Witness Artistic Rendition and its Impacts on Visual Memory for Forensic Facial Composite Creation

Structured Abstract

PURPOSE: In the absence of photographic or other identifying evidence, composites provide crucial intelligence in police investigations, though their accuracy depends on a witness's facial memory and recall. The current study investigated a novel technique aimed at increasing face recall and composite effectiveness.

APPROACH: In this study, one group of participants (control) viewed a facial photograph, recalled the face using a cognitive interview and created a composite with a forensic artist. A second (experimental) group of participants did the same except that they sketched the face themselves prior to the cognitive interview. The impact of participants sketching the face was measured by assessing the number of "units of information" produced during free recall of the face, as well as the identifiability of the composites, evaluated by an additional group of participants who attempted to name the sketched composites. Participants also rated their general ability to draw and their general level of observation.

FINDINGS: Results showed, relative to the control group, that the experimental group provided more detailed descriptions of the face and that this improvement to memory led to creation of more identifiable composites. Therefore, our findings suggest that this artistic rendition technique enhances both the cognitive interview and the accuracy of forensic facial composites. It was also found that participants' self-rated measures of drawing and observant behaviour were positively related to the accuracy of the participants' composites.

PRACTICAL IMPLICATIONS: This simple technique of asking witnesses to sketch the face themselves could be implemented by police forces with minimal effort and impact to budget. It presents a straightforward and budget-efficient way to increase the identifiability of composite images without the need for additional lengthy training for forensic practitioners.

VALUE: Results suggest that the witness artistic rendition technique represents a novel, low-cost, and simple method that could be utilized to increase composite accuracy.

KEYWORDS: forensic art, facial identification, forensic psychology, forensic sketch, police sketch, facial composite, composite sketch, cognitive interview, facial memory, memory recall, face recognition

TYPE: Research paper

INTRODUCTION

A facial composite is an image created by a forensic artist or computer system in response to witnesses' recollection of a face they saw (also known as a "target" identity). Facial composites can be constructed either traditionally by a forensic sketch artist, or by using a variety of facial composite systems that have been developed over the last five decades. It is normal practice for a forensic practitioner to take witnesses through a process known as a cognitive interview to help them to remember and describe the face of the person they saw prior to working further with them to create a composite of the face. Facial composites are most useful in cases where eyewitness testimony is the only record of an offender's appearance—that is, in cases where camera footage or other means of recording an event is not available, or is of poor quality. After a composite has been produced, it can be disseminated to law enforcement and / or the general public in order to facilitate identification of the target (Wilkinson, 2015).

Composite construction has traditionally involved trained forensic sketch artists, professionals who use pencil and paper to produce face sketches based on the memory of the witnesses they interview. Since the mid twentieth century, forensic practitioners began to use 'mechanical' composite systems as a way to streamline the composite creation process, as a lack of access to trained personnel had long been a barrier to the production of a sketch. The first of these systems were feature based, asking witnesses to select individual features (eyebrows, nose, etc.) that matched their memory of the target from a bank of pre-selected features. The earliest of these systems were Photofit and Identikit, in the 1960s and 70s, which used physical images of features, either solid images or transparent slides, which witnesses would select and adjust on the face to achieve the best match to the face (McDonald, 1960; Penry, 1971). However, these systems resulted in very poor rates of identification (Davies et al., 2000; Ellis et al., 1978; Frowd et al., 2005). Eventually these systems were computerized, resulting in digital systems such as E-FIT, FACES, Mac-a-Mug Pro or PRO-fit, which followed the same essential concept, just in a digital workspace

(Koehn & Fisher, 1997). Recently, a new type of system has emerged, a holistic system, that creates a composite by asking witnesses to select faces for an overall match to their memory instead of individual features, 'evolving' the face by blending features and attributes. Examples include EvoFIT and E-Fit 6 (Frowd et al., 2004; VisionMetric, 2025). These more recent systems have resulted in higher identification rates, and are commonly used today in the UK and Europe (Brace et al., 2006; Davies et al., 2000; Frowd et al., 2005; Frowd et al., 2015).

Research has greatly improved the efficacy of composite systems. One technique, for example, has utilized a whole-face approach, with external features (hair, ears and neck) de-emphasized (blurred) during face construction (cf. showing external features intact). This method has been shown to increase identifiability of the internal features, the area that includes the eyes, nose and mouth that is important for later recognition of the composite face (Frowd et al., 2011; Skelton et al., 2015). Also, averaged composites – formed by combining independently produced images of the same face from different witnesses – increase the visual match of the internal features and are generally more identifiable than individual composites (Bruce et al., 2002; Frowd et al., 2014a). In addition, certain digital image manipulations, such as vertically stretching the image or by showing the face horizontally misaligned, reduce the perceived error in the face and improve composite recognition (Frowd et al., 2013; McIntyre et al., 2016). These studies, along with many others, have dramatically increased the efficacy of forensic composites (Frowd et al., 2021).

When creating a composite by any system, be it by sketching or using a feature- or holistic-based system, a forensic practitioner first takes a witness through a cognitive interview. Developed by forensic psychologists, this interview protocol helps a witness remember important details about the face by following a four-step procedure (see summary in Memon & Bull, 1991). This interview is invaluable for aiding composite creation, and research has continued to improve the process. For instance, encouraging witnesses to retrieve memories earlier can lead to more accurate composites (Brown et al., 2017), as can describing the scene of the crime in detail (Fodarella et al., 2021), or reflecting on the perceived character of a face, to promote holistic face processing (Frowd et al., 2012a; Frowd et al., 2015). Further studies investigating new techniques to potentially improve the cognitive interview and face construction, are of course valuable.

In one such study, Dando et al. (2009) developed a procedure known as *sketch plan mental reinstatement of context*, or *sketch MRC*. After building rapport with a witness, reinstatement of context is the next step of a cognitive interview that encourages witnesses to imagine themselves at the scene of the crime when they first saw the offender, 'reinstating' themselves in the original

setting. This exercise improves memory recall. Dando et al. asked witnesses to sketch a plan of the scene before the cognitive interview, as replacement for the usual MRC procedure (a simple visualization of the scene). Utilizing the manual process of drawing encourages the brain to take a systematic approach to memory. For example, witnesses remembering the placement of a corner of a wall might then recall a window in the wall, which might then lead to recalling a view of the street outside, etc. Dando et al. (2020) found that the sketch MRC technique was as effective as the standard MRC, but easier for police officers to administer. Additionally, the technique is also easier to use for certain populations, such as older witnesses (Dando et al., 2020). More generally, several psychological studies have investigated the relationship between drawing and memory, finding a positive influence on recall (Fernandes et al., 2018; Wammes et al., 2016, 2019). However, no studies seem to have used this drawing technique with witnesses to help face recall and, in doing so, potentially facilitate face construction.

The current study therefore considers this self-sketch technique for the novel application of helping a witness to recall more information about a previously-seen ("target") face. As such, the face should be remembered better; consequently, there should be a beneficial effect: construction of an ensuing composite should be more effective. Using this approach, witnesses would be asked, prior to the cognitive interview, to draw their own facial image of the person they remember seeing. Similar to Sketch MRC, this technique is based upon the psychological theory that drawing can be used as a methodical exercise that connects details to one another in witness memory (Dando et al, 2009). In other words, as witnesses sketch one part of the face (e.g., the eyes), their memory of the surrounding area should be improved (e.g., shape and colour of the eyebrows, forehead, etc.). The research invited one group of participants, the 'witnesses', to each complete a cognitive interview and produce a single composite, with half of the participants performing the self-sketch technique (the experimental group) and the other half not (the control group). For the experimental group, two sketches were thus produced from each witness, one from the witnesses themselves (which were set aside, and not used for the remainder of the research) and the other from the forensic artist.

The objective is thus to determine how this technique when used prior to the cognitive interview impacts facial recall and the recognizability of the finished composite. The research also considered two potential mediating factors (see discussion below), referred to as *Artistic Ability* and *Observant Behavior*, both of which had the potential to be positively related to face recall and the identifiability of the witness's composite. We hypothesized that the experimental group

(witnesses drawing the target face) would remember more detail about the face (compared to control group witnesses who did not draw the face) and that, as the memory of the face would now be relatively better, the resulting composites would also be more effective.

EXPERIMENT

Ethical approval for the study was obtained from the University of Dundee, Centre for Anatomy and Human Identification (CAHID) School of Science and Engineering Research Ethics Committee in April of 2024 (UOD_SSEREC_CAHID_MSc_2024_01). The study involved two stages of participation: Stage 1, where 'witness' participants were shown an image of a target face, completed a cognitive interview, and later constructed a composite sketch with the primary investigator; and Stage 2, where 'namer' participants were shown the resulting composite sketches and asked to name them.

Stage 1: Composite Creation (Witnesses)

METHOD

Design

The witness phase of the research was conducted in person, with each participant meeting the primary investigator (CD) individually. The cognitive interviews given by the investigator followed the standard four-phase format (see Memon & Bull, 1991), with context reinstatement, free recall, open questions, and probing questions phases. Participants were also asked, during the open questions phase, to describe the 'character' of the target face in a few words, an interview technique thought to increase holistic processing and enhance visual memory which has shown positive results in composite studies (Frowd et al., 2013; Frowd et al., 2015).

After signing their consent form, participant witnesses were given a question sheet, which contained basic demographic questions as well as a series of Likert scale questions about the participant's interests, skills, and experiences. The questions (see Appendix 1) were coded according to two scores, *Artistic Ability* (AA) and *Observant Behavior* (OB), both of which were expected to positively correlate with face recall and construction. AA is a measure of participants' self-reported tendency to rely on their visual brain. It includes any visual art experience they have, and their predisposition to enjoy artistic pursuits. For example, if participants select a high score on the Likert scale for statements such as "I have a lot of experience with art," or "I like to draw,"

they will have a relatively higher AA. OB is a measure of participants' self-reported memory, as well as their desire to observe others. For example, if participants select a high score on the Likert scale for statements such as "I pay a lot of attention to small details," or "I have a good memory," they will have a higher OB. See Appendix 1 for more details.

Participants

The number of participants required for each stage (face construction and composite naming) was based on existing research and checked by computer simulation (Appendix 2). For good statistical power, $1 - B > .8$, this exercise revealed that a minimum of 10 participants per group were required for each stage of the experiment, an estimate that we exceeded.

For face construction, a total of 22 witness participants were recruited through advertisements displayed around the University of Dundee campus. Participants were between 18 and 69 years, with a mean age of 26.2 (SD = 10.7). Eight participants were biologically male and 14 female; 10 identifying as women, 8 as men, and 4 nonbinary/other. Participants were randomly allocated in equal groups of 11 to the two levels of the between-subjects variable, *Interview Type*.

Materials

Target identities used in the study were well-known American celebrities (footballers, musicians, television personalities, etc.) that would be easily recognizable to the planned American 'namer' participants, but generally not to Scottish / international 'witness' participants. The third author (CF) selected neutral-expression images for each of the identities and prepared a random order so that the primary investigator, who conducted the interviews and created the composites, would not see the images. Eleven targets were used, all White European, aged 32-69 years, with 6 males and 5 females. Each identity was viewed twice in the first stage of the study, once by an experimental witness and once by a control witness. The target identities were Kristen Bell, Tom Bergeron, Tom Brady, Kelly Clarkson, Miley Cyrus, Jimmy Kimmel, Brad Paisley, Rachael Ray, Aaron Rodgers, Jon Stewart, and Chrissy Teigen.

The Samantha Steinberg catalogue (Steinberg, 2006) and FBI catalogue were used during the final phase of the cognitive interview to give witnesses visual references to describe their memories of the target. The Steinberg catalog provided image references for male faces and the FBI catalogue provided references for female faces. Witnesses were encouraged to request

adjustments periodically throughout the process, as necessary, until the composite resembled, to the highest degree possible, the witness' memory of the previously-seen face.

Procedure

Witness participants were tested individually. They were greeted by the primary investigator, who answered any questions. After signing the consent form, participants completed a questionnaire that gathered information on their age, gender, and the two AA and OB measures.

Witness participants were then shown a photograph image, blind to the primary investigator and randomly selected without replacement from the set of 22 target photographs prepared by the third author. Witnesses were asked if they recognized the target face; if they answered yes, a different image was given. This was to ensure that witnesses viewed an unfamiliar face, the normal situation for real witnesses. Witnesses reporting that the face was familiar were shown a different photograph, randomly selected as described. For the first unfamiliar face seen, participants were instructed to study the image for 60 seconds while imagining themselves as witnesses to a crime, anticipating the need to describe the suspect to law enforcement.

After an elapsed time of 3 to 4 hours, the witness participants returned to the study room for the second phase of their participation. Participants assigned to the experimental condition were given drawing materials (pencils, sharpener, eraser, A4 paper) and asked to draw an image of the person they had seen earlier; drawing time averaged around five minutes. After completion, the same as for participants in the baseline control condition, they proceeded to the cognitive interview, following the previously described four-phase format (Memon & Bull, 1991), and worked with the investigator to create a sketch of the face (following the procedure described in Fodarella et al., 2015). The face was drawn using graphite pencils on A3 paper.

Once witnesses were satisfied with their composite, the witness was debriefed as to the aims of the study. The free recall phase (which was recorded) lasted from 54 seconds to 5 minutes 43 seconds, with a mean of 2.8 minutes (*SD* =1.2), while construction of the sketch ranged from 45 minutes to 90 minutes, with a mean of 65 minutes (*SD* = 9.2). All participant information was recorded in an anonymized format.

Stage 2: Composite Identification (Namers)

METHOD

Design

The experiment involved two dependent variables (DVs). While one DV assessed the amount of information recalled by witnesses during free recall of the face (see Results), the other considered identifications for the composites made by the namer participants, which could be correct or incorrect. It was hypothesized that artistic rendition (where witnesses sketched the face) would lead to relatively greater face recall, and, for the composites, more correct and fewer incorrect identifications.

Participants

Forty namer participants were recruited through advertisements on social networking and paper flyers in various US cities. Namers received participant consent and information sheets via email, followed by a scheduled video call with the primary investigator to provide verbal consent and answer eligibility questions. Participants were required to be eighteen years or older and have spent no more than two of the last twenty years living outside the United States. Participants took part from Alabama, California, Indiana, Kentucky, Massachusetts, Michigan, New York, North Carolina, Ohio, Rhode Island, and Wisconsin.

Materials

Materials used were digital assets: photographs of the target identities and scanned images of the completed composites from Stage 1. See Figure 1 for example images.

Figure 1. Example images created in Stage 1 for Tom Bergeron

Note. Shown are materials involved in the face construction of Tom Bergeron (far right). From left to right: experimental witness-produced image, artist composite from experimental group witness, and artist composite from control witness. For reasons of copyright, the image shown here was

obtained from Wikimedia (https://commons.wikimedia.org/wiki/File:TomBergeronApr09.jpg); the image used in the research involved a more front-on view of the face.

Procedure

Participants were tested individually. Each person was asked to name a set of composites of famous faces. Participants were informed that there might be repeats, and it was also acceptable not to give a name if unsure. Images were presented sequentially and participants provided a name for each where possible. Following presentation of the composites, participants were shown the 11 original target photographs and were asked to name them. Participants received a different random order of composites and target pictures. Responses were recorded in an anonymized form. The procedure took about 15 minutes to complete, including debriefing on the aims of the experiment.

RESULTS

Two main analyses were conducted, the first on verbal recall of the face from witness participants collected in Stage 1, and the second on the effectiveness of each composite, as assessed by namer participants in Stage 2.

Verbal Recall

The first analysis assessed the total recall produced by witness participants, a measure sometimes referred to as 'completeness' (Wells, 1995). It was expected that the total amount of unique information recalled about the face should be greater for participants who sketched the face relative to those who did not.

The free recall section of each cognitive interview was transcribed from the audio recordings, and Units of Information (UOI) were derived. For example, if a participant reported that a target's "nose was long and straight, with a mole on the right nostril," this would be counted as three units of information: long nose, straight nose, mole on right nostril. The procedure was conducted such that the scorer was blind to both the identity of the target and the condition from which the description had been produced. The veracity of the resulting total scores per description was checked by an independent person.

Witness participants recalled from 11 to 34 units of information. As illustrated in Table 1, total recall was much greater overall from participants who sketched the face (Experimental) than those who did not (Control), an increase (*MD*) in 4.0 units of information on average per person.

Table 1. Units of Information (UOI) recalled for the target face from witness participants by Interview Type.

Interview Type[1]	Units of Information (UOI) Recalled				
	Total	*M*	*SE*	*95CI-*	*95CI+*
Control	223	20.3	1.4	17.8	23.1
Experimental	267	24.3	1.5	21.5	27.4

Note. The mean (*M*), standard error of the mean (*SE*) and 95% Confidence Intervals of the mean (*95CI-* and *95CI+*) were calculated by Estimated Marginal Means (EMMEANs) using GEE based on UOI recalled from witness participants and two covariates (for Artistic Ability and Observant Behavior). See text for more details. [1] $p < .05$.

In Table 2, UOI has been organized by facial area and *Interview Type*. For example, the category for *Eyes* include recall for eye colour, eye shape, and information about crow's feet and eye bags. *Face shape* includes recall relating to the shape of the face, chin, jaw and forehead. For *General*, we have included basic observations such as age, race, gender, character and skin. For convenience, we have grouped areas into *Internal Features*, *External Features*, and *Other*.

Table 2. Units of Information recalled by area of the face and Interview Type.

Interview Type	Internal Features					External Features			Other	
	Brows	Eyes	Nose	Mouth	Facial hair	Hair	Face shape	Ears	General	Neck, clothing
Conrtol	12	26	15	24	7	41	24	8	38	28
Experimental	15	35	27	27	14	38	33	13	32	33
Difference	3	9	12	3	7	-3	9	5	-6	5

Note. The row called Difference is UOI for Experimental – UOI for Control: positive values indicate more information recalled after witnesses sketched the face (Experimental) than not (Control).

As can be seen, witnesses who sketched the face recalled more information consistently for areas in the internal features, and, for external features, for face shape and ears but not hair. These data indicate that recall is generally more consistent following witness sketches for the internal features, the area (in particular the eyes) that is important for recognition of an ensuing composite face (e.g., Ellis et al., 1979; Fodarella et al., 2017; Frowd et al, 2011, 2012b, 2013).

The UOI recalled by each witness participants was analyzed using Generalized Estimating Equations (for a review of GEE, see Everitt & Howell, 2005). There was a single categorical IV, *Interview Type* (coding *0* = Normal and *1* = Self Sketch). To model the data appropriately, a Poisson probability distribution was selected with a Log link function. Two sources of random error were included, the effect of witness participants (coded as 1 to 22), a subject effect; and the stimulus identities, or *items* (coded from 1 to 11), a within-subject effect. Given that items were repeated across *Interview Type*, an exchangeable structure was selected for the Working Correlation Matrix.

A GEE model was constructed that contained the categorical IV *Interview Type* (coded as above) and two covariates derived from witness participants, each one a continuous variable, and adjusted to their overall mean (see *M* values below) by *Interview Type*. The first covariate was *Artistic Ability*. This variable had good range across the scale, from 6 to 23 (possible range from 0 to 24) (M = 17.2, SD = 4.6). The second was *Observant Behavior*, with a fairly good range, from 14 to 30 (possible range from 0 to 32) (M = 21.0, SD = 4.1). A point-serial (Pearson) correlation was conducted between these two covariates, an exercise that revealed a medium-sized correlation [$r(20)$ = .29, p = .19), indicating that there should not be an issue with collinearity in the model (e.g., Reed & Wu, 2013). The same as for all analyses presented in this paper, emerging model parameters were checked to be within sensible bounds, values that were neither too low nor too high, a situation that would otherwise indicate a problem with model stability.

The model (Table 3) revealed that the odds of face recall were higher for the experimental group (cf. control) with a small effect [B = 0.18, $SE(B)$ = 0.09, $Exp(B)$ = 1.2, *95%CI* (1.00, 1.43)]; it also revealed that neither covariate impacted face recall [$Exp(B) \leq 1.01$]. As a check, since the presence of more than one covariate can impact on statistical power[1], covariates were introduced separately. This exercise did not change the conclusion drawn for the effect of *Interview Type* (as $p \leq .047$), but one model yielded an increasing odds that was marginal for *Artistic Ability* [B = 0.02, $SE(B)$ = 0.01, p = .07, $Exp(B)$ = 1.02, *95%CI* (0.998, 1.04)], while the other had a clear effect for *Observant Behavior* [B = 0.02, $SE(B)$ = 0.01, p = .045, $Exp(B)$ = 1.02, *95%CI* (1.00, 1.04)]. These covariates represent a small positive increase in face recall, with *Observant Behavior* being relatively stronger.

Table 3. GEE for Face Recall (UOI).

[1] Here, $SE(B)$ values for these two covariates increased when both were included in the model (cf. when each covariate was contained in a separate model).

	$X^2(1)$	*p*
Interview Type	3.97	.046
Artistic Ability	1.74	.19
Observant Behaviour	2.41	.12

Note. GEE were conducted using the GENLIN function in SPSS version 29. Parameter settings not mentioned in the text remained at their default values. Other details of the model were the Intercept [*B* = 2.76, *SE(B)* = 0.20] and Goodness of Fit (*QIC* = 50.3, *QICC* = 45.5).

Face Construction Accuracy

Two measures were used to determine the accuracy of the composites constructed by witness participants. The first was an analysis of correct names given to composites, and the second, an analysis of incorrect (mistaken) names.

A. Correct Composite Naming

The most forensically relevant measure is the extent to which composites are named correctly. This outcome is particularly valuable, giving criminal investigations accurate identification for composites that have been constructed from eyewitnesses. In this section, responses from namer participants were scored for accuracy, with a numeric value of 1 assigned when either the intended identity's name or a sufficiently detailed, accurate description of the identity was provided by the namer. In all other cases, a value of 0 was assigned. For a relatively small number of cases (*N* = 95, 23.8% overall), the target photograph (presented to namer participants after the composites had been seen) was not named correctly. In these cases, since the associated composite could also not have been named correctly, composite responses were removed.

Analysis of the resulting data revealed that the majority of namer participants correctly named at least one of the composites, with nearly half (*M* = 42.5%) correctly named one composite, and four being the maximum (*M* = 10.8%). By group, seven of the 11 composites were correctly recognized by at least one person (*M* = 63.6%) in the Control group, but this increased to 10 identities (*M* = 90.9%) in the Experimental group. As shown in Table 4, composites created after a witness sketched the face received many more correct responses both in total as well as on average per witness participant (*M*) relative to witness participants who did not sketch the face.

Table 4. Correct Responses to Composites for each Interview Type.

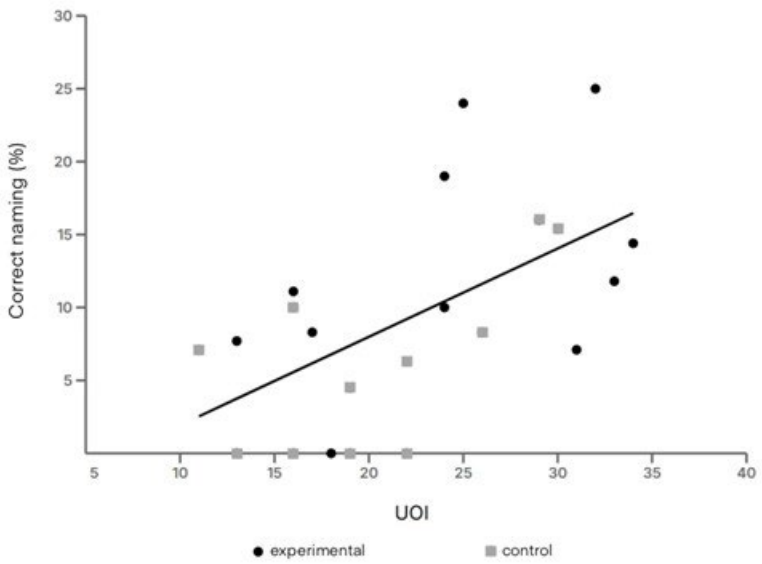
Interview Type[1]	Correct responses to composites				
	Total	*M*	*SE*	*95CI-*	*95CI+*
Control	11	4.8	1.5	2.5	8.8
Experimental	26	11.2	2.5	7.2	17.1

Note. For a definition of these variables, see Table 1, Note. Values presented in the table were calculated by EMMEANs using GLMM based on correct responses from namer participants (N = 191 responses at each level of Interview Type) plus three covariates included (for Face Recall (UOI), Artistic Ability and Observant Behavior). See text for details. [1] *$p < .05$.*

For these data, as there were now multiple responses (i.e., resulting from the set of composites presented to each namer participant) compared to single (total) responses from the previous analysis, the regression technique Generalized Linear Mixed Models was used (e.g. analyses, see Erickson et al., 2022; Frowd et al., in press). GLMM have the advantage of creating a single model that takes into account the two sources of random error, as before, along with the potential effects of the namer participants (coded as 1 – 40) and the composite *set* presented to namer participants (coded as 1 – 2). Responses were scored (described above) to create nominal-level data, and a Binomial probability distribution was selected with a Logistic link function. Also, as the three random sources of error de-correlate participant responses in the analysis, an independent structure was used for the Working Correlation Matrix (by selecting "variance components").

The model included the same two covariates as before, but also total face recall (UOI), as this variable should be positively related to correct naming; as before, values by *Interview Type* were adjusted to the overall mean of this variable. Figure 2 illustrates the anticipated positive relationship between the two main DVs, UOI and correct naming. Prior to model construction, point-serial correlations were conducted between these three covariates; again, all correlations were medium sized [$r(20)$ = .21-.31, ps ≥ .16], suggesting that collinearity should not be an issue.

Figure 2. UOI versus correct naming, coded by interview type, with best fit line



So GLMM contained *Interview Type*, *Artistic Ability*, *Observant Behavior* and *Face Recall (UOI*). It was first assessed for composition of random effects, which were planned to be 'maximal' according to the design (Barr et al., 2013). This meant that the model could include (i) random intercepts for each source of variance: witness participants, namer participants, items and presentation set (to namer participants), and, (ii) random slopes (to cater for repeated responses) for namer participants and items. When built, it was apparent that the random effects structure could be simplified, in this case to contain random slopes for namer participants only (σ^2 = 0.59, *SE* = 0.45); that is, the impact of all other random effects was negligible (σ^2 < 0.001).

The resulting model (Table 5) indicated that *Interview Type*, *Artistic Ability* and *Observant Behavior* influenced Correct Naming. The resulting parameter estimates (Table 5) indicated that all of these variables increased the odds of a correct response to a composite (as *B values* are positive). That is, that there was greater odds of a correct response from a composite created after witness participants in the experimental group sketched the face (cf. those in the control group)[2]. Also, ratings for *Artistic Ability* and *Observant Behavior* from witness participants both predicted an increase in odds of a correct response. However, while *Face Recall* trended in the anticipated

[2] As a further check for model stability, we conducted GLMM without any covariates. The same conclusion was reached for the effect of *Interview Type* on correct naming [*p* = .030, *B* = 0.90, *SE(B)* = 0.41, *Exp(B)* = 2.47].

direction (i.e., the value of *B* was positive), it did not influence the odds of a correct response from a witness participant's composite.

Table 5. GLMM for Correct Naming.

	DF	*F*	*p*
Corrected Model	4,377	8.04	< .001
Interview Type	1,377	4.23	.040
Artistic Ability	1,377	7.34	.007
Observant Behaviour	1,377	3.95	.048
Face Recall (UOI)	1,377	0.23	.629

Note. GLMM were conducted using the GENLINMIXED function in SPSS version 29. The same as before, parameter settings not mentioned remained at default values. Other details of the model were Information Criteria (*AICC* = 2049.7 and *BIC* = 2053.6), Coefficients of Determination (*Pseudo* R^2: *Marginal* = .06 and *Conditional* = .10), Intraclass Correlation (Overall ICCs: *Adjusted* = .05 and *Conditional* = .05) and Overall Correct Classification (90.3%).

Table 6. Parameter Estimates of the GLMM for Correct Naming.

	B	*SE(B)*	*t(377)*	*p*	*Exp(B)*	*95CI-*	*95CI+*
Fixed Effects							
Intercept	-2.97	0.33	-8.97	< .001	0.05	0.03	0.10
Interview Type	0.92	0.45	2.06	.040	2.52	1.04	6.08
Artistic Ability	0.08	0.04	1.99	.048	1.09	1.04	6.08
Observant Behaviour	0.11	0.04	2.71	.007	1.12	1.03	1.21
Face Recall (UOI)	0.01	0.03	0.48	.63	1.01	1.03	1.21

B. Mistaken Composite Naming

While a correct name given for a composite is valuable information to a criminal investigation by providing a lead that is accurate, sometimes a composite is named as another person—that is, the identity is incorrect. In a criminal investigation, mistaken names provide a mechanism that allows law enforcement to eliminate a person from an enquiry, ideally reducing the number of potential

suspects that could have committed the offence. Theoretically, a composite that gives rise to frequent mistaken names suggests that the constructed face is less accurate, by virtue of looking like another, albeit non-intended, person. Such an outcome may occur, for example, if details portrayed in the face are minimal or the naming procedure is demanding (e.g., Frowd et al., 2015; Portch et al., in press).

To assess composite effectiveness using this supplementary measure, responses from namer participants were re-scored: an incorrect identity for a composite was given a value of 1, and 0 otherwise. As before, responses to composites were removed from the analysis when a namer participant could not correctly name a target photograph, but we also removed correct responses (*N* = 37). This additional step is necessary when assessing this metric, otherwise group means for mistaken naming will tend to reduce as correct responses increase (and vice versa).

As can be seen in Table 7, the number of mistaken names was much higher overall and per participant (*M*) than for correct names, but this is a usual outcome, as sketches often contain less detail (cf. composites from computerized systems) and so tend to match more identities (e.g., Portch et al., in press). However, differences by *Interview Type* were minimal. Accordingly, when conducted in the same way as described above, GLMM revealed that the odds of a mistaken response was unaffected either by *Interview Type* [$X^2(1,225)$ = 0.09, p = .76, *Exp(B)* = 1.10] or by any of the three covariates [ps ≥ .42, *Exp(|B|)* ≤ 1.02-1.03] [3].

Table 7. Mistaken Responses to Composites for each Interview Type.

Interview Type	Correct responses to composites				
	Total	*M*	*SE*	*95CI-*	*95CI+*
Normal	61	51.9	6.3	39.7	63.9
Self Sketch	59	54.2	6.4	41.6	66.3

Note. See Table 4, Note, and *text* for more details. Other model details were Random Effects (random intercepts for both namer participants σ^2 = 0.36 and set σ^2 = 0.01, and random slopes for *Interview Type* for both namer participants σ^2 = 0.06 and items σ^2 = 0.04), Information Criteria (*AICC* = 1014.8 and *BIC* = 1028.3), Coefficients of Determination (*Pseudo* R^2: *Marginal* = .01 and *Conditional* = .11), Intraclass Correlation (Overall ICCs: *Adjusted* = .11 and *Conditional* = .10) and Overall Correct Classification (71.3%).

[3] To save space for presenting results for this supplementary measure, and as effects are very small, we summarise results from the GLMM concisely.

DISCUSSION

This research investigated a novel addition to the cognitive interview for witnesses creating forensic composites. Research on improving the cognitive interview, as well as adjustments to the process for creating facial composites, has been ongoing for decades. However, with the exception of Dando et al. (2009, 2020), researchers have yet to investigate the full potential of a witness's own drawing, here for face recall and construction. We found that asking witnesses to draw their own depiction of the target face before the cognitive interview and creation of the composite with the forensic practitioner had a positive impact on facial recall and composite identifiability.

The difference in identification rates and number of units of information between the experimental and control groups was statistically significant with a positive effect, suggesting that utilizing the witness artistic rendition method helps moderately improve both face recall (a *small* effect) and correct naming of a composite (a *medium* effect). Observant Behavior and Artistic Ability also had an independent, positive effect on correct naming of a composite, implying that more artistically-minded and more observant individuals produce better composites, independent of the novel method described in this study. In addition, total face recall and correct naming were positively correlated, but not statistically, suggesting that recollection of the face may lead to a more identifiable composite, but this effect is not consistent. However, the experimental interview type had a significant medium effect for both UOI and correct naming, and so the witness artistic rendition method influences both of these factors, if not directly with each other. More specifically, UOI was consistently higher for the internal features of the face (cf. external features) following witness sketching. Internal features are relatively more important than external features for familiar-face recognition (e.g., Ellis et al., 1979), and thus improved recognition of the composite is a result of improved memory for, and thereby improved construction of, this central facial area. Thus, these data provide promising evidence for the potential of the technique to improve the composite creation and identification process. The findings bridge psychological research on the benefits of drawing as a memory aid with longstanding efforts aimed at improving the efficacy of forensic composite creation.

As mentioned in the Introduction, sketch is only one of the ways facial composites are produced in modern forensics. While the project suggests that the new technique should improve the effectiveness of sketch composites created with the assistance of an artist, composite systems in general are likely to benefit from improvement in face recall (e.g., Frowd, 2021). To confirm this expectation, further research could usefully explore the potential positive impact of

the witness artistic rendition technique on composites produced using different systems (e.g., E-Fit 6, FACES and EvoFIT). Other studies could also investigate other witness demographics, as this study was conducted on a UK university campus, with the majority of constructor participants being predominantly White (with 82% identifying as White European or White Middle Eastern) and more highly educated than the average population. However, there appears to be no real reason why the sketching technique should not benefit witness recall (and thereby the ensuing composite) in general.

This study also developed two simple scales to assess a witness's artistic ability (AA) and observant behaviour (OB). Both of these self-reported measures are straightforward to administer and provide estimates of face recall and composite effectiveness, but again it would be valuable to assess their generalizability in future research. Another potential limitation pertains to the selection of target images used. While celebrity images are unlikely to be constructed or named in a criminal investigation, there is good evidence that this type of image leads to similar performance to non-celebrity images (e.g., Frowd et al., 2015). That said, assessing generalizability is always good practice and so any further research in the area should consider use of non-celebrity images.

This research was conducted with a retention period of 3 to 4 hours, both for the convenience of witness participants and to mimic the retention period in cases where face construction takes place on the same day as the crime, which has been seen to occur in several police forces about 10% of the time (Portch et al., in press). At other times, police investigations experience delays of one or two days (or longer) between an event and face construction (Frowd et al., 2015). Given that face recall decreases with extended retention periods (e.g., Ellis et al., 1980; Portch et al., in press), limiting the effectiveness of facial composites, the novel sketching technique is likely to be even more valuable following longer delays to promote an identifiable composite; indeed, the impact of the technique with increasing delay would be a worthwhile focus of future research. More generally, continued investigation of this technique will augment our understanding of its effectiveness for different demographics and forensic applications.

In conclusion, this research aimed to investigate the effect on witness memory and composite identifiability as a result of witness artistic rendition (i.e. asking witnesses to create their own drawn image of the remembered face before a cognitive interview with a forensic artist). Twenty-two participants witnessed a target face and created a composite in collaboration with a forensic artist, half of them creating their own image first. Improved witness memory was quantified by the number of units of information produced during free recall and composite

identification. Results reveal that both total face recall and correct naming were improved for the experimental composites than for the control composites. If used forensically, the research suggests that the technique could usefully improve correct identification of criminal suspects.

IMPLICATIONS FOR PRACTICE

- Both witness recall and composite identification rates were positively benefitted by asking witnesses to draw their own image of the remembered face prior to the normal cognitive interview procedure
- Witnesses who self-reported to be more artistic had a tendency to remember more facial details and created more identifiable composites
- Witnesses who self-reported to be overall more observant in the world and had a better memory remembered more facial details and created more identifiable composites
- The witness artistic rendition technique is quick, simple, low cost and requires no additional training for forensic practitioners to perform
- If implemented by police, the technique could improve composite identification rates with little impact upon budgets and time constraints

REFERENCES

Brace, N., Pike, G,. Kemp, R., Turner, J., & Bennett, P. (2006). Does the presentation of multiple facial composites improve suspect identification? *Applied cognitive psychology 20*(2), 213–226.

Brown, C., Frowd, C., & Portch, E. (2017). Tell me again about the face: Using repeated interviewing techniques to improve feature-based facial composite technologies. https://doi.org/10.1109/EST.2017.8090396

Bruce, V., Ness, H., Hancock, P. J. B., Newman, C., & Rarity, J. (2002). Four Heads Are Better Than One: Combining Face Composites Yields Improvements in Face Likeness. *Journal of Applied Psychology, 87*(5), 894–902. https://doi.org/10.1037/0021-9010.87.5.894

Dando, C., Wilcock, R., & Milne, R. (2009). The Cognitive Interview: The Efficacy of a Modified Mental Reinstatement of Context Procedure for Frontline Police Investigators. *Appl. Cognit. Psychol. 23*, 138–147.

Dando, C., Gabbert, F., & Hope, L. (2020). Supporting older eyewitnesses' episodic memory: the self-administered interview and sketch reinstatement of context. *Memory*. *28*(6), 712-723. https://doi.org/10.1080/09658211.2020.1757718

Davies, G., van der Willik, P., & Morrison, L. J. (2000). Facial Composite Production: A Comparison of Mechanical and Computer-Driven Systems. *Journal of Applied Psychology, 85*(1), 119–124. https://doi.org/10.1037/0021-9010.85.1.119

Davis, J. P., Simmons, S., Sulley, L., Solomon, C., & Gibson, S. (2015). An evaluation of post-production facial composite enhancement techniques. *Journal of Forensic Practice, 17*(4), 307–318. https://doi.org/10.1108/JFP-08-2015-0042

Ellis, H.D., Davies, G.M., & Shepherd, J.W. (1978). A critical examination of the photofit system for recalling faces. *Ergonomics, 21*, 297–307.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1980). The deterioration of verbal descriptions of faces over different delay intervals. *Journal of Police Science and Administration*, *8*, 101-106.

Erickson, W. B., Brown, C., Portch, E., Lampinen, J. M., Marsh, J. E., Fodarella, C., Petkovic, A., Coultas, C., Newby, A., Date, L., Hancock, P. J. B., & Frowd, C. D. (2024). The impact of

weapons and unusual objects on the construction of facial composites. *Psychology, Crime & Law, 30*(3), 207–228.

Everitt, B. S., & Howell, D. (2005). *Encyclopedia of Statistics in Behavioral Science*. John Wiley & Sons, Ltd. ISBN: 0-470-86080-4.

Federal Bureau of Investigation, a.k.a. FBI. Facial composite catalog.

Fernandes, M., Wammes, J., & Meade, M. (2018). The surprisingly powerful influence of drawing on memory. *Current Directions in Psychological Science, 27*(5), 302-8.

Fodarella, C., Kuivaniemi-Smith, H. J., Gawrylowicz, J., & Frowd, C. D. (2015). Forensic procedures for facial-composite construction. *Journal of Forensic Practice, 17*(4), 259-270.

Fodarella, C., Marsh, J. E., Chu, S., Athwal-Kooner, P., Jones, H. S., Skelton, F. C., Wood, E., Jackson, E., & Frowd, C. D. (2021). The importance of detailed context reinstatement for the production of identifiable composite faces from memory. *Visual Cognition*, *29*(3), 180–200. https://doi.org/10.1080/13506285.2021.1890292

Fodarella, C., Frowd, C. D., Warwick, K., Hepton, G., Stone, K., Date, L., & Heard, P. (2017). Adjusting the focus of attention: helping witnesses to evolve a more identifiable composite. *Forensic Research & Criminology International*, 5(1), DOI: 10.15406/frcij.2017.05.00143.

Frowd, C. D., Carson, D., Ness, H., Richardson, J., Morrison, L., Mclanaghan, S., & Hancock, P. (2005). A forensically valid comparison of facial composite systems. *Psychology, Crime & Law, 11*(1), 33–52. https://doi.org/10.1080/10683160310001634313

Frowd, C. D., Erickson, W. B., Lampinen, J. M., Skelton, F. C., McIntyre, A. H., & Hancock, P. J. B. (2015). A decade of evolving composites: regression- and meta-analysis. *Journal of Forensic Practice*, 17(4), 319–334. https://doi.org/10.1108/JFP-08-2014-0025

Frowd, C. D., Hancock, P. J. B., Bruce, V., Skelton, F. C., Atherton, C. J., Nelsz, L., McIntyre, A. H., Pitchford, M., Atkins, R., & Webster, D. (2011). Catching more offenders with EvoFIT facial composites: lab research and police field trials. *Global Journal of Human Social Science, 11*, 46-58.

Frowd, C. D., Hancock, P. J. B., & Carson, D. (2004). EvoFIT: A holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on Applied Perception (TAP), 1*(1), 19–39.

Frowd, C. D., Nelson, L., Skelton, F. C., Noyce, R., Atkins, R., Heard, P., Morgan, D., Fields, S., Henry, J., & McIntyre, A. (2012a). Interviewing techniques for darwinian facial-composite systems. *Applied Cognitive Psychology, 26*(4), 576–584

Frowd, C. D., Jones, S., Fodarella, C., Skelton, F., Fields, S., Williams, A., Marsh, J. E., Thorley, R., Nelson, L., Greenwood, L., Date, L., Kearley, K., McIntyre, A. H., & Hancock, P. J. B. (2014b). Configural and featural information in facial-composite images. *Science & Justice, 54*(3), 215–227. https://doi.org/10.1016/j.scijus.2013.11.001

Frowd, C. D., Portch, E., Estudillo, A. J., Ford, C. J., Purcell, A., & Brown, C. (in press). The value of whole-face procedures for the construction and naming of identifiable likenesses for recall-based methods of facial-composite construction. *Applied Cognitive Psychology*.

Frowd, C. D., Skelton, F. C., Atherton, C., Pitchford, M., Hepton, G., Holden, L., McIntyre, A. H., & Hancock, P. J. B. (2012b). Recovering faces from memory: The distracting influence of external facial features. *Journal of Experimental Psychology: Applied, 18*(2), 224-238.

Frowd, C. D., Skelton, F. C., Butt, N., Hassan, A., & Fields, S. (2011). Familiarity effects in the construction of facial-composite images using modern software systems. *Ergonomics*, *54*, 1147-1158.

Frowd, C. D., Skelton, F., Hepton, G., Holden, L., Minahil, S., Pitchford, M., McIntyre, A., Brown, C., & Hancock, P. J. B. (2013). Whole-face procedures for recovering facial images from memory. *Science & Justice*, *53*(2), 89–97. https://doi.org/10.1016/j.scijus.2012.12.004

Frowd, C. D., White, D., I. Kemp, R., Jenkins, R., Nawaz, K., & Herold, K. (2014a). Constructing faces from memory: the impact of image likeness and prototypical representations. *Journal of Forensic Practice, 16(*4), 243–256. https://doi.org/10.1108/JFP-08-2013-0042

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

Koehn, C. E., & Fisher R. P. (1997). Constructing facial composites with the Mac-a-Mug Pro system. *Psychology, Crime & Law, 3*, 215-224

McDonald, H. (1960). The identi-kit manual. Townsend Company (Identi-Kit Division), Santa-Ana, California

McIntyre, A. H., Hancock, P. J. B., Frowd, C. D., & Langton, S. R. H. (2016). Holistic Face Processing Can Inhibit Recognition of Forensic Facial Composites. *Law and Human Behavior, 40*(2), 128–135. https://doi.org/10.1037/lhb0000160

Memon, A. and Bull, R. (1991), The cognitive interview: Its origins, empirical support, evaluation and practical implications. *J. Community. Appl. Soc. Psychol., 1*, 291-307. https://doi.org/10.1002/casp.2450010405

Ness, H., Hancock, P. J. B., Bowie, L., Bruce, V., & Pike, G. (2015). Are two views better than one? Investigating three-quarter view facial composites. *Journal of Forensic Practice, 17*(4), 291–306. https://doi.org/10.1108/JFP-10-2014-0040

Penry, J. (1971). Photofit kit. John Waddington of Kirkstall Ltd, Leeds

Portch, E., Brown, C., … Frowd, C. D. (in press). The impact of forensic delay: facilitating facial composite construction using an early-recall retrieval technique. *Ergononics*.

Reed, P. & Wu, Y. (2013). Logistic regression for risk factor modelling in stuttering research. *Journal of Fluency Disorders, 38*, 88–101.

Skelton, F. C., Frowd, C. D., & Speers, K. E. (2015). The benefit of context for facial-composite construction. *Journal of Forensic Practice, 17*(4), 281–290. https://doi.org/10.1108/JFP-08-2014-0022

Steinberg, S. (2006). *Steinberg's facial identification catalogue*. Ninth Printing, Miami, FL.

VisionMetric (2025). E-fit 6. https://visionmetric.com/efit6/

Wammes, J., Jonker, T., & Fernandes, M. (2019). Drawing improves memory: The importance of multimodal encoding context. *Cognition, 191*, 103955.

Wammes, J., Meade, M., & Fernandes, M. (2016). The drawing effect: Evidence for reliable and robust memory benefits in free recall. *Quarterly Journal of Experimental Psychology, 69*(9), 1752-76.

Wells, G.L. (1985). Verbal descriptions of faces from memory: are they diagnostic of identification accuracy? *Journal of Applied Psychology*, 70, 619-626.

Wilkinson, C. (2015). A review of forensic art. *Research and Reports in Forensic Medical Science, 5*, 17-24. https://doi.org/10.2147/RRFMS.S60767

TABLES:

1. Units of Information recalled for the target face from witness participants by Interview Type.
2. Units of Information recalled by area of the face and Interview Type.
3. GEE for Face Recall (UOI).
4. Correct Responses to Composites for each Interview Type.
5. GLMM for correct naming.
6. Parameter Estimates of the GLMM for Correct Naming.
7. Mistaken Responses to Composites for each Interview Type.

FIGURES:

1. Example images created in Stage 1 for Tom Bergeron
2. UOI versus correct naming, coded by interview type, with best fit line

APPENDIX 1: Supplementary Measures of Participant Witness Ability

For each measure, witness participants were asked to rate their responses to the following eight questions, from 0 (nothing like me) to 4 (exactly like me). Responses were summed to calculate a total score for Artistic Ability (AA) and Observant Behavior (OB). Total possible for AA was 24, and for OB, 32.

AA:

I am more creative and free-thinking than I am practical and logical

I like to draw

I am a visual learner

I have had a lot of experience with art

I take a lot of photos

I can visualize things easily and completely in my mind's eye

OB:

I find it easy to remember birthdays

I am an observant person

I pay a lot of attention to small details

I have a good memory

I like to people watch

I don't find it difficult to make eye contact

I like meeting new people

I never have trouble recognizing people after a haircut or other dramatic change

APPENDIX 2: Estimation of Sample Size

Computer simulation was used to assess the number of participants required for face construction (witness participants) and composite naming (namer participants). This exercise considered a medium effect [*Exp(B)* ≈ 2.5] for the between-subjects IV, *Interview Type* (Normal vs. Self Sketch), an advantage that (if found) should have practical importance in the real world. We based the estimated sample sizes on existing research that suggested that, to achieve good power, a minimum of 10 participants were necessary per group for both face construction and composite naming. The approach generates a series of sets of data (10 in this case) for the IV, simulating the DV based on the estimated number of witness and namer participants. Then, these data are analysed as conducted in the paper, using GLMM for correct responses from namer participants. The fraction of cases that the IV is maintained in the model (given the usual value for alpha of .1) provides an estimate of statistical power.

Referring to Equation 1, composite naming in the control condition (Y_{ij}) was set to an average of 18%, typical for a feature system (Frowd et al., 2005), resulting in B_0 = -1.52, performance that was anticipated to increase with a medium effect, to 32%, B_1 = -0.75. Both of these *B* values were drawn from a random Normal distribution with *SD* = 0.1, to provide good variability. Residual error (e_{ij}) was added to each response from a namer participant, again drawn

from a random Normal distribution (M = 0.0), with SD = 0.5 to give suitably variable responses (e.g., *MD* changed at baseline from -10% to +20%). Finally, we modelled the situation that namer participants are not always familiar with (and so do not correctly name) all of the relevant target identities, the result of which typically reduces statistical power (as there are fewer responses available to analyse). Here, we assumed that 1 in 20 responses, selected at random, would be unfamiliar. These cases, as conducted in the paper, were removed prior to analysis.

Based on these data, GLMM revealed that *Interview Type* would be retained in the model (given α = .1) for 90% of simulations. This result indicates that there is good power (i.e. $1 - B \geq .8$) for the proposed design, sample size and method of analysis.

Equation 1

Model for the predictor *Interview Type* in the linear Regression Equation:

$y_{ij} = B_0 + (x_1 * B_1) + e_{ij}$

Where Predictor x_1 = *Interview Type* (*0* = Normal, *1* = Self Sketch). B_0 is the model's intercept. The value for B_1 was modelled with positive values (to give an increase in y_{ij}). The term e_{ij} represents residual error. For GLMM, the equation is subject to the Sigmoidal (logistic) function, $Y_{ij} = Exp (y_{ij} / (1 + Exp (y_{ij})))$, to give nominal responses.