx.doi.org/10.1073/pnas.1706891114>
11. Meissner CA, Sporer SL, Schooler JW. Person descriptions as eyewitness evidence. In: Lindsay RCL, Ross DF, Read JD, Toglia MP, editors. *Handbook of eyewitness psychology.* New York: Routledge; 2007. p. 3–34. <https://doi.org/10.4324/9781315805535>
12. Wells GL, Hasel LE. Facial composite production by eyewitnesses. *Curr Dir Psychol Sci.* 2007;16(1):6–10. <http://dx.doi.org/10.1111/j.1467-8721.2007.00465.x>
13. Hurley G. The trouble with eyewitness identification testimony in criminal cases [document on the Internet]. c2017 [cited 2021 Oct 23]. Available from: <http://ncsc.contentdm.oclc.org/cdm/ref/collection/criminal/id/280>
14. Charman SD, Gregory AH, Carlucci M. Exploring the diagnostic utility of facial composites: Beliefs of guilt can bias perceived similarity between composite and suspect. *J Exp Psychol Appl.* 2009;15(1):76–90. <http://dx.doi.org/10.1037/a0014682>
15. Fahsing IA, Ask K, Granhag PA. The man behind the mask: Accuracy and predictors of eyewitness offender descriptions. *J Appl Psychol.* 2004;89(4):722–729. <http://dx.doi.org/10.1037/0021-9010.89.4.722>
16. Meissner CA, Sporer SL, Susa KJ. A theoretical review and meta-analysis of the description-identification relationship in memory for faces. *Eur J Cogn Psychol.* 2008;20(3):414–455. <http://dx.doi.org/10.1080/09541440701728581>
17. Frowd CD. Forensic facial composites. In: Smith AM, Toglia PT, Lampinen JM, editors. *Methods, measures, and theories in eyewitness identification tasks.* New York: Routledge; 2021. p. 34–64. <https://doi.org/10.4324/9781003138105-5>
18. R v Shekelele (1953) 1 SA 636 (T).
19. Sporer SL. Psychological aspects of person descriptions. In: Sporer SL, Malpass RS, Koehnken G, editors. *Psychological issues in eyewitness identification.* New York: Psychology Press; 1996. p. 53–86. <https://doi.org/10.4324/9781315821023>
20. Milne B, Bull R. *Investigative interviewing: Psychology and practice.* Chichester, UK: Wiley; 1999.
21. Shepherd JW, Ellis HD. Face recall – methods and problems. In: Sporer SL, Malpass RS, Koehnken G, editors. *Psychological issues in eyewitness identification.* New York: Psychology Press; 1996. p. 87–115. <https://doi.org/10.4324/9781315821023>
22. Mcleod SA. The cognitive interview [webpage on the Internet]. c2019 [cited 2022 Oct 23]. Available from: <https://www.simplypsychology.org/cognitive-interview.html>



23. Kebbell MR, Milne R, Wagstaff GF. The cognitive interview: A survey of its forensic effectiveness. *Psychol Crime Law*. 1999;5(1-2):101-115. <http://dx.doi.org/10.1080/10683169908414996>
24. Frowd CD, Bruce V, Smith AJ, Hancock PJB. Improving the quality of facial composites using a holistic cognitive interview. *J Exp Psychol Appl*. 2008;14(3):276-287. <http://dx.doi.org/10.1037/1076-898X.14.3.276>
25. Demarchi S, Py J. A method to enhance person description: A field study. In: Bull R, Valentine T, Williamson T, editors. *Handbook of psychology of investigative interviewing: Current developments and future directions*. Oxford, UK: Wiley-Blackwell; 2009. p. 241-256. <https://doi.org/10.1002/9780470747599>
26. Gabbert F, Hope L, Fisher RP. Protecting eyewitness evidence: examining the efficacy of a self-administered interview tool. *Law Hum Behav*. 2009;33(4):298-307. <http://dx.doi.org/10.1007/s10979-008-9146-8>
27. Odinet G, Wolters G, Van Koppen PJ. Eyewitness memory of a supermarket robbery: A case study of accuracy and confidence after 3 months. *Law Hum Behav*. 2009;33(6):506-514. <http://dx.doi.org/10.1007/s10979-008-9152-x>
28. Woolnough PS, MacLeod MD. Watching the birdie watching you: Eyewitness memory for actions using CCTV recordings of actual crimes. *Appl Cogn Psychol*. 2001;15(4):395-411. <http://dx.doi.org/10.1002/acp.717>
29. Van Koppen PJ, Lochun SK. Portraying perpetrators: The validity of offender descriptions by witnesses. *Law Hum Behav*. 1997;21(6):661-685. <http://dx.doi.org/10.1023/a:1024812831576>
30. Frowd C, Bruce V, McIntyre A, Hancock P. The relative importance of external and internal features of facial composites. *Br J Psychol*. 2007;98(1):61-77. <http://dx.doi.org/10.1348/000712606x104481>
31. Frowd CD, Erickson WB, Lampinen JM, Skelton FC, McIntyre AH, Hancock PJB. A decade of evolving composites: Regression- and meta-analysis. *J Forensic Pract*. 2015;17(4):319-334. <http://dx.doi.org/10.1108/jfp-08-2014-0025>
32. Frowd CD, Hancock PJB, Carson D. EvoFIT: A holistic, evolutionary facial imaging technique for creating composites. *ACM Trans Appl Percept*. 2004;1(1):19-39. <http://dx.doi.org/10.1145/1008722.1008725>
33. Gibson SJ, Solomon CJ, Maylin MIS, Clark C. New methodology in facial composite construction: From theory to practice. *Int J Electron Secur Digit Forensics*. 2009;2(2):156-168. <http://dx.doi.org/10.1504/ijesdf.2009.024900>
34. Tredoux C, Nunez D, Oxtoby O, Prag B. An evaluation of ID: An eigenface based construction system. *S Afr Comput J*. 2006(37):90-97. <https://journals.co.za/doi/10.10520/EJC28017>
35. Fodarella C, Kuivaniemi-Smith H, Gawrylowicz J, Frowd CD. Forensic procedures for facial-composite construction. *J Forensic Pract*. 2015;17(4):259-270. <http://dx.doi.org/10.1108/jfp-10-2014-0033>
36. Zahradnikova B, Duchovicova S, Schreiber P. Facial composite systems: Review. *Artif Intell Rev*. 2018;49(1):131-152. <http://dx.doi.org/10.1007/s10462-016-9519-1>
37. Frowd CD, Pitchford M, Skelton F, Petkovic A, Prosser C, Coates B. Catching even more offenders with EvoFIT facial composites. In: *Third International Conference on Emerging Security Technologies*; 2012 September 5-7; Lisbon, Portugal. IEEE; 2012. p. 20-26. <http://dx.doi.org/10.1109/EST.2012.26>
38. Christie DF, Ellis HD. Photofit constructions versus verbal descriptions of faces. *J Appl Psychol*. 1981;66(3):358-363. <http://dx.doi.org/10.1037/0021-9010.66.3.358>
39. McQuiston-Surrett D, Topp LD. Externalizing visual images: Examining the accuracy of facial descriptions vs. composites as a function of the own-race bias. *Exp Psychol*. 2008;55(3):195-202. <http://dx.doi.org/10.1027/1618-3169.55.3.195>
40. Lech AM, Johnston R. The relative utility of verbal descriptions and facial composites in facial identifications. *Int J Bio-Sci Bio-Technol*. 2011;3(3):1-16.
41. Ness H. *Improving composite images of faces produced by eyewitnesses*. Stirling: University of Stirling; 2003. <http://hdl.handle.net/1893/12555>
42. Meissner CA, Brigham JC. Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychol Public Policy Law*. 2001;7(1):3-35. <http://dx.doi.org/10.1037/1076-8971.7.1.3>
43. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp; 2017.
44. Field A. *Discovering statistics using IBM SPSS statistics*. 5th ed. London: SAGE Publications; 2017.
45. Aguinis H, Gottfredson RK, Joo H. Best-practice recommendations for defining, identifying, and handling outliers. *Organ Res Methods*. 2013;16(2):270-301. <http://dx.doi.org/10.1177/1094428112470848>
46. Buzzi-Ferraris G, Manenti F. Outlier detection in large data sets. *Comput Chem Eng*. 2011;35(2):388-390. <http://dx.doi.org/10.1016/j.compchemeng.2010.11.004>
47. Templeton GF. A two-step approach for transforming continuous variables to normal: Implications and recommendations for IS research. *Commun Assoc Inf Syst*. 2011;28(1), Art. #4. <http://dx.doi.org/10.17705/1cais.02804>
48. Shapiro PN, Penrod S. Meta-analysis of facial identification studies. *Psychol Bull*. 1986;100(2):139-156. <http://dx.doi.org/10.1037/0033-2909.100.2.139>
49. Wogalter MS, Bradley Marwitz D. Face composite construction: In-view and from-memory quality and improvement with practice. *Ergonomics*. 1991;34(4):459-468. <http://dx.doi.org/10.1080/00140139108967329>
50. Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2019 June 15-20; Long Beach, CA, USA. IEEE; 2019. p. 4396-4405. <https://doi.org/10.1109/CVPR.2019.00453>
51. Chen X, Qing L, He X, Luo X, Xu Y. FTGAN: A fully-trained generative adversarial network for text to face generation [preprint]. *arXiv [cs.CV]*. 2019. <http://arxiv.org/abs/1904.05729>
52. Dhariwal P, Nichol A. Diffusion models beat GANs on image synthesis [preprint]. *arXiv [cs.LG]*. 2021. <http://arxiv.org/abs/2105.05233>
53. Ramesh A, Dhariwal P, Nichol A, Chu C, Chen M. Hierarchical text-conditional image generation with CLIP latents [preprint]. *arXiv [cs.CV]*. 2022. <http://arxiv.org/abs/2204.06125>
54. Ho J, Chan W, Saharia C, Whang J, Gao R, Gritsenko A, et al. Imagen Video: High definition video generation with diffusion models [preprint]. *arXiv [cs.CV]*. 2022. <http://arxiv.org/abs/2210.02303>
55. Reed S, Akata Z, Yan X, Logeswaran L, Schiele B, Lee H. Generative adversarial text to image synthesis. In: Balcan MF, Weinberger KQ, editors. *Proceedings of the 33rd International Conference on Machine Learning*; 2016 June 19-24; New York, USA. New York: PMLR; 2016. p. 1060-1069.
56. Xu T, Zhang P, Huang Q, Zhang H, Gan Z, Huang X, et al. AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2018 June 18-23; Salt Lake City, UT, USA. IEEE; 2018. p. 1316-1324.
57. Ding M, Yang Z, Hong W, Zheng W, Zhou C, Yin D, et al. CogView: Mastering text-to-image generation via transformers [preprint]. *arXiv [cs.CV]*. 2021. <http://arxiv.org/abs/2105.13290>