Applied Cognitive Psychology

WILEY

RESEARCH ARTICLE OPEN ACCESS

The Value of Whole-Face Procedures for the Construction and Naming of Identifiable Likenesses for Recall-Based Methods of Facial-Composite Construction

Charlie D. Frowd[1] | Emma Portch[2] | Alejandro J. Estudillo[2] | Claire J. Ford[2] | Amy Purcell[2] | Melanie Pitchford[3] | Charity Brown[4]

[1]School of Psychology and Humanities, University of Lancashire, Preston, UK | [2]Department of Psychology, Bournemouth University, Bournemouth, UK | [3]School of Psychology, University of Bedfordshire, Luton, UK | [4]School of Psychology, University of Leeds, Leeds, UK

Correspondence: Charlie D. Frowd (cfrowd1@uclan.ac.uk)

Received: 21 February 2024 | **Revised:** 13 October 2024 | **Accepted:** 19 November 2024

Funding: The UK Economic and Social Research Council provided support for this work [ES/I002022/1].

Keywords: facial composite | perceptual stretch | retention interval | trait attribution

ABSTRACT

Traditional methods of facial-composite construction rely on an eyewitness recalling features of an offender's face. We assess the value of the addition of a trait–recall mnemonic to a cognitive-type interview, and perceptually stretching presented composites, to aid image recognition. Participant-constructors intentionally or incidentally encoded a target face, were interviewed about its facial features 3–4 h or 2 days later, made a series of trait attributions (or not) about the face and constructed a feature-based composite. Regardless of encoding manipulation, faces constructed after 3–4 h were twice as likely to be correctly named (cf. after 2 days) both when the trait–recall mnemonic was applied and composites were viewed stretched. Thus, the research indicates that benefit should be afforded when trait–recall mnemonics are employed for feature composites constructed on the same day as the crime and when composites are presented to potential recognisers with instruction to view the face as a perceptual stretch.

1 | Introduction

A facial composite is a visual representation of a face, usually constructed of an offender, by a witness to or victim of crime. Typically, an eyewitness will construct a composite of an unfamiliar offender, a person that a witness has seen just once, at the time of the crime. Traditionally, composite systems have relied predominantly upon a witness recalling and describing featural details of the face (e.g., eyes, nose, mouth, hair). The witness's facial description is used to select a sub-set of relevant visual feature exemplars from a large photographic database. While by design, feature systems (e.g., E-FIT, PRO-fit, FACES and Identikit 2000) necessitate a witness recalling individual features of the face, it is generally agreed that directing attention to global/holistic information (e.g., the spatial relationships between features) is best suited for facilitating the process of face recognition (see Peterson and Rhodes 2003, for a review). Feature composite systems incorporate this concept to some extent, as a witness first selects an appropriate face shape from the system and works within this whole-face context to view, exchange and edit individual facial features (e.g., Frowd, Carson, Ness, McQuiston, et al. 2005; Portch et al. 2017; Skelton et al. 2015). Nevertheless, a general emphasis on feature recall has the likely knock-on effect that a witness is less able to recognise when a composite-under-construction best resembles his or her memory of the previously-seen face (e.g., Brown et al. 2020; Frowd and Fields 2011).

Abbreviations: CI, Cognitive Interview; CoI, Composite Interview (pre-construction interview involving witnesses recalling features of the face); GEEs, Generalised Estimating Equations; H-CI, Holistic Cognitive interview; H-CoI, Holistic Composite interview (as CoI, but then asking witnesses to focus on the character of the face).

Indeed, recognition rates are significantly higher when composites are instead constructed using contemporary holistic systems. These newer techniques harness a whole-face (cf. feature-based) focus to better accommodate recognition processes (e.g., see meta-analysis by Frowd et al. 2015). Here, witnesses repeatedly select whole faces (or whole-face regions) from face arrays based on their resemblance to the offender (e.g., E-FIT-V, EvoFIT and ID). As such, holistic composites tend to hold a global recognition advantage. In contrast, feature systems usually require attention to fine-level feature details; indeed, typically well-recognised feature composites suffer recognition decrements following application of low levels of Gaussian blurring, a procedure that obscures this fine-level detail (Frowd et al. 2014). Thus, identifying factors within forensic settings that affect the encoding and retention of information about facial features may help our understanding of when such composites are likely to be more or less effective. Such information also has practical significance since feature composites are still constructed by eyewitnesses in the UK, Europe, the USA and Australia (e.g., Tredoux et al. 2023).

Retrieval of both verbal and visual facial information is clearly important for composite construction; however, these types of information may decay at different rates. Decreasing access to visual information may specifically hamper face recognition processes, with recognition accuracy diminishing as the delay increases between encoding and a subsequent recognition attempt (e.g., Deffenbacher et al. 2008; Shapiro and Penrod 1986). However, forgetting demonstrably follows a negative exponential-type decay curve (e.g., Deffenbacher et al. 2008; Ebbinghaus 2013); face recognition memory rapidly reaches an asymptote, with little further decline occurring between 48-h and 1 month after encoding (e.g., Chance et al. 1975; Laughery et al. 1974; Shepherd and Ellis 1973).

In contrast, access to verbal information (i.e., the typically feature-based information recalled for a face) is less enduring. Participants recall few facial descriptors even when retrieval is invited immediately after encoding under optimal conditions (e.g., an average of 4.46 descriptive items about a target person, Sporer 2007). Further decreases occur with increasing retention interval, with significantly fewer accurate face descriptors reported at 1 day compared to 1 h, and at 1 week compared to 1 day and 1 h (Ellis et al. 1980). As it is current practice for witnesses to recall information about the previously seen face—in particular, to allow example features to be located within a feature-type composite system—fewer details recalled may thus contribute to a decline in composite effectiveness over time. Indeed, research has shown correct naming to substantially drop for feature composites constructed following a 2-day delay (a mean of ~5% or less) compared to delays of up to a few hours (~20%; e.g., see Frowd et al. 2015; Portch et al. under revision).

Self-reports by laboratory participants further indicate that less attention is paid to individual facial features of an offender (cf., holistic information) if encoding of the face is incidental (Olsson and Juslin 1999). Overall, face recognition tends to be less successful when participants are unaware of an impending memory test, and attention is diffused across the scene (compared to intentional focus on a target; see meta-analysis by Shapiro and Penrod 1986). While it is common practice within laboratory research to model instances where a witness is aware that a crime is taking place (i.e., via intentional encoding), these manipulations fail to capture some real-world circumstances (i.e., distraction burglary or fraud), where incidental encoding is sometimes involved. When a target face is encoded incidentally (cf., intentionally) memory strength for featural information about a target face is likely to be weaker. We may thus surmise that attempts to induce incidental encoding conditions will also lead to the construction of less identifiable modern feature composites. This is because feature-based construction is facilitated at encoding when witnesses focus on individual facial features of a target identity rather than when they adopt a comparatively global focus, through attribution of character (Frowd, Bruce, Ness, et al. 2007; Wells and Hryciw 1984).

However, laboratory-based findings reveal that composite effectiveness can actually be facilitated when face constructors make such global judgements (e.g., the degree to which an individual might be regarded as honest and intelligent) after they have freely recalled the features of the face (e.g., Frowd et al. 2008, 2012, 2015). This approach has proved advantageous, with a meta-analysis of results from seven experiments showing that composites from feature, sketch and EvoFIT systems are 2.5 times (95% CI [2.1, 3.1]) more likely overall to be correctly named compared to only using a face-recall procedure (Frowd et al. 2015). Further research confirms that the advantages associated with this technique generalise across systems, apparent when (i) holistic systems are used for face construction and a nominal 24-h post-encoding delay is imposed (Frowd et al. 2013; Skelton et al. 2020) and (ii) feature systems are used and construction is conducted both after relatively short post-encoding delays, of 3–4 h (Frowd et al. 2008), and longer delays of 24 h, when memory strength for the face is arguably diminished (Skelton et al. 2020).

The mechanism for this facilitation is proposed to occur (e.g., Skelton et al. 2020) as recalling a face activates memory traces for facial features, while character attribution organises these memories into a format that is congruent with processing for the subsequent composite task—specifically, recognition of individual facial features or faces from face arrays, operations that rely predominantly on holistic processing. Pre-construction interviewing procedures must thus be carefully negotiated (for discussion of these procedures, see Frowd 2021). Indeed, as part of best practice for around four decades, eyewitnesses have been interviewed using a Cognitive Interview (CI, e.g., Geiselman et al. 1985, 1986). This interview has particular application for eyewitnesses to recall information related to a crime, and has gone through various revisions of its constituent memory-enhancing techniques (e.g., Fisher et al. 1987; for discussion, see Dando et al. 2009; Milne and Bull 1999). A truncated version, specifically used to elicit face recall prior to composite construction, typically requests eyewitnesses to mentally reinstate the environment, and freely recall the face in detail, without guessing. When involving sketch or a feature system (e.g., Frowd, Carson, Ness, McQuiston, et al. 2005; Skelton et al. 2020), to increase overall recall, face constructors may also be invited to attempt additional (cued) recall following a free recall attempt (as was the case in the current experiment).

In the current project, to avoid confusion with the original CI (Geiselman et al. 1985) or its subsequent enhancements (e.g., Fisher et al. 1987), we use the term Composite Interview (CoI) to refer to the just-described pre-construction procedure. We also use the term Holistic Composite Interview (H-CoI) when this interview involves the standard face-recall components (i.e., the CoI) followed by holistic recall. As briefly alluded to previously, holistic recall procedures involve asking witnesses to reflect silently on the global aspects of the face and then make a series of global ratings (for intelligence, extraversion, pleasantness, etc.) based only upon the face's visual appearance (i.e., when other cues, such as behaviour, are ignored).

For recognition (naming) of a completed composite, global face processing techniques can also be applied to improve processing of holistic cues (e.g., Frowd et al. 2008, 2014; Skelton et al. 2020). For example, physically stretching a composite has been found to improve recognition (naming) of faces constructed from holistic (e.g., EvoFIT) and feature (e.g., PRO-fit) systems (e.g., Frowd et al. 2013, 2014; Skelton et al. 2020). Similarly, when a composite is viewed side-on, the result creates the illusion that the image is stretched on its vertical axis—a 'perceptual' stretch—as the side of the face furthest away from the viewer appears elongated (although, due to perspective, the face is also perceived with an affine transformation of shear). As vertically-stretched images (both facial photographs and facial composites) remain recognisable even when obscuring feature information via the application of high-level visual blur (e.g., Frowd et al. 2014; Hole et al. 2002), the notion is that stretching an image increases the salience of holistic cues in the face; for a composite, this seems to reduce the perception of visual error between features in the composite and the familiar face stored in memory. While physical and perceptual image stretch facilitates recognition of holistic composites (Frowd et al. 2013, 2014), the advantage is restricted for feature composites. Here, side-on viewing of the composite has only been found to facilitate naming (cf. front-on) following an H-CoI (cf. CoI; Skelton et al. 2020), and so holistic cues do not appear to be sufficiently rendered in a feature composite (when created following a non-H-CoI procedure).

In the current experiment, we explored how best, or indeed worst, these feature composites could be constructed. We involved a typical modern system of this type, PRO-fit, and assess whether the effectiveness of its composites could be facilitated when applying both the addition of the trait–recall mnemonic (the H-CoI vs. the standard face-recall [CoI] procedure alone) and side-on (vs. front-on) naming. The efficacy of both techniques was compared when composites were produced following either incidental or intentional encoding of a target face. Further, the influence of these variables was modelled under one of two forensically relevant delays: When the composite was constructed 2 days following the crime, a delay typically experienced by witnesses, and on the same day as the crime, specifically 3–4 h later, a scenario that occurs when the opportunity arises (e.g., in ~10% of cases in several forces in the UK and Europe, according to our conversations with police practitioners).

Based on the aforementioned research, we anticipated that the three between-subjects predictors (encoding, delay and interview) would each facilitate face construction. Specifically, more effective composites, images that result in higher naming rates, should be produced when memory for the face is stronger—that is, when we adopt levels of the encoding (intentional) and retention interval (3–4 h) variables that support preservation of, and access to, this trace. We further predicted that the influence of these variables would be independent, and so additive rather than interactive effects would emerge. Also, while the H-CoI has been found to be effective (cf. CoI) for short (up to 3–4 h) and long (1 day) delays under intentional encoding (e.g., Frowd et al. 2015; Portch et al. 2017; Skelton et al. 2020), there was no good theoretical reason to predict that it should not remain effective following incidental encoding. In addition, following on from the above predictions, it was further expected that interview would also have an additive benefit on face construction. However, while the benefit of the within-subjects predictor, view at naming, does not seem to be restricted by type of interview for a holistic system, for PRO-fit, this predictor was expected to interact with interview (as found in Skelton et al. 2020). Here, correct naming was anticipated to benefit from side-on (cf. front-on) viewing of composites constructed following an H-CoI, with no such benefit anticipated following a CoI.

2 | Methods

A two-stage sequential experimental procedure was administered (e.g., Frowd, Carson, Ness, McQuiston, et al. 2005). In Stage 1, participants viewed a single video clip from the long-running UK TV soap EastEnders, a sequence that depicted an interaction between two people. Participants subsequently returned to the laboratory to provide a description of the target face they had seen, before constructing a single composite of this individual. Crucially, all Stage 1 participants confirmed that they did not follow the soap, to be *unfamiliar* with the sampled EastEnders's identities. In Stage 2, a second set of participants attempted to name a sub-set of these composites. These participants were recruited on the basis of being regular viewers of EastEnders, so as to be *familiar* with the relevant identities. Using target-unfamiliar participants as composite constructors and target-familiar participants as recognisers models the typical forensic situation under which composites are usually constructed and recognised.

2.1 | Stage 1: Composite Construction

2.1.1 | Participants

We aimed to recruit sufficient participants to Stages 1 and 2 of the experiment to be able to detect a medium, and thus a forensically useful, effect. This equates to a minimum difference in means (MD) of ~15% correct in composite naming. As such, we aimed to be able to detect an odds ratio [Exp(*B*)] of at least 2.5 (as calculated by Sporer and Martschuk 2014) for Generalised Estimating Equations (GEE), a frequent method of analysis in the field (e.g., Brown et al. 2020; Frowd et al. 2013; Martin et al. 2017; Portch et al. 2017). While not pre-registered, the study specified the described sample size, method and approach for analyses prior to commencement of the project.

Based on previous experience of conducting considerable similar, multi-predictor experiments (Frowd 2021; ibid.), we

estimated that ~100 participant naming responses per condition were required, an estimate that we operationalised for each between-subjects factor (encoding, delay and interview) as two groups of 12 participants (per cell of each IV) for face construction and two groups of six participants for composite naming. For the three between-subjects predictors, this resulted in a sample size of 96 participants ($Ns = 2 \times 2 \times 2 \times 12$) for face construction and 48 participants ($Ns = 2 \times 2 \times 2 \times 6$) for naming.

We also assessed statistical power for this proposed sample using a series of computer simulations based on an effect size appropriate for the planned statistical analyses (see Appendix A).

Based on this design, for face construction, 96 target-unfamiliar staff and students from the University of Leeds were recruited (77 females, 19 males; M = 24.1, SD = 10.0, range: 18–69 years). Participants received either course credit or £5 for their participation, and were allocated in equal groups of 12 to the eight individual conditions of the three between-subjects predictors.

2.1.2 | Materials

The target identities were six male and six female characters from the BBC TV soap, EastEnders. These identities were presented via six non-violent video clips, each lasting between ~30–60 s, and each portraying a social interaction between a different male and female character. In each clip, both characters were visible in a largely frontal pose for approximately equal proportions of time, with the sequence ending at a natural break in the interaction (i.e., at the end of a sentence).

2.1.3 | Design and Procedure

Participants were randomly assigned to a 2 (Encoding: Intentional vs. Incidental) × 2 (Delay: 3–4 h vs. 2 days) × 2 (Interview: CoI vs. H-CoI) between-participants design. Note that the fourth predictor, View (Front-on vs. Side-on), is relevant to Stage 2 of the experiment, composite naming. For face construction, the 12 target identities were each constructed by different participants in each of these eight conditions, producing a total of 96 composites.

The decision to employ a single experimenter reduced the potential for differences in interviewing expertise to impact composite construction (e.g., Davies et al. 1983). This person (the second author) was trained to use the PRO-fit composite system in-house, and practiced face construction extensively. She was responsible for all interactions with participants, presenting stimuli, conducting the relevant interview (CoI or H-CoI) and controlling the composite software. Her role was to facilitate face construction with the aim of allowing participants to create the best likeness possible. So that she could assist in the process of face construction, but not influence the identity under-construction, she did not view any of the target videos until all composites had been constructed.

Participants were tested individually. In the first experimental session, one of the 12 video clips was randomly selected and shown to the participant, in the absence of the experimenter. Those given incidental encoding instructions were told that they would later need to recall details about the social interaction (e.g., the dialogue) and were not made aware of the impending composite construction task. Participants given intentional encoding instructions were asked to attend to the facial features of one specific target face, (i.e., either the male or female character), as they would later be asked to construct a composite of this person. We note that, while participants in the incidental encoding condition had their attention directed to the social interaction, they may still have encoded the target face to some extent; however, as desired, participants in the intentional condition were expected to have a qualitatively better memory of the face. On five occasions, a check revealed that a character in the video clip was reported to be familiar to the participant and, in these cases, a new video clip was randomly selected (from the same experimental condition) and presented similarly (with these participants then reporting that the second face presented was unfamiliar).

Participants returned for a second session either 3 to 4 h or 2 days later. At this time, all participants were informed that they would be required to describe and construct a composite of one of the two identities seen in the video. Participants first recalled the face using a CoI. The experimenter asked the participant to think back to when the 'target' had been seen and to form a visual image of the face (context reinstatement). Then, the participant was asked to freely recall as many details as possible about the face, without guessing. In a subsequent cued recall stage, the experimenter repeated back the participant's description of each facial feature, pausing each time to ask whether any further information could be recalled. Facial feature information was prompted in the following order: overall appearance, face shape, hair, eyebrows, eyes, nose, mouth and ears.

Half of the participants then received the trait–recall instruction (as per the H-CoI procedure of Frowd et al. 2008). Participants were given 60 s to visualise the face and think silently about the personality conveyed by the target face. Afterwards, participants were asked to make seven trait judgements about the face, in their own time. Judgements were requested to be made solely on the basis of the face's appearance, ignoring knowledge participants may have gained about the person's character in the video. Participants were prompted to rate the face on a scale of 'low', 'medium' or 'high' in order of intelligence, friendliness, kindness, selfishness, arrogance, distinctiveness and aggressiveness.

To construct a composite, the experimenter entered the participant's description of the face into the PRO-fit system, to locate approximately 20 appropriate examples per facial feature, which created an 'initial' composite (i.e., a face whose appearance matched the description). Under the guidance of the participant, the experimenter exchanged features in this face with other appropriate examples, and made changes to a feature's size, position, brightness and contrast, until the participant indicated that the face could not be improved upon. The artwork package in PRO-fit was offered, to enhance the face if the participant felt this was necessary, for example, by adding wrinkles or stubble. Composite face construction, including debriefing, took approximately 50 min, per person. The holistic procedure increased session duration by 5 min.

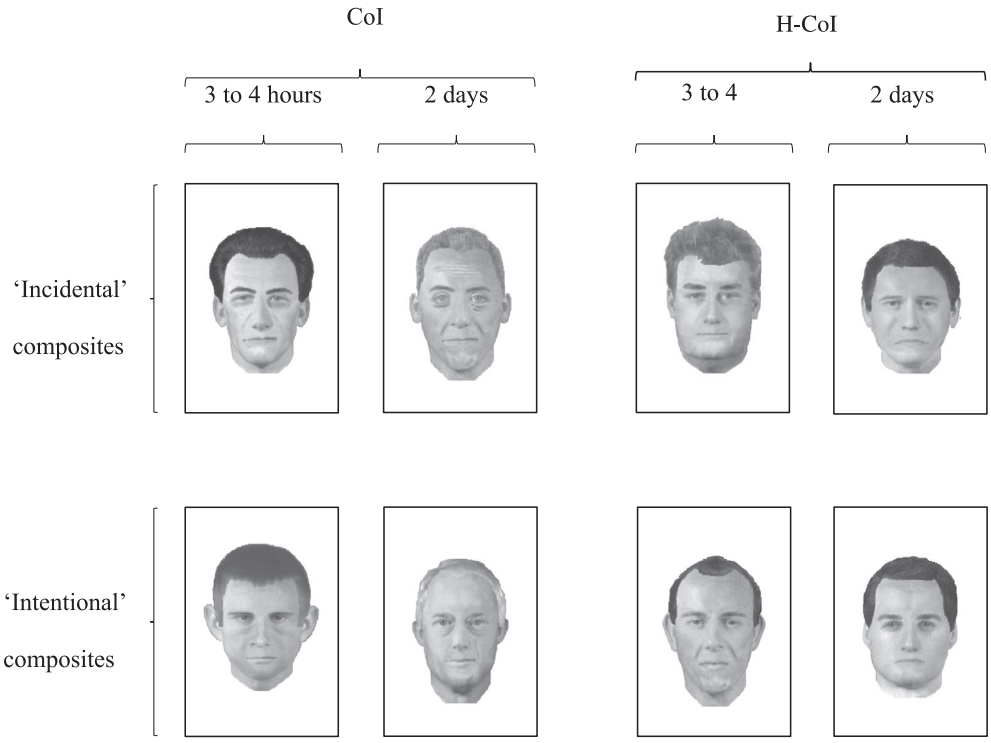

FIGURE 1 | Composites constructed to resemble Billy Mitchell from the BBC TV programme EastEnders. Composites were constructed by (i) encoding condition (intentional vs. incidental), (ii) post-encoding delay (3–4 h vs. 2 days) and (iii) interview (CoI vs. H-CoI). These composites (along with other composites produced in the study) were given to reported-to-be regular viewers of EastEnders to name: These participants looked at the face from the front (e.g., by looking at this page normally) and then from the side (which readers could perhaps try for themselves by turning the page to the side so that the face appears to be long and thin). Note that the photograph of Billy Mitchell used in the study cannot be reproduced here for reasons of copyright, but readers may like to view this face via a simple internet search.

2.2 | Stage 2: Composite Evaluation

2.2.1 | Participants

Forty-eight staff and student volunteers from the University of Leeds were recruited on the basis that they reported to be regular viewers of EastEnders (45 females, 3 males; M = 26.2, SD = 9.6, range: 18–56 years).

2.2.2 | Materials

The composites were printed on A4 paper in greyscale, the image format of the facial-composite system, one per page (10 cm wide × 15 cm high) (see Figure 1 for examples). There were eight composite sets, each including the 12 composites constructed in a single condition during Stage 1, along with six additional 'foil' composites (three male and three female), also constructed using PRO-fit, repeated per condition. Inclusion of foils parallels the real-world situation wherein a recogniser must first decide if a composite is familiar before attempting to name it. Foil composites were representative of the age range sampled within the target set and did not share any obvious features with any of the experimental composites (e.g., none had the same hairstyle). Colour photographs, showing head and shoulder frontal views of the 12 targets, were also printed, one per page (10 cm wide × 15 cm high).

2.2.3 | Design and Procedure

Six participants were randomly assigned, with equal sampling, to inspect one of the eight individual sets of composites in a Mixed Factorial 2 (Encoding: Incidental vs. Intentional) × 2 (Delay: 3–4 h vs. 2 days) × 2 (Interview: CoI vs. H-CoI) × 2 (View: Front-on vs. Side-on) design. All factors were manipulated between-subjects, as in Stage 1, except View, which was also manipulated here, within-subjects. View always followed a fixed order, rather than being counterbalanced, with the faces presented front-on and then side-on. This design reflects the order followed by police practitioners[1].

Participants were tested individually, and the task was self-paced. They were recruited on the basis of being regular viewers of the TV soap EastEnders, and so may have expected that these identities would be involved. However, to avoid potential differences in expectation, all participants were told that, while some of the composites were constructed to resemble EastEnders's characters, others resembled people who would be unfamiliar (i.e., the foils). Each participant viewed the composites belonging to a single set sequentially in the normal front-on perspective, and attempted to provide identifying information for each (real or stage names, or sufficient individuating semantic details) or gave a 'don't know' response. Participants then attempted to name each composite for a second time, wherein they were instructed to turn each page

so that the face could be viewed from the side, having been informed that this presentation method might prompt recognition. Lastly, to check that participants were familiar with the targets to which the composites corresponded, they were asked to name a photograph of each EastEnders's character involved in the study. Composite and target stimuli were presented in a different random order for each participant. The naming task, including debriefing, was completed in about 15 min.

2.2.4 | Results

Participant responses to composites and target photographs were initially scored for accuracy: a numeric value of 1 was assigned when the correct identity was given, and 0 otherwise. The target photographs were correctly named at 97.2% (SD = 5.0%), and so participants, in general, were highly familiar with the relevant identities. More specifically, participants did not correctly name a target photograph on 16 occasions. In these instances, the response to the corresponding composite was removed from the dataset prior to inferential analysis. An additional 12 missing data points occurred due to one participant not completing the side-on section of the naming task. As composites were presented twice for naming, front-on and then side-on, a total of 44 responses ($N = 2 \times 16 + 12$) were removed.

2.2.4.1 | Correct Composite Naming. Mean correct naming for the 96 composites was low, at 12.2% (SD = 32.7%). We estimate chance naming to be around 1%, or less, as indicated by other research using a feature system following a long retention interval (e.g., Frowd, Carson, Ness, McQuiston, et al. 2005; Frowd, McQuiston-Surrett, et al. 2007). Thus, while mean correct naming was somewhat higher than chance, it was much lower than correct naming of the target photographs. This outcome is expected given that, unlike photographs, composites do not represent a veridical image of the person, making them difficult to recognise.

Individual condition means ranged from 4.5% to 29.0% (see Table 1), the latter rate attained in the condition predicted to be most effective, specifically when construction occurred 3–4 h after a target was intentionally encoded, a H-CoI was applied, and composites were viewed side-on at naming. More specifically, the four individual predictors each led to an overall increase in correct naming in the predicted direction, but differences between means were at best small. Thus, composites emerged with relatively higher naming: (i) for intentional than incidental encoding (MD = 4.6%), (ii) when the delay was short (3–4 h) than long (2 days) (MD = 3.4%), (iii) following H-CoI than CoI (MD = 2.2%) and (iv) when the face was viewed side-on than front-on (MD = 2.2%).

Inferential analysis was conducted in SPSS (version 29) on the participant responses (coded as above) using GEE. As responses for correct naming were dichotomous, a logistic 'link' function was selected with a binomial probability distribution. Also, as participants attempted to name a series of composites, 12 in total, the related nature of each person's responses was taken into account by selecting an 'exchangeable' correlation matrix. In terms of composition of a model that best describes the influence of the predictors, we followed the principle of parsimony (Field 2018). Here, as predictors usually influence the DV to some extent, and to facilitate interpretation, the approach includes all predictors that have explanatory value on the DV, except when an interaction is involved. Then, as the individual predictors involved in the interaction tend to influence the interaction itself, and vice versa, these individual predictors were always included (i.e., even if they themselves do not influence the DV). Finally, it is important to avoid making a Type II error. This situation can occur if individual predictors are assessed in isolation to each other (e.g., Reed and Wu 2013), an issue that was avoided by considering predictors in a 'combined' model.

The approach proceeded with the largest model and then, if necessary, considered successively smaller models, following a 'stepwise', backward-type method for selection of variables. As such, an iterative process is usually required to determine model composition. However, to lessen the chance of making a Type II error when variables are removed, a conventional, evidence-supported and SPSS-default alpha value of 0.1 was used to retain predictors, and interactions between predictors, in the model (Field 2018; Harrell 2015). In the case involving two predictors, for example, a full-factorial model (i.e., one containing the two individual predictors and their interaction) would be constructed first. If the interaction emerged with a *p*-value less than alpha, this GEE would be taken as the 'final' model (and the interaction explored). If not, the interaction would be removed and both individual predictors assessed in a combined model. If both predictors emerged less than alpha, the result is a final, null-predictor model; if not, the individual predictors would be assessed separately.

We followed this approach for our factorial design involving four predictors, each time checking that coefficients (B) and their standard error [SE(B)] remained within sensible limits,

TABLE 1 | Correct naming by Delay (3–4 h vs. 2 days), Encoding (incidental vs. intentional), Interview (CoI vs. H-CoI) and View (front-on vs. side-on).

		Presentation at naming			
		Front-on		Side-on	
Delay	Encoding	CoI	H-CoI	CoI	H-CoI
3–4 h	Incidental	9.9 (7/71) [30.0]	10.4 (7/67) [30.8]	6.8 (4/59) [25.4]	13.4 (9/67) [34.4]
	Intentional	13.9 (10/72) [34.8]	14.5 (10/69) [35.5]	12.5 (9/72) [33.3]	29.0 (20/69) [45.7]
2 days	Incidental	12.9 (9/70) [33.7]	4.5 (3/67) [23.9]	12.9 (9/70) [33.7]	6.0 (4/67) [23.9]
	Intentional	8.3 (6/72) [27.8]	13.9 (10/72) [34.6]	11.1 (8/72) [31.6]	13.9 (10/72) [34.8]

Note: Values are correct-naming scores calculated by dividing responses shown in parentheses and expressed as a percentage. Underneath, parenthesised values are summed correct responses (numerator) of total responses (denominator: correct + mistaken + no-name). SD of the means are presented in square brackets.

TABLE 2 | Final model for correct naming for the predictors: Delay (3–4 h vs. 2 days), Interview (CoI vs. H-CoI) and View (front-on vs. side-on).

Tests of model effects	$X_1^2(1)$	p_1	$X_2^2(1)$	p_2
Intercept	283.03	< 0.001	104.11	< 0.001
Delay	1.48	0.22	0.31	0.58
View	2.16	0.14	1.91	0.17
Interview	1.75	0.19	0.02	0.89
Delay × View	0.26	0.61	0.19	0.66
Delay × Interview	1.75	0.19	1.01	0.32
View × Interview	2.60	0.11	2.81	0.094
Delay × View × Interview	3.74	0.062	3.92	0.048

Note: X_1 and p_1 refer to the analysis by-participants, with the model's goodness of fit: $QIC = 825.35$ and $QICC = 825.48$. X_2 and p_2 refer to the analysis by-items: $QIC = 843.34$ and $QICC = 826.29$.

since values that are too low or too high indicate an issue with model fit. This approach revealed that the robust (or 'sandwich') estimator for the Covariance Matrix was preferable (see Huber 1967), since GEE resulted in much lower SE(B) estimates compared to selecting a Model-based estimator, and so all analyses involved this method of estimation.

Commonly, inferential analyses initially examine the influence of participant responses on the DV using a traditional by-participants analysis (e.g., Frowd, Bruce, Ross, et al. 2007), one that essentially assesses whether results generalise to other participants. As participants attempt to name multiple composites (here, one composite for each of 12 identities), it is important to check that results generalise to other identities, to avoid the risk of making a stimulus-as-a fixed-effect fallacy (Clark 1973; Lewis 2023). Therefore, a second analysis by-items[2] was conducted. However, when multiple predictors are involved, the by-participants analysis tends to be more powerful, given the usual case that there are more participants than there are items in an experiment—here, there are 48 participant-namers and 12 items. Therefore, it is better to reverse the order of analyses, conducting by-items first and then checking that results generalise to other participants[3]. We followed this approach.

For best generation, three major sources of random error were included. These were the 96 participant-constructors (coded from 1 to 96), the 48 participant-namers (coded from 1 to 48) and the 12 item identities (coded from 1 to 12) involved in the experiment. For the analysis by-participants, participant-namers were specified as a between-subject variable and items as a within-subjects variable; the order of these variables was reversed in the by-items analysis. For both analyses, participant-constructors were specified as a between-subjects variable.

We therefore proceeded with a full-factorial model, by-items. The predictors were Encoding (coded as *1* = Incidental, *2* = Intentional), Delay (*1* = 3–4 h, *2* = 2 days), Interview (*1* = CoI, *2* = H-CoI) and View (*0* = Front-on, *1* = Side-on). All predictors were between-subjects except View, and predictors and DV (coded as above) were arranged in descending numerical order. For this four-factor model, GEE suggested that the four-way interaction should not be retained ($p = 0.34$, $\text{Exp}(|B|) = 2.54$)[4]. When removed, the model was run again, to allow assessment of the three-way interactions. GEE indicated that Delay × Interview × View should be retained ($p = 0.032$, $\text{Exp}(|B|) = 2.65$), unlike the remaining three-way interactions ($ps > 0.36$, $\text{Exp}(|B|) = 1.13–3.54$). The result was a full-factorial model for Delay, Interview and View, and the remaining single predictor, Encoding. For this model, GEE indicated that Encoding should also be removed ($p = 0.19$, $\text{Exp}(|B|) = 1.58$). The result was a final, full-factor model comprising Delay, Interview and View (Table 2).

TABLE 3 | Correct naming by Delay (3–4 h vs. 2 days), Interview (CoI vs. H-CoI) and View (front-on vs. side-on).

	View			
	Front-on		Side-on	
Delay	CoI	H-CoI	CoI	H-CoI
3–4 h	11.9 (17/143) [32.5]	12.5[a] (17/136) [33.2]	9.9[b] (13/131) [30.0]	21.3[a,b] (29/136) [41.1]
2 days	10.6 (15/142) [30.8]	9.4 (13/139) [29.2]	12.0 (17/142) [32.6]	10.1 (14/139) [30.2]

Note: See Table 1, *Note*, for derivation of values. In the final model, by-items and by-participants, Delay × Interview × View ($p < 0.1$): [a]$p < 0.05$; [b]$p < 0.1$.

A summary of means by Delay, Interview and View are presented in Table 3. A simple-main effects analysis was conducted (by-items) for Delay × Interview × View. This analysis (Table 4) revealed that, following a 3–4 h delay, composites were more identifiable (a) for side-on naming, when composites were created following a H-CoI than CoI ($p = 0.049$) and, as a marginal effect, (b) for an H-CoI, when composites were viewed side-on rather than front-on ($p = 0.052$). In the associated analysis by-participants, the analysis (Table 2) retained Delay × Interview × View ($p = 0.062$) in a full-factor model involving these three predictors; the two aforementioned contrasts involved in this interaction (Table 4) also had explanatory value ($ps < 0.02$).

TABLE 4 | Summary of GEE model for correct naming for the three-way interaction between Delay (3–4 h vs. 2 days), Interview (CoI vs. H-CoI) and View (front-on vs. side-on).

Fixed effects	*B*	SE(*B*)	$X^2(1)$	*p*	Exp(*B*)	95% CI(−)	95% CI(+)
Intercept							
By-participants	−2.00	0.26	60.10	<0.001	0.14	0.09	0.21
By-items	−2.01	0.32	39.53	<0.001	0.13	0.08	0.23
Interaction (following a 3–4 h delay)							
(i) H-CoI > **CoI** (for side-on)							
By-participants	0.87	0.35	6.14	0.013	2.39	1.20	4.77
By-items	0.88	0.45	3.87	0.049	2.40	1.00	5.76
(ii) Side-on > **front-on** (for H-CoI)							
By-participants	0.64	0.24	7.35	0.007	1.90	1.19	3.01
By-items	0.65	0.34	3.77	0.052	1.92	0.99	3.71

Note: The reference category for each contrast is shown bolded. Interaction (i) (ii) other *p*s > 0.1.

We also compared composites created under the best combined condition (M = 21.3%, H-CoI, 3–4 h delay and side-on naming) with traditional practice (M = 10.6%, 2-day delay, face-recall CoI and front-on naming). Correct naming doubled over these conditions (29/136 vs. 15/142 correct responses, respectively): the advantage of the best combined condition was of medium size in the simple-main effects analysis, by-items [$B = 0.90$, SE(B) = 0.41, $X^2(1) = 4.80$, $p = 0.029$, Exp(B) = 2.45 95% CI (1.10, 5.48)] and by-participants [$B = 0.83$, SE(B) = 0.34, $X^2(1) = 5.83$, $p = 0.016$, Exp(B) = 2.30 95% CI (1.30, 4.04)].

To summarise, the GEE analysis revealed, contrary to expectation, that none of the between-subjects predictors (Encoding, Delay and Interview) exerted a significant overall effect on correct naming of feature composites. Also, while an interaction between Interview and View was predicted, it emerged qualified by Delay: composites attracted significantly higher correct naming following an H-CoI (cf. CoI), when they had been constructed after 3–4 h, and were named side-on. Thus, the benefit of a side-on (cf. front-on) view (see Table 3: MD = 11.4%) was reliant not only upon a H-CoI having been conducted (as predicted), but when there existed a short delay between encoding and construction (3–4 h), with predicted benefits absent at 2 days (MD = −1.9%).

As considered in the General Discussion, more effective composites were created when the memory of the constructor was relatively stronger (i.e., after 3–4 h cf. 2 days), when constructors' face recognition had been enhanced (i.e., using an H-CoI cf. CoI), and when the face was presented to encourage holistic processing (i.e., when participant-namers viewed the face side-on cf. front-on). Note that these effects were independent of how a constructor encoded a target face—that is, incidentally or intentionally.

2.2.4.2 | Mistaken Composite Naming. Composites may be recognised as an identity that is different to that intended by the person constructing the face. Such 'mistaken' names occur sometimes when a witness unknowingly creates a likeness that shares facial characteristics with another identity, a situation more likely to occur when memory for the target identity is weak, perhaps as a result of a longer post-encoding delay or incidental encoding. While an inaccurate name put forward for a composite might seem problematic, it can actually be beneficial in the context of good policing and forensic practice. In these fields, where sufficient and accurate evidence is essential to support a reliable conviction, mistaken names can help the police eliminate a person from an investigation—specifically, someone who was not the identity intended to be portrayed in the composite. From a theoretical perspective, examining both correct and mistaken names provides a more comprehensive assessment of composite accuracy.

For this second measure of composite effectiveness, participant data were rescored, this time, for cases where the given name was of the wrong identity (coded as *1*) relative to all other responses (*0* = correct name or 'don't know' response). We again removed responses to composites ($N = 44$) for which the target identity had not been correctly named. Note that it is a common occurrence for feature composites to be misnamed frequently (e.g., Frowd et al. 2015), as was the case here ($N = 595/1108$, M = 53.7%, SD = 49.9%). As for correct naming, this DV changed little across levels of each predictor: Encoding (MD = 1.5%), Delay (MD = 6.4%), Interview (MD = 1.2%) and View (MD = 3.5%).

By mean condition, mistaken naming ranged from 38.6% to 67.8% (see Table 5). However, while the individual predictors made little difference, it is worth noting that one of the *lowest* means for mistaken naming in the experiment (M = 39.1%, an outcome that is indicative of relatively *superior* composites) emerged in the condition that was predicted to produce the most effective composites by correct naming (i.e., following intentional encoding, 3–4 h delay, H-CoI and side-on naming). This indicates that faces created in this condition were overall more accurate: visually closer to the intended identities (i.e., based on higher correct naming) and also further away from non-intended identities (i.e., based on lower mistaken naming).

TABLE 5 | Correct naming by Delay (3–4 h vs. 2 days), Encoding (incidental vs. intentional), Interview (CoI vs. H-CoI) and View (front-on vs. side-on).

		Presentation at naming			
		Front-on		Side-on	
Delay	Encoding	CoI	H-CoI	CoI	H-CoI
3–4 h	Incidental	57.8 (41/71) [49.7]	55.2 (37/67) [50.1]	67.8 (40/59) [47.1]	56.7 (38/67) [49.9]
	Intentional	62.5 (45/72) [48.8]	52.2 (36/69) [50.3]	65.3 (47/72) [47.9]	39.1 (27/69) [49.2]
2 days	Incidental	38.6 (27/70) [49.0]	56.7 (38/67) [49.9]	52.9 (37/70) [50.3]	52.2 (35/67) [50.3]
	Intentional	41.7 (30/72) [49.6]	51.4 (37/72) [50.3]	50.0 (36/72) [50.4]	61.1 (44/72) [49.1]

Note: Values are mistaken-naming scores calculated by dividing responses shown in parentheses and expressed as a percentage. Underneath, parenthesised values are summed mistaken responses (numerator) of total responses (denominator: correct + mistaken + no-name). SD of the means are presented in square brackets.

TABLE 6 | Final model for mistaken naming for the predictors: Delay (3–4 h vs. 2 days), Interview (CoI vs. H-CoI) and View (front-on vs. side-on).

Tests of model effects	$X_1^2(1)$	p_1	$X_2^2(1)$	p_2
Intercept	4.18	0.041	2.29	0.13
View	2.70	0.10	2.70	0.10
Delay	2.98	0.084	2.50	0.11
Interview	0.15	0.70	0.01	0.95
Delay × interview	8.96	0.003	6.02	0.014
View × interview	5.36	0.021	5.34	0.021

Note: X_1 and p_1 refer to the analysis by-participants, with the model's goodness of fit: $QIC = 1520.83$ and $QICC = 1518.95$. X_2 and p_2, by-items: $QIC = 1531.12$ and $QICC = 1519.36$.

The same approach, as described above, was followed for analysing mistaken responses. Thus, a full-factorial model was constructed comprising the four predictors, by-items. This GEE indicated that the four-way interaction ($p = 0.12$, $\text{Exp}(|B|) = 3.00$) should not be retained. When removed, the subsequent model indicated that Delay × Encoding × View should be retained ($p = 0.083$, $\text{Exp}(|B|) = 1.87$), while the other three-way interactions should not ($ps > 0.42$, $\text{Exp}(|B|) = 1.20–1.67$). In a revised model, however, Delay × Encoding × View ($p = 0.11$, $\text{Exp}(|B|) = 1.81$) failed to reach the necessary alpha and was also removed[5]. When removed, GEE indicated that both Delay × Interview ($p = 0.021$, $\text{Exp}(|B|) = 2.48$) and View × Interview ($p = 0.022$, $\text{Exp}(|B|) = 1.53$) should be retained (other $ps > 0.13$, $\text{Exp}(|B|) = 1.24–1.31$). Next, a combined model was assessed comprising these two-way interactions and their constituent predictors, and the remaining single predictor, Encoding. GEE suggested that Encoding ($p = 0.58$, $\text{Exp}(|B|) = 1.11$) should also be removed. The resulting, final model (Table 6) comprised Delay × Interview and View × Interview and their three associated individual predictors.

TABLE 7 | Mistaken naming by Interview (H-CoI vs. CoI) and Delay (3–4 h vs. 2 days).

	Delay	
Interview	3–4 h	2 days
CoI	63.1[a] (173/274) [48.3]	45.8[a] (130/284) [49.9]
H-CoI	50.7 (138/272) [50.1]	55.4 (154/278) [49.8]

Note: See Table 5, *Note*, for derivation of values. In the final model, by-items and by-participants, Delay × interview ($p < 0.05$): [a]$p < 0.01$.

TABLE 8 | Mistaken naming by Interview (H-CoI vs. CoI) and View (front-on vs. side-on).

	View	
Interview	Front-on	Side-on
CoI	50.2[a] (143/285) [50.1]	58.6[a] (160/273) [49.3]
H-CoI	53.8 (148/275) [49.9]	52.4 (144/275) [50.0]

Note: See Table 5, *Note*, for derivation of values. In the final model, by-items and by-participants, Interview × view ($p < 0.02$): [a]$p < 0.01$.

A summary of means is presented in Table 7 for Delay × Interview and in Table 8 for View × Interview. A simple-main effects

TABLE 9 | Summary of GEE model for mistaken naming for predictors: Delay (3–4 h vs. 2 days), Interview (CoI vs. H-CoI) and View (front-on vs. side-on).

Fixed effects	*B*	SE(*B*)	$X^2(1)$	*p*	Exp(*B*)	95% CI(−)	95% CI(+)
Intercept							
By-participants	0.36	0.16	5.14	0.023	1.43	1.10	1.85
By-items	0.37	0.22	2.87	0.090	1.45	1.01	2.08
Interaction (following CoI)							
(i) **3–4 h** > 2 days							
By-participants	0.70	0.21	11.37	< 0.001	2.01	1.34	3.02
By-items	0.80	0.29	7.87	0.005	2.23	1.27	3.89
(ii) Side-on > **front-on**							
By-participants	0.34	0.13	7.15	0.008	1.41	1.10	1.82
By-items	0.37	0.14	7.27	0.007	1.45	1.11	1.89

Note: The reference category for each contrast is shown bolded. Interaction (i) (ii) other *p*s > 0.1.

analysis (Table 9) revealed that the two interactions emerged due to differences caused when face construction followed a CoI. By-items, (i) Delay × Interview was retained in the model as there was higher mistaken naming when construction occurred 3–4 h (cf. 2 days) after encoding ($p = 0.005$); all other comparisons were ns (*p*s > 0.36, Exp(|*B*|) = 1.18–1.32), and (ii) View × Interview was also retained as mistaken names were higher for side-on than front-on naming ($p = 0.007$); other comparisons were ns (*p*s > 0.36, Exp(|*B*|) = 1.06–1.32). Conclusions were the same, by-participants (see Table 9).

The results revealed that the number of mistaken names were generally not influenced by the manipulations in the experiment, except when a CoI was involved. Then, mistaken names were more prevalent at the shorter (3–4 h) than the longer (2 day) delay, and also when the face was viewed side-on than front-on. The same as for the other DV, mistaken naming was also not influenced by type of encoding (incidental vs. intentional). These intriguing results are considered in the Discussion.

3 | Discussion

We evaluated the effectiveness of two techniques previously shown to increase composite naming: adding a trait–recall mnemonic to a face-recall composite interview (CoI) typically used in police practice (to form the H-CoI), and perceptual stretch. We considered whether the advantage of using these techniques would be maintained for feature composite systems (used within Europe, the USA and Australia) across conditions typically encountered within forensic settings. When composite construction took place on the same day as viewing the target face (i.e., within 3–4 h), adding the trait–recall mnemonic increased correct naming (i) compared to the standard face-recall CoI, for side-on viewing, and (ii) for side-on compared to front-on naming. There was no corresponding increase in incorrect naming following joint application of the trait–recall mnemonic and side-on naming (for either type of encoding or retention interval). Thus, applying the trait–recall mnemonic led to construction of a more accurate visual likeness following the short retention interval; however, this diagnostic information was not readily extracted from a composite until a technique was applied that increased a recogniser's sensitivity to this information (i.e., via a side-on view of the face).

It is proposed that directing a witness's attention to holistic information, via engaging in trait–recall, facilitates accurate selection and placement of facial features, likely because holistic processing encourages individual features to be considered within the context of a whole face (e.g., Tanaka and Farah 1993). However, to be of measurable benefit, holistic recall needs to be elicited on the same day as encoding (here, 3–4 h later), rather than 2 days later (i.e., where there was no/little benefit for the H-CoI vs. CoI). Indeed, the addition of the trait–recall mnemonic to the CoI has been found to facilitate the accurate construction of both external (hair, ears and neck) and internal features (eyes, brows, nose and mouth; Frowd et al. 2008), with the latter being particularly important for the recognition of familiar faces (e.g., Ellis et al. 1979). However, the ability to recall feature information about a face (i.e., using a CoI) also appears to be important. Frowd et al. (2012) found that asking participants only to attribute trait characteristics to the target face, compared to recalling the face using a CoI, led to *less* effective composites. As mentioned, this implies that to benefit from attending to the face in a holistic manner, a witness first needs to have effectively brought to mind the features of the face. Having done that, facial features can then be organised in a global way, one that favours the ensuing task, face recognition.

When potential recognisers view a composite side-on, this may increase the perceived accuracy of feature placement (Frowd et al. 2014), as well as encouraging these observers to process the face as a whole. This happens as side-on viewing likely requires the cognitive system to transform (i.e., normalise) the stretched image in order to extract its face-like properties (Hole et al. 2002). As mentioned, this process of transformation may reduce the appearance of error between the individual features within the composite and the target face, giving rise to perception of an image that more successfully matches the representation of the face in memory. Thus, when the memory of the

face was sufficient (i.e., 3–4 h after encoding), it is likely that the trait–recall mnemonic and perceptual stretch techniques worked in harmony: Trait–recall reduced error in the selection and placement of features within the composite, and this error was perceived to be further reduced when the composite was viewed from the side. It is perhaps worth mentioning that, as our naming participants always viewed composites side-on after the initial front-on naming stage, side-on composites may have simply attracted higher naming rates as a function of repeated viewing. However, if this were the case, side-on (cf. front-on) viewing would have attracted higher naming in all conditions, which it did not, just some.

Following a 2-day post-encoding delay, addition of trait–recall to the CoI and side-on (vs. front-on) naming techniques did not lead to an increase in correct naming. This contrasts with Frowd et al. (2013), who did find an additive effect of these two techniques when participants constructed composites using the holistic system, EvoFIT, 24 h after viewing the target face. Contrasting results may arise as holistic systems (e.g., EvoFIT, EFIT-V or ID c.f., feature systems) place greater emphasis upon the importance of face recognition. Recognition is relatively more stable over time (cf. face recall), and thus the processes involved in holistic composite construction (recognising and selecting whole faces that resemble the target face) may increase the likelihood that diagnostic information is recreated, even at longer retention intervals (e.g., Frowd et al. 2012; Hancock et al. 2011). In contrast, participants have been found to show a rapid deterioration in the ability to recall information about features of the face, with notably fewer facial details recalled following a 24-h retention interval (Ellis et al. 1980) and, relatedly, construction of a less effective composite (e.g., Frowd et al. 2015; Portch et al. under revision). For a modern feature system such as PRO-fit, the ability to construct an identifiable composite that accurately represents diagnostic featural information will rely heavily on witnesses' ability to effectively recall fine-level feature information about a face. Thus, we may anticipate that even following a 24-h delay (shorter than the 2-day delay used here), use of the trait–recall mnemonic (the H-CoI) and perceptual stretch may similarly fail to provide a consistent benefit to feature-based face construction. Importantly, at such a retention interval, our data imply that neither the H-CoI, nor side-on naming, effectively compensates for the decay of facial information in memory over time.

Our results are only partially consistent with Frowd et al. (2008), however, who also found a benefit of incorporating the trait–recall mnemonic within the CoI when PRO-fit composites were constructed 3–4 h after intentionally encoding a target face. Unlike here, where the benefit was only found when naming composites side-on, their study found an increase in correct naming when composites were viewed front-on. There are procedural differences that explain the less robust effect of applying trait–recall for the present data. First, during the naming task, we presented composites intermixed with foils and participants were warned that not all composites were constructed to resemble characters from the target pool (i.e., EastEnders's characters). The inclusion of foils has been found to suppress correct naming (Frowd et al. 2015). Further, while Frowd et al. (2008) used video footage (similar to the format used here), their video presentation ended with a 5-s freeze frame on the target face, a format that resembles photographic presentation. Indeed, higher naming rates tend to be attracted by composites constructed from memory of photographs, as opposed to video footage (Frowd et al. 2015). Arguably, fine-level feature detail about a face may be more effectively encoded from photographs and, when constructing feature composites, accurate construal of this type of information can effectively cue identification. Indeed, with facial photographs as targets, recent research indicates that an encoding duration as short as 10 s can be sufficient to allow participants to create composites that are as effective as those following a longer, 30-s exposure (Erickson et al. 2022). This result suggests that shorter encoding times than those in the current experiment promote suitable encoding; it may also indicate why intentional encoding did not lead to more effective composites overall (since sufficient time was available for face encoding in the social interaction of the incidental condition), a result that would be worthy of further exploration. More generally, taken together with Frowd et al. (2008), our data demonstrate that incorporating trait–recall within the CoI confers a benefit on composite effectiveness when composite construction is undertaken on the same day as viewing the person of interest. However, in some contexts, this benefit will be too weak to detect unless an additional technique is applied to enhance a potential recogniser's sensitivity to diagnostic information within the composite; here, applying the perceptual stretch technique during composite viewing (but other techniques may be applicable; e.g., Frowd et al. 2008, 2014).

The perceptual stretch technique did not provide a general benefit to correct naming, differing from the significant benefit observed by Frowd et al. (2013) wherein EvoFIT was used for construction. However, in Frowd et al.'s (2013) study, EvoFIT composites constructed under conditions typical of best forensic practice, and following a CoI with front-on naming, were correctly named at a higher rate (37%, c.f., 11% here). Thus, EvoFIT composites evidently contain a higher proportion of identity-diagnostic information and naming rates may be further improved when viewing conditions appropriately increase potential recogniser's sensitivity to this information (i.e., via side-on naming). More recently, Skelton et al. (2020) found benefit for side-on (cf. front-on) naming for composites created from a feature system, but again participants encoded static photographs of target faces in this work, with this format associated with creation of a more robust memory trace (similar to the shorter, 3–4 h, retention interval used here).

Typically, in a real-world context, a composite is normally constructed with a witness 1 or 2 days after the event of interest. Our data indicate that when using a recall-based construction method, addition of either a trait–recall mnemonic or a perceptual stretch technique at these longer post-encoding delays does not improve composite naming rates. In fact, we found that the application of perceptual stretch led to a general increase in mistaken naming for composites constructed after a face-recall CoI (Table 8); there was also higher mistaken naming for CoI under the shorter (cf. longer) retention interval (Table 7). In both cases, the CoI tends to have a focus of attention on facial features, an effect that carries over to face construction, yielding a face that is recognised to resemble other identities when viewed side-on, or when face construction is conducted on the same day as the event. For side-on viewing, it would appear that perceptual stretch upregulates 'recognition'

experiences more generally (i.e., whether the proffered name is correct for that identity, see Table 3, or not, Table 8) by concealing inaccuracies; for construction following a short retention interval, constructors might experience over-confidence in face construction, prompting them to adjust the face beyond optimal representation (leading to a closer match with another identity). In either case, follow up research could be of value—although, use of the trait–recall mnemonic eliminates this outcome, irrespective of whether a composite is created after a short or long post-encoding delay and whether the composite face is viewed front-on or side-on.

4 | Conclusions

In sum, our data indicate that bringing forward the process of constructing a feature-based composite to the same day as the witnessed event yields good results when adding the trait–recall mnemonic to the face-recall interview (to give the H-CoI) as well as when asking potential recognisers to view the face from the side. Here, compared with standard police procedures (i.e., a CoI, a 2-day post-encoding delay, and front-on view at naming), participants were twice as likely to correctly name composites when an H-CoI was involved (cf. CoI) for composites created 3–4 h after encoding and after 3–4 h (cf. 2 days) following an H-CoI. This indicates that use of a conjunction of techniques can have substantive positive impact for policing (Frowd et al. 2015; Morris and Fritz 2013). There will of course be circumstances when face construction on the same day is not appropriate or feasible. For example, in cases where witnesses have experienced trauma, they may not be immediately ready to engage with the process of building a composite face (e.g., Frowd, Carson, Ness, McQuiston, et al. 2005). Nevertheless, when appropriate, it is clear from the current work that there is worthwhile forensic benefit to undertaking composite construction earlier in an investigation, ideally on the same day as the crime. In these cases, best practice involves both use of the H-CoI prior to the witness constructing a composite, and a prompt to view the face side-on when showing the resulting composite to potential recognisers.

Author Contributions

The presented experiment was conceptualized by the first and last author, with the majority of data collection carried out by the second author. These three authors had primary responsibility for drafting and re-drafting the manuscript. All authors assessed the paper, each providing substantial, constructive feedback on all aspects of the work presented here.

Acknowledgments

We would like to thank Kate Herold for her assistance with aspects of data collection.

Ethics Statement

All experimental work received ethical approval from the Ethics Committee at the University of Leeds and the research was conducted in accordance with the ethical code of the British Psychological Society.

Consent

Informed consent to participate was provided by all participants.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The dataset supporting the conclusions of this article is available in the UK Data Service repository, http://doi.org/10.5255/UKDA-SN-850883.

Endnotes

[1] We considered manipulating the order of this factor across participants. This would mean that, ideally for half of the time, a potentially more effective representation would be presented first (side-on), then a less effective one (front-on). As there is no practical advantage for doing this, as indicated elsewhere (Brown et al. 2019), and to avoid a large increase in sample size, levels for this factor were always presented in the same order: front-on and then side-on.

[2] To ensure that correct naming responses were not specifically tied to a few highly identifiable composites, we examined the distribution of correct names. Seventy-four percent of the composites were named correctly by at least one person, and so the majority of items constructed were identifiable to some extent. Identifiable composites were distributed across conditions: each of our eight conditions contained between two and seven such composites.

[3] Otherwise, the by-participants analysis tends to reveal differences that are smaller than the planned medium-size effect (i.e., indicating differences that are not reflected in the by-items analysis).

[4] The odds ratio is expressed throughout this paper as a value greater than 1. This standardisation is advised for convenience of interpretation (Osborne, 2017). It is achieved by taking the absolute value of B for the exponential function, as indicated by vertical bars around B (i.e., the result is always a postive number).

[5] This situation has arisen since the size of the effect is beyond statistical power for the experiment, and also that smaller models are inherently less powerful than larger models (e.g., Reed and Wu 2013). Here, the regression model is more powerful with three rather than with one or two three-way interactions. Thus, Delay × Encoding × View was removed from the analysis, and a combined model considered comprising all two-way interactions.

References

Beers, S. R., and M. D. De Bellis. 2002. "Neuropsychological function in children with maltreatment-related posttraumatic stress disorder" *The American Journal of Psychiatry* 159: 483–486. doi: https://10.1176/appi.ajp.159.3.483.

Brown, C., E. Portch, L. Nelson, and C. D. Frowd. 2020. "Reevaluating the Role of Verbalization of Faces for Composite Production: Descriptions of Offenders Matter!" *Journal of Experimental Psychology: Applied* 26: 248–265.

Brown, C., E. Portch, F. C. Skelton, et al. 2019. "The Impact of External Facial Features on the Construction of Facial Composites." *Ergonomics* 62: 575–592.

Chance, J., A. G. Goldstein, and L. McBride. 1975. "Differential Experience and Recognition Memory for Faces." *Journal of Social Psychology* 97, no. 2: 243–253.

Clark, H. H. 1973. "The Language-As-Fixed-Effect Fallacy: A Critique of Language Statistics in Psychological Research." *Journal of Verbal Learning and Verbal Behavior* 12: 335–359.

Dando, C. J., R. Wilcock, and R. Milne. 2009. "The Cognitive Interview: The Efficacy of a Modified Mental Reinstatement of Context Procedure for Frontline Police Investigators." *Applied Cognitive Psychology* 15: 679–696.

Davies, G. M., A. Milne, and J. W. Shepherd. 1983. "Searching for Operator Skills in Face Composite Reproduction." *Journal of Police Science and Administration* 11, no. 4: 405–409.

Deffenbacher, K. A., B. H. Bornstein, E. K. McGorty, and S. D. Penrod. 2008. "Forgetting the Once-Seen Face: Estimating the Strength of an Eyewitness Memory Representation." *Journal of Experimental Psychology: Applied* 14, no. 2: 139–150.

Ebbinghaus, H. 2013. "Memory: A Contribution to Experimental Psychology." *Annals of Neurosciences* 20, no. 4: 155–156. https://doi.org/10.5214/ans.0972.7531.200408 (Originally published, 1885).

Ellis, H. D., J. W. Shepherd, and G. M. Davies. 1979. "Identification of Familiar and Unfamiliar Faces From Internal and External Features: Some Implications for Theories of Face Recognition." *Perception* 8, no. 4: 431–439.

Ellis, H. D., J. W. Shepherd, and G. M. Davies. 1980. "The Deterioration of Verbal Descriptions of Faces Over Different Delay Intervals." *Journal of Police Science and Administration* 8: 101–106.

Erickson, W. B., C. Brown, E. Portch, et al. 2022. "The Impact of Weapons and Unusual Objects on the Construction of Facial Composites." *Psychology, Crime & Law* 30, no. 3: 207–228.

Field, A. 2018. *Discovering Statistics Using SPSS*. 5th ed. Sage.

Fisher, R. P., R. E. Geiselman, D. S. Raymond, L. M. Jurkevich, and M. L. Warhaftig. 1987. "Enhancing Enhanced Eyewitness Memory: Refining the Cognitive Interview." *Journal of Police Science and Administration* 15: 291–297.

Frowd, C. D. 2021. "Forensic Facial Composites." In *Methods, Measures, and Theories in Forensic Facial-Recognition*, edited by M. Toglia, A. Smith, and J. M. Lampinen, 34–64. Taylor and Francis.

Frowd, C. D., V. Bruce, H. Ness, et al. 2007. "Parallel Approaches to Composite Production." *Ergonomics* 50: 562–585.

Frowd, C. D., V. Bruce, D. Ross, A. McIntyre, and P. J. B. Hancock. 2007. "An Application of Caricature: How to Improve the Recognition of Facial Composites." *Visual Cognition* 15: 1–31.

Frowd, C. D., V. Bruce, A. Smith, and P. J. B. Hancock. 2008. "Improving the Quality of Facial Composites Using a Holistic Cognitive Interview." *Journal of Experimental Psychology: Applied* 14: 276–287.

Frowd, C. D., D. Carson, H. Ness, et al. 2005. "Contemporary Composite Techniques: The Impact of a Forensically-Relevant Target Delay." *Legal and Criminological Psychology* 10: 63–81.

Frowd, C. D., D. Carson, H. Ness, et al. 2005. "A Forensically Valid Comparison of Facial Composite Systems." *Psychology, Crime & Law* 11: 33–52.

Frowd, C. D., W. B. Erickson, J. M. Lampinen, F. C. Skelton, A. H. McIntyre, and P. J. B. Hancock. 2015. "A Decade of Evolving Composites: Regression- and Meta-Analysis." *Journal of Forensic Practice* 17, no. 4: 319–334.

Frowd, C. D., and S. Fields. 2011. "Verbalisation Effects in Facial Composite Production." *Psychology, Crime & Law* 17: 731–744.

Frowd, C. D., S. Jones, C. Fodarella, et al. 2014. "Configural and Featural Information in Facial-Composite Images." *Science & Justice* 54, no. 3: 215–227.

Frowd, C. D., D. McQuiston-Surrett, S. Anandaciva, C. E. Ireland, and P. J. B. Hancock. 2007. "An Evaluation of US Systems for Facial Composite Production." *Ergonomics* 50: 1987–1998.

Frowd, C. D., L. Nelson, F. C. Skelton, et al. 2012. "Interviewing Techniques for Darwinian Facial Composite Systems." *Applied Cognitive Psychology* 26, no. 4: 576–584.

Frowd, C. D., F. Skelton, G. Hepton, et al. 2013. "Whole-Face Procedures for Recovering Facial Images From Memory." *Science & Justice* 53: 89–97.

Geiselman, R. E., R. P. Fisher, D. P. MacKinnon, and H. L. Holland. 1985. "Eyewitness Memory Enhancement in the Police Interview: Cognitive Retrieval Mnemonics Versus Hypnosis." *Journal of Applied Psychology* 70: 401–412.

Geiselman, R. E., R. P. Fisher, D. P. MacKinnon, and H. L. Holland. 1986. "Eyewitness Memory Enhancement With the Cognitive Interview." *American Journal of Psychology* 99: 385–401.

Hancock, P. J. B., K. Burke, and C. D. Frowd. 2011. "Testing Facial Composite Construction Under Witness Stress." *International Journal of Bio-Science and Bio-Technology* 3: 65–71.

Harrell, F. E. 2015. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. 2nd ed. Springer.

Hole, G. J., P. A. George, K. Eaves, and A. Rasek. 2002. "Effects of Geometric Distortions on Face-Recognition Performance." *Perception* 31, no. 10: 1221–1240.

Huber, P. J. 1967. "The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions." In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 221–233. University of California Press.

Laughery, K. R., P. K. Fessler, D. R. Lenorovitz, and D. A. Yoblick. 1974. "Time Delay and Similarity Effects in Facial Recognition." *Journal of Applied Psychology* 59, no. 4: 490–496.

Lewis, M. B. 2023. "Fixing the Stimulus-As-a-Fixed-Effect Fallacy in Forensically Valid Face-Composite Research." *Journal of Applied Research in Memory and Cognition* 13, no. 2: 306–314.

Martin, A. J., P. J. B. Hancock, and C. D. Frowd. 2017. "Breath, Relax and Remember: An Investigation Into How Focused Breathing Can Improve Identification of EvoFIT Facial Composites." In *Proceedings of IEEE 2017 Seventh International Conference on Emerging Security Technologies, 4th–8th September*, edited by G. Howells et al. University of Kent.

Milne, R., and R. Bull. 1999. *Investigative Interviewing: Psychology and Practice*. John Wiley & Sons, Ltd.

Morris, P. E., and C. O. Fritz. 2013. "Effect Sizes in Memory Research." *Memory* 21, no. 7: 832–842.

Olsson, N., and P. Juslin. 1999. "Can Self-Reported Encoding Strategy and Recognition Skill Be Diagnostic of Performance in Eyewitness Identifications?" *Journal of Applied Psychology* 84, no. 1: 42–49.

Osborne, J. W. 2017. "Regression & Linear Modeling: Best Practices and Modern Methods. Simple Linear Models With Categorical Dependent Variables: Binary Logistic Regression. Ch. 5." https://dx.doi.org/10.4135/9781071802724.

Peterson, M. A., and G. Rhodes. 2003. *The Perception of Faces, Objects, and Scenes: Analytic and Holistic Processes*. Oxford University Press.

Portch, E., C. Brown, C. Fodarella, et al. under revision. "The Impact of Forensic Delay: Facilitating Facial Composite Construction Using an Early-Recall Retrieval Technique." *Ergonomics*.

Portch, E., K. Logan, and C. D. Frowd. 2017. "Interviewing and Visualisation Techniques: Attempting to Further Improve EvoFIT Facial Composites." In *Proceedings of the 2017 Seventh International Conference on Emerging Security Technologies (EST)*, 97–102. Institute of Electrical and Electronics Engineers.

Reed, P., and Y. Wu. 2013. "Logistic Regression for Risk Factor Modelling in Stuttering Research." *Journal of Fluency Disorders* 38: 88–101.

Shapiro, P. N., and S. D. Penrod. 1986. "Meta-Analysis of Facial Identification Rates." *Psychological Bulletin* 100: 139–156.

Shepherd, J. W., and H. D. Ellis. 1973. "The Effect of Attractiveness on Recognition Memory for Faces." *American Journal of Psychology* 86, no. 3: 627–633.

Skelton, F. C., C. D. Frowd, P. J. B. Hancock, et al. 2020. "Constructing Identifiable Composite Faces: The Importance of Cognitive Alignment

of Interview and Construction Procedure." *Journal of Experimental Psychology: Applied* 26: 507–521.

Skelton, F. C., C. D. Frowd, and K. E. Speers. 2015. "The Benefit of Context for Facial-Composite Construction." *Journal of Forensic Practice* 17, no. 4: 281–290. https://doi.org/10.1108/JFP-08-2014-0022.

Sporer, S. L. 2007. "Person Descriptions as Retrieval Cues: Do They Really Help?" *Psychology, Crime & Law* 13, no. 6: 591–609.

Sporer, S. L., and N. Martschuk. 2014. "The Reliability of Eyewitness Identifications by the Elderly: An Evidence-Based Review." In *The Elderly Eyewitness in Court*, edited by M. P. Toglia, D. F. Ross, J. Pozzulo, and E. Pica, 3–37. Psychology Press.

Tanaka, J. W., and M. J. Farah. 1993. "Parts and Wholes in Face Recognition." *Quarterly Journal of Experimental Psychology: Human Experimental Psychology* 46A: 225–245.

Tredoux, C. G., C. D. Frowd, A. Vredeveldt, and K. Scott. 2023. "Construction of Facial Composites From Eyewitness Memory." In *Biomedical Visualisation: Advances in Experimental Medicine and Biology*, edited by L. Shapiro and P. M. Rea, vol. 1392. Springer.

Wells, G. L., and B. Hryciw. 1984. "Memory for Faces: Encoding and Retrieval Operations." *Memory & Cognition* 12, no. 4: 338–344.

Appendix A

Statistical Power Analysis

We assessed statistical power for the proposed design using computer simulation. This method simulates participant naming responses and assesses the frequency that the manipulated factors achieve statistical significance when repeated (i.e., to indicate statistical power). As power depends on whether predictors are between- or within-subjects, we assessed the effect of these variables separately. First, we considered a single model containing the three between-subjects predictors, encoding, delay and interview. This approach was preferred (for reasons of statistical power, as mentioned in the Results for Correct naming) over computation of three separate models. We then included the fourth predictor, view of composite, within-subjects.

Baseline performance was defined as intentional (cf. incidental) encoding of the target, a short (3–4 h) delay from encoding to interview and construction, use of a CoI and front-on view for naming. Several studies suggest that composites created using the PRO-fit system are named with a mean of 18% correct when presented front-on (e.g., Frowd et al. 2008; Frowd, Carson, Ness, Richardson, et al. 2005), baseline performance that we copied. Based on a medium effect—specifically, a mean Exp(*B*) of 2.5—we followed previous research that suggested an *increase* in correct naming following the between-subjects predictor H-CoI (e.g., Frowd et al. 2008), but a *decrease* for (i) incidental encoding (e.g., Frowd, Bruce, Ness, et al. 2007) and (ii) a long (2-day) delay (e.g., Frowd et al. 2015). Settings for the GEE were as specified in the Results (e.g., use of a Robust Covariance Matrix).

With reference to Equation (1), to achieve a baseline performance of 18%, the models' intercept (B_0) was drawn randomly from a Normal distribution centred on −1.52, with SD = 0.1 specified to provide variability in the range 15%–21% correct (i.e., for 95% of observations). Values of Beta for the three between-subjects predictors (B_1–B_3) were also drawn from a random Normal distribution; these were centred on an absolute value of 0.92, to give mean Exp(|*B*|) = 2.5, with SD = 0.1 to give sensible variability of Exp(|*B*|) in range 2.0–3.0. For the fourth predictor, view, we modelled this within-subjects variable x_4 (based on Skelton et al. 2020) to give consistent naming responses for a second presentation of composites to participants, except that correct responses per participant were set to randomly increase at a probability of 0.08, but previously correct responses to decrease at a probability of 0.01. Residual errors (e_{ij}) were added to each participant response, again using a random Normal distribution (M = 0.0), with SD = 0.5 to give suitably variable responses (e.g., at baseline, MD changed between −10% and +20%). Finally, we modelled the usual situation where the target identities (facial photographs) were sometimes not correctly named (typically 1 in 20), since their associated cases are removed prior to analyses, increasing SE(*B*) and impacting statistical power. As such, 5% of cases were selected by chance to be an unfamiliar target identity, and then responses to composites were processed in this way. Included in the simulation were three random effects: stimulus items (coded 1–12), and participants who (i) constructed composites (1–96) and (ii) named composites (1–48).

Model for each Predictor in the linear Regression Equation:

$$y_{ij} = B_0 + (x_1 \times B_1) + (x_2 \times B_2) + (x_3 \times B_3) + (x_4 \times B_4) + e_{ij} \quad (1)$$

where Predictor x_1 = Interview (*0* = CoI, *1* = H-CoI), x_2 = Delay (*0* = 3–4 h, *1* = 2 days), x_3 = Encoding (*0* = Intentional, *1* = Incidental) and x_4 = View (*0* = Front-on, *1* = Side-on). B_0 is the model's intercept. Values for B_1 and B_4 were modelled as positive values (to give an increase in *y*), while B_2 and B_3 were negative (to give a decrease in *y*). The term e_{ij} represents residual error. For the analysis of nominal responses, the equation was subject to the Sigmoidal (logistic) function, $Y_{ij} = \text{Exp}(y_{ij})/(1 + \text{Exp}(y_{ij}))$.

A total of 100 repetitions were conducted in SPSS using Generalised Estimating Equations (GEE) for the proposed sample size. The three between-subjects predictors were significant in the by-participants and by-items analyses ($p < 0.05$, SE(*B*) in the range.23–.31) for 90–97 of the 100 repetitions. This indicates suitable statistical power (i.e., power ≥ 90%). The fourth predictor, view, within-subjects, was then included. For these simulations, the four predictors in a combined model (for both types of analysis) were significant ($p < 0.05$, with SE(*B*) for x_4 in range 0.06–0.10) at or above 83% of occasions, again indicating suitable power.

Nevertheless, we acknowledge that un-estimated sources of variance may make higher-order interactions harder to detect. However, we applied the above simulation procedure for an anticipated two-way interaction between interview (between-subjects) and view (within-subjects), representing a small-to-medium benefit of view (Exp(*B*) of 1.9 from Skelton et al. 2020) for side-on (cf. front-on) presentation of composites constructed following an H-CoI. We modelled this situation by removing the benefit of front- to side-on naming for the CoI. This interaction effect ($p < 0.05$, with SE(*B*) in range.06–.11) was observed by-participants and by-items for 99% of cases, again indicating suitable statistical power.