Ergonomics

The impact of external facial features on the construction of facial composites

Charity Brown, Emma Portch, Faye C. Skelton, Cristina Fodarella, Heidi Kuivaniemi-Smith, Kate Herold, Peter J. B. Hancock & Charlie D. Frowd

Accepted author version posted online: 07 Dec 2018.

Submit your article to this journal

View Crossmark data

The impact of external facial features on the construction of facial composites

Charity Brown (1) psccbr@leeds.ac.uk

Emma Portch (2) eportch@bournemouth.ac.uk

Faye C. Skelton (3) F.Skelton@napier.ac.uk

Cristina Fodarella (4) CFodarella3@uclan.ac.uk

Heidi Kuivaniemi-Smith (4) HKuivaniemi-smith@uclan.ac.uk

Kate Herold (4) k8_herold@hotmail.com

Peter J. B. Hancock (5) pjbh1@stir.ac.uk

Charlie D. Frowd (4) cfrowd1@uclan.ac.uk

(1) School of Psychology, University of Leeds, Leeds LS2 9JT

(2) Department of Psychology, University of Bournemouth, Poole BH12 5BB

(3) School of Life, Sport and Social Sciences, Edinburgh Napier University, Edinburgh EH11 4BN

(4) School of Psychology, University of Central Lancashire, Preston PR1 2HE

(5) Department of Psychology, University of Stirling, Stirling FK9 4LA

*Corresponding author: Charity Brown, Associate Professor, School of Psychology, University of Leeds, Leeds LS2 9JT, UK. psccbr@leeds.ac.uk (0113) 34 35748.

The impact of external facial features on the construction of facial composites

Abstract

Witnesses may construct a composite face of a perpetrator using a computerised interface. Police practitioners guide witnesses through this unusual process, the goal being to produce an identifiable image. However, any changes a perpetrator makes to their external facial-features may interfere with this process. In Experiment 1, participants constructed a composite using a holistic interface one day after target encoding. Target faces were unaltered, or had altered external-features: (i) changed hair, (ii) external-features removed or (iii) naturally-concealed external-features (hair, ears, face-shape occluded by a hooded top). These manipulations produced composites with more error-prone internal-features: participants' familiar with a target's unaltered appearance less often provided a correct name. Experiment 2 applied external-feature alterations to composites of unaltered targets; although whole-face composites contained less error-prone internal-features, identification was impaired. Experiment 3 replicated negative effects of changing target hair on construction and tested a practical solution: selectively concealing hair and eyes improved identification. (150 words)

Practitioner summary

The research indicates that when a target identity disguises or changes hair, this can lead to a witness (or victim) constructing a composite that is less readily identified. We assess a practical method to overcome this forensic issue.

(38 words)

Forensic and security settings make use of a number of tasks relying upon face perception and recognition. The field of ergonomics has a history of examining the interaction of humans and human interfaces to better understand how these real world tasks are administered and the factors underpinning successful performance. Examples have included eliciting witness descriptions of offenders (e.g., Wogalter, 1996), identification of suspects from mugshots (e.g., Ellis, Davies, & Shepherd, 1978; Lee & Whalen, 1996; Lee et al., 2004), verification of photo-IDs by border control officials (e.g., Chiller-Glaus, Schwaninger, & Hofer, 2007) and use of face blurring as a cyber-security privacy-protection technology (Li, Vishwamitra, Hu, Knijnenburg, & Caine, 2017).

One task where an ergonomics approach has made particular in-roads is in facilitating the construction and subsequent identification of facial composites (a visual likeness of an offender). Traditionally, composites have been poorly recognised (e.g., Frowd et al., 2005) and work has focused upon modifying the human interfaces (or systems) that assist memory in this situation, or the contexts in which these interfaces are used (e.g., Frowd, Bruce, Ness et al., 2007; Frowd, Skelton, Butt, Hassan, & Fields, 2011; Gibson, Solomon, Maylin, & Clark, 2009; Skelton, Frowd, & Speers, 2015). Typically, a witness will construct a composite of an unfamiliar offender, someone they have seen once or perhaps a few times, before. This is an unusual task for most people to have to undertake, but construction of an identifiable image can be very valuable to law enforcement, and so a specialist police practitioner is usually involved to guide the witness through the various procedures to externalise the face.

Traditionally computerised 'feature' or manual 'sketch' systems have been used to construct composites, where individual facial features are selected (eyes, nose, hair, etc.). However, intact faces are usually processed as complete entities, with both individual features (eyes, nose, mouth, etc.) and their relative position on the face (configuration)

3

processed holistically (Bruce & Young, 1998; Maurer, Le Grand, & Mondloch, 2002; Tanaka & Farah, 1993). Modern feature systems have incorporated this knowledge and now a witness selects individual facial features within a whole face context. Other, alternative modern 'holistic' human interfaces have also been developed that require witnesses to repeatedly select from arrays of whole faces, or whole face regions, with selected items being 'bred' together, allowing a composite to be 'evolved'.

However, exposure to the whole face context may also hinder face recognition. Tanaka and Sengco (1997) reported worse identification of facial features when the configuration of those features was different (vs. the same) at encoding and test. Also, recognition of an individual feature (a mouth) was disrupted when another part of the face was changed (eye spacing). Thus, alterations in the appearance or position of a single feature can change how other features are perceived. A powerful demonstration of this is the 'split-face effect' (also known as the 'composite-face effect'). When novel faces comprised the upper and lower halves taken from photographs of different faces, identification of individual halves were impaired unless the two sections were misaligned horizontally, or holistic processing itself was disrupted via stimulus inversion (Baudouin & Humphreys, 2006; Calder, Young, Keane, & Dean, 2000; Hole, 1994; Young, Hellawell & Hay, 1987; see also Diamond & Carey, 1986).

Further, during face identification, the utility of some face regions may vary. For unfamiliar face recognition, best recognition occurs for the complete face, but external-features are particularly important (e.g., hair, ears and head shape; Ellis, Shepherd, & Davies, 1979). For example, when participants were trained to identify previously unfamiliar facial photographs, recognition declined when disguises were applied to predominantly alter/conceal external-features (e.g., by addition of a wig, or beard: Patterson & Baddeley, 1977; cap: Cutler, Penrod & Martens, 1987; or changes to hairstyle: Chan & Ryan, 2012).

Similarly, when matching unfamiliar faces to arrays of alternatives, choices are often based on hair (Bruce et al., 1999) which, if concealed, can reduce identification and confidence judgements (Henderson, Bruce, & Burton, 2001). Thus, exposure to accurate external-features facilitates unfamiliar face processing. When applied to a composite interface this principle is found to increase the effectiveness with which a witness interacts with the system. For the EvoFIT holistic interface, participants create composites with more identifiable internal-features if external-features (e.g., hair) in presented face arrays better match the target (Frowd & Hepton, 2009).

Ensuring exposure to "accurate" external-features within the composite interface is difficult to implement – in particular, to provide enough example features within the system to cater for all cases. A more practical approach has been to minimise exposure to external-features themselves. Frowd, Skelton et al. (2012) found superior composites emerged when face arrays within EvoFIT presented internal-features (with external-features masked), a technique that is now standard police practice (with external-features selected and added to the face towards the end of construction).

Once constructed by a witness, a facial composite is shown to police officers and members of the public to identify: a task requiring familiar-face recognition. Ellis et al. (1979) found that the complete face was best for the recognition of a familiar identity, but internal-features, composed of the eyes, brows, nose and mouth, were relatively more important than external-features for correct naming (see also Bonner, Burton, & Bruce, 2003; Clutterbuck & Johnston, 2002; Endo, Takahashi, & Maruyama, 1984; Young, Hay, McWeeny, Flude, & Ellis, 1985). Nevertheless, although external-features are often considered of secondary importance for familiar face recognition, they can increase false identification when inaccurate. In the 'Presidential Illusion', whereby internal-features of Al Gore were swapped with internal-features of Bill Clinton, some observers required a prompt

to recognise that Gore's face had changed (Sinha & Poggio, 1996; see also Andrews & Thompson, 2010; Carbon, 2008; Sinha & Poggio, 2002).

The present research investigated the importance of external-features for face construction, focusing on their veracity. An offender may change or disguise their hair (and / or other external-features) at the time of the crime and a composite that incorporates such a change should inhibit identification by those familiar with the offender (e.g., Patterson & Baddeley, 1977). Recognition may be inhibited for two reasons: first, changes made to external-features of a target should lead the constructor to produce more error-prone internal-features, and second, the composite when shown to a potential recogniser would now display hair different to how it would usually appear, altering the perception of its internal-features. Both effects are expected to occur due to the interactive nature of face perception, which likely has a strong influence upon how witnesses utilise information within the composite interface and how potential recognisers utilise information within the resulting composite. We present three experiments designed to investigate these issues.

In Experiment 1, we first establish that changing a target identity's hairstyle influences the way in which the internal-features of an ensuing composite are constructed. Here, we asked participants to construct a composite of an unfamiliar target. To do this, participants viewed an unaltered target photograph or a target photograph where we had changed the hair (i.e., exchanged the hair with that of another person). We then asked participants familiar with the target faces to attempt to name the internal region of these composites. It is found that the internal-features are more error-prone (i.e., less successfully identified) when constructed from targets with altered versus unaltered external-features.

Experiment 2 then manipulated the external-features of already constructed composites, and presented these images to potential recognisers. We used composites constructed of unaltered targets, as the internal-features of these faces were less prone to error

(as found in Experiment 1). We expected modifications to external-features to alter how internal-features would be perceived. In keeping with this suggestion, changing the hair, and so increasing the dissimilarity of hair to the original target, reduced correct naming rates (cf. presenting hair as originally constructed).

As part of Experiments 2 and 3, we also introduced manipulations to potentially overcome detrimental effects of inaccuracies in hair. Here, we selectively concealed aspects of the composite face known to provide prominent cues for successful face recognition: specifically, hair and eyes. We hypothesised that for composites constructed from targets with changed (vs. unaltered) hair, hair and eyes should be represented less accurately. Consistent with this idea, it was found that concealing these regions provided protection against decreases in correct naming when hair had been changed (cf. unaltered). Experiment 3 also reveals a novel method for increasing identification of composites for which target hair had not changed.

EvoFIT was used for the project since this system produces composites from human memory that attract modest naming rates after a 1-2 day forensically-relevant target-encoding delay (M = 30 - 50% correct, e.g., Frowd, Skelton et al., 2012; Frowd et al., 2013), thus facilitating a design that copied forensic procedures.

Experiment 1:

Experiment 1 was conducted in three stages. In Stage 1, participants ('constructors') each created a single EvoFIT composite of an unfamiliar target face from memory. Target faces were either presented unaltered or modified in one of three ways. First, hair was changed. Second, a hooded top was added that occluded hair, ears and face-shape. Third, external-features were removed (by cropping the image) to present the internal-features, including for these stimuli, face-shape. In Stage 2, participants who reported to be familiar

with the target identities (evaluators) were recruited to name the composites from their internal-features (which also included face-shape). In Stage 3, another group of participants (also 'evaluators') rated composites for likeness in the presence of target identities. It was expected that alterations to the targets' external-features, whereby hair was changed or a hooded top added would result in construction of less accurate/identifiable internal-features.

The inclusion of the internal-features condition was theoretically driven. The theory of encoding specificity (e.g., Tulving & Thomson, 1973) suggests that performance should be facilitated when a task presents overlap in the way that stimuli are encoded and later processed. For targets where the external-features were removed, processing overlap should now be maximal when constructing EvoFIT composites. This is because the EvoFIT system similarly presents faces with external-features removed during the early stages of construction, thereby requiring participants to focus on overall face shape and internal-features during selection from face arrays. Facial shape, visible on these additional target stimuli, is also a factor known to be important for face perception and construction (e.g., Ellis, 1986). It was therefore anticipated that composite construction would be most accurate in this internal-features condition.

Stage 1: Face construction

Method

  Constructors. Forty participants, recruited on the basis of being *unfamiliar* with international-level football, were opportunity sampled from the University of Central Lancashire and compensated with course credit (30 female, ages from 18-35 years, $M_{age} =$ 24.2; $SD_{age} = 7.9$). Ten participants were allocated to each of the four conditions in the experiment.

Materials. In the unaltered condition, participants were presented at encoding with a good quality, front-facing facial photograph of one of 10 footballers who play at an international level in the UK. All targets had a near-neutral expression, minimal stubble, fairly short hair and did not wear glasses or jewellery. All photographs were presented on a white background.

For the three remaining conditions, Adobe Photoshop was used to carefully modify the target photographs in three ways to create the hair changed, hooded top added and internal-features conditions. First, for the 'hair changed' condition, the target photographs were randomly paired and editing tools were used to plausibly exchange hairstyles within each pair. Second, a hooded top was added to each target, which occluded hair, shape of face, and other external-features.

To test the contribution of the match between stimuli at encoding and test (i.e., encoding-specificity) in facilitating composite construction, a third set of targets were created by removing/cropping external-features, leaving face-shape and internal-features intact. Presenting targets with internal-features (incl. face shape) would now match presentation of face stimuli within the first stage of EvoFIT construction, during which participants' view arrays of internal-features and select overall matches to the target face in memory.

The resulting 40 target photographs (10 unaltered; 30 edited) were printed in greyscale at 6cm (wide) x 8cm (high).

Composites were created using EvoFIT (software version 1.3), deployed on a PC. Currently, EvoFIT is used by many police forces within the UK and has been subject to formal police trials (e.g., Frowd, Hancock et al., 2011; Frowd, Pitchford et al., 2012). We note that EvoFIT relies upon the development of a greyscale composite image, whilst other available recognition-based systems can produce colour-based images (e.g., EFIT-V/EFIT-6, EigenFIT and ID). However, past research reveals that the use of colour does not produce

extra benefit to composite identification overall, but leads to slower production of the face (Frowd et al., 2006).

Design and Procedure.

Participants were randomly assigned to a 4(target type: unaltered image, internal-features, hooded top added, and hair changed) between-participants design. The ten unique identities were each constructed by a single participant in each of the four conditions, giving rise to 10 unique composites per condition (40 in total; see Fig. 1 for example composites).

Participants (constructors) were tested individually and worked at their own pace. They viewed a target photograph for 60 seconds in the knowledge that they would construct a composite of this face 22-26 hours later. A second target was selected if participants, when asked, indicated that they recognised the first identity. Photograph selection was random, without replacement, for target type (unaltered, hair changed, hooded top added and internal-features). Following the retention interval, the researcher (who was unaware of the targets' identity) worked with each participant to construct a single EvoFIT composite.

The procedure for creating a composite when using the EvoFIT system is described in detail elsewhere (Fodarella, Kuivaniemi-Smith, Gawrylowicz, & Frowd, 2015), and so we provide a summary here. The procedure mimics standard procedures adopted within real-world witnessing situations. Adhering to Cognitive Interviewing techniques (e.g., Geiselman, Fisher, MacKinnon, & Holland, 1986), participants were invited to visualise the target face and then freely describe it in detail. Next, participants were asked to select overall matches (to the target) from arrays of internal-features, first for facial shapes and then for facial textures, with selected items 'evolved' together over two generations. Participants then attempted to improve the likeness by changing (i) age, weight, health and other global facial properties, and (ii) the shape and placement of facial features. External-features were added

to the face, except for participants who had viewed internal-features targets and targets where a hooded top was added, and any additional global or featural adjustments made. The resulting composite (both with and without external-features present) was saved. Interviewing and composite construction took approximately 45 minutes, including debriefing.

Stage 2: Naming of composites' internal-features

Method

 Evaluators. Forty participants were recruited from the local area in Lancashire, UK (29 males; aged 18-56 years, $M_{age} = 26.7$, $SD_{age} = 7.7$). All self-reported to be fans of international-level football.

 Materials, Design and Procedure.

 Internal-features composites were printed in greyscale sized to 6cm x 8cm, and printed one per page. As participant testing took place in a public setting, the task was made faster to complete: the 40 constructed composites (10 per condition) were distributed over five booklets. Participants were then randomly assigned with equal sampling to each booklet (8 participants per booklet). Each booklet contained eight composites, and each of these eight composites corresponded to a unique target identity. Within each booklet, two composites were sampled from each of the four conditions (with composite identities and conditions rotated around booklets). Thus, here we applied a 4(target type: unaltered image, internal-features, hooded top added, and hair changed) within-participants design.

Fig. 1 about here

Participants (evaluators) were tested individually and tasks were self-paced. Each participant completed three separate tasks in the following order: spontaneous composite naming, target photograph naming and cued composite naming. First, participants were told that composites had been constructed to resemble footballers playing at international level in the UK. Similarly, in real-world situations, when composites are circulated they are usually accompanied by contextual information that may help aid identification (e.g., modus operandi, offender's accent, height and build; Frowd, Bruce, Ness et al., 2007). Participants were instructed to give a name to a composite where possible, even if unsure (i.e., guessing was encouraged), and that "don't know" responses were acceptable. After spontaneously naming their assigned set of composites, participants named photographs of the corresponding target footballers (i.e., the unaltered targets). This second task confirmed familiarity with the relevant identities. A third task was used to boost naming rates, as composites presented without their external-features tend to be poorly named (Frowd, Skelton et al., 2011). In this third task, participants attempted to name the composites for a second time, presented in the same item order as before. On this second occasion, having just completed the task of target-photograph naming, participants were cued as to the relevant target identities to which the composites corresponded. Cued naming is a good proxy for spontaneous naming (e.g., Frowd, Bruce, Ross, McIntyre, & Hancock, 2007) and is somewhat representative of a situation in which a police officer views a composite in the context of accompanying information (such as modus operandi) which serves to bring a selective pool of known offenders to mind.

Each participant received a different random presentation order for composites and target photographs. The task was completed in about 15 minutes, including debriefing.

Results

Target photograph naming was scored for accuracy: a numeric value of '1' for an accurate name and '0' otherwise. Target photographs were consistently named correctly (*M* = 89.0%, *SD* = 9.0%), indicating appropriate levels of familiarity amongst participants. The same scoring procedure was applied to composites, across the spontaneous- and cued-naming tasks. When a target photograph was not correctly named (*N* = 33), participants would have been unable to correctly name the associated composite; and so responses to such associated composites were treated as missing data and not analysed further.

From Table 1, it can be seen, in both naming tasks, that accurate (correct) naming responses were by far the highest overall for composites constructed of unaltered targets.

Table 1 about here.

Correct naming responses were analysed using the Generalized Estimating Equations' (GEE) function within IBM SPSS Statistics. GEE is a regression-type approach that accommodates the presence of a within-participant factor. This approach is more statistically powerful than ANOVA given that it provides a combined by-participants and by-items model (Ballinger, 2004). As the DV (correct naming) is dichotomous, a binary logistic link function was used, and cell frequencies were initially checked as meeting the appropriate assumptions for this Chi-Square type of analysis [*f(observed)* > 0 and *f(expected)* ≥ 5 for at least 20% of cells]. We chose a regression analysis that could detect variables whose influence is determined by the presence of other variables (suppressor variables). Therefore, we built a saturated model with predictors subject to backward sequential elimination; at each step, the variable possessing the smallest partial correlation with correct naming (the DV) was removed (based on the established standard criteria of *p* > .1 and lowest X^2). The model was built containing two factors, target type (coded as 1 = unaltered, 2 = internal-features, 3 =

hooded top added, and 4 = hair changed) and naming task (1 = spontaneous and 2 = cued naming) and their interaction. For the Working Correlation Matrix, an Exchangeable structure was selected to model repeated observations for the within-participants factor, naming task. When built, standard errors (SE) of Beta (*B*) coefficients were inspected for markers of model instability (of which none were present).

The analysis indicated that the two main predictors were reliable (*p* < .1); however, the interaction between the two was not (*p* = .34), indicating consistent effects occurred for both spontaneous and cued naming, and so this term was removed. The resulting model emerged reliable for target type, and parameter estimates relative to the unaltered condition revealed that correct naming was significantly lower (*p* < .05) for composites constructed from targets in all other conditions: internal-features [*B* = -1.7, *SE(B)* = 0.4, 1/*Exp(B)* = 5.3], hooded top added [*B* = -0.7, *SE(B)* = 0.3, 1/*Exp(B)* = 2.1] and changed hair [*B* = -1.5, *SE(B)* = 0.4, 1/*Exp(B)* = 4.5]. Naming task was also reliable as correct naming was higher (as expected) in cued than spontaneous tasks [*B* = 0.7, *SE(B)* = 0.3, *p* = .011, *Exp(B)* = 2.1].

For the spontaneous naming task, we also examined the prevalence of incorrect names (i.e., when a composite gives rise to a wrong name). Inaccurate naming rates are indicative of a tendency to guess (an increase in response bias) and lower values indicate more effective composites. The DV was coded as '1' for an incorrect name and '0' for "don't know" responses. The analysis focused on incorrect names, and thus both correct responses and responses to composites for which the target was not correctly named were excluded. There was a high level of inaccurate naming (*M* = 77.4%; 205 incorrect names given out of 265 possible attempts), likely caused by the absence of external-features, which themselves provide cues to identity (and thus reduce guessing). A Binary Logistic Regression model (applicable as no within-participants factor was involved) emerged reliable for target type: composites constructed from targets with changed hair attracted higher incorrect names than

those constructed from unaltered targets [$B = 1.1$, $SE(B) = 0.5$, $p = .043$, $Exp(B) = 2.9$]; there were no other reliable contrasts relative to the unaltered condition.

Stage 3: Likeness ratings of composites' internal-features

Method

 Evaluators. Twenty staff and student volunteers at the University of Central Lancashire provided likeness ratings (13 female; aged 20-47 years; $M_{age} = 25.8$; $SD_{age} = 8.6$). Participants reported being unfamiliar with the footballer identities.

  Materials, Design and Procedure.

      We applied a 4(target type: unaltered image, internal-features, hooded top added, and hair changed) x 16(scale type) within-participants design. Participants rated internal-features composites for likeness to the corresponding target photograph (1 = very poor likeness .. 7 = very good likeness) on 16 separate rating scales (e.g., eye and eyebrow area, mouth; see Table 2 for the full list). To keep the task tractable, each participant only rated a small subset of composites, four in total. To achieve this objective, for each participant one composite was taken from each of the four between-participants conditions in the experiment, but of the same target identity, randomly selected without replacement. For the 20 participants recruited, this design resulted in two complete sets of ratings for the stimulus set. Prior to giving likeness ratings, participants indicated whether they were familiar with the target identity and, if so, were shown composites constructed to resemble a different target identity. The procedure took about 20 minutes per person, including debriefing. Participants were tested individually and the task was self-paced.

Results

Likeness ratings are displayed in Table 2. Individual participant ratings (four sets of ratings per target) were subject to Generalized Estimating Equations (GEE). As for the analyses of composite naming in Stage 2, an Exchangeable Working Correlation Matrix was used (to model repeated observations); the link function was again logistic, but ordinal (to model Likert scale data); and here, suitably low emerging B and $SE(B)$ values meant that a model-based estimator was appropriate.

Table 2 about here.

Again, a saturated model was built with predictors subject to backward sequential elimination (removal based on $p > .1$ and lowest X^2). The model was built containing two factors, target type (coded as 1 = unaltered, 2 = internal-features, 3 = hooded top added, and 4 = hair changed, as in Stage 2) and scale type (consisting of the 16 unique rating scales, coded from 1 to 16; see Table 2), and their interaction. The analysis indicated that the two main predictors were reliable ($ps \leq .001$); however, the interaction between the two was not ($p = .12$), indicating consistent effects occurred across scale types, and so the term was removed. Simple contrasts (collapsing over scale type) revealed that, relative to the unaltered condition, likeness ratings were reliably lower for composites constructed of a target with changed hair [$B = -0.4$, $SE(B) = 0.1$, $p = .001$, $1/Exp(B) = 1.5$] and hooded top added [$B = -0.8$, $SE(B) = 0.1$, $p < .001$, $1/Exp(B) = 2.2$]. There was no reliable effect for composites constructed to resemble internal-feature targets ($p = .12$).

The eye region has been shown to be important for familiar face recognition (e.g., see Frowd, Bruce, Smith, & Hancock, 2008; Roberts & Bruce, 1988). To check whether construction of the eye region (vs. other facial regions) was more sensitive to the effects of our external-feature manipulations, we considered two scales in more detail, the eye region

(scale 9) versus the mouth region (scale 12). The model was built containing two factors, target type (coded from 1 to 4, as above) and scale type (eyes vs. mouth), and their interaction. The interaction term was again removed from the resulting GEE model, as was scale type (eye and mouth ratings did not differ significantly). Target type was a reliable predictor [$X^2(3) = 12.1$, $p = .007$]; relative to composites constructed from unaltered targets, worse likeness ratings were given to composites constructed from targets with changed hair [$B = -0.8$, $SE(B) = 0.3$, $p = .039$, $1/Exp(B) = 2.2$], hooded top added [$B = -0.8$, $SE(B) = 0.3$, $p = .013$, $1/Exp(B) = 2.1$] and internal-feature targets [$B = -1.0$, $SE(B) = 0.3$, $p = .002$, $1/Exp(B) = 2.6$]. Thus, any changes made to the appearance of a target face at the time of encoding led to the construction of composites receiving measurably lower likeness ratings for both an upper and a lower constructed facial feature.

Discussion

In Experiment 1, modifying target appearance interfered with participants' ability to construct the internal-feature region (cf. unaltered target condition). Consequently, these partial composites attracted consistently lower correct naming rates, as measured by both spontaneous and cued naming. Targets with changed hair also attracted a significantly higher proportion of incorrect names. Likeness ratings provided further support, showing that the eye and mouth regions of composites in the experimental conditions were less accurately constructed.

Findings concerning composites constructed from internal-feature targets require further discussion. Based on the theory of encoding specificity, best performance was expected in this condition due to the match in internal-features focus during encoding and face construction. However, correct naming rates and likeness ratings relating specifically to the eye and mouth regions for these composites were inferior. Arguably, removing external

feature cues (i.e. hair) changes perception of the target's internal-features, presumably in ways similar to the two other experimental conditions, and composite construction therefore suffered.

Indeed, previous research has shown the success of unfamiliar face recognition to be influenced by the context in which a face is perceived. For example, changes in facial appearance, such as angle-of-view and facial expression, impair recognition (e.g., Davies & Milne, 1982). External-features (hair, forehead, ears and neck) have similarly been highlighted as important contextual cues, particularly in terms of supporting our ability to successfully recognise (e.g., Ellis, Shepherd, & Davies, 1979) or match (Bruce et al., 1999) internal-features. Our findings indicate that exposure to the undisguised external-features of a perpetrator can provide useful contextual cues for another type of task: face construction. Here, exposure to undisguised external-features allowed participant-constructors to more accurately construct the internal region of the face, making it more likely to be recognised by those familiar with the typical appearance of that person.

When composites are constructed from targets (with visible external-features) a further issue emerges. If an offender has changed the appearance of his or her hair, then exposing potential recognisers to this inaccurately constructed feature may also negatively affect identification of a composite as a whole. We explore this possibility in Experiment 2.

Experiment 2:

Experiment 2 measures recognition (naming) rates when external alterations are applied to composites that themselves are constructed from unaltered target photographs. Correct naming rates in Experiment 1 showed these composites to possess less error-prone central regions compared to those constructed from targets with altered external-features.

Even so, we anticipate that correct naming would suffer when these internal-features are presented with changes to hair, as facial features are processed interactively and changes to external-features influence how internal-features are perceived (cf. Sinha & Poggio, 1996; 2002). To test this proposal, we compare recognition rates of these composites (i.e., unaltered composites including hair as constructed) with composites with changed hair. We also include an 'exact hair' condition, where hair from the corresponding original target photograph is copied onto the composite. This condition is of theoretical value: if reducing the accuracy of the composite's hair decreases correct naming, then improving the likeness should increase correct naming compared to the unaltered composite.

We also examine whether realistic occlusion techniques could effectively conceal inaccuracies in the construction of hair (i.e. as portrayed by the unaltered composite). We include two separate conditions whereby external-features of the composites are concealed by use of either a hat or hooded top. Here, occlusion may reduce the distracting influence of any inaccuracies in hair and allow cues to identity to be more effectively extracted from the composites' internal-features.

In comparison to composites, which are naturally imperfect facial representations, photographs represent a veridical image of the person. Despite this, previous research has found that modifying external-features on photographs of familiar faces leads internal-features to be misrecognised (e.g., Sinha & Poggio, 1996). This finding is a powerful demonstration of how face recognition is a holistic process, whereby individual features and the relationships between features are perceived interactively (e.g., Tanaka & Farah, 1993): changing a feature can influence how other properties of a face are perceived. Here, we applied the same external-feature modifications to photographs of the target identities as we had done to their corresponding composites. As for composites, we anticipate that modifications leading to the inaccurate portrayal of hair (i.e., hair changed) would impair

19

correct naming rates. For these stimuli, modifications to external-features would alter the perception of internal-features, and so both external- and internal-features would now less adequately match information in memory about the target individual. However, unlike the unaltered composites, unaltered photographs present accurate representations of hair. Thus, we expect that occluding hair within the photographs (i.e., by adding a hat or hooded top) will impair rather than benefit correct naming rates. In this case, occlusion of accurate representation of hair likely removes helpful cues to identity. These outcomes would provide further evidence that external-features alter how internal-features are perceived.

Method

  Evaluators. Forty football fans volunteered from the University of Central Lancashire (35 male; aged from 18-50 years; $M_{age} = 26.5$; $SD_{age} = 8.9$). We assigned participants, with equal sampling, to one of five between-participant conditions.

  Materials and Design. We used a 5(external-feature presentation: unaltered image, hair changed, exact hair, hooded top added, hat added) x 2 target type (composite vs. photograph) design. External-feature presentation was a between-participants variable with eight participants viewing images within each of these five conditions.

To produce the composite images, we used the 10 complete composites constructed in the 'unaltered target' condition in Experiment 1. The unaltered composites formed materials for our first condition. The composites were then modified to produce four further conditions (see examples in Fig. 2). For the second condition, hair was exchanged within composite pairs; here, we re-used the identity pairings from the hair changed condition in Experiment 1. For the third, the exact hair from the original target photograph was copied onto the corresponding composite. For the fourth condition, a hooded top was added to conceal all

external-features, and for the fifth, a dark-grey woolly hat was added to conceal almost all of the hair.

Participants also attempted to name images as photographs. We took the 10 target photographs corresponding to the composite identities as our unaltered images. We then applied the same modifications to these facial photographs as we had for the composite images. The task involving modified photographs was called 'initial target naming' and included the same five external-feature presentation conditions as for composite naming. Note, however, that for photographs, unlike composites, the images within the exact hair condition replicated those of the unaltered condition. Participants only named sets of composites and targets that had been modified in the same way (i.e., participants who named composites edited with a hat would also see facial photographs with a hat).

Fig. 2 about here.

Procedure. Participants individually completed a self-paced, three-stage naming task. First, participants attempted to name complete composites in one of the five external-feature presentation conditions. Participants were informed that (i) each composite had been constructed to resemble an international level footballer playing for the UK, and (ii) guesses and "don't know" responses were acceptable. Second, participants were asked to name target photographs (the 'initial target naming' task). Third, to check target familiarity, all participants saw the 10 unaltered target photographs and attempted to name them. Here, participants assigned to unaltered and exact hair conditions in the 'initial target naming' task would see duplicate target photographs; these participants were asked to name the targets again in the third stage of the task. Each participant encountered composites and target

stimuli presented in a different random sequential order. The task took 15 minutes to complete, including debriefing.

Results

The three sets of responses (composite naming, initial target naming and unaltered target naming) were scored for accuracy (Table 3). Participants scored well on the third task, unaltered target naming ($M = 94.8\%$, $SD = 5.5\%$), again suggesting appropriate levels of target familiarity.

Correct responses to composites and photographs presented in the initial target naming task were analysed separately. Thus, for the purposes of analysis, target type was not treated as repeated measures factor and Binary Logistic Regression was applied. The analysis included the between-participants variable external-feature presentation, consisting of five levels. Cases were again screened out for inaccurately named (unaltered) target identities ($N = 21$), and checks conducted for observed and expected response frequencies, and model parameter values. External-feature presentation was reliable, and simple contrasts revealed that unaltered composites attracted significantly more correct names than composites with hair changed [$B = 0.8$, $SE(B) = 0.4$, $p = .048$, $Exp(B) = 2.1$] and hooded top added [$B = 1.5$, $SE(B) = 0.4$, $p = .001$, $Exp(B) = 4.4$], but there were no reliable differences when compared to composites with a hat added ($p = .21$) or presented with exact hair ($p = .70$).

Further regression analyses specifically addressed differences in naming performance for unaltered composites, and those with a hat added or hair changed (conditions entered in this order). The analysis emerged with polynomial contrasts reliable as a linear [$B = 0.5$, $SE(B) = 0.3$, $p = .048$, $Exp(B) = 1.7$], but not as a quadratic trend ($p = .81$). In sum, naming rates for composites with a hat added fell midway between the other two types of composite,

suggesting that a hat might protect somewhat against a decrease in naming when hair had been changed.

Rates of inaccurate naming were also subject to Binary Logistic Regression (Table 3). External-feature presentation was reliable, and simple contrasts revealed that incorrect naming rates were higher for composites with hair changed relative to unaltered stimuli [$B = 0.9$, $SE(B) = 0.4$, $p = .024$, $Exp(B) = 2.2$].

Last, we analysed responses in the initial target photograph naming task. External-feature presentation emerged as reliable: unaltered targets were correctly named more often than targets with a hat added [$B = 2.5$, $Exp(B) = 11.8$], hooded top added [$B = 3.4$, $Exp(B) = 31.7$] or with hair changed [$B = 2.9$, $Exp(B) = 17.7$]; also, as expected, there was no reliable difference between unaltered targets and those with exact hair, as both stimulus-sets used the same images ($p = .54$; see Table 3 for mean naming rates). A polynomial trend analysis (as above) was reliable both as a linear ($p < .001$) and quadratic ($p < .001$) trend. Thus naming levels for targets with a hat added were no longer simply midway between the unaltered and hair changed conditions (as observed for composites).

Table 3 about here.

Discussion

Contrary to expectation, concealing a composite's external-features using a hooded top led to a sizeable *decrease* in correct naming. A smaller decrease was observed when a hat was added, suggesting that maintaining other external cues (available with this item of clothing) does not greatly interfere with recognition. It may be that adding a hat could help if hair has been changed, a manipulation itself found here to have a negative impact upon correct naming. There was no reliable increase in correct naming when composites were

presented with 'exact' hair. This suggests that constructors in Experiment 1 were able to accurately recreate this external feature 22-26 hours after viewing an unaltered target photograph (a forensically-relevant delay; see also Frowd, Bruce, McIntyre, & Hancock, 2007). Replicating previous research, external-feature modifications decreased correct naming for photographs, that is, veridical images of the target identities (e.g., Sinha & Poggio, 1996, 2002).

Experiment 3:

Experiment 2 demonstrated that alterations to hair reduce correct naming rates for whole-face composites, likely because they impair the effectiveness with which a composite's internal-features can cue identification. Experiment 1 further demonstrated that when a target has altered external-features at the time of encoding, the resulting composite a witness (or victim) produces likely contains more error-prone internal-features. Experiment 3 attempts to marry and extend these findings. We expected that whole-face composites constructed from targets with changed (cf. unaltered) hair at encoding would attract lower correct naming rates due to the higher inclusion of error-prone external- and internal-features. However, given our findings in Experiment 2, we anticipated that recognition rates may actually improve if inaccurate features of composites are realistically occluded. This was examined in two ways. Firstly by covering inaccurate hair, with a hat. Findings from Experiment 2 suggest that a hat might somewhat protect against a decrease in naming when hair has been changed (cf. unaltered). Secondly, as internal-features construction suffers when participants construct a composite from a target with changed (vs. unaltered) hair, we examine whether occlusion of an internal-feature, known to be important for familiar face recognition, would mitigate against a corresponding decrease in naming. The eye region is considered of primary importance for familiar face recognition (e.g., Frowd, Bruce, Smith, &

Hancock, 2008; Roberts & Bruce, 1988). Recognition deficits ensue when eyes are concealed

with (sun) glasses, both when a face is presented as a photograph (e.g., Patterson & Baddeley,

1977; Righi, Peissig, & Tarr, 2012; Yarmey, 2004) and composite (McIntyre et al., 2010).

Given the importance of the eye region for face perception, we anticipated that concealing

inaccuracies within the eye region, with a pair of sunglasses, might also protect against the

decrease in naming associated with composites constructed of hair changed (cf. unaltered)

targets.

Stage 1: Face construction

Method

  Constructors. Sixteen Psychology students, unfamiliar with the BBC TV soap EastEnders,

from the University of Winchester constructed composites for course credit (12 female, aged

18-42 years, M_{age} = 21.9; SD_{age} = 5.6). We assigned participants, with equal sampling, to one

of two between-participant conditions.

  Design, Materials and Procedure. We used a 2(accuracy of hair: unaltered vs. changed)

between-participants design. Eight male EastEnders's characters were selected as new target

identities. The set comprised (i) unaltered photographs, and (ii) these same images with

changed hair (exchanged within pairs of identities, as in Experiment 1). As before, target

photographs were good quality frontal images located on the Internet; hair varied in style and

colour, but length was consistently short. Images were printed as before.

       The 16 target images were constructed as EvoFIT composites using the procedure

described in Experiment 1. Constructors reported to be unfamiliar with their (randomly)

assigned target identity.

Stage 2: Naming of whole face composites

  Evaluators. Thirty-two Psychology students, familiar with EastEnders, from the University of Winchester attempted to name composites for course credit (24 females; aged from 18-42 years, $M_{age} = 21.6$; $SD_{age} = 5.0$). Participants were assigned, with equal sampling, to one of two between-participant conditions.

  Design, Materials and Procedure. A 2(accuracy of hair: unaltered vs. changed) $\times$ 2 (image format: single image vs. array) design was used, with accuracy of hair as a between-participants factor. The 16 composites produced in Stage 1 were allocated to each of two between-participants conditions: 8 composites constructed from an unaltered target photograph, and 8 composites constructed from a hair changed target photograph. The single image condition presented these original composites. For the image array condition, these composites were modified using Adobe Photoshop in three different ways, by adding (i) a woollen hat (see Experiment 2), (ii) a pair of opaque sunglasses, and (iii) both hat and sunglasses.   Sunglasses were obtained from a facial-composite system and edited to conceal eyes, eye sockets and the lower extent of the eyebrows. A four-image array was then created, per target identity, to include: the original composite (constructed from an unaltered or hair changed target photograph), and the three newly created images (the original composite wearing a hat, sunglasses, or both a hat and sunglasses; see Fig. 3).

Fig. 3. about here

        Evaluators viewed composites presented in two different formats, as a single, complete image and within a four-image array (array construction described above). Array images were expected to differ in effectiveness dependent on whether the original composite

had been constructed from an unaltered or hair changed target image in Stage 1. For composites constructed from targets with changed (inaccurate) hair, the three edited images should contain less error overall (cf. the original composite), facilitating naming. However, when composites had been constructed from unaltered target images, the original composite was expected to be the most effective image in the array, containing less construction error. Here, selective concealments may hide accurate featural information and reduce identifiability (see related findings with photographic stimuli: Badham, Wade, Watts, Woods, & Maylor, 2013; Henderson, Bruce, & Burton, 2001, for unfamiliar faces and for familiar faces see Sadr, Jarudi, & Sinha, 2003; and our findings in Experiment 2). To explore separate hypotheses for unaltered and altered target arrays, participants were asked to indicate which item in the array appeared to help them recognise the face.

Presentation order of the single image and array format was also manipulated. All evaluators viewed composite identities as a single image, and an array, but half of participants first saw the single image, followed by the array, and the other half viewed the reversed presentation.

Participants familiar with EastEnders's characters were tested individually in a self-paced task. They were randomly assigned, with equal sampling, to view a set of composites, either constructed from targets with unaltered or changed hair. Composites were presented to participants both as a single image and as part of an array, order randomised across participants. Participants attempted to name each composite sequentially; guesses were encouraged and "don't know" responses permitted. For array presentation, participants were informed that the target may have changed appearance and that one of the four images might be helpful for recognition. Participants were asked to indicate which array image had been most useful for identification. After viewing composites in both formats, participants attempted to name unaltered target photographs. Participants received a different random

presentation order for composites and target photographs. Due to anticipated good naming of composites, cued naming did not ensue.

The task took around 15 minutes to complete, including debriefing.

Results

Composite naming

Participant responses to composites and target photographs were scored for accuracy. Target naming was high overall ($M = 94.9\%$, $SD = 8.9\%$), and responses to composites were screened for cases ($N = 13$) where the associated target had not been named correctly.

Correct naming rates are presented in Table 4. We first analysed participant responses to the first set of composites that they saw (whether presented as single images or within arrays). Here, we initially explored the influence of accuracy of target hair (unaltered vs. changed) and image format (single image vs. array), treating both as between-participant factors. We were able to do this legitimately given that presentation order was counterbalanced: half of participants first saw single image composites and half arrays, as we described above. For composites constructed from targets with changed hair, we expected exposure to the original composite plus accompanying concealments (i.e. the array) to improve identification rates compared to the original composite alone (single image). Scored responses to composites were analysed by GEE (to accommodate subsequent analyses, which include image format as a within-participant factor, see below). As before, a binary logistic link function was used to model naming responses ('0' = incorrect and '1' = correct).

Table 4 about here.

The analysis proceeded with a saturated model, containing the accuracy of target hair (unaltered vs. changed) and image format (single images vs. arrays) and their interaction. GEE was reliable for (i) accuracy of target hair [$X^2(1) = 21.0$, $p < .001$], with more identifiable composites constructed from target photographs with unaltered versus changed hair [$B = 1.2$, $SE(B) = 0.2$, $p < .001$, $Exp(B) = 3.2$], and (ii) image format [$X^2(1) = 14.0$, $p < .001$], with composites more frequently identified from arrays versus single images [$B = 1.0$, $SE(B) = 0.2$, $p < .001$, $Exp(B) = 2.8$]. The interaction between these two predictors was also significant [$X^2(1) = 6.3$, $p = .012$]: while there was a reliable array benefit (cf. single images) for composites constructed from targets with changed hair [$B = 1.0$, $SE(B) = 0.2$, $p < .001$, $Exp(B) = 2.8$], no such benefit emerged when target hair was unaltered ($p = .34$). Further, composites constructed from targets with changed hair were less identifiable as single images (cf. unaltered) [$B = -1.0$, $SE(B) = 0.2$, $p < .001$, $1/Exp(B) = 3.2$], but the two types of composite were recognised at a similar rate when presented within arrays ($p = .11$). Additional planned contrasts revealed no reliable differences for composites of unaltered targets, presented as a single image, and composites of targets with changed hair, presented as an array ($p = .51$). Importantly, this result suggests that the problems caused by inaccurate construction of target hair can be overcome when using the array presentation.

In subsequent analyses, we reran the analysis for all responses including image format (single image vs. array) and presentation order (first vs. second; see composite naming rates, by presentation order, in Table 4) as within-participant factors. The analysis assessed various predictions concerning array presentation: (i) that this presentation format would increase recognition rates when composites with inaccurate (changed) hair were seen first, (ii) that the array would offer no recognition benefit when unaltered composites were seen first, and (iii) that irrespective of target type (unaltered vs. changed hair), overall naming rates would not differ for participants who first saw arrays, then single images, as the second presentation

provided no additional information. The analysis proceeded with a saturated model, containing the three factors and their interaction. As predicted, the three-way interaction was significant [$X^2(7) = 30.6$, $p < .001$] and was explored by analyses of parameter estimates.

For composites constructed of targets with changed hair, there was a reliable naming benefit when single images were presented first, followed by arrays [$B = 1.0$, $SE(B) = 0.2$, $p = .001$, $Exp(B) = 2.7$]. However, a similar, albeit smaller benefit unexpectedly extended to composites constructed from unaltered targets [$B = 0.5$, $SE(B) = 0.2$, $p = .013$, $Exp(B) = 1.6$]. Irrespective of target type, naming did not improve at second presentation when arrays were presented first ($p = .73$ in both cases). These (null) contrasts suggest that, for unaltered targets, the benefit of seeing a single composite, followed by an array, is unlikely to be due to attempting composite naming more than once.

An analysis of inaccurate names revealed no significant differences. Therefore differences in accurate naming are *unlikely* to result from participants adopting different response criterion, due to (a) multiple image presentation, or (b) warnings that the face may have changed appearance.

Array items responsible for recognition

Participants indicated which array item led to recognition: the composite itself, the composite with hat, with sunglasses, or both hat and sunglasses (see Table 5). As expected, when composites were constructed from unaltered targets, participants reported that the original composite was most useful, irrespective of whether the array was presented first, or second. However, when composites were constructed from changed targets, composite images with a hat were most useful. For incorrect array selection, item choices were evenly distributed over the four composite types (Overall $M = 1.0$, $SD = 0.8$).

Table 5 about here.

Chi Square tests analysed correct responses attributed to the composite itself, and composite with hat (incl. Yates correction for continuity). First, differences in overall counts for composites constructed from unaltered (48 | 12) and changed targets (10 | 30) were reliable [$X^2(1) = 27.6$, $p < .001$, Odds Ratio $OR = 12.0$]: observers benefit from viewing the composite itself when target hair had not been altered, and composite with hat when altered. Also, for composites made from unaltered targets, the pattern of responses differed depending on whether the array was seen first (18 | 9) or second (30 | 3) [$X^2(1) = 4.0$, $p = .044$, $OR = 5.0$]. Here, the composite itself triggered recognition more often (cf. composite and hat) for an array seen second, presumably as the composite had just been seen. However, no reliable difference (4 | 14 vs. 6 | 16) emerged for composites constructed from changed targets ($p = 1.0$).

Discussion

Correct naming rates again demonstrate the negative impact of inaccurate (changed) target hair for composite construction. However, array presentation, including images wearing a hat, sunglasses, or hat and sunglasses, improved naming rates for composites constructed from changed targets. In particular, participants indicated that composites wearing a hat were most useful for recognition (but the benefit of adding sunglasses appeared to be fairly limited). Importantly, when arrays were presented first, correct naming was similar to that obtained for single composites that had been constructed from unaltered targets. For composites constructed of unaltered targets, we also found a reliable improvement in correct naming from seeing a single composite, followed by the array.

General Discussion

The current work extends research showing that facial features and their configuration are perceived interactively in upright, intact stimuli (e.g., Tanaka & Farah, 1993; Yasuda, 2005). This knowledge has previously driven development of human interfaces that allow construction to occur in the context of a whole face, or whole face regions. Here, we demonstrate a specific problem with this approach and highlight the complex interaction between witness and composite interface: When an offender has changed or disguised his or her external-features this alters how a witness utilises information within the interface. In Experiment 1, participants constructed composites from target photographs that were either (a) unaltered, or (b) subject to external-feature modifications (hair changed, hooded top added and internal-features, with external-features removed). The resulting internal-feature composites, constructed from modified targets, were less successfully named (i.e., across both spontaneous and cued tasks); these modifications reduced accurate composite construction. Likeness ratings, obtained for the composites' internal-features, support this pattern of results showing that eye and mouth regions were less accurately constructed.

The data further supports the powerful influence of external-features upon recognition of familiar face identities (e.g., Sinha & Poggio, 2002). Experiments 2 and 3 found a reduction in identification rate when changes to hair were applied to composites that were constructed from unaltered target photographs. These composites possessed less error-prone central regions (cf. those constructed from targets with altered external-features) and our findings extend past similar work with facial photographs (e.g., Chan & Ryan, 2012; Patterson & Baddeley, 1977). Critically, recognition rates were improved by realistically occluding inaccurate features, a technique that appears to have altered how potential recognisers interacted with visual information presented within the resulting composite.

Before committing a crime, offenders may deliberately alter their appearance to avoid later detection (e.g., Patterson & Baddeley, 1977). This may be a useful strategy: after a forensically-relevant delay, participants were worse to construct the internal-features region of a composite when the target's hair had been changed (Experiment 1), and this detriment extended to construction of whole-face composites, including both external and internal-features (Experiment 3). Reductions in correct identification were obtained even when modest modifications were made to the hair (i.e., hairstyles remained short and unremarkable, with no changes in hair colour).

Findings confirm a two-fold impact when target hair is modified: resulting composites are constructed with more error-prone (i) hair and (ii) internal-features. In Experiment 2, two manipulations attempted to overcome the former effect by concealing changes to hair. Addition of a hooded top markedly interfered with naming, perhaps because it also conceals important facial-shape information (Shepherd, Davies, & Ellis, 1981), but addition of a hat seemed to produce some naming benefit.

Experiment 3 extended this idea. Here, participants benefitted from seeing multiple representations of a composite that had been constructed to resemble an altered target. Composites in the resulting array featured selective concealments to hair, eyes, or both. When participants first viewed this composite array, they showed similar naming rates to those who viewed a single composite, or array, constructed from an unaltered target. So, while Experiment 2 revealed some benefit from presenting a composite with a hat, Experiment 3 showed that presenting an array of images that included concealments to the eyes and hair could overcome the deficit to correct naming (with no change to incorrect naming). Participants further indicated that the composite with hat was most beneficial in this case for naming. Thus it seems important to conceal the feature that is most likely to be inaccurately constructed, due to modification of the target photograph (hair). While this type

of target modification may also reduce the constructed likeness of the eyes, results suggest that the eye region may still provide some useful recognition cues, and that complete concealment of this feature is unnecessary.

Experiment 3 also assessed naming when an observer first saw single composites, then arrays, as might occur in a forensic setting. Here, the impact of prior exposure to a single composite can be assessed by comparing the benefit of arrays when measured as a between- and a within-participants' factor (see Table 4). Results suggest that arrays offer very similar benefit for same and different observers [*Exp(B)* = 2.8 vs. 2.7, respectively], and may facilitate identification whether or not the composite itself had been previously seen.

Further, for composites constructed of unaltered targets, an unexpected situation arose: when individual composites were presented first, they promoted a reliable (correct) naming *benefit* when arrays were presented second, emerging as a small effect size [*Exp(B)* = 1.6]. This result has two facets. First, if the participant had correctly named the single composite then they could simply repeat that name during a second presentation (a situation that occurred frequently; $N = 25 / 27$). Second, in the event that the participant had not correctly named the single composite, he or she could search the array for additional cues to identification[1].

Image manipulations were carried out carefully to extract and exchange hairstyles (using photographic elements in Adobe Photoshop) and to add hats or hooded tops (accessories available from a composite system). There is evidence that this was done acceptably, and that the manipulations conducted were plausible. First, some manipulations to composites did not change correct naming rates (see Experiment 2, exact hair vs. unaltered composite conditions). Second, some manipulations to target photographs did not alter

---

[1] Although not tested here, it is also conceivable that the array should promote better recognition of offenders who change their hair *after* a crime has taken place. In this situation, the composite's hair would be presumed accurate at the time of the crime, but not when identification is attempted. Thus, showing the composite with hat would usefully hide the feature that had been changed after face construction.

correct naming of the ensuing composite (e.g., see Experiment 3, first presentation: hair changed vs. unaltered arrays) or levels of incorrect naming (e.g., Experiment 1, hooded top added vs. unaltered composite conditions).   Further, self-report data from our participants in Experiment 3 suggest that adding a hat and (to a lesser extent) sunglasses can *sometimes* improve composite identification. This suggests that participants viewed these manipulated composites as plausible representations of familiar identities.

Hair and eyes are features particularly important for familiar face recognition, thus using selective concealments to hide any constructed inaccuracies may be helpful (e.g., O'Donnell & Bruce, 2001). This mechanism has potential to explain why the array works well when target hair is inaccurate: for some people, sometimes, the edited representations simply look better. Indeed, within Experiment 3, for changed targets, when seeing the array first, participants indicated that the unmodified composite was most helpful on only four occasions. In contrast, on 20 occasions adding a concealment (i.e., hat, sunglasses, or hat plus sunglasses) was reported to help recognise the face.

Our results may extend to other holistic human interfaces which operate similarly to EvoFIT (e.g., EFIT-V/EFIT6 and ID; Gibson et al., 2009; Tredoux, Nunez, Oxtoby, & Prag, 2006), although EvoFIT composites are estimated to be correctly named at a much higher percentage overall (Frowd et al., 2015), presumably due to the focus on internal-features at construction. Similarly, modern 'feature' methods of construction are also influenced by context effects (e.g., Skelton et al., 2015), and thus results should extend to these interfaces. However, this suggestion may be difficult to verify experimentally as feature composites are often poorly named when constructed after a forensically-relevant retention interval ($M \approx$ 5%; for a review, see Frowd et al., 2015). However, as hair and eyes are likely to be inaccurately constructed when using these more traditional systems, array presentation could still be a particularly useful format even when target hair is accurate (research in progress).

Our results clearly have implications for forensic practice: as it is unlikely to be known whether offenders have changed their hair (i.e., whether construction is based on inaccurate target hair), it would seem best to present the resulting composite in an array whenever possible. Hat and sunglasses are usual accessories in composite systems, and simple editing techniques can be used (as here) to apply them to a composite face. We advise accompanying the array with wording such as, "The alternatives shown may help you to recognise the person depicted". Findings from this project have been disseminated to UK and overseas police for use in criminal investigations. In a recent case, an EvoFIT composite was presented in different ways (incl. hat, beard and sunglasses) to brief officers hunting a murder suspect; this person has since been apprehended and charged (Farell-Roig, 2018).

The results also suggest a better procedure for law enforcement to identify composites, be it by identification within a police force or by carrying out a public appeal for information: to present the single composite initially, followed by the array. In this way, when an offender's hair has not changed (i.e., construction is based on accurate hair), naming should be facilitated (cf. presenting just the array). Conversely, if target hair is inaccurate, this double-stage procedure should prove as effective as presenting the array alone. The 'single image – array' combination is suitable for web browsers and smart phones, allowing a composite to be seen first and then observers (police and members of the public) invited to "click the face" to reveal alternatives. This method is straightforward for web administrators to implement; alternatively, it could easily be prepared and distributed using content presentation programs (e.g., PowerPoint).

References

Andrews, T.J., & Thompson, P. (2010). Face-to-face coalition. *i-Perception*, *1*, 28-30.

Badham, S. P., Wade, K. A., Watts, H. J. E., Woods, N. G., & Maylor, E. A. (2013). Replicating distinctive facial features in lineups: Identification performance in young versus older adults. *Psychonomic Bulletin & Review, 20*, 289-295.

Ballinger , G.A. (2004). Using Generalized Estimating Equations for Longitudinal Data Analysis. *Organizational Research Methods*, 7, 127-150.

Baudouin, J.Y., & Humphreys, G.W. (2006). Configural information in gender categorisation. *Perception*, *35*, 531-540.

Bonner, L., Burton, A.M., & Bruce, V. (2003). Getting to know you: how to learn new faces. *Visual Cognition*, *10*, 527-536.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P., Burton, A.M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied, 5* (4) , 339-360.

Bruce, V., & Young, A. (1998). *In the eye of the beholder: The science of face perception*. New York: Oxford University Press.

Calder, A.J., Young, A.W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception and performance*, *26*, 527–551.

Carbon, C.C. (2008). Famous faces as icons: the illusion of being an expert in the recognition of famous faces. *Perception, 37*, 801-806.

Chan, J.P.K., & Ryan, J.D. (2012). Holistic representations of internal and external face features are used to support recognition. *Frontiers in Psychology*, *87*, 1-10.

Chiller-Glaus, S.D., Schwaninger, A., & Hofer, F. (2007). Identity verification from photographs in travel documents: The role of display duration and orientation on performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 51* (19), 1327 – 1330.

Clutterbuck, R. & Johnston, R.A. (2002). Exploring Levels of Face Familiarity by Using an Indirect Face-Matching Measure. *Perception*, *31*, 985-994.

Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). The reliability of eyewitness identifications: The role of system and estimator variables. *Law and Human Behavior*, *11*, 223-258.

Davies, G. M., & Milne, A. (1982). Recognizing faces in and out of context. *Current Psychological Research, 2,* 235–246.

Diamond, R., & Cary, S. (1986). Why faces, are and are not special: an effect of expertise. *Journal of Experimental Psychology: General*, *11*, 107-117.

Ellis, H. D. (1986). Face recall: A psychological perspective. *Human Learning*, *5*, 1-8.

Ellis, H.D., Davies, G.M., & Shepherd, J.W. (1978). A critical examination of the photofit system for recalling faces, *Ergonomics*, 21, 4, 297-307.

Ellis, H.D., Shepherd, J.W., & Davies, G.M. (1979). Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. *Perception*, *8*, 431-439.

Endo, M., Takahashi, K., & Maruyama, K. (1984). Effects of observer's attitude on the familiarity of faces: using the difference in cue value between central and peripheral facial elements as an index of familiarity. *Tohoku Psychologica Folia*, *43*, 23-34.

Farell-Roig, E. (2018). Fifth man charged in connection with death of Simon Clark in Pendine appears in court. *Wales Online*, retrieved from www.walesonline.co.uk/news/wales-news/fifth-man-charged-connection-death-15352247 on 3rd November 2018.

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

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.

Frowd, C.D., Bruce, V., Ness, H., Bowie, L., Thomson-Bogner, C., Paterson, J., McIntyre, A., & Hancock, P.J.B. (2007). Parallel approaches to composite production. *Ergonomics*, *50*, 562-585.

Frowd, C.D., Bruce, V., Plenderleith, Y., & Hancock, P.J.B. (2006). Improving target identification using pairs of composite faces constructed by the same person. *Institution of Engineering and Technology Conference on Crime and Security,* London, pp. 386-395.

Frowd, C.D., 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*, 1-31.

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.

Frowd, C.D., Carson, D., Ness, H., McQuiston, D., Richardson, J., Baldwin, H., & Hancock, P.J.B. (2005). Contemporary Composite Techniques: the impact of a forensically-relevant target delay. *Legal & Criminological Psychology*, *10*, 63-81.

Frowd, C. D., Erickson, W. B., Lampinen, J. M., Skelton, F. C., McIntyre, A. H., & Hancock, P. J. (2015). A decade of evolving composites: regression-and meta-analysis. *Journal of*

Forensic Practice, *17*(4), 319–334.

Frowd, C.D., Hancock, P.J.B., Bruce, V., Skelton, F.C., Atherton, C.J., Nelson, L…&
Sendrea, G. (2011). Catching more offenders with EvoFIT facial composites: lab research
and police field trials, *Global Journal of Human-Social Science, 11 (3)*, 35-46.

Frowd, C.D., & Hepton, G. (2009). The benefit of hair for the construction of facial
composite images. *British Journal of Forensic Practice*, *11*, 15-25.

Frowd, C.D., Pitchford, M., Skelton, F.C., Petkovic, A., Prosser, C. & Coates, B. (2012),
"Catching even more offenders with EvoFIT facial composites", in Stoica, A., Zarzhitsky, D.,
Howells, G., Frowd, C., McDonald-Maier, K., Erdogan, A. and Arslan, T. (Eds), IEEE
Proceedings of 2012 Third International Conference on Emerging Security Technologies.
doi: 10.1109/EST.2012.26, pp. 20-6.

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

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

Frowd, C.D., Skelton F., Hepton, G., Holden, L., Minahil, S., Pitchford, M., McIntyre, A., Brown, C., & Hancock, P.J.B. (2013). Whole-face procedures for recovering facial images from memory. *Science & Justice*, *53*, 89-97.

Geiselman, R.E., Fisher, R.P., MacKinnon, D.P., & Holland, H.L. (1986). Eyewitness memory enhancement with the cognitive interview. *American Journal of Psychology, 99*, 385-401.

Gibson, S.J., Solomon, C.J., Maylin, M.I.S., & Clark, C. (2009). New methodology in facial composite construction: from theory to practice. *International Journal of Electronic Security and Digital Forensics*, *2*, 156-168.

Henderson, Z., Bruce, V., & Burton, A.M. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology*, *15*, 445-464.

Hole, G.J. (1994). Configurational factors in the perception of unfamiliar faces. *Perception*, *23*, 65-74.

Lee, E.S., and Whalen, T. E., (1996). Feature approaches to suspect identification: The effect of multiple raters on system performance, *Ergonomics, 39*, 17-34.

Lee, E., Whalen, T., Sakalauskas, J., Baigent, G., Bisesar, C., McCarthy, A., Reid, G., & Wotton, C. (2004) Suspect identification by facial features, *Ergonomics, 47*(7), 719-747.

Li, Y., Vishwamitra, N., Hu, H., Knijnenburg, B., & Caine, K. (2017). Effectiveness and users' experience of face blurring as a privacy protection for sharing photos via online social networks. *Proceedings of the Human Factors and Ergonomics Society, 61*, 803 – 807.

Maurer, D., Le Grand, R., & Mondloch, C.J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*, 255-260.

McIntyre, A., Frowd, C.D., Bruce, V., & Hancock, P.J.B. (2010). Looking at people who are wearing glasses: the impact of glasses on matching photographs of unfamiliar faces and recognizing familiar composite faces. In G. Howells, K. Sirlantzis, A. Stoica, T. Huntsberger and A.T. Arslan (Eds.) *2010 IEEE International Conference on Emerging Security Technologies* (pp. 36-43). ISBN 978-0-7695-4175-4.

O'Donnell, C., & Bruce, V. (2001). Familiarisation with faces selectively enhances sensitivity to changes made to the eyes. *Perception, 30*, 755-764.

Patterson, K.E., & Baddeley, A.D. (1977). When recognition fails. *Journal of Experimental Psychology: Human Learning & Memory*, *3*, 406-417.

Righi, G., Peissig, J. J., & Tarr, M. J. (2012). Recognizing disguised faces. *Visual Cognition, 20*, 143-169.

Roberts, T. & Bruce, V. (1988). Feature saliency in judging the sex and familiarity of faces. *Perception, 17*, 475-481.

Sadr, J., Jarudi, I., & Sinha, P. (2003). The role of eyebrows in face recognition. *Perception, 32*, 285-293.

Shepherd, J.W., Davies, G.M., & Ellis, H.D. (1981). Studies of cue saliency. In G.M. Davies, H.D. Ellis, J.W. Shepherd (Eds), *Perceiving and Remembering Faces*. (pp. 105-132). London: Academic Press.

Sinha, P., & Poggio, T. (1996). I think I know that face. *Nature*, *384*, 404.

Sinha, P., & Poggio, T. (2002). United we stand. *Perception*, *31*, 133-133.

Skelton, F.C., Frowd, C.D., & Speers, K. (2015). The benefit of context for facial-composite construction, *Journal of Forensic Practice, 17*, 281-290.

Tanaka, J.W., & Farah, M.J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *46A*, 225-245.

Tanaka, J.W., & Sengco, J.A. (1997). Features and their configuration in face recognition. *Memory & Cognition*, *25*, 583-592.

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.

Tulving, E., & Thomson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352-373.

Wogalter, M. S. (1996). Describing faces from memory: Accuracy and effects on subsequent recognition performance. *Proceedings of the Human Factors and Ergonomics Society*, 40, 536–540.

Yarmey, A. D. (2004). Eyewitness recall and photo identification: a field experiment. *Psychology, Crime & Law, 10*, 53-68.

Yasuda, T. (2005). Effects of configural information on face shape perception. *Japanese Journal of Psychology*, *76*, 131-138.

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.

Young, A.W., Hellawell, D., & Hay, D.C. (1987). Configurational information in face perception, *Perception*, *16*, 747-759.

List of figures and tables

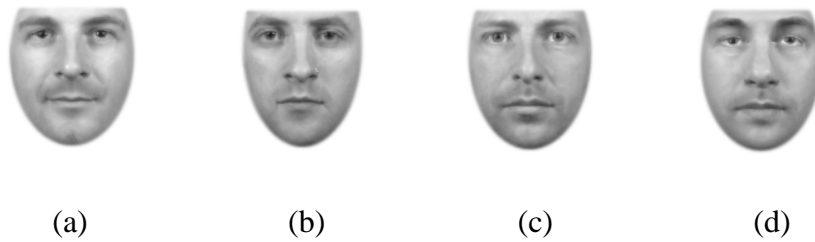   (a)   (b)   (c)   (d)

Fig. 1. Example internal-features composites constructed in Stage 1, 22-26 hours after the participant had seen: (a) an unaltered target, (b) a target wearing a hooded top, (c) a target with hair changed, and (d) an internal-features target. Composites were constructed to resemble UK football player, John Terry, and were used as part of the identification task in Stage 2. For reasons of copyright, we are unable to reveal the target photograph used, but a simple image search on the internet should easily reveal his appearance.

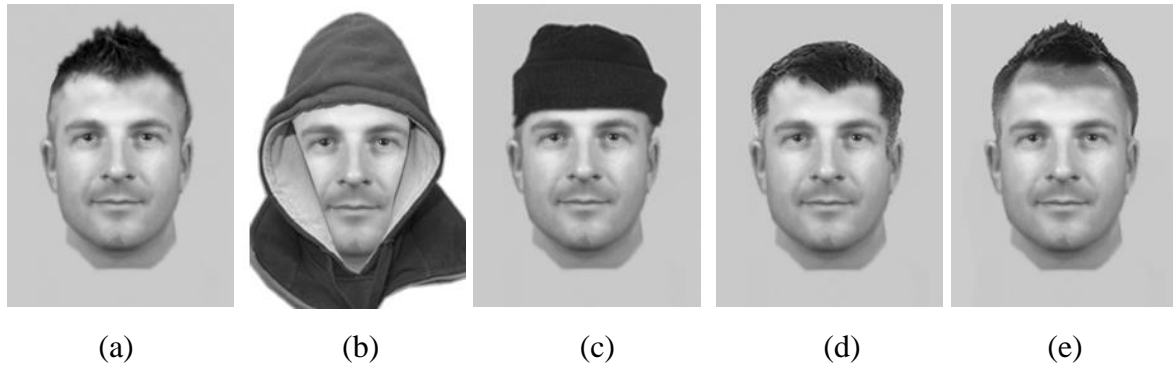|				
(a)	(b)	(c)	(d)	(e)

Fig. 2. Example whole-face composites used in the spontaneous-naming task in Experiment 2. Composite (a) was constructed from a target photograph of John Terry in Experiment 1, Stage 1; it is the same image of Fig. 1, far left, after the constructor had added hair. For Experiment 2, this composite was edited to create another composite image, wearing: (b) hooded top, (c) woolly hat, (d) hair from another composite in this set and (e) 'exact' hair (copied from the footballer's target photograph). The target photographs were manipulated in the exact same way and presented to participants to name after they had attempted to name the composites.

Fig. 3. Example composite constructed from memory by a participant in Experiment 3 of EastEnders' character Alfie Moon (played by actor Shane Richie), top left, and different representations, edited with selective concealments (woolly hat and sunglasses). Images were presented together as an array. In this example, the composite constructor initially saw a target photograph with inaccurate hair (a bald head in this case).

Table 1. The effect of target type on naming rates for internal-feature composites.

	Unaltered image	Hair changed	Hooded top added	Internal-features
Naming task				
Accurate				
Spontaneous	14.5 (10 / 69)	5.7 (4 / 70)	5.6 (4 / 72)	5.3 (4 / 76)
Cued	29.0 (20 / 69)	4.3 (3 / 70)	18.1 (13 / 72)	6.6 (5 / 76)
Mean	21.7 (30 / 138)	5.0** (7 / 140)	11.8* (17 / 144)	5.9** (9 / 152)
Inaccurate				
Spontaneous	78.0 (46 / 59)	93.9* (62 / 66)	72.1 (49 / 68)	66.7 (48 / 72)

Note. Figures for 'accurate' naming scores were calculated by dividing responses shown in parentheses, and expressed as a percentage. In parentheses, we present summed correct responses (numerator) and total (correct and incorrect) responses (denominator) for composites for which participants correctly named the relevant target photograph ($N = 287$ correct responses out of 320 attempts). For 'inaccurate' naming scores, values shown in parenthesis are summed incorrect responses (numerator) and total responses (incorrect and no name; denominator): note, here correct responses and responses to composites for which the target was not correctly named were excluded ($N = 265$ out of 320). Regression models for Spontaneous and Cued Accurate naming, and Spontaneous Inaccurate naming tasks [$X^2 > 6.5$, $p < .05$]: Presentation and the Constant ($p < .001$) were significant. Significantly different from Unaltered image: * $p < .05$, ** $p < .001$. See text for details.

Table 2. Likeness rating scales used to assess the composites' internal-features region.

Rating scale	Target type			
	Unaltered image	Hair changed	Hooded top added	Internal-features
1. Overall likeness (entire face)	3.3 (0.3)	2.8 (0.4)	2.6 (0.3)	3.0 (0.3)
2. Outline (incl. shape of jaw, width of face and [as far as possible] top of head)	3.5 (0.4)	3.3 (0.3)	2.7 (0.4)	4.2 (0.4)
3. Width of face	3.8 (0.4)	3.2 (0.4)	2.9 (0.4)	4.5 (0.4)
4. Internal area of the face (inner region including eyes, brows, nose and mouth)	3.8 (0.3)	3.3 (0.3)	3.0 (0.4)	3.6 (0.4)
5. Eye and brow area (internal area)	3.8 (0.4)	3.4 (0.4)	2.9 (0.3)	3.3 (0.4)
6. Nose and mouth area (internal area)	3.7 (0.3)	2.8 (0.3)	3.2 (0.3)	3.9 (0.3)
7. Facial hair (stubble on lower portion of face)	3.9 (0.3)	4.0 (0.4)	3.3 (0.4)	3.8 (0.4)
8. Brows	3.2 (0.4)	3.4 (0.4)	3.1 (0.4)	3.2 (0.4)
9. Eyes (including any eye bags)	3.3 (0.3)	2.8 (0.4)	3.0 (0.4)	2.4 (0.3)
10. Cheeks	3.7 (0.4)	3.6 (0.3)	2.6 (0.3)	3.3 (0.3)
11. Nose	3.8 (0.4)	2.9 (0.3)	2.8 (0.3)	3.2 (0.4)
12. Mouth	3.7 (0.3)	3.2 (0.4)	2.8 (0.3)	3.2 (0.4)
13. Chin region (chin and jaw but not mouth)	3.8 (0.4)	3.4 (0.3)	2.5 (0.3)	3.9 (0.4)
14. Spacing / proportions between facial features (relative distances in the internal area)	4.4 (0.3)	3.8 (0.4)	2.9 (0.3)	4.1 (0.3)
15. Character or personality	3.7 (0.3)	3.8 (0.4)	3.1 (0.4)	3.6 (0.4)
16. Quality of skin	3.7 (0.3)	3.4 (0.4)	3.1 (0.3)	3.1 (0.3)
Mean	3.7 (0.3)	3.3† (0.4)	2.9† (0.2)	3.5 (0.5)

Note. Values are mean ratings (1 = very poor likeness .. 7 = very good likeness), calculated from 1280 responses (20 participants x 4 conditions x 16 rating scales); SE for ratings are presented in parentheses. See text for more details. Mean likeness less than Unaltered image: † *p* < .002.

Table 3. Naming of complete composites and targets by external feature presentation.

	Unaltered image	Hat added	Hair changed	Hooded top added	Exact hair
Naming task					
Spontaneous: composites					
Accurate	33.3† (25 / 75)	24.0† (18 / 75)	18.9*† (14 / 74)	10.3** (8 / 78)	36.4 (28 / 77)
Inaccurate	24.0 (12 / 50)	24.6 (14 / 57)	45.0* (27 / 60)	15.7 (11 / 70)	40.8 (20 / 49)
Spontaneous: targets					
Accurate	94.7 (71 / 75)	60.0** (45 / 75)	50.0** (37 / 74)	35.9** (28 / 78)	92.2 (71 / 77)

Note. Values are naming scores expressed as a percentage (see Table 1, *Note*)—Accurate (*N* = 379 out of 400) and Inaccurate naming (*N* = 286 out of 400). The regression models for spontaneous naming of composites (accurate and inaccurate) and targets (accurate) [$X^2(4) > 17$, $p < .002$, $R^2 > .05$ (Cox & Snell) and $R^2 > .08$ (Nagelkerke)]. Presentation of composite images (external-features) and the Constant were significant (*p*s < .002). The Unaltered image category contained either the composites constructed by participants in Experiment 1, or their unaltered target images, while targets and composites in the other conditions were subject to external-feature modifications. Less than Unaltered image: * $p < .05$, ** $p < .002$. †Significant linear trend in the order shown, $p < .05$.

Table 4. Correct naming rates by accuracy of target hair, image format and presentation order.

		Presentation (first / second) x Image format (single image / array) at naming				
		First			Second	
Accuracy of target hair						
Changed	Single image	20.3 [a,b,c] (12 / 59)		Array	40.7 [c] (24 / 59)	
	Array	39.3 [a] (24 / 61)		Single image	37.7 (23 / 61)	
Unaltered	Single image	43.5 [b,d] (27 / 62)		Array	54.8 [d] (34 / 62)	
	Array	47.5 (29 / 61)		Single image	49.2 (30 / 61)	

See Table 1, *Note*, and text for more details. [a,d] $p < .05$. [b,c] $p = .001$.

Table 5. Count of items that participants indicated led to recognition in the presented arrays

| Accuracy of target hair | Presentation order of array format | | Key for order of array |
	First	Second	
			[1] | [2]
			[3] | [4]
Changed	4 | 14 3 | 3	6 | 16 2 | 0	1 = Composite 2 = Composite + hat
Unaltered	18 | 9 2 | 0	30 | 3 1 | 0	3 = Composite + sunglasses 4 = Composite + hat + sunglasses