

Central Lancashire Online Knowledge (CLoK)

Title	Constructing identifiable composite faces: the importance of cognitive alignment of interview and construction procedure
Type	Article
URL	https://clock.uclan.ac.uk/id/eprint/29932/
DOI	https://doi.org/10.1037/xap0000257
Date	2019
Citation	Skelton, Faye Collette, Hancock, Peter J.B., Jones, Helen orcid iconORCID: 0000-0002-2716-051X, Battersby, Kirsty, Fodarella, Cristina, Logan, Karen, Jones, Ben and Frowd, Charlie (2019) Constructing identifiable composite faces: the importance of cognitive alignment of interview and construction procedure. <i>Journal of Experimental Psychology: Applied</i> , 26 (3). pp. 507-521. ISSN 1076-898X
Creators	Skelton, Faye Collette, Hancock, Peter J.B., Jones, Helen, Battersby, Kirsty, Fodarella, Cristina, Logan, Karen, Jones, Ben and Frowd, Charlie

It is advisable to refer to the publisher's version if you intend to cite from the work.
<https://doi.org/10.1037/xap0000257>

For information about Research at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clock.uclan.ac.uk/policies/>

Abstract

We investigated the impact of congruency between the witness interview and method used to construct a composite face. Experiment 1, using a typical feature-by-feature composite method, revealed that aligning cognitive processes during interview and face construction enhanced the effectiveness of composites compared with composites produced following unaligned (incongruent) procedures. Experiment 2 revealed that incorporating character judgements in the witness interview substantially enhanced identification of feature-based composites when constructing the central (internal) features first, suggesting that such judgements focus attention on this region of the face. Experiment 3 explored alignment of processes using an approach based on an evolutionary algorithm, a method requiring witnesses to create a composite by selecting from arrays based on the eye-region. A combination of character judgements, first for the whole face and then for the eye region, led to best-identified composites. Overall, results indicate that more effective composites are produced when both interview and construction procedures are aligned cognitively. Results are discussed with relevance to the theory of transfer-appropriate processing (Morris, Bransford, & Franks, 1977).

Public Significance statement: This experimental study reveals that facial composites are much more effective as an identification aid if the witness interview used to elicit a description of the culprit draws on the same cognitive processes as the method used to construct the composite face. Findings are valuable to developers of facial composite systems, and are also relevant to police and forensic practitioners, all of whom should ensure that methods of face production are compatible with interviewing procedures used with witnesses and victims of crime.

Keywords: Facial composites, interview, police, witnesses;

Constructing identifiable composite faces: the importance of cognitive alignment of interview and construction procedure

Witnesses and victims of crime have an important role to play in criminal investigations. They are interviewed to gather information about the event they have witnessed and in addition may be asked to take part in a police line-up (identity parade) in order to identify the perpetrator. In cases where there is no suspect, a witness (who may also be a victim) may also be interviewed to provide a free-recall description of the perpetrator's face and construct a facial composite, a pictorial likeness, from memory. Composite images are distributed within police circles and via the media to generate lines of enquiry.

In the past, composites were created using manual systems (e.g. Photofit) with features printed onto jigsaw-style pieces or transparencies. A witness would work with a trained police officer, selecting from catalogues of features (eyes, nose, mouth etc.) to create a likeness of the face. Nowadays, specialist software (such as PRO-fit and E-FIT in the UK, and FACES and Identikit in the US) operates on the same principle of selecting individual features to build a likeness. In spite of differences in how composite systems operate, all require eyewitnesses to focus on individual facial features. These 'feature' composites, when constructed from memory either immediately or a few hours after viewing a face, are named correctly approximately 20% of the time; however, after a forensically-appropriate two-day delay, correct naming typically drops to around 5% (e.g., Frowd, Carson, Ness, et al., 2005; Frowd et al., 2010).

Recent years have seen substantial improvements to the effectiveness of facial composites using software based on evolutionary algorithms (e.g., EFIT-V, EvoFIT and ID; Davis, Thorniley, Gibson, & Solomon, 2016; Frowd et al., 2013; Tredoux, Nunez, Oxtoby, & Prag, 2006). Rather than selecting individual features, witnesses are presented with screens of whole faces (or whole-face regions) from which they choose based on resemblance to the target face.

Over the course of several screens, multiple faces are selected and then ‘bred’ together. Despite differences in software systems and exact procedures, these newer ‘holistic’ systems attempt to capitalise on recognition memory, which decays at a slower rate than face recall (Davies, 1983), and to be closely aligned with humans’ natural tendency to process faces as complete or holistic¹ entities rather than as a series of isolated features (e.g., Tanaka & Farah, 1993; Tanaka & Sengco, 1997). Indeed, while ineffective initially (Frowd, Hancock, & Carson, 2004), these systems now enable witnesses to construct a composite face after a 1-2 day delay that other people name at around 40% correct (e.g., Frowd, Erickson, Lampinen, Skelton, McIntyre, & Hancock, 2015) or, as discussed later, even better when used in conjunction with other techniques (Frowd et al., 2013).

As mentioned above, recovering an unfamiliar face from memory usually involves a witness reflecting on the appearance of this face, and is a fundamental component of composite construction. This procedure, itself a memory task, precedes composite construction, and we argue that retrieval during interview serves to re-encode a target face. Indeed, the ‘testing effect’ (Roediger & Karpicke, 2006) is a phenomenon whereby retrieval may enhance subsequent memory for recalled information; the act of recall modifies items such that they are much better remembered in subsequent tests compared with items not recalled (Brewer, Marsh, Meeks, Clark-foos, & Hicks, 2010). Thus, the interview itself should impact upon face construction.

One of the best-known methods for information recall is the Cognitive Interview (CI; see e.g. Geiselman, Fisher, MacKinnon, & Holland, 1986). The technique is frequently used with

¹ We use the term ‘holistic’ to mean a representation containing information on facial features and their interrelationships: both featural and configural, both of which can be accessed. This is in contrast to Searcy and Bartlett (1996) who used this term to mean a ‘gestalt’, whereby individual components could not be accessed. Both featural and configural information are important to unfamiliar face recognition, and although adults typically struggle to recognise isolated face parts (Tanaka & Farah, 1993), we do seem to access them in memory (Cabeza & Kato, 2000; Towler & Eimer, 2016).

witnesses and victims to obtain an account of events and a description of the perpetrator's face. The CI has been adapted for composite construction and the recommended procedure (e.g., Fodarella, Kuivaniemi-Smith, Gawrylowicz, & Frowd, 2015) involves a witness mentally reinstating the context in which the face was seen, followed by free and uninterrupted recall of the face. An interviewer would then read aloud the description provided for each facial feature and prompt for further information (cued recall). The CI thus requires detailed recall of individual features and, as such, does not appear to be well-aligned with holistic face construction (see Memon, Meissner, & Fraser, 2010 for a review).

Research has also demonstrated that whole-face attribution conducted after initial face recall can enable a person to create a more identifiable composite. Following a CI, a witness is invited to reflect silently on the perceived character of the target face, and then to judge the face as low, medium or high on characteristics such as health, masculinity and extraversion (outgoing). Frowd, Bruce, Smith, and Hancock (2008) revealed that accurate naming of composites was four times higher using this Holistic Cognitive Interview (H-CI) compared with a face-recall CI. An additional opportunity for recall is one possible reason for this advantage. However, evidence suggests that simply offering a second recall attempt does not in itself enhance composites: Skelton, Frowd, and Andrews (2011) compared a single face-recall CI with interviews prompting recall of relational information (i.e., distance between features) prior or subsequent to recall of featural information. The relational interview had a positive impact only when used after featural recall; when used before, accuracy was equivalent to the CI alone. The H-CI advantage is therefore unlikely to be due to an additional recall attempt.

Alternatively, attribution of character may focus attention on the internal, expressive facial features (including eyes, nose and mouth) that are more relevant for such judgements than are external features (including face shape and hair). Eye-tracking data support this hypothesis: Kanan, Bsesio, Ray, Hsaio, and Cottrell (2015) tracked participants' scan patterns whilst

engaged in trait judgements for complete faces. When individuals' scan-paths were combined, a 'T' shaped pattern emerged incorporating eyes, nose and mouth. This suggests that the central portion of the face is important when making character judgements, and therefore that considering the personality of a previously-seen face focuses attention on internal-features relative to external-features. It is also known that low composite identification rates may arise in part from poorly-constructed internal-features, which are potentially more important to identification than external-features (e.g., Frowd, Skelton, Butt, Hassan, & Fields, 2012), presumably due to the internal-features' relative stability over time (Ellis, Shepherd, & Davies, 1979; Young, Hay, McWeeny, Flude, & Ellis, 1985). Internal-features of a composite must therefore be a good match to those in a target face; however, internal-features are typically constructed less accurately than external-features (particularly hair; Frowd, Bruce, McIntyre, & Hancock, 2007). Finally, presenting witnesses with only internal-features in face arrays for composite construction, with external-features added at the end of the procedure, greatly facilitates ensuing identification of EvoFIT composites (e.g., Frowd, Skelton, Atherton et al., 2012). In contrast to the CI approach requiring only recall of facial features—that is, with no particular emphasis on features that might be useful for recognition—there is therefore evidence that focusing witnesses on internal features during composite construction results in this region being more accurately constructed, and that this change contributes to a naming advantage for complete composites.

Another possible explanation for the H-CI advantage is that character attribution rehearses information on how individual features interrelate. This process may strengthen a holistic face representation, ultimately aiding recognition for the whole-face. In turn, it may help a witness to assess when a good visual likeness has been achieved during composite construction (Frowd, Skelton, Atherton et al., 2012). It is also well-aligned with the process involved in newer, holistic systems (See Frowd et al., 2015). These two explanations are not mutually exclusive,

and thus holistic judgements could improve memory for internal-features as well as improving the whole-face representation by strengthening memory for configural information. Recent evidence also provides support: judgements of human characteristics such as masculinity are disrupted by facial inversion, suggesting that they rely to a large extent on configural processing (Wilson, Young, Rule, & Hugenberg, 2018). Character judgements may therefore serve to rehearse and strengthen both featural and configural information.

The advantage of the H-CI might even be explained by deeper *re-encoding* of a target face. Established research reveals that personality attribution enhances subsequent facial memory (e.g. Bower & Karlin, 1974; Shapiro & Penrod, 1986; Wells & Hryciw, 1984), suggesting a degree of re-encoding has taken place. Some work (e.g., Bloom & Mudd, 1991; Bower & Karlin, 1974) favours a ‘feature quantity’ account, as better performance arises due to scanning of more facial features during encoding. However, Wells and Hryciw (1984) found that judgements about facial features facilitated performance only on a feature-based memory task (face construction using a feature system), whereas trait-judgements were more effective for a holistic (recognition) retrieval task. If the trait-encoding advantage was due to a greater number of features being encoded, then trait-encoding would have shown an overall benefit across task types. As it was not, an account based on re-encoding seems more likely.

Results are thus consistent with an *encoding-specificity* explanation (Tulving & Thomson, 1973): success at retrieval depends on the extent to which retrieval operations match those at encoding. Indeed, Bornstein, Deffenbacher, and Penrod’s (2012) meta-analysis revealed a face-recognition advantage when trait judgements were used at encoding. They argue that trait judgements, like recognition, involve holistic encoding, and that alignment of cognitive operations facilitates performance. Of relevance here is transfer-appropriate processing (TAP; Morris, Bransford & Franks, 1977). This theory proposes that task performance benefits when there is overlap of mental operations required at encoding and those required at retrieval. If

one considers the interview as re-encoding, a composite should then be constructed more accurately when processes at interview are aligned with processes at construction.

We assessed the relevance of the above arguments in three experiments. Experiment 1 investigated two of these possibilities. Using a typical feature-based composite system, PRO-fit, we examined whether composite effectiveness was enhanced when interview-to-construction processes were aligned versus unaligned. We also investigated whether focusing attention on the internal-features during construction promotes more effective composites than those constructed with the whole-face visible throughout.

In more detail, participants undertook a regular face-recall CI (free-order recall), which acted as a control condition, or a novel two-stage interview whereby internal-features were recalled prior to recall of external-features (internals-first recall). Within these groups, half constructed a composite in the usual way with the whole-face visible throughout (full-face construction), while half constructed the composite in a novel way: first for internal-features with external-features masked prior to creating external-features with internal-features masked ('internals-first' construction).

If process alignment is important, then the CI should work best with full-face construction, and internals-first recall should work best with internals-first construction; also, we expect these congruent conditions to outperform incongruent conditions. If focus on internal-features contributes more to the benefit, then the congruent internals-first condition should lead to more accurate composites than the congruent full-face condition.

All three experiments follow the 'gold standard' procedure for facial-composite research (see Frowd, 2015): target faces were selected to be unfamiliar to participant-witnesses constructing the face, and a forensically-relevant delay introduced prior to interview. Participants who later named the composites were recruited on the basis of being familiar with the target identities.

As described in the ensuing experiments, the design, number of participants, items, procedure and method of analysis involved was such that each experiment would have sufficient power to be able to detect at least a medium-to-large effect size, should one exist. The approach would appear to be valuable, to give results that should have real-world importance.

EXPERIMENT 1

Two stages were required, described below, each involving a different group of participants.

These stages are composite construction (Stage 1) and composite naming (Stage 2).

Stage 1: Composite construction

Design

A between-participants 2 (face recall: full-face vs. internal-features first) x 2 (construction type: full-face vs. internal-features first) design was used.

Participants

An opportunity sample of 32 students from the University of Winchester volunteered based on being unfamiliar with the BBC TV programme *EastEnders*. Fifteen females and 17 males took part, with an age range from 18 to 26 ($M = 20.3$, $SD = 2.3$) years.

Materials

Eight colour photographs (8 cm x 10 cm) of actors from *EastEnders* were sourced from the Internet. At the time of testing, the BBC website listed 47 characters in the soap. Characters who were not white Caucasian were omitted, to avoid identities that might otherwise be distinctive, and this yielded a potential target pool size of approximately 40. The four male and four female target images depicted the face in a front view, without glasses, and with a

neutral expression; male targets were clean-shaven. The experimenter used a verbal-description sheet to note the description of the target recalled by participants. This includes sections for each individual feature (hair, eyes, nose, mouth, ears and eyebrows), and is similar to the format used by police practitioners. PRO-fit software version 3.5 was used to construct the composites. Each target was constructed four times, once in each condition.

Procedure

Ethical approval was granted by the appropriate university committee. After providing informed consent, participants were randomly assigned, with equal sampling, to the four cells of the design. Participants viewed a photograph of an unfamiliar face for one minute in the knowledge that they would return the following day to construct a composite. Participants were asked if they recognised their target face; there were no reports of such familiarity.

After a forensically-relevant delay of 22 to 28 hours, each participant returned to the laboratory to construct a composite of the face they had seen the previous day. They worked with a single trained experimenter to recall the face and construct a composite. We note that the experimenter did not see any of the target images previously presented to participants, and so was 'blind' to target identity under construction. All sessions started with rapport-building to put participants at ease, with the aim of facilitating face recall. Participants were given an overview of the general method used to construct a face using PRO-fit. The procedure varied according to assigned condition, as illustrated in Figure 1. See Fodarella et al. (2015) for a detailed description of the usual PRO-fit procedure.

Participants assigned to full-face recall underwent a CI. During the CI, they were asked to visualise the face, and then freely recall as much information about the face as possible, without guessing, taking as much time as needed. During this process the experimenter made written notes on a verbal-description sheet and did not interrupt. The description was then read back,

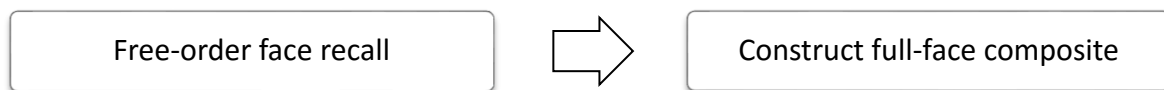
feature-by-feature, and the experimenter noted any further information that the participant could remember (cued recall). As order of participant-witness recall was not central to this research, we did not record sessions. Next, the experimenter entered the information recalled into PRO-fit to identify around 20 examples per face feature.

For full-face construction (Conditions A and B), all facial features were visible throughout and participants were able to focus on features in any order. Features were viewed in-context; for example, a pair of eyes was seen in the context of all other features in the face. Participants were asked which facial feature they would like to work on first. Examples were duly shown for that feature, with the participant prompted to select the example that best matched memory, and then adjust its size and the distance between features, as necessary. This procedure was repeated for each feature in the face. Participants were encouraged to create the best likeness possible in their own time.

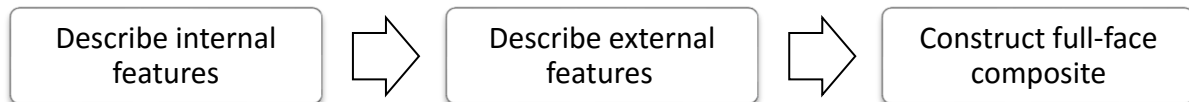
For participants assigned to construct internal-features first following full-face recall (Condition C), the procedure was the same except that they selected and sized features for the internal-features region first (with external-features masked), then external-features (with internal-features masked) and finally made any additional adjustments with the whole-face in view.

Those assigned to recall and construct internal-features first (Condition D) took part in a two-stage interview. This followed the same basic procedure as above, except that they recalled only internal features to begin with, and then constructed a composite of these features. Following this, external features were recalled and subsequently constructed. As is normal for face construction, testing sessions lasted for between 30 and 60 minutes per person. The resulting composite face was saved to disk, and participants were debriefed.

A. Full-face recall – Full-face construction



B. Internals-first recall – Full-face construction



C. Full-face recall – Internals-first construction



D. Internals-first recall – Internals-first construction

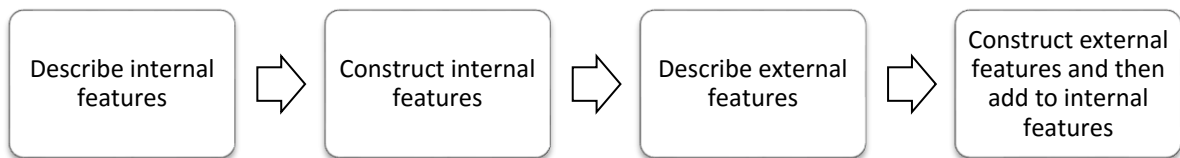


Figure 1. Process diagram illustrating the procedure for conditions in Experiment 1.

Stage 2: Composite naming

Design

A 2 (interview: full-face vs. internal-features first) x 2 (construction type: full-face vs. internal-features first) between-subjects design was used. Participants named composites constructed in Stage 1 from one of the four experimental conditions.

Participants

Nineteen female and 5 male students from the University of Central Lancashire (UCLan) participated for course credit. Their age ranged from 19 to 24 ($M = 20.5$, $SD = 1.4$) years. They were recruited based on being fans of EastEnders, to be familiar with the target identities.

Materials

The 32 composites constructed in Stage 1 and the eight colour target photographs (all printed as before) were used for the naming task. One male and one female ‘foil’ composite were included in each set of composites presented to participants, to limit guessing and to make the task more realistic. The experimenter created these additional composites of unfamiliar targets, which were not part of the target set, using the normal procedure for the white male and white female databases. These unfamiliar faces appeared within the same age range as the EastEnders’ characters and had plausible hairstyles. Each of the four testing booklets contained composites from a single experimental condition, plus the same two foils.

Procedure

Participants were tested individually and the task was self-paced. They were randomly assigned, with equal sampling, to one of the four testing booklets. Participants were told that they would see composites, some of which had been constructed of male and female characters from EastEnders, and that they should try to name as many as possible; they were also told that ‘don’t know’ responses were acceptable, and we accepted as correct character or actor names. The assigned set of composites were presented sequentially, along with the same two foil composites, and participants provided a name or not for each. Next, to check that participants were familiar with the relevant identities, the target photographs were presented likewise for

them to name. Composites and target photographs were presented in a different random order for each person. The naming task took about 10 minutes to complete including debriefing.

Results

Correct naming

Responses to composites were checked for missing data, of which no cases were detected, and then scored for accuracy: a value of 1 was assigned if a correct name had been given, and 0 otherwise. Targets were named correctly 174 times out of a possible 192 ($M = 90.6\%$), indicating that participants appropriately had high familiarity with the target set. For any target named incorrectly, the response to the associated composite was coded as missing data, as such responses could not have been correct. Taking these 18 missing cases into account, there were 49 correct responses in total to composites out of a possible 174 ($M = 28.2\%$). The resulting correct naming scores for composites are summarised in Table 1.

Table 1. Correct naming for face recall and face construction type

<i>Face construction</i>	<i>Face recall</i>	
	Complete face recalled	Internal-features recalled first
Full-face	42.9 (18 / 42)	13.6 (6 / 44)
Internal-features constructed first	16.3 (7 / 43)	40.0 (18 / 45)

Note. Figures are percentage-correct calculated from responses in parentheses: summed

correct responses (numerator) and total (correct and incorrect) responses (denominator). The analysis revealed a significant interaction between face recall and construction type: all pairwise comparisons are significant, $p < .05$. See text for more details.

Correct naming scores were subjected to Logistic Regression, a frequency-type analysis that is appropriate for this between-subjects design and dichotomous (0 and 1) responses. Cell frequencies and model variables were checked for appropriateness. A full-factorial (saturated) model was built and estimators were successively removed (using the standard level for elimination, $p > .1$) based on the Likelihood Ratio: face construction was removed at Step 1 and face recall at Step 2. The final model, $X^2(1) = 15.5$, $p < .001$, $R^2(\text{Cox and Snell}) = .09$, $R^2(\text{Nagelkerke}) = .12$, thus contained a single estimator, face recall x construction, Gradient $B = 2.8$, $SE(B) = 0.7$, $p < .001$, Odds Ratio $Exp(B) = 16.2$, $CI(-) = 3.8$, $CI(+) = 69.2^2$. Fisher's Exact Tests (two tailed) indicated that all pairwise comparisons were reliable, $p < .05$. An additional test revealed that the two congruent recall and construction conditions were equivalent, $p = .83$.

Incorrect Naming

We next assessed responses to incorrectly-named composites, an analysis that is independent to correct (accurate) naming. This is a useful exercise as a composite that has a higher level of mistaken names, *per se*, indicates a representation that is more similar to another identity: a face that is less accurate. Analysis of mistaken names also allows an assessment of whether a manipulation led to increased inaccurate naming without a corresponding increase in correct naming, a result that would indicate a shift to a more liberal response criterion.

² We note that CIs are large, but such variability is expected using experimental stimuli (facial composites) that have inherent error due to being constructed from memory.

Composite responses were re-scored for instances where a participant had given an incorrect (mistaken) name. Again, participant responses were coded as missing data for incorrectly-named targets; this time, however, responses that had been previously scored as correct were also omitted (also coded as missing data), to allow an analysis of mistaken names relative to the number incorrect (i.e. sum of mistaken and no names; $M = 30.0\%$ overall). A logistic-regression analysis conducted as above revealed no reliable effects for either of the two estimates or their interaction (all $p > .1$).

Discussion

Results indicated that more effective composites emerged from congruent than incongruent conditions. Also, the lack of additional benefit from conditions that focused participant-witnesses on internal-features suggests that alignment of cognitive processes is more important for the effectiveness of composites than a focus on internal-features.

We acknowledge that the four conditions involved different change in or interruption to the procedure for face recall and construction. Three of the conditions—those involving internals-first-recall and internals-first construction—involved switching between recall and construction during the process, which witnesses would not ordinarily experience. It is not possible to tell whether this impacted upon the effectiveness of the composites, and so results should be interpreted with some caution. That said, participants in the condition involving interruption of both interview and construction (internal first recall and internal first construction – Condition D) performed equally well to those in the condition involving no interruption (full-face recall and construction – Condition A), suggesting that experimental interruptions do not have a negative impact in comparison to the usual procedure.

Experiment 1 involved the standard CI, involving free recall of the whole face, and a novel presentation (Conditions C and D) whereby internal- and external-features were constructed

separately. Recent evidence with the EvoFIT holistic method (Frowd, Nelson et al., 2012) demonstrates substantial improvement in composite effectiveness using an internal-features' approach after a H-CI relative to a CI. Experiment 2 investigates whether this effect would extend to a feature system. In a fully-crossed design, we compare type of interview (CI vs. H-CI) and type of PRO-fit construction (full-face vs. internals-first). The subsequent experiments also include likeness ratings of composites as an additional measure of performance.

EXPERIMENT 2

The H-CI requires witnesses to freely recall facial features, as in the CI, and then (i) to think silently about the character of the target for 60 seconds and (ii) to make a series of character judgements about the face. We argue that, in doing so, the H-CI procedure rehearses memory traces for individual facial features (as part of the CI) and additionally rehearses configural information (the HI part), strengthening the whole-face memory representation. We further suggest that making character judgements is likely to focus attention on the expressive internal-facial features.

Although Experiment 1 revealed that focus on internal-features by itself does not facilitate composite effectiveness, the deeper processing of internal-features required for character judgements (e.g., Frowd et al., 2013) may be beneficial. The H-CI is thus likely to promote a better memory for the face both via this deeper encoding and because judging character (the HI component) likely involves visualising multiple, particularly internal, features as well as their interrelationships (Kanan et al., 2018; Wilson et al., 2018). We expect that more effective composites will be created following the H-CI than the CI. If the H-CI does indeed focus attention on the internal-features then more effective likenesses should be produced when the H-CI is followed by internal-features first rather than full-face construction, due to the

alignment of cognitive processes and in accordance with TAP (Morris et al., 1977). As the CI is unlikely to result in more attention to internal-features than to external-features, we expect that this type of recall would align better with full-face construction than with internal-features first construction.

Methods to facilitate recognition of completed composites also exist (for a review of techniques, see Frowd, 2017). Frowd et al. (2013) invited observers to look at the page on which the composite was printed from one side to make the face appear long and thin. This procedure improved correct naming of composites compared to viewing the face front-on, a development that was originally adopted for use by the police (Astbury, 2011; Frowd et al., 2014). This effect may arise because facial composites, by their nature, are error-prone and thus some degree of flexibility in our perception must be present to allow us to identify people from these images. A similar effect appears to occur when attempting to recognise facial photographs (i.e., not composite faces) when presented in a vertically- or horizontally-stretched format (e.g., Hole, George, Eaves, & Rasek, 2002). So, presenting a face side-on may reduce the appearance of error in distances between features, to the benefit of face recognition (for discussion on this issue, see Frowd et al., 2013, 2014).

The H-CI, external-feature masking, and sideways-on naming have each demonstrated independent improvements to the recognition of feature and holistic composites. Combining these techniques has proved very effective with EvoFIT (Frowd et al., 2013), and so Experiment 2 investigated whether this finding would extend to feature-based composites. Although correct naming is likely to be lower with the PRO-fit feature system than with EvoFIT, if this reasoning is correct then all conditions should enjoy some benefit from side-on naming (as the appearance of error should be reduced, facilitating recognition). As this method of naming is a fairly new development, its inclusion here is largely exploratory; however, highest levels of naming were anticipated for the H-CI internals-first condition, where

composites should already be accurate and where further reduction of error may nudge a representation over the threshold for recognition.

Experiment 2 comprised three stages. Stage 1 and 2 were as before; for Stage 3, to provide a supplementary measure of performance, a separate group of participants were invited to assign ratings of likeness to the constructed composites.

Stage 1: Composite construction

Design

A 2 (interview: CI vs. H-CI) x 2 (construction type: full-face vs. internal-features first) between-subjects design was used. The impact of viewing angle was explored in Stage 2.

Participants

An opportunity sample of 32 undergraduates volunteered from UCLan based on being unfamiliar with rugby-league players (targets in the study). There were 20 females and 12 males, aged between 18 and 23 ($M = 19.0$, $SD = 1.1$) years who were randomly assigned with equal sampling to the four cells of the design. None had participated in Experiment 1.

Materials

Eight photographs (8cm x 10cm) of international-level rugby-league players were sourced from the Internet. Each photograph depicted the face front-facing and clean-shaven, without glasses, and with a neutral expression. Four copies of each target photograph were printed (as before). Verbal description sheets and PRO-fit software were used as in Experiment 1. Each target was constructed once in each condition.

Procedure

Participant-witnesses were tested individually and followed the procedure for full-face recall described in Experiment 1. Both interview procedures were identical up until the verbal description had been completed. At this point, those assigned to CI moved directly to construct a composite, following the full-face or internal-first construction procedure in Experiment 1, while those assigned to the H-CI undertook a holistic interview (HI). For this HI component, participants were asked to think about the perceived personality of their target face, with one minute given for silent reflection. They were then asked to make judgments (in the following order) for intelligence, distinctiveness, pleasantness, masculinity, coldness, caring and competence on a scale of 'Low', 'Medium' and 'High' (as per Frowd et al., 2008). The CI took a maximum of 10 minutes and the H-CI up to three minutes more.

Participants moved on to construct a single composite each, as described in Experiment 1, Stage 1, Procedure. Thus, in the whole-face condition, the entire face was visible throughout construction; in the internal-features first condition, the face was constructed in separate regions, internal and then external.

Stage 2: Composite naming

Design

A 2 (interview: CI vs. H-CI) x 2 (construction: full-face vs. internal-features first) x 2 (face view: front vs. side) mixed-factorial design was used, with interview and construction type as between-subjects factors. Participants named composites from one of four individual conditions (CI, whole; CI, internals-first; H-CI, whole; H-CI, internals-first), randomly assigned with equal sampling; participants first inspected all composites from their assigned condition front-on and then sideways-on.

Participants

Thirty-two undergraduate students (18 males and 14 females) from the same university were recruited based on being a fan of rugby league, to be familiar with the target identities. These volunteers were aged between 18 and 27 ($M = 21.8$, $SD = 2.7$) years and were randomly allocated in equal groups to one of the four testing booklets.

Materials

Four testing booklets were prepared for naming: each booklet contained two greyscale composites (12 cm x 18 cm) from each condition. The same four foil composites were added to each booklet (constructed as before in Experiment 1 using PRO-fit of young adult males). Each person thus saw eight composites of rugby-league players and four foils. Target images were subsequently presented to participants to verify familiarity with the relevant identities, printed as before to the same standard.

Procedure

Participants were tested individually and the task was self-paced. They were informed that some of the composites were of rugby-league players and that they should try to provide a name where possible. Participants were also informed that “don’t know” responses were acceptable. Composites were presented sequentially from the front, and participants duly provided a name or not for each one. Afterwards, participants were informed that looking at a composite sideways can sometimes improve naming. They were shown how to look at the composite in this way, by turning the page to the side, and asked to have another attempt at naming. Composites were shown for a second time in the same order. Participants were informed that it was acceptable to change their answer from their first attempt. Finally, participants were shown the target photographs, presented sequentially, and asked to attempt

to name those images. As before, composites and targets were presented in a different random order of presentation to each person. Testing sessions took about 10 minutes per person, including debrief.

Stage 3: Composite likeness ratings

Design

A 2 (interview: CI and H-CI) x 2 (construction: full-face and internal-features first) x 3 (composite face type for rating: internal-features, external-features and complete) mixed-factorial design was used. Participants rated all 32 images for likeness in the presence of the relevant target photograph. We were interested in how separate regions of the face had been constructed, and so each person viewed composite images either as a full face or in one of two part-face conditions. Note that as the sideways view of a composite is really only relevant to face recognition, it was not considered here.

Participants

There were 30 participants, 21 males and 9 females, comprising of staff and students from UCLan, none of whom had participated elsewhere in the experiment. This opportunity sample comprised non-rugby fans aged between 19 and 53 ($M = 30.0$, $SD = 9.7$) years. Participants were randomly allocated with equal sampling to the three levels of the between-subjects factor, rating face type.

Materials

Each of the eight target photographs and their associated composites were printed in the same way as in Stage 2. Internal- and external-features versions of the composites were prepared using Adobe Photoshop. Internal-features comprised the central region of the face just above

the brows and followed the shape of the face; external-features comprised the area remaining once internal-features had been extracted.

Procedure

Participants were tested individually and the task was self-paced. They were asked to rate the likeness (1 = very poor likeness ... 7 = very good likeness) of a set of composites of rugby players. For the relevant face type, participants were presented with the four images for each target and asked to rate each on how similar it was to the target photograph. The order of identities was randomised for each person, as was the order of presentation for the four composites within each identity. The task took about 15 minutes including debriefing.

Results

Correct Naming

When subsequently presented with target photographs to name, the vast majority of participants ($N = 26$ out of 32) correctly named all eight targets, with the remaining participants offering a 'don't know' response to either one or two items. The overall naming accuracy across participants was very high ($M = 96.9\%$, $SD = 7.1\%$), indicating very good familiarity with the relevant identities. For the few cases where the target was incorrectly named, the associated responses to composites were coded as missing data, as in Experiment 1.

Composite naming was analysed in terms of accurate responses (correct names) and then inaccurate responses (mistaken names) given by participants. These analyses are presented below followed by an analysis of composite likeness ratings.

Responses to composites viewed from the front and from the side were checked for missing data (none detected), and scored for accuracy as before. There were 204 correct responses in total out of a possible 512 ($M = 40.6\%$) for both front-on and sideways-on naming. Correct naming data are summarised in Table 2.

Table 2. Correct composite naming: the impact of type of interview, method of face construction and view of composite face at naming.

<u>Construction Type</u>		<u>View at naming</u>		<u>Mean</u>
		Front	Side	
<u>Interview Type</u>				
Complete face	CI	27.9 (17/61)	27.9 (17/61)	27.9 (34/122)
	H-CI	44.6 (28/63)	55.6 (35/63)	50.0 (63/126)
Internals-first	CI	6.5 (4/62)	8.1 (5/62)	7.3 (9/124)
	H-CI	71.0 (44/62)	87.1 (54/62)	79.0 (98/124)

Note. Figures are percentage-correct calculated from responses in parentheses (see Table 1, *Note*). These data are for composites for which participants correctly named the relevant target ($N = 248$ out of 256).

Naming data were analysed using Generalised Estimating Equations (GEE), a multiple regression based analysis that is appropriate for correlated data (Barnett, Koper, Dobson, Schmiegelow & Manseau, 2009; Hanley, Negassa, Edwardes & Forrester, 2003)—to accommodate both multiple responses per participant as well as the within-subjects factor, view

at naming. The approach creates what is essentially a combined by-participants and by-items analysis of the data (for a similar analysis, see Frowd et al., 2013). A model-based estimator was used for the covariance matrix and an exchangeable structure for the working correlation matrix. The model's response (link) function was binary logistic (for dichotomous responses). Naming data were checked for appropriateness for analysis by GEE ($f(\text{observed}) > 0$ and $f(\text{expected}) > 5$ for $\geq 80\%$ of cells), and the resulting models were checked for appropriate variables (e.g., $SE(B)$ values were not too low or too high relative to B).

Three predictors were included in a full-factorial model: interview (1 = CI and 2 = H-CI), construction type (1 = internal-features first and 2 = full-face) and view at naming (1 = front-on and 2 = sideways-on). Using sequential elimination using a standard alpha of .1, three terms emerged as not significant (construction, construction x view, and the three-way interaction), and were removed. The resulting model contained reliable predictors for interview, Model fit $X^2(1) = 53.5$ $p < .001$, with H-CI superior to CI, $B = 1.3$, $SE(B) = 0.4$, $Exp(B) = 3.5$, $CI(-) = 1.7$, $CI(+) = 7.3^{(3)}$; and view, $X^2(1) = 11.2$ $p = .001$, with side view superior to front view, $B = 0.7$, $SE(B) = 0.1$, $Exp(B) = 1.9$, $CI(-) = 1.5$, $CI(+) = 2.5$. There were two reliable interaction terms involving interview, analysed below.

Interview interacted with construction type, $X^2(2) = 20.3$, $p < .001$ (see first column of data, Table 2). The effect of interview (H-CI > CI) was reliable *both* for construction using full-face, $B = 1.2$, $SE(B) = 0.4$, $p = .001$, $Exp(B) = 3.5$, $CI(-) = 1.7$, $CI(+) = 7.3$, and internal-features, $B = 3.5$, $SE(B) = 0.5$, $p < .001$, $Exp(B) = 33.3$, $CI(-) = 11.4$, $CI(+) = 97.2$, indicating consistency for this type of interview. However, the CI was detrimental to composites constructed with internal-features first (cf. full-face method), $B = -1.6$, $SE(B) = 0.5$, $p = .003$, $1/Exp(B) = 4.9$,

³ Confidence intervals presented here and elsewhere are +/- 95% values of $Exp(B)$.

$CI(-) = 1.7$, $CI(+) = 13.9^{(4)}$ but the H-CI was beneficial for composites constructed in this way (i.e., internal-features first > complete face), $B = 1.3$, $SE(B) = 0.4$, $p = .001$, $Exp(B) = 3.5$, $CI(-) = 1.7$, $CI(+) = 7.4$.

Interview also interacted with view, $X^2(1) = 7.7$, $p = .006$ (Table 3), as side view led to reliably better correct naming than front view just for composites constructed with a H-CI, $B = 0.7$, $SE(B) = 0.1$, $p < .001$, $Exp(B) = 1.9$, $CI(-) = 1.5$, $CI(+) = 2.5$; using a CI, there was no reliable difference by view ($p = .71$). The effect of interview (H-CI > CI) was again consistent for both front-, $B = 3.5$, $SE(B) = 0.5$, $p < .001$, $Exp(B) = 33.3$, $CI(-) = 11.4$, $CI(+) = 100.0$, and for side-view naming, $B = 4.1$, $SE(B) = 0.6$, $p < .001$, $Exp(B) = 58.8$, $CI(-) = 20.4$, $CI(+) = 166.7$.

Table 3. Percentage correct front- and side-view naming for composites constructed with a holistic-cognitive interview and standard face-recall interview.

<i>Interview type</i>	<i>View at naming</i>	
	Front	Side
CI	17.1 ^(ns) (21 / 123)	17.9 ^(ns) (22 / 123)
H-CI	57.6 (72 / 125)	71.2 (89 / 125)

Note. Figures are percentage-correct calculated from responses in parentheses (see Table 1, *Note*). Cell differences are significant, $p < .001$, except between the contrasts labelled (ns).

⁴ The inverse of the Odds Ratio is presented, $1/Exp(B)$, sometimes referred to as the Risk Ratio, as this permits an easier interpretation (a whole number multiple) than the Odds Ratio itself (which is less than 1.0 here).

Incorrect Naming

Incorrect names were scored as described in Experiment 1 and the analysis followed a procedure similar to that described above. There were 173 mistaken responses out of a possible 292. While this level of incorrect responses may seem high ($M = 59.3\%$), this situation is not unusual for feature-based composites (e.g., Frowd et al., 2014).

The final GEE model, carried out as before, emerged with two reliable terms. The main predictor of view was reliable, $X^2(1) = 5.1$, $p = .024$, as mistaken names were higher from the side (cf. front), $B = 0.5$, $SE(B) = 0.2$, $Exp(B) = 1.7$, $CI(-) = 1.1$, $CI(+) = 2.6$. This predictor also interacted with interview, $X^2(1) = 6.5$, $p = .038$, as illustrated in Table 4, since the increase in mistaken names for side-on (cf. front-on) naming was restricted to naming of composites constructed after an H-CI, $B = 0.5$, $SE(B) = 0.2$, $p = .025$, $Exp(B) = 1.7$, $CI(-) = 1.1$, $CI(+) = 2.6$: there was no reliable difference after a CI ($p = .51$). The interaction term was also significant as construction following an H-CI led to fewer incorrect names (cf. CI) when viewing the face from the front, $B = -0.8$, $SE(B) = 0.3$, $p = .016$, $1/Exp(B) = 2.3$, $CI(-) = 1.2$, $CI(+) = 4.5$; there was no significant difference for side-on naming ($p = .26$).

Table 4. The interactive effect of interview and view on inaccurate naming.

<i>Interview type</i>	<i>View at naming</i>	
	Front	Side
CI	62.7 ^a (64 / 102)	65.3 (66 / 101)
H-CI	41.5 ^{a,b} (22 / 53)	58.3 ^b (21 / 36)

Note. Figures are inaccurate naming scores calculated by dividing responses shown in parentheses and expressed as a percentage. In parentheses, these are summed mistaken (numerator) and total (mistaken and no name) responses (denominator) for composites for which participants correctly named the relevant target and for composites that had not been named correctly ($N = 292$ out of 512). ^{a,b} Difference significant, $p < .05$.

Likeness Ratings

Likeness rating data were analysed using GEE; the model and analysis used were the same as above except that the link function was ordinal logistic (appropriate for Likert ratings). Two within-subjects predictors were included in a full-factorial model: interview (1 = CI and 2 = H-CI) and face-construction type (1 = internal-features first and 2 = full-face).

The final model was significant for both interview type, $X^2(1) = 4.9$, $p = .027$, with H-CI superior to CI, $B = 1.3$, $SE(B) = 0.2$, $Exp(B) = 3.7$, $CI(-) = 2.6$, $CI(+) = 5.1$; and face-construction type, $X^2(1) = 5.7$, $p = .017$, with internal-feature construction having higher likeness than full-face construction, $B = 0.8$, $SE(B) = 0.2$, $Exp(B) = 2.2$, $CI(-) = 1.6$, $CI(+) = 3.0$. There was one significant interaction, between interview and face-construction type, $X^2(1) = 78.8$, $p < .001$ (Table 5). All four contrasts involved were significant ($p < .001$): (i) H-CI was better than CI for internals-first, $B = 1.3$, $SE(B) = 0.2$, $Exp(B) = 3.7$, $CI(-) = 2.6$, $CI(+) = 5.1$, but the reverse was true (CI > H-CI) for full-face construction, $B = 0.8$, $SE(B) = 0.2$, $Exp(B) = 2.2$, $CI(-) = 1.6$, $CI(+) = 3.0$, and (ii) full-face was better than internals-first for the CI, $B = 0.8$, $SE(B) = 0.2$, $Exp(B) = 2.1$, $CI(-) = 1.6$, $CI(+) = 3.0$, but internals-first were better than full-face for the H-CI, $B = 1.3$, $SE(B) = 0.2$, $Exp(B) = 3.7$, $CI(-) = 2.7$, $CI(+) = 5.2$.

Table 5. The interaction between interview and construction type for likeness ratings.

<i>Face Construction</i>	<i>Interview</i>	
	CI	H-CI
Internals-first	3.7 (0.3)	5.4 (0.3)
Full-face	4.7 (0.3)	3.6 (0.3)

Note. Likeness rating scale (1 = very poor likeness ... 7 = very good likeness); SE of mean likeness are in parentheses. All four contrasts were significant, $p < .001$.

Discussion

Consistent with our predictions and with previous research (Frowd et al., 2008; Frowd, Nelson, et al., 2012) results from correct naming and ratings of likeness revealed that the H-CI was superior to the CI overall and that presenting a complete face at construction after an H-CI led to sub-optimal performance. Also, although H-CI outperformed CI overall, and internals-first construction was better than full-face construction overall, results confirm that CI is well-aligned with full-face construction, and H-CI with internals-first construction. This result supports the proposal that the H-CI benefit is underpinned at least partially by the alignment of interview and construction, and that this alignment is important for producing effective composites. The result concurs with TAP (Morris et al., 1977) and with previous studies showing that recall can affect subsequent task performance (e.g., Roediger & Karpicke, 2006). The result also adds weight to our argument that the witness interview acts as re-encoding and so should be carefully aligned with the cognitive operations required in subsequent tasks. Further, as H-CI is optimally paired with internals-first construction, this suggests that H-CI particularly activates memory traces for internal-features. Furthermore, results for correct

naming indicate benefit of side-on (cf. front-on) for H-CI, and for incorrect naming, there was overall benefit of H-CI (lower incorrect names cf. CI), both indicating construction of a more accurate face for H-CI—although note that side-on did lead to more inaccurate names (cf. front-on) for this interview. Findings are considered in more detail in the General Discussion.

Prior to that, we apply the theory of aligning interview and construction to a holistic system. Using this method, constructors repeatedly select from face arrays, with choices combined, to “evolve” a likeness (Davis et al., 2016; Frowd et al., 2013; Tredoux et al., 2006). We used EvoFIT, largely because evidence indicates it produces identifiable faces as part of a realistic design (e.g., Frowd, Skelton, Atherton et al., 2012), as used here, although our intention is that results should be applicable to similar methods to create a face. Research using the gold standard and 24 hr retention interval indicates that the H-CI is very effective at facilitating correct naming (cf. CI), as is viewing composites sideways-on (cf. front-on) for naming (Frowd et al., 2013).

The most recent EvoFIT procedure aligns with research on the importance of the region around the eyes for face processing. Eyes are the most attended-to feature, irrespective of face familiarity (Heisz & Shore, 2008) and task demands (Schyns, Jentzsch, Johnson, Schweinberger, & Gosselin, 2003). Further, eye fixations increase with expertise in face processing (Vinette, Gosselin, & Schyns, 2004) and the N170 face-sensitive ERP component is stronger for eyes than for other facial features (Nemrodov, Aston, Preston, & Itier, 2014), again demonstrating the importance of the eye-region for face processing. For EvoFIT, witnesses are now invited to focus on this area during selection of faces from arrays, as the ensuing composites have been found to be more effective than when constructors are asked to focus on the overall face (Fodarella et al., 2017). However, in spite of considerable research indicating consistent benefit of the H-CI (cf. CI) for both holistic and feature systems (see Frowd, Erickson et al., 2015 for a meta-analysis), Portch, Logan, and Frowd (2017) found the

H-CI to be ineffective when constructors were asked to focus on the eye-region prior to constructing an EvoFIT. We propose that the H-CI promotes configural processing of the internal-features and therefore face selection by eye-region is now not optimally aligned with the interview. If this explanation is correct, a shift in focus to the eye-region at the end of the interview should augment the H-CI for the newer construction method that also focuses on the eye-region. We consider this novel proposal in the following experiment.

EXPERIMENT 3

The original H-CI requires whole-face character judgements to be made. In the current experiment, we assess the effectiveness of a version of the H-CI whereby constructors are asked to make such judgements based on the eye-region rather than the whole face. Based on results so far, we anticipated that, due to process alignment, an overall consequence of the interview is that composites would be more identifiable when constructed using this modified interview than using either the face-recall CI or the original H-CI. However, we were concerned that it might be too challenging for constructors to make character judgements for the eye-region alone; as such, we also assessed a “combined” H-CI comprising the usual H-CI followed by further character attention that focused on the eye-region. These four individual conditions were operationalised in a 2 x 2 design, as described below.

Stage 1: Composite construction

Design

A between-participants 2 (interview: CI vs. H-CI) x 2 (secondary eye-region interview: absent vs. present) design was used. Thus, there were four individual conditions: (A) face-recall CI, (B) H-CI Standard, (C) face-recall CI followed by character attribution of the eye-region (H-

CI ER) and (D) H-CI Standard followed by character attribution of the eye-region (H-CI Combined).

Participants

Twenty-seven male and 13 female staff and students from the University of Glasgow were recruited based on being unfamiliar with the TV programme EastEnders. All participants were over 18 years of age, although age and gender information were not recorded. They received a small honorarium as compensation for participation, and had not participated elsewhere in this research. Participants were randomly allocated with equal sampling to the four conditions.

Materials

The targets comprised five male and five female characters from EastEnders. They were sourced via the Internet, had the same characteristics as for the other two experiments (e.g., as far as possible, good quality, front-facing, neutral expression), and were printed in the same way to a size of 8cm x 10cm.

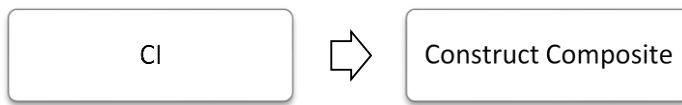
Procedure

The method was the same as for Experiment 2, except that EvoFIT (version 1.6) was used for face construction. Each participant viewed one of the 10 target images for 30 seconds in the knowledge that they would return to construct a composite the following day. Between 20 and 28 hours later, participants were interviewed in the same way as in Experiment 2 for those assigned to either CI or H-CI (what we refer to here as H-CI Standard). Participants assigned to Condition C did the same as H-CI Standard except were instructed to make judgements for the eye-region only, following an initial Face-recall CI; those assigned to H-CI 'Combined' underwent H-CI Standard and then repeated this for the region around the eyes. So, all

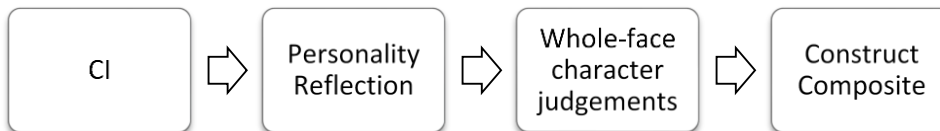
participants freely recalled the face and, in the three experimental conditions, underwent one of three holistic-type interviews. Figure 2 depicts the procedure used in each of the conditions.

Participants constructed a single composite. The experimenter described the general procedure and controlled the software. The procedure itself is fairly involved and is detailed in previous research (e.g., Fodarella et al., 2015). In brief, participants repeatedly selected from arrays of internal-features faces, with a prompt to focus on the eye-region. Participants first selected for smooth faces, then for texture (colouring), combinations of smooth and texture, and overall best match. This selection procedure was repeated for a second iteration ('generation'), whereby faces contained characteristics of those selected previously. The best likeness was then carried forward into 'holistic' tools for overall enhancement of the internal-features for width, weight, health and 12 other overall characteristics of the face. Participants were given the opportunity to enhance the brightness of the features' texture, and to alter size and position of features using a 'shape' tool. They then added external features (hair, ears and neck) and were given the opportunity to enhance the face further using holistic and shape tools. Construction sessions lasted for approximately 45 minutes including debrief.

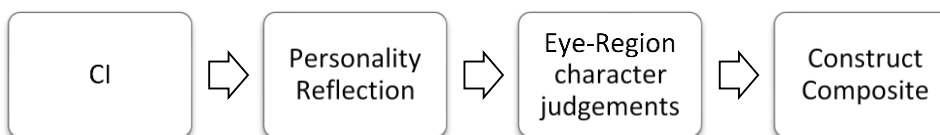
A. CI



B. H-CI Standard



C. H-CI Eye-Region



D. H-CI Standard + Eye-Region (H-CI Combined)

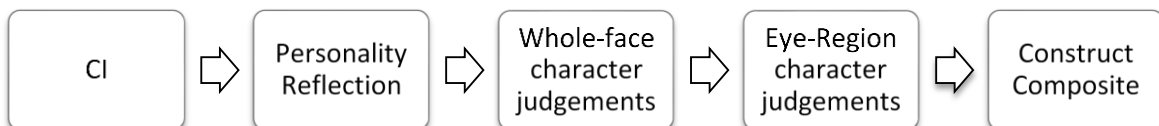


Figure 2. Process diagram illustrating the procedure for Experiment 3 conditions.

Stage 2: Composite naming

Design

The design was mixed-factorial: 2 (interview, between subjects) x 2 (eye-region interview, between subjects) x 2 (view at naming [front-on vs. sideways-on] within subjects).

Participants

An opportunity sample of 42 females and 8 males, different to before, volunteered from the same university. They were recruited based on being familiar with characters from EastEnders. Participant age ranged from 21 to 67 ($M = 38.0$, $SD = 11.9$) years, and participants were randomly assigned in two groups of 13 to CI and H-CI Standard, and in two groups of 12 to H-CI Eye-Region and H-CI Combined.

Materials and Procedure

Materials and procedure to name the composites and target photographs were the same as before, except that participants were recruited based on being familiar with characters from EastEnders. Participants thus named 10 composites (from one of the four conditions) first front-on, and then from the side. Finally, participants named the 10 target photographs.

Stage 3: Composite likeness ratings

Participants

Fifteen female and two male volunteers rated the composites for likeness. Participants were staff and students at UCLan, and their ages ranged from 18 to 61 ($M = 32.5$, $SD = 10.2$) years. None had participated elsewhere in these experiments.

Design, Materials and Procedure

Apart from the fact that target identities were characters from the EastEnders and that complete composite faces were presented throughout, the same design and procedure as Experiment 2 were used to rate the composites for likeness.

Results

Correct Naming

All participants correctly named all of the target photographs, and so familiarity of the relevant identities was 100%.

Responses to composites were scored and checked for missing data in the same way as in Experiment 2 (naming). There were 378 correct responses in total out of a possible 1000 ($M = 37.8\%$) for both front-on and side-on composite naming (Table 6).

Table 6. Percentage correct naming for composites constructed with a holistic-cognitive interview and standard face-recall interview, either followed by an eye-region interview or not.

<i>Eye-Region</i>	<i>Interview</i>	
	CI	H-CI
Absent	35.4 (92 / 260)	35.4 (92 / 260)
Present	37.1 (89 / 240)	45.4 [†] (109 / 240)

Note. Data have been collapsed over view at naming since this factor did not emerge as reliable. [†]Significantly different to CI, $p < .05$. Intercept of model, $B = 0.6$, $SE(B) = 0.2$, $p < .001$, $Exp(B) = 0.5$. See *Note* for Table 1, and text, for more details.

The accurate naming data were checked for validity and analysed in the same way as in Experiment 2 (naming) using Generalised Estimating Equations (GEE). This time, an independent (cf. exchangeable) structure was selected for the covariance matrix as a more accurate model was produced of the data (i.e., emerging $SE(B)$ values were lower, presumably due to the now much smaller effect for the within-subjects predictor, View at naming).

Based on the outcome of the previous experiment, the analysis first considered the three-way interaction between interview, eye-region and view (at naming). This interaction turned out to be non-significant, $X^2(7) = 7.7$, $p = .63$, and so the two-way interactions were entered as a factorial model. The two terms involving view at naming were also non-significant ($p > .23$) and were removed. However, based on the usual $\alpha = .1$, the interaction between interview and eye-region was significant, $X^2(3) = 7.3$, $p = .063$, and parameter estimates indicated that composites from the H-CI Combined condition received significantly more correct names than those from the CI, $B = 0.4$, $SE(B) = 0.2$, $p = .023$, $Exp(B) = 1.5$, $CI(-) = 1.1$, $CI(+) = 2.2$.

Incorrect naming

The data were re-scored for inaccurate responses and checked for missing data in the same way as before. Participants provided a name on most occasions (Table 7).

Inferential analysis proceeded in the same way as for accurate naming, with neither view nor the interactions involving view significantly affecting inaccurate naming ($p > .1$). The resulting model for GEE was reliable for both the effect of interview, $X^2(1) = 2.8$, $p = .092$, with lower incorrect naming under H-CI Standard (cf. CI), $B = 2.3$, $SE(B) = 0.6$, $p = .008$, $Exp(B) = 5.6$, $CI(-) = 1.6$, $CI(+) = 19.9$, and eye-region, $X^2(7) = 11.9$, $p = .001$, with lower incorrect naming when eye-region judgements were present (cf. absent), $B = 1.7$, $SE(B) = 0.6$, $p < .001$, $Exp(B) = 10.5$, $CI(-) = 3.1$, $CI(+) = 35.3$. The interaction between these two factors was also reliable, $X^2(7) = 9.9$, $p < .001$, as the H-CI Combined produced incorrect naming that was equivalent to

the two individual H-CI conditions—in other words, incorrect naming was reliably higher in the CI relative to the three other conditions.

Table 7. Combining holistic interview types (Interview and Eye-Region) for inaccurate composite naming using a holistic system

<u>Eye-Region</u>	<u>Interview</u>	
	CI	H-CI
Absent	97.6 [†] (164 / 168)	79.9 (135 / 169)
Present	88.0 (132 / 150)	87.0 (114 / 131)

Note. Analysis of inaccurate responses revealed a significant interaction (shown here) between interview and eye-region. Intercept $B = 1.8$, $SE(B) = 0.4$, $p < .001$, $Exp(B) = 6.2$.

[†]Significantly different to other conditions, $p < .005$. Also see *Note* for Table 4, and text.

Likeness Ratings

Likeness ratings are summarised in Table 8. They were analysed in the same way as Experiment 2.

Table 8. Mean rated likeness of composites: Interview x Eye-Region (ER).

<u>Eye-Region</u>	<u>Interview</u>	
	CI	H-CI
Absent	3.9 (0.1)	2.9 (0.1)

Present	3.2 (0.1)	3.8 (0.1)
---------	--------------	--------------

Note. Likeness rating scale (1 = very poor likeness ... 7 = very good likeness); SE of mean likeness are in parentheses. All pairwise comparisons were significant, $p < .002$; CI and H-CI Combined was not significant, $p = .43$.

Three predictors were included in a full-factorial model: interview (1 = CI and 2 = H-CI Standard), eye-region (1 = Absent and 2 = Present) and their interaction. GEE revealed that neither predictor was reliable ($p > .18$), and both were subsequently removed. The resulting model was reliable for the interaction, $X^2(3) = 31.3, p < .001$. Analyses of parameter estimates suggested that all pairwise comparisons were reliable, $p < .002$. Of particular note, relative to H-CI Combined, likeness ratings were worse following H-CI ER, $B = -0.5, SE(B) = 0.2, 1/Exp(B) = 1.7, CI(-) = 1.3, CI(+) = 2.4$ and H-CI Standard, $B = -0.7, SE(B) = 0.2, 1/Exp(B) = 2.1, CI(-) = 1.4, CI(+) = 3.1$. Ratings were equivalent between composites constructed for CI and H-CI Combined, $B = -0.1, SE(B) = 0.2, p = .43$.

In summary, for the latest construction procedure with EvoFIT, relative to CI, the naming data indicate equivalence between the original H-CI and H-CI ER, but benefit from H-CI Combined. The likeness ratings revealed the same pattern of results except that CI emerged equivalent to H-CI Combined.

Discussion

Using the Combined H-CI reliably increased correct names and reduced incorrect names compared with the CI. The H-CI ER was not effective in promoting more identifiable composites; as suggested earlier, we suspect this may stem from difficulty in making character

judgements based solely on the eye-region. However, when the procedure follows the standard H-CI (which involves full-face character judgements), the ER component provides alignment with the eye-region focused construction procedure. Also, while there was little difference for correct naming, there was some evidence that H-CI Standard was effective by producing fewer *inaccurate* composites (cf. CI composites), confirming that the H-CI currently used by some police forces results in fewer incorrect names, thus saving police time in pursuing false leads. In this experiment, side-on (cf. front-on) naming did not lead to a reliable change in either correct or incorrect naming.

Finally, although likeness ratings did not entirely follow the same pattern as naming, they do reinforce superiority of the H-CI Combined over H-CI Std and H-CI ER, indicating differences in face processing. Interpretation of these data will be considered in the following section.

General Discussion

A facial composite is a visual representation produced from witness memory with the intention that someone will recognise the face, name it to the police, and thereby assist in bringing an offender to justice. Over the last decade, research has made substantial contributions to improving recognition of these images: by developing the initial interview (H-CI), composite software (feature systems such as E-FIT, PRO-fit and FACES, and holistic systems such as EvoFIT, ID and EFIT-V/6) and methods used to facilitate recognition of a finished composite (e.g., dynamic caricature and image stretch). The current project sought to establish whether findings from a holistic system (Frowd et al., 2013) would extend to composites constructed from a feature-based system. We also wanted to gain insight into the underpinnings of the H-CI advantage and explore whether side-on naming would benefit feature-based composites.

Experiment 1 demonstrated that alignment of initial interview and face construction is more beneficial than focusing witness attention on internal-features during interview. Further, the congruent conditions performed equivalently: the two incongruent conditions resulted in similar quality composites that were much less identifiable than those constructed from aligned processes, consistent with TAP (Morris et al., 1977). We propose that an internal-feature focus may contribute to the H-CI advantage, however an internal-feature focus alone did not provide benefit.

In Experiment 2, the H-CI was more effective than the CI across all construction techniques and viewing conditions. Frowd et al. (2008) revealed that the H-CI dramatically improved correct naming compared with the CI when composites were created using a full-face procedure in PRO-fit. Our results extend their findings to internals-first face construction, providing good evidence that the H-CI is more effective than an unmodified CI for composite construction using a feature-based system.

Experiments 1 and 2 manipulated construction methods in the same way, prompting participant-witness to create a face either in the full-face context or by constructing the internal-features first with the external-features masked. However, the methods used for aligning interview (re-encoding) to construction conditions differed across experiments. Experiment 2 involved the HI—a series of character judgements, which appear to focus attention on the internal-features and also promote deep encoding of those features and their interrelations (e.g., Bower & Karlin, 1974; Wells & Hyrciw, 1984). Experiment 1 only required a description of features, which is a more ‘surface’ level task, presumably requiring shallower processing (Craik & Lockhart, 1972). The HI was shown to be more effective than recalling internal-features first demonstrating that the H-CI benefit is not simply a function of focus on the internal features. Where the CI and H-CI both involve detailed feature recall, we propose that the H-CI’s additional HI component serves to rehearse and re-encode configural (feature

spacing) information such that the whole-face representation is of better quality (cf. CI). Trait encoding thus assists witnesses when selecting features (as in PRO-fit) or selecting wholes (internal-face portions, as in EvoFIT; Frowd, Nelson et al., 2012) because the memory representation is more accurate. The H-CI was particularly effective here when combined with internals-first construction, suggesting that character judgements involve focus on internal-features. In fact, internals-first construction may enjoy an additional advantage over full-face construction because there are no external-features to distract witnesses from constructing the face's important internal-features (Frowd, Skelton, Atherton et al., 2012; Frowd et al., 2013).

Overall, correct naming of composites from the CI-internals-first condition was largely at floor level ($M = 7.3\%$). Relative to other conditions, this condition involved incongruent processes and did not benefit from deeper encoding afforded by the H-CI. This likely resulted in a representation of features, encoded at a surface level, without reinforcement of inter-relations. Correct naming of PRO-fit composites created likewise after a CI may be as low as 3% (e.g., Bruce, Ness, Hancock, Newman, & Rarity, 2002), with other feature-based systems performing at a similar level (Frowd, McQuiston-Surrett, Anandaciva, Ireland, & Hancock, 2007).

Having established the importance of process-alignment, Experiment 3 sought to extend this idea to one of the holistic systems, EvoFIT, involving selection of faces now based on the eye-region. The H-CI Combined resulted in superior facial composites as compared with the CI condition, but the H-CI Standard and H-CI ER conditions were no more effective by correct naming than the CI. The CI did, however, result in higher inaccurate naming than these two conditions. Naming analyses thus confirm that for eye-region selection for a holistic system, composites are constructed best after character attribution for the full-face and then for the eye-region. This result also supports Portch et al. (2017) who found no H-CI advantage when asking participants to focus on the eye-region when evolving a face in EvoFIT.

We propose that whole-face character attribution helps to integrate individual feature traces recalled during the CI, and also allows for re-encoding of relationships between features (configural information) that are important for recognition (Tanaka & Sengco, 1997). These processes help to form a holistic face memory that represents individual features and how these fit together in the whole-face. The eye-region component of the interview then focuses witnesses on the eye-region, which is well-aligned with the face construction procedure now recommended for this holistic system. Taken together with the findings of Experiments 1 and 2, there is strong evidence that good quality facial composites—those that are frequently named correctly—arise from careful alignment of the interview and the construction process.

Viewing composites side-on promoted higher correct naming compared with front-on presentation, but only for PRO-fit, and only for composites produced after a H-CI. Side-on naming also increased the number of mistaken names compared with front-views for H-CI composites. When considering the summed number of mistaken names and instances where no name was provided, the frequency of mistaken names was similar across both CI conditions and both H-CI conditions with little change from front-on to side-on naming. However, because the number of ‘no name’ responses decreases in the H-CI conditions, the *proportion* of mistaken names increases. In Experiment 3 side-on naming did not increase the number of correct or mistaken names offered. The technique therefore does not consistently inflate the proportion of mistaken names, nor does it show a consistent benefit. In these experiments, side-on naming is confounded with a second naming attempt; however, there is no consistent effect across conditions and so any benefit is unlikely to be due to simply having another attempt at naming. Likewise, if side-on naming reduces the threshold for naming for higher quality composites on a second attempt, we would expect an advantage for the H-CI Combined composites in Experiment 3. Future studies could examine view-at-naming as a between-participants factor, which would establish whether any benefit persists for a first naming

attempt and further help to elucidate the locus of any benefit. Our findings suggest that side-on naming only assists in conditions involving holistic processing of internal-features, as in the H-CI. The new eye-region focused H-CI, now part of regular police practice for EvoFIT (Frowd, Portch, Killeen, Mullen, Martin, & Hancock, 2019), may improve the quality of the eye-region at the expense of other internal features. If the other internal-features are constructed less accurately then side-on naming may be less helpful. Although our findings in respect of view at naming are inconsistent, side-on naming does not increase the number of inaccurate names, and so we are cautiously optimistic that this technique can potentially assist the police for little additional time or effort.

Likeness ratings are a proxy to naming, and so one would expect differences between measures. A good composite should trigger a familiarity response in somebody familiar with the depicted identity, whereas likeness ratings require matching between composite and target photograph. In Experiment 2, likeness ratings reflected the accurate naming results, indicating overall superiority of the H-CI (cf. CI) and internal-features first construction (cf. full-face construction), and highlighting the effectiveness of the aligned over the unaligned conditions. In Experiment 3, however, likeness ratings for CI composites were equivalent to those for composites from the H-CI Combined condition. This was unexpected, as CI composites were named least accurately.

The CI requires detailed recall of individual features, and if these features are then represented quite accurately in the resulting composite, this should bear out in good likeness ratings—as such ratings require detailed comparisons between the composite and the target image. The CI, however, does not appear to promote a good quality holistic face representation. The H-CI Standard does this but is not well-aligned with eye-region focused construction; the H-CI ER interview focuses attention on the correct face region for construction, but presumably excludes focus on other internal-features which are then less well-constructed. This explanation may

account for why CI composites were rated as good likenesses, but were not well-named, because they do not accurately convey relationships between features. We conclude that alignment of cognitive processes is therefore important for the ecologically-valid task of composite naming, but also for the detailed scrutiny of features required for likeness ratings.

In Experiment 3, there is an apparent confound of recall attempts, and time spent on recall, with the CI offering one attempt and the H-CI Combined offering three. The time spent on recall was 2 - 3 minutes longer for H-CI Standard and ER than for CI, and the H-CI Combined lasted about a minute longer compared to the other H-CI conditions (i.e., a maximum of 4 - 5 minutes longer compared to the CI). However, as the H-CI Combined was no more effective vis-à-vis the H-CI Standard and H-CI ER, and these were no more effective than the CI, the benefit is not explained by recall attempts or recall time (although there was a reduction in incorrect names). In support of this, Campos and Alonso-Quecuty (1999) found that multiple free-recall attempts were less effective than using different CI mnemonics, and Skelton et al. (2011) found that the *order* in which interviews were conducted was instrumental in improving recall. While the possibility cannot be ruled out that additional recall opportunities may have a beneficial effect on memory, we argue that it is the nature of the recall that is important. Specifically, cognitive operations should be congruent from recall to construction, consistent with TAP (Morris et al., 1977). Future research could usefully investigate whether multiple CI attempts, or the combined interview in reverse order, produce similar results to ours.

We also note that two-dimensional photographs are quite different to moving, expressive faces to which we are accustomed. Although the vast majority of face-processing research uses static rather than dynamic stimuli, we acknowledge that this may affect the data. Some previous composite studies (Frowd, Nelson et al., 2012; Frowd et al., 2013) have used video clips and found similar effects to studies using static images (Frowd et al., 2010; Frowd, Skelton,

Atherton, et al., 2012); nevertheless, future studies would benefit from using more ecologically-valid stimulus exposure conditions wherever possible.

In conclusion, our findings suggest that the H-CI is the most effective interview to use for producing effective facial composites with feature and holistic composite systems. This outcome supports our proposal that the CI component activates facial feature traces, which the HI then integrates into a holistic representation through rehearsal of configural information via character judgements. This process benefits recognition and enhances feature-based composites produced using full-face or internal-features first construction. Its particular benefit in combination with the latter adds weight to the suggestion that the HI encourages processing of internal-features. However, the key to producing highly identifiable composites is careful alignment of cognitive processes from interview to construction. For a system such as EvoFIT that involves focus on the eye-region, it would appear advantageous to shift attention to this face region during interview. These findings should be considered as part of advising police practitioners on interviewing techniques when working with witnesses and victims of crime.

References

- Astbury, R. (2011). Another side to fighting crime. *Lancashire Evening Post*, 02/11/2011, 17.
Retrieved from <http://www.lep.co.uk/news/education/another-side-to-fighting-crime-1-3927910>
- Barnett, A. G., Koper, N., Dobson, A.J., Schmiegelow, F., & Manseau, M. (2009). Selecting the correct variance–covariance structure for longitudinal data in ecology : a comparison of the Akaike, quasi-information and deviance information criteria, <http://eprints.qut.edu.au/19195/>
- Bloom, L.C. & Mudd, S.A. (1991). Depth of processing approach to face recognition: A test of two theories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(3), 556-565. DOI: 10.1037/0278-7393.17.3.556
- Bornstein, B. H., Deffenbacher, K., & Penrod, S. (2012). Effects of exposure time and cognitive operations on facial identification accuracy : A meta-analysis of two variables associated with initial memory strength. *Psychology, Crime and Law*, 18(5), 473–490. <https://doi.org/10.1080/1068316X.2010.508458>
- Bower, G.H. & Karlin, M.B. (1974). Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology*, 103(4), 751-757. DOI: 10.1037/h0037190
- Brewer, G. A., Marsh, R. L., Meeks, J. T., Clark-foos, A., & Hicks, J. L. (2010). The effects of free recall testing on subsequent source memory. *Memory*, 18(4), 385–393. <https://doi.org/10.1080/09658211003702163>

- 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, 894-902. DOI: 10.1037/0021-9010.87.5.894
- Cabeza, R. & Kato, T. (2000). Features are also important: contributions of featural and configural processing to face recognition. *Psychological Science*, 11(5), 429-433. DOI: 10.1111/1467-9280.00283
- Campos, L., & Alonso-Quecuty, M. L. (1999). The cognitive interview: Much more than simply “try again.” *Psychology, Crime & Law*, 5(1), 47–59. DOI: 10.1080/10683169908414993
- Craik, F.I.M. & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, 11(6), 671-684. DOI: 10.1016/S0022-5371(72)80001-X
- Davies, G.M. (1983) Forensic face recall: the role of visual and verbal information. In S.M.A. Lloyd-Bostock, and B.R. Clifford, (Eds.), *Evaluating Witness Evidence* (pp. 103-123), Chichester, England:Wiley.
- Davis, J.P., Thorniley, S., Gibson, S., & Solomon, C. (2016). Holistic facial composite construction and subsequent lineup identification accuracy: Comparing adults and children. *Journal of Psychology: Interdisciplinary and Applied*, 150(1), 102-118. DOI: 10.1080/00223980.2015.1009867
- Ellis, H., Shepherd, J., & Davies, G. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception*, 8(4), 431-439. DOI: 10.1068/p080431

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.

Fodarella, C., Kuivaniemi-Smith, H., Gawrylowicz, J., & Frowd, C.D. (2015). Forensic procedures for facial-composite construction. *Journal of Forensic Practice*, 17(4), 259 – 270. DOI: 10.1108/JFP-10-2014-0033

Frowd, C.D. (2015). Facial recall and computer composites. In C.Wilkinson and C.Ryann (Eds). *Facial Identification*. Cambridge University Press, Cambridge:UK.

Frowd, C. D. (2017). Facial composite systems: Production of an identifiable face. In M. Bindemann and A. Megreya (Eds.) *Face Processing: Systems, Disorders and Cultural Differences*. Nova Science.

Frowd, C.D., Bruce, V., McIntyre, A. & Hancock, P.J.B. (2007). The relative importance of external and internal-features of facial composites. *British Journal of Psychology*, 98, 61-77. DOI: 10.1348/000712606X104481

Frowd, C., Bruce, V., Ross, D., McIntyre, A. & Hancock, P.J.B. (2007). An application of caricature: How to improve the recognition of facial composites. *Visual Cognition*, 15(8), 954-984. DOI: 10.1080/13506280601058951

Frowd, C.D., Bruce, V., Smith, A., & Hancock, P.J.B. (2008). Improving the quality of facial composites using a holistic cognitive interview. *Journal of Experimental Psychology: Applied*, 14, 276-287. DOI: 10.1037/1076-898X.14.3.276

- Frowd, C.D., Carson, D., Ness, H., Richardson, J., Morrison, L., McLanaghan, S. & Hancock, P.J.B. (2005). A forensically valid comparison of facial composite systems. *Psychology, Crime & Law*, *11*, 33-52. DOI: 10.1080/10683160310001634313
- Frowd, C.D., Erickson, W.B., Lampinen, J.L., Skelton, F.C., McIntyre, A.H., & Hancock, P.J.B. (2015). A decade of evolving composite techniques: regression- and meta-analysis. *Journal of Forensic Practice*, *17*, 319-334. DOI: 10.1108/JFP-08-2014-0025
- Frowd, C.D., Hancock, P.J.B., & Carson, D. (2004). EvoFIT: A holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on Applied Psychology (TAP)*, *1*, 1-21.
- Frowd, C.D., Jones, S., Fodarella, C., Skelton, F.C., Fields, S., Williams, A., Marsh, J., Thorley, R., Nelson, L., Greenwood, L., Date, L., Kearley, K., McIntyre, A., & Hancock, P.J.B. (2014). Configural and featural information in facial-composite images. *Science & Justice*, *54*, 215-227.
- Frowd, C.D., McQuiston-Surrett, D., Anandaciva, S., Ireland, C.G., & Hancock, P.J.B. (2007). An evaluation of US systems for facial composite production. *Ergonomics*, *50*(12), 1987-1998. DOI: 10.1080/00140130701523611
- Frowd, C.D., Nelson, L., Skelton F.C., Noyce, R., Atkins, R., Heard, P., ... Hancock, P.J.B. (2012). Interviewing techniques for Darwinian facial composite systems. *Applied Cognitive Psychology*, *26*, 576-584. DOI: 10.1002/acp.2829
- Frowd, C.D., Pitchford, M., Bruce, V., Jackson, S., Hepton, G., Greenall, M., ... Hancock, P.J.B. (2010). The psychology of face construction: giving evolution a helping hand. *Applied Cognitive Psychology*, *25*, 195–203. DOI: 10.1002/acp.1662

Frowd, C. D., Portch, E., Killeen, A., Mullen, L., Martin, A. J., & Hancock, P. J. B. (2019). EvoFIT facial composite images: a detailed assessment of impact on forensic practitioners, police investigators, victims, witnesses, offenders and the media. In A. Stoica, G. Howells, K. McDonald-Maier, A. Erdogan, and T. Arslan (Eds.) *Proceedings of IEEE International Conference on Emerging Security Technologies*, July, University of Essex.

Frowd, C.D., Skelton F., Atherton, C., Pitchford, M., Hepton, G., Holden, L., ... Hancock, P.J.B. (2012). Recovering faces from memory: the distracting influence of external facial features. *Journal of Experimental Psychology: Applied*, 18, 224-238. DOI: 10.1037/a0027393

Frowd, C.D., Skelton F.C., Butt, N., Hassan, A., & Fields, S. (2012). Familiarity effects in the construction of facial composite images using software systems. *Ergonomics*, 54, 1147-1158. DOI: 10.1080/00140139.2011.623328

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

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., Holland, H. L. (1986). Enhancement of eyewitness memory with the cognitive interview. *American Journal of Psychology*, 99(3), 385-401. DOI: 10.2307/1422492

Hanley, J. A., Negassa, A., Edwardes, M. D. d. B., & Forrester, J. E. (2003). Statistical analysis of correlated data using generalized estimating equations: An orientation. *American Journal of Epidemiology*, 157(4), 364–375. DOI: 10.1093/aje/kwf215

- Heisz, J.J., Shore, D.I. (2008). More efficient scanning for familiar faces. *Journal of Vision*, 8(1):1–10. DOI: 10.1167/8.1.9
- Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception*, 31, 1221-1240. DOI:10.1068/p3252
- Kanan, C., Bsesio, D.N.F., Ray, N.A., Hsiao, J.H. & Cottrell, G.W. (2015). Humans have idiosyncratic and task-specific scanpaths for judging faces. *Vision Research*, 108, 67-76. DOI: 10.1016/j.visres.2015.01.013
- Memon, A., Meissner, C.A. & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy and Law*, 16(4), 340-372. DOI: 10.1037/a0020518
- Morris, C.D., Bransford, J.D. & Franks, J.J. (1977). Levels of processing versus transfer-appropriate processing. *Journal of Verbal Learning and Verbal Behaviour*, 16, 519-533. DOI: 10.1016/S0022-5371(77)80016-9
- Nemrodov, D., Anderson, T., Preston, F. F., & Itier, R.J. (2014). Early sensitivity for eyes within faces: A new neuronal account of holistic and featural processing. *NeuroImage*, 97, 81–94. DOI:10.1016/j.neuroimage.2014.04.042
- Portch, E., Logan, K., & Frowd, C.D. (2017). Interviewing and visualisation techniques: attempting to further improve EvoFIT facial composites. In G. Howells et al. (Eds.) *Proceedings of IEEE 2017 Seventh International Conference on Emerging Security Technologies*, 4th – 8th September, University of Kent, Canterbury, UK. DOI: 10.1109/EST.2017.8090406

Roediger, H.L. III., & Karpicke, J.D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249-255. DOI: 10.1111/j.1467-9280.2006.01693.x

Schyns, P. G., Jentzsch, I., Johnson, M., Schweinberger, S. R., & Gosselin, F. (2003). A principled method for determining the functionality of ERP components. *NeuroReport*, 14, 1665–1669

Shapiro, P.N. & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, 100(2), 139-156. DOI: 10.1037/0033-2909.100.2.139

Tanaka, J., & Farah, M. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, 85, 397–405. DOI: 10.1080/14640749308401045

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition*, 25(5), 583–92. DOI: 10.3758/BF03211301

Tredoux, C. G., Nunez, D. T., Oxtoby, O., & Prag, B. (2006). An evaluation of ID: an eigenface based construction system. *South African Computer Journal*, 37, 1-9.

Towler, J. & Eimer, M. (2016). Electrophysiological evidence for parts and wholes in visual face memory. *Cortex*, 83, 246-258. DOI: 10.1016/j.cortex.2016.07.022

Vinette, C., Gosselin, F., & Schyns, P. G. (2004). Spatio-temporal dynamics of face recognition in a flash: It's in the eyes. *Cognitive Science*, 28(2), 289–301. DOI:10.1016/j.cogsci.2004.01.002

Wells, G.L. & Hyrciw, B. (1984). Memory for faces: Encoding and retrieval operations. *Memory and Cognition*, 12(4), 338-344. DOI: 10.3758/BF03198293

Wilson, J. P., Young, S. G., Rule, N. O., & Hugenberg, K. (2018). Configural processing and social judgments: Face inversion particularly disrupts inferences of human-relevant traits. *Journal of Experimental Social Psychology*, 74, 1–7. DOI:

10.1016/j.jesp.2017.07.007

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M. & Ellis, A. W. (1985) Matching familiar and unfamiliar faces on internal and external-features. *Perception*, 14, 737-746. DOI: 10.1007/BF00309318