Frowd, C.D., Bruce, V., Smith, A., & Hancock, P.J.B.

Improving the quality of facial composites using a holistic cognitive interview.

Journal of Experimental Psychology: Applied. Vol 14 (3), Sep 2008, 276-287

http://dx.doi.org/10.1037/1076-898X.14.3.276

Charlie D. Frowd (1*)
Vicki Bruce (2)
Ashley J. Smith (3)
Peter J.B. Hancock (3)

(1) School of Psychology
University of Central Lancashire, PR1 2HE
* Corresponding author: Charlie Frowd, School of Psychology, University of Central Lancashire, Preston PR1 2HE, UK. Email: cfrowd@uclan.ac.uk. Phone: (01772) 893439.
(2) College of Humanities and Social Science
University of Edinburgh, EH8 9JU
(3) Department of Psychology
University of Stirling, FK9 4LA

Improving the quality of facial composites using a holistic cognitive interview
Running head: Holistic Composite Construction

Journal of Experimental Psychology: Applied

Abstract

Witnesses to and victims of serious crime are normally asked to describe the appearance of a criminal suspect, using a Cognitive Interview (CI), and to construct a facial composite, a visual representation of the face. Research suggests that focussing on the more global aspects of a face, as opposed to its facial features, facilitates recognition and improves composite quality; also, that the CI enables more effective use of a composite system. The current study evaluated a novel 'holistic' Cognitive Interview (H-CI). This comprised a descriptive phase, using a CI, followed by a recognition-enhancing phase, involving the attribution of seven holistic properties. Participant-witnesses watched a video of a target, then 3-4 hours later received either a CI or an H-CI and constructed a single composite with a standard system, PRO-fit. Composites constructed after the H-CI were correctly named more than four times as often as those after the CI, attributable to an improvement in the quality of both the internal and external parts of the face. In police work, the H-CI offers the possibility of substantially improving the identification of criminal suspects.
(180 words)

Acknowledgement
This research was supported by a grant from the Engineering and Physical Sciences Research Council (EP/C522893/1). The authors would like to thank Cindy Frowd, for her insightful comments on an early draft of the paper, as well as three anonymous reviewers.

Keywords: facial composite, trait attribution, holistic interview, cognitive interview, PRO-fit, recall, recognition.

Witnesses to and victims of serious crime, such as rape or murder, normally carry out a number of important tasks in order to bring a criminal to justice. They are initially asked to describe the events of the crime along with the physical and facial characteristics of those involved. Witnesses (and victims) may also be asked to try to identify the criminal from a mugshot album or construct a facial composite, a visual likeness of the face normally achieved by the selection of individual facial features (e.g. hair, eyes, brows, nose and mouth). Later, they may participate in a police line-up, another form of identification.

There are clearly two types of processes involved with these tasks: recall and recognition (e.g. Davies, 1983). The former is concerned with verbalising information, such as events and facial appearance; the latter, with comparing whether an image or person being presented is the same as that seen previously (i.e. at the scene of the crime). Facial composite construction traditionally involves a mixture of the two, since a description is used to locate facial features within a large set of alternatives, and recognition is required to identify when the best facial likeness has been achieved; face recognition is also engaged later when other people attempt to recognise the composite.

The recall and recognition of visual information are known to be largely separate mental processes (e.g. Sporer, 1989; Woodhead & Baddeley, 1981); the underlying neural mechanisms also reside in separate lobes of the brain (e.g. Baddeley, 1990). While recognition tends to be a fast, accurate, automatic process, and is reasonably stable over time, the serial recall of information is effortful, takes much longer, and decays considerably more rapidly (e.g. Bruce, 1982; Burton, Wilson, Cowan & Bruce, 1999; Davies, 1983; Ellis, 1975; Ellis, Shepherd & Davies, 1980; Reinitz, Morrisey & Demb, 1994; Sporer, 1989).

Face recognition is believed to be holistic in nature, emerging from the features of a face being processed in the context of other features (e.g. Bruce & Young, 1998; Davies & Christie, 1982; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). Indeed, the recognition of a face tends to be enhanced if learned (or encoded) holistically, for example by attributing personality traits to a face; conversely, recognition is suppressed when a face is encoded by its physical attributes (e.g. Berman & Cutler, 1998; Shapiro & Penrod, 1986; Wells & Hryciw, 1984). Face recognition may also be enhanced by a global Navon task (Macrae & Lewis, 2002; Navon, 1977). As part of a modern approach to interviewing witnesses, known as the Cognitive Interview (for a review, see Wells, Memon & Penrod, 2007), mentally reinstating the context in which the face was originally seen improves recognition (e.g. Malpass, 1996). Further, the identification of individual facial features is also facilitated by presenting these features in the context of a complete face (Davies & Christie 1982; Tanaka & Farah 1993; Tanaka & Sengco 1997), a finding which is incorporated into modern composite systems.

The recall of information can also be improved. For example, verbal description production is facilitated when a face is encoded by its physical features, rather than as one or more personality judgements (e.g. Finger & Pezdek, 1999; Wells & Turtle, 1988); a description has been found to be more accurate and complete following an exhaustive, unhindered recall, and by the use of several recall attempts (e.g. Wells et al., 2007) – both of which are a key part of the Cognitive Interview. However, processes that benefit recall may hinder recognition, and vice versa, and the method of face encoding described above is an example. In fact, it has been demonstrated that the act of describing a face can itself interfere with the recognition of a face (e.g. Dodson, Johnson & Schooler, 1997; Meissner & Brigham, 2001; Schooler & Engstler-Schooler, 1990), a process which is sometimes referred to as the Verbal Overshadowing Effect (VOE).

The implication of the above research is that improvements could be made to the procedure for constructing facial composites. It is known that even under favourable conditions composites are normally named only about 20% of the time (Brace, Pike & Kemp, 2000; Bruce, Ness, Hancock, Newman & Rarity, 2002; Davies, van der Willik & Morrison, 2000; Frowd, Hancock & Carson, 2004; Frowd et al., 2005b, 2007b; Frowd, Bruce, McIntyre & Hancock, 2007a). Currently, witnesses undergo a recall phase (description) followed by a recognition phase (composite construction), and therefore a VOE might be induced. In this case, a witness's ability to judge when the most recognisable face has been reached might be suppressed. Instead, it would appear possible to employ procedures that might actively enhance recognition. Berman and Cutler (1989) found that recognition ability was improved following the attribution of personality traits, such as rating for intelligence or attractiveness, relative to rating of facial features, such as length of nose or eye spacing. There is also some evidence that character attribution may be of value for composite production: Shepherd, Ellis, McMurran and Davies (1978), Wells and Hryciw (1984), and Davies and Oldman (1999) found that personality attribution at encoding can influence composite quality.

Frowd et al. (2007b) designed an alternative to the CI, which they referred to as a Holistic Interview, or HI. Participant-witnesses watched a video of an unfamiliar target face, then both described and rated the personality of the face before constructing a facial composite. The personality traits used were honesty, intelligence, friendliness, kindness, excitability, selfishness and arrogance. A second group of participant-witnesses watched the video and underwent a CI. Composite quality was assessed by a sorting task, whereby further participants matched the composites to the target photographs, and this indicated an approaching significant benefit for composites constructed after the HI. In Frowd, McQuiston-Surrett, Kirkland, and Hancock (2005c), participant-witnesses looked at a photograph of an unfamiliar face and two days later were given a CI, an HI or no interview. Composites were evaluated by matching them to a list of written names, which indicated that the region of the face containing the eyes, brows, nose and mouth – the so-called 'internal facial features' which are important for recognising a familiar face (e.g. Campbell et al., 1999; Ellis, Shepherd & Davies, 1979; Frowd et al. 2007a; Young, Hay, McWeeny, Flude & Ellis, 1985) – were of significantly better quality when constructed following an HI; the effect size was also large, $d = 0.98$. An advantage also emerged when a CI was administered relative to no interview. In general, the work suggested that the HI was effective by enhancing the recognition ability of the composite constructors: they were better able to identify when an optimal likeness had been achieved. Results of the latter study also suggested that asking a participant-witness to describe a face was valuable since it allowed facial features to be more effectively located within a composite system.

A drawback of the HI is that it does not work well with current facial composite systems. Systems such as PRO-fit and E-FIT allow a face to be constructed by the selection of individual facial features: a witness selects a hairstyle, a face shape, a pair of eyes, a nose, a mouth, etc. To be effective, however, these systems contain several hundred examples per facial feature, but this number is considerably more than would be shown to a witness, which is normally up to about twenty, otherwise fatigue and/or interference is likely to occur. For example, if a suspect is said to have a narrow nose, then only narrow noses would be shown. The role of the CI, therefore, is to obtain the best description of the face, which in turn can be used to pre-select suitable sets of features within the system. When using the HI in our previous work, we had either to accept a default face within PRO-fit, which tended not to look like the target, or ask for a description after the HI, neither of which is likely to be optimal. Despite this drawback, the HI outperformed the CI.

The aim of the current work was to evaluate a novel hybrid interview, comprised of a CI followed by a HI, which we refer to as a 'holistic' Cognitive Interview, or H-CI. We expected it to be especially effective at composite production by capitalising on the benefits of both interview types. Thus the initial CI would provide context reinstatement and obtain a detailed description of the face, for initialising PRO-fit. The HI component would then switch the person constructing the composite into a more holistic mode of processing, allowing better decisions to be made about the face presented to them. The hybrid interview might also help to overcome a VOE, a so-called ‘release’ from verbal overshadowing (e.g. Finger & Pezdek, 1999), as discussed later.

Specifically, we compared the quality of composites constructed using a CI with those constructed using an H-CI. We did not seek to include composites constructed with just an HI, since we already know that the HI is better than a standard CI: there was a large, significant benefit in Frowd et al. (2005c) and an approaching benefit in Frowd et al. (2007b); we note that including this condition would have also introduced the methodological issue raised above. The design coupled more naturalistic target stimuli than Frowd et al. (2005c, 2007b) and involved police-type construction procedures in an attempt to produce stimuli that were representative of those constructed in real crimes. The main assessment of composite quality was naming, and was expected to be better for constructions made following the H-CI.

Evaluating a Holistic Cognitive Interview

Two stages were required to evaluate the effectiveness of the Holistic Cognitive Interview (H-CI). In the first stage, participants watched a video containing a target face, then three to four hours later received a traditional Cognitive Interview (CI) to elicit a verbal description of the target. Half the participants then constructed a composite of this face; the other half immediately received a holistic interview with composite construction thereafter. Thus, each person constructed a single composite after either a CI or an H-CI. In the second stage, the resulting composites were evaluated, initially by asking other people to name them. Since naming levels tend to be quite low (e.g. Davies et al., 2000; Frowd et al., 2005a, 2005b), two supplementary tasks were administered. In the first, known as a *sorting* task, participants attempted to match the composites to photographs of the targets; in the second, to rate them along a number of potentially useful dimensions. The sorting task was also carried out using just the internal region of the face, known to be important for familiar face recognition (e.g. Ellis et al., 1979; Frowd et al., 2007a; Young et al., 1985), and the external part, known to be important for the perception of an unfamiliar face (Bruce et al., 1999; Ellis et al., 1979; Gibling, Ellis, Shepherd & Shepherd, 1987; Hancock, Bruce & Burton, 2000; Young et al., 1985).

Frowd et al. (2007b) designed their Holistic Interview to broadly match the CI. Participants first provided a free description of the personality of the face and then rated along the following dimensions: honesty, intelligence, friendliness, kindness, excitability, selfishness and arrogance. As it was not considered sensible to rate the honesty or the excitability of a criminal face, these attributes were replaced with facial distinctiveness and aggressiveness here. We have no reason to believe that the particular scales used are important in themselves to enhance face recognition ability, though they should be practical for a witness. Berman and Cutler (1989) used intelligence, attractiveness and height. What is perhaps more important is the number of attributions made: Berman and Cutler’s data reveal a trend such that better recognition was found when six ratings were made compared to two, suggesting that recognition ability and the number of attributions made may be positively related.

Stage 1: Composite Construction

Participants

Participants who constructed the composites were 24 students from the University of Stirling, 11 male and 13 female, aged 19 to 22 years (M = 20.0, SD = 0.9). None of these reported watching the UK soap Eastenders and therefore they constructed a composite of an unfamiliar face, the norm for real witnesses. One additional person reported knowing their target and was replaced. Each person received a course credit for participation.

Design

The design was between participants: half created their composite directly after being given a cognitive interview, the other half had the same CI followed by an HI and then created their composite. Each of the 12 target identities was used once for each group, producing 24 composites in total.

Materials

Target stimuli were non-violent video clips of six male and six female characters from the UK TV soap Eastenders. Each video clip contained edited footage from the TV programme, depicted an interaction between the target and another person and lasted for about 15 to 45 seconds. The targets spanned a wide age range for both genders, from twenty to sixty years. At the end of each clip, the video froze on a front-face view of the target's face for about 5 seconds. PRO-fit software version 3.1m running on a laptop was used to construct the composites.

Procedure

Participants were tested individually. They made two visits to the laboratory, first to inspect a target video, and 3 to 4 hours later to construct a composite of this person.

In the first visit to the laboratory, participants watched one of the 12 target video clips. Each person was told the approximate length and nature of the clip, as above, and given headphones to listen to the dialog. Video clips were watched in the knowledge that a composite would be required of the target's face. Afterwards, each person was asked whether the target was recognised. Only one person reported to be familiar with the face: while a composite was constructed for this person, it was not used and another person was recruited. Each clip was watched by a total of two people, one who was later given a Cognitive Interview (CI), and one, a Cognitive plus a Holistic Interview (H-CI); assignment of participants to both target videos and interview type was randomized. This part was carried out by the lead author, so the Experimenter (the third author) was unaware of the identity of the targets until all composites had been constructed.

Participants returned to the laboratory after 3 to 4 hours. It was explained that each person would first describe their target face and then construct a single composite using the PRO-fit system; also, that the description was necessary in order to locate facial features within PRO-fit. A Cognitive Interview was administered, with each person asked to freely recall as much as possible of the target's appearance in his or her own time. While this was being carried out, the Experimenter took notes. Next, the Experimenter repeated details of each feature recalled sequentially – in the order: overall appearance, face shape, hair, brows, eyes, nose, mouth and ears – and asked participants to attempt further recall.

Those assigned to the H-CI condition then received a holistic interview. Each person was asked to think to themselves about the personality of the face, for which a minute was allowed, and then to make a series of overall, or holistic judgements about the face on a three-point scale (low / medium / high). The holistic scales were then read aloud sequentially and participants gave a rating as requested for each in their own time. The scales were given in the following order: intelligence, friendliness, kindness, selfishness, arrogance, distinctiveness and aggressiveness.

Once the interview part was complete, all participants were informed that the session would move on to composite construction. The Experimenter provided a brief overview of the construction procedure and introduced the PRO-fit composite system. The adult white male database was selected in PRO-fit for a male target, the female equivalent for a female target. She then provided an overview of how facial features could be selected, resized and positioned as required within PRO-fit. Participants were also made aware that, in spite of many examples available for each feature, only an approximate likeness may be possible, but an artwork program was available within PRO-fit to improve the likeness. This additional program could, for example, add bags under the eyes, add wrinkles, or provide shading for any feature. It was also explained that such additions were normally applied as a final stage to avoid having to rework them were a feature to be changed.

The Experimenter entered the given description of each facial feature (from the CI) into PRO-fit and prepared an 'initial' composite, a face whose appearance matched the description. This complete face was then presented to the participant for improvement. Under the guidance of the participant, the experimenter exchanged features in the face (which was normally hair and face shape first) with other examples that matched the description to achieve the best likeness possible as described above. The experimenter offered to enhance the face using the artwork program as a final stage before saving the composite to disk.

Composites took about an hour to construct including the time taken for interviewing; the holistic procedure added an additional five minutes.

Stage 2: Composite Evaluation

In this part, the composites constructed in Stage 1 were evaluated using three tasks. These are described below and included naming, the primary task, plus the supplementary measures of sorting and rating. All tasks were expected to suggest that constructions following the H-CI would be of better quality.

Naming

Participants

Twelve male and 24 female students from the University of Stirling volunteered. All reported to be regular viewers of Eastenders and watched the soap at least twice a week (or the omnibus edition at the weekend). Their age was 17 to 53 years (M = 24.6, SD = 8.1).

Design

The effect of target gender on composite quality was investigated, which to our knowledge is untested for a modern composite system. While past research has suggested that target gender does not strongly influence face recognition (e.g. Shapiro & Penrod, 1986), the range of facial features available within a composite system might; certainly such a limitation was reported in an older non-computerised system called Photofit (Davies, 1983). To this end, a 2 (interview type) x 2 (gender of composite) within-subjects design was employed.

Materials

Composites from the construction stage were printed in greyscale (as PRO-fit does not display colour faces) one per page at approximately 6cm (width) x 8cm (high) on a good quality printer. Example composites are presented in Figure 1. These images were divided into two testing booklets: Book A contained 6 male composites constructed following the CI and 6 female composites constructed following the H-CI; Book B, the reverse. Front view colour portraits of the targets were obtained on the Internet and similarly printed, albeit in colour this time, one per page; they were used to verify that participants were familiar with the target faces.

Figure 1. Example composites constructed of the actress Wendy Richards (stage name, Pauline Fowler), top row, and John Bardon (Jim Branning), bottom row. The composite produced following the Cognitive Interview is on the left, the Holistic Cognitive Interview, on the right.

Procedure

Participants were told that they would be shown composites constructed of characters from the TV soap Eastenders and were asked to name them using real or stage names. Composites were presented sequentially from one testing booklet, then the other, with a counterbalanced order, and each person provided a name where possible in their own time. When all the composites had been inspected, the naming procedure was repeated for the target photographs. The order of presentation of composites and targets was randomised for each person.

Results

Participants correctly named the target photographs 99.0% (*SD* = 7.4) of the time. Composites were named by participants at 8.6% (*SD* = 7.3) in the CI condition but 41.2% (*SD* = 16.3) in the H-CI. All items enjoyed better naming following the holistic procedure, although performance was higher overall for female (*M* = 30.8, *SD* = 11.6) than male composites (*M* = 19.0, *SD* = 11.0). These data are presented in Figure 2 and show an advantage for both the H-CI and the female composites.

The participant correct naming scores were subjected to a 2 (interview type) x 2 (composite gender) repeated-measures Analysis of Variance (ANOVA). This produced significant main effects of interview, $F(1,35) = 135.4$, $p < .001$, $\eta^2 = .80$, indicating the benefit of the H-CI procedure, and of gender, $F(1,35) = 32.1$, $p < .001$, $\eta^2 = .48$, as female composites were named overall better than male ones. The interaction between these factors was not significant, $F(1,35) = 3.0$, $p = .094$, $\eta^2 = .08$. The weaker items analysis was similarly significant for interview, $F(1,10) = 12.0$, $p = .006$, $\eta^2 = .55$, but not gender, $F(1,10) = 3.1$, $p = .107$, $\eta^2 = .24$; the interaction remained non-significant, $F(1,10) = 0.2$, $p = .685$, $\eta^2 = .02$.

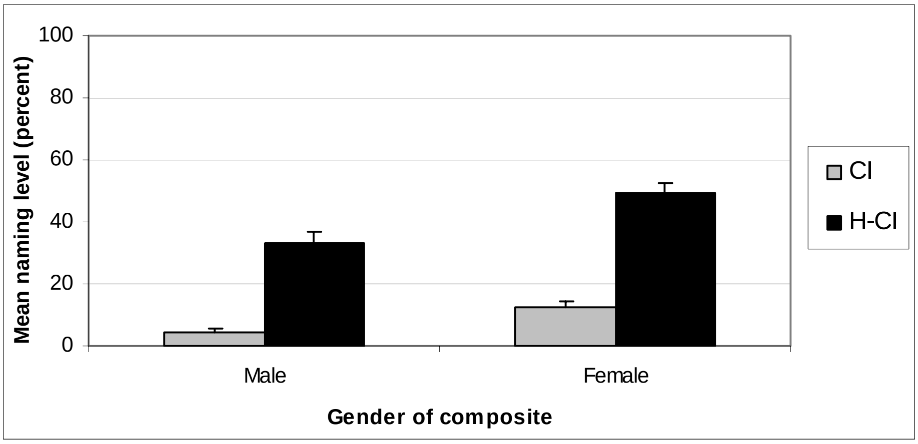

Figure 2. Improvement in correct naming for the Holistic Cognitive Interview (H-CI) relative to the Cognitive Interview (CI). Error bars are standard errors of the means.

An analysis by incorrect naming is included to provide a further indication of composite quality; note that lower values indicate more accurate composites. As illustrated in Figure 3, incorrect names substantially reduced overall from 63.4% (*SD* = 23.3) in the CI to 38.4% (*SD* = 19.9) in the H-CI. Incorrect names were also somewhat higher for male (*M* = 56.9, *SD* = 23.5) than for female composites (*M* = 44.9, *SD* = 20.0).

An ANOVA indicated that the H-CI produced significantly fewer incorrect names than the CI, $F(1,35) = 63.0$, $p < .001$, $\eta^2 = .64$, and that female composites were incorrectly named significantly less often than male composites, $F(1,35) = 13.4$, $p < .001$, $\eta^2 = .28$; these factors did not interact with each other, $F(1,35) = 0.1$, $p = .804$, $\eta^2 = .00$. The items analysis was significant for interview, $F(1,10) = 17.2$, $p = .002$, $\eta^2 = .63$; gender, $F(1,10) = 2.5$, $p = .148$, $\eta^2 = .197$, and the interaction, $F(1,10) = 0.02$, $p = .881$, $\eta^2 = .02$, were non-significant.

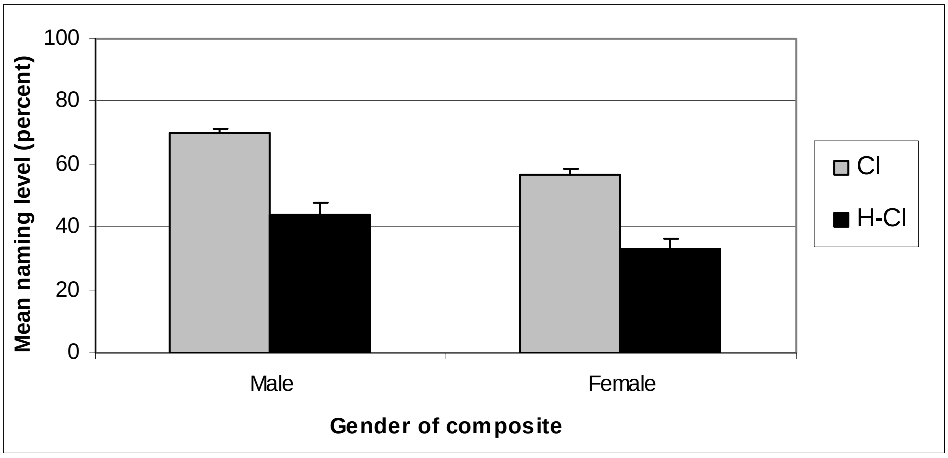

Figure 3. Reduction in incorrect name production for composites constructed after the Holistic Cognitive Interview (H-CI). Error bars are standard errors of the means.

Sorting

Composite sorting involved participants matching the composite faces to the target photographs. Three versions of this supplementary task were administered, as carried out elsewhere (e.g. Frowd et al. 2007a), to explore the quality of the complete composites; the so called ‘external facial features’, the region including the hair, ears and neck; and the ‘internal facial features’, the region including the eyes, brows, nose and mouth. While the inner face is known to be particularly important for familiar face processing (e.g. Ellis, Shepherd &

Davies, 1979; Young, Hay, McWeeny, Flude & Ellis, 1985), the external parts of the face have a higher salience when the face is not known (e.g. Bruce et al., 1999; Gibling, et al., 1987; Young et al., 1985). Since the aim is recognition by someone familiar with the face, the quality of the internal feature composites is of particular importance: this was expected to mirror the composite naming data and thus be better when produced under the H-CI.

Participants

Fifty-four students from the University of Stirling volunteered, 17 males and 37 females, aged 19 to 39 years (*M* = 23.7, *SD* = 3.7). They reported being regular viewers of Eastenders, as for the naming task. None had participated in the above tasks.

Design

Participants inspected composites of one type (internal / external / complete), but from both interview conditions (CI / H-CI), and therefore the design was between-participants for composite type but within-participants for interview.

Materials

Three sets of 24 composites were prepared, one each for complete, internal and external features composites; examples are presented in Figure 4. In the complete condition, composites were printed at approx. 6cm (wide) x 8cm (high) and cut to size with a 1 cm border. Other composites were prepared using Adobe Photoshop. For internal composite features, the elliptical marquee tool in Photoshop was used to highlight the region just above the jaw line and included the eyes, brows, nose and mouth. This region was then truncated just above the brows, so that hair was excluded, and printed to the same dimensions as above. External feature composites were the part of the composite remaining when the internal features had been removed.

The 12 Eastenders target photographs used for naming were similarly printed in colour. In this sorting task, and also in the following rating task, the goal was to remind participants of the facial appearance of the targets. The use of a set of images obtained from a source other than video stills from the sequence shown to witnesses is important. It has the advantage of encouraging an analysis by identity (what Bruce & Young, 1986, refer to as a *structural* code) rather than by the idiosyncratic aspects of the face: specific head pose, lighting, expression, and so forth (a *pictorial* code).

Figure 4. Example stimuli used in the Sorting task: a complete composite (left), an internal features composite (centre) and an external features composite (right). The target is of the actor, John Bardon (Jim Branning).

Procedure

Participants were tested individually and informed that they would be evaluating a set of composites constructed of Eastenders characters. The 12 target photographs were introduced and laid out on the table in front of the participant. A set of 24 composites was then given, comprised of either (a) complete composites, (b) internal feature composites or (c) external feature composites. Assignment of participants to composite type was random with equal sampling. Participants were requested to match the composites to the target photographs in their own time; they were made aware that there were two composites to each target picture and that composites could be exchanged at any time. The order of presentation of target and composite faces was randomised for each person.

Results

Composites were sorted overall much more accurately when constructed after the H-CI (*M* = 38.3%, *SD* = 13.4) than the CI (*M* = 23.3%, *SD* = 14.7); as for naming, female composites (*M* = 37.2%, *SD* = 15.7) performed overall better than male composites (*M* = 24.9%, *SD* = 11.2). As illustrated in Figure 5, the benefit of the H-CI was consistent for complete (*M* = 41.1%, *SD* = 9.3), external (*M* = 31.1%, *SD* = 8.5) and internal feature composites (*M* = 20.8%, *SD* = 8.9).

There was a significant effect of interview type, $F(1,51) = 61.0$, $p < .001$, $\eta^2 = .55$, and of composite type, $F(2,51) = 23.3$, $p < .001$, $\eta^2 = .48$; these factors did not interact, $F(2,51) = 0.9$, $p = .882$, $\eta^2 = .01$. There was also a main effect of target gender, $F(1,51) = 56.9$, $p < .001$, $\eta^2 = .53$, and a significant interaction between gender and interview type, $F(1,51) = 6.2$, $p = .016$, $\eta^2 = .11$, but not between gender and composite type, $F(2,51) = 2.9$, $p = .063$, $\eta^2 = .10$. The significant interaction is interpretable by floor effects for composites constructed in the CI condition. Simple contrasts of the ANOVA revealed that complete composites were sorted better than those of external features, $p = .001$, which in turn were sorted better than those of internal features, $p = .001$. The items analysis indicated the exact same pattern of significant effects.

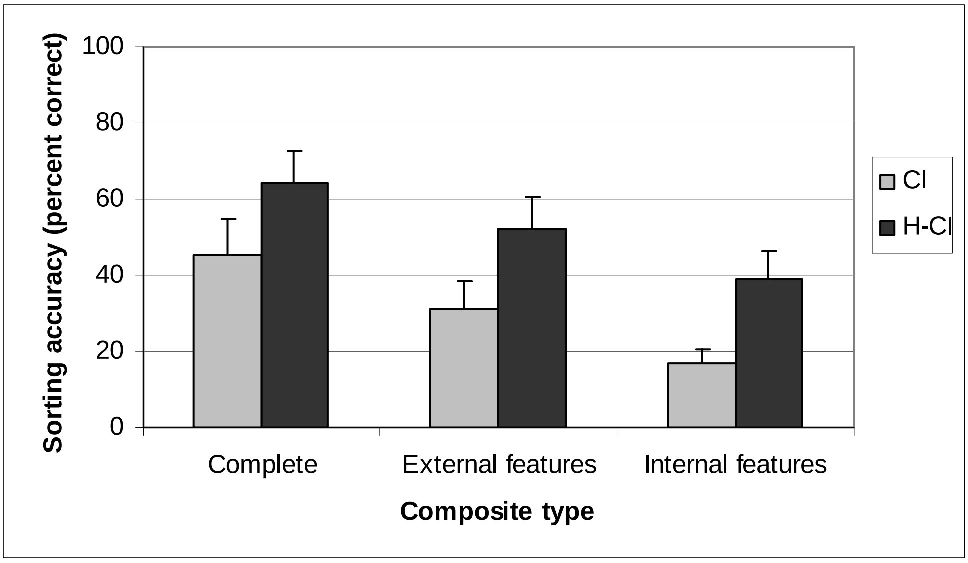

Figure 5. The benefit of the Holistic Cognitive Interview (H-CI) as indicated by a composite sorting task. Error bars are standard errors of the means.

Rating

The composites were further assessed by a rating task. In this part, two main questions were asked. Firstly, whether the composites improved in quality along the holistic dimensions requested as part of the H-CI? To do this, fresh participants simply rated the degree of match between composites and targets along the seven dimensions. And secondly, whether there was evidence for an improvement in quality along six further dimensions? These were age and hair, important for unfamiliar face perception (e.g. Ellis, 1986; Shepherd, Ellis & Davies, 1977); eye and brow region, important for familiar face recognition (e.g. Fisher & Cox, 1975; O'Donnell & Bruce, 2001; Sadr, Jarudi & Sinha, 2003); relational information (distances) between the internal facial features, especially important familiar face perception (e.g. Leder & Carbon, 2006; Mantyla, 1997; O'Donnell & Bruce, 2001); the internal features (eyes, brows, nose and mouth); and the overall likeness of the face. It was expected that the quality of the composites would improve along these 13 dimensions as a result of the H-CI.

Participants

Six male and four female staff and students from the University of Stirling participated, aged 22 to 48 years (M = 32.1, SD = 8.5). Similar to those involved in the naming and sorting parts, all were regular watchers of Eastenders and had not participated previously in the study. Each person was paid £2.

Design

To provide good experimental power, participants were shown two composites at a time, one constructed in the CI condition and one of the same target face but in the H-CI, and provided a rating for each in the presence of a target photograph; thus, relative judgements could be made when rating. Participants were shown all combinations of composites and rating scales, and this was carried out in two blocks with a short break in between to limit fatigue. In the first block, ratings for the seven ‘holistic’ scales used in the H-CI were collected. Participants provided a rating for each scale which represented the match between the given composite and the target face. In the second block, ratings were similarly collected for the additional five scales mentioned above. The design was therefore within-subjects for rating scale (aggressiveness / arrogance / distinctiveness / friendliness / intelligence / kindness / selfishness / age / hair / internal individual features / eye and brow region / relational / overall likeness) and interview type (CI / H-CI). It was considered inappropriate to follow the more feature-based ratings (i.e. generally the scales used in the second block) with the more holistic ratings (i.e. in the first block), as this might cause interference (Macrae & Lewis, 2002), and therefore block order was not counterbalanced.

Materials

Materials were the 24 (complete) composites and the 12 target photographs. Seven of the targets were the same as those used for the naming and sort tasks, but new photographs were located (via the Internet) for five of them to more clearly show a front face view of the face, which was believed to help when rating, especially for the relational match. A bespoke computer program was written to present the composites and targets, and record subject responses.

Procedure

Participants were tested individually and were asked to evaluate the quality of a set of facial composites by providing ratings. They were informed that this would be carried out in their own time in two testing blocks. A short demo of the computer program used was run which illustrated the scales featured in the first task block and contained composites taken of a famous face from an unrelated study. The two composites were presented below a photograph of the famous face. Participants were asked to first provide a rating match of kindness for the composite on the left hand side: they should estimate the level of kindness in both the target and the composite, then rate according to how close was the kindness match (1 = poor match / 7 = good match). Note that this is not a rating of the perceived kindness of the composite, but how well it matches the target. A high rating might mean that both look kind, or both unkind, or anything in between, a low rating that one looks kind and the other does not. Participants were then asked to do this a second time, for the composite on the right hand side, but provide a rating relative to the first. That is, they should think about whether the match between target and the second composite was better or worse than the first and answer accordingly. Participants then worked through the other scales for practice: aggressiveness, arrogance, friendliness, intelligence, selfishness and distinctiveness. The program was then run again for the Eastenders' composites and participants worked through these in their own time. As for the demo trials, the program presented each pair of composites constructed from each target face, a total of 84 trials (12 pairs x 7 scales). The order of presentation of composites and scales was randomised for each person, as was the left-right order of each presented pair.

Next, the practice trials for the second block of tests were run. Participants were asked to rate the composites and target along the given scale. The presented scales were: overall likeness; individual internal features (eyes, brows, nose, mouth and ears, but not hair); hair; age; eye and brow region; and relational information (distances between eyes, brows, nose and mouth). As before, participants rated using practice composites and then moved on to the Eastenders' set. The entire procedure took approximately 45 minutes per person.

Table 1. The effect of a Holistic Cognitive Interview (H-CI) on composite ratings relative to a Cognitive Interview (CI). Values are mean participant ratings (1 = poor match / 7 = good match).

Rating scale	CI	H-CI
Age	4.0	4.7**
Aggressiveness	4.0	4.1
Arrogance	3.8	4.2
Distinctiveness	3.9	4.6*
eye/brow area	3.5	3.8
Facial features	3.5	4.1**
Friendliness	4.2	4.2
Hair	3.4	4.2**
Intelligence	4.1	4.4
Overall likeness	2.9	3.8*
Relational	4.0	4.3
Selfishness	4.0	4.3

Note * represents a significant increase by participants from CI to H-CI, at $p < .05$; ** at $p < .001$.

Results

The mean participant ratings were 3.8 ($SD = 0.4$) for composites constructed following a CI, and this increased to 4.2 ($SD = 0.3$) following an H-CI. Cronbach's Alpha was .82, thus indicating good reliability of rating scores across participants. Average ratings by scale and interview are presented in Table 1; ratings by target gender are not shown due to their non-significant influence, as detailed below.

The participant rating data from the two testing blocks were combined and subjected to a repeated-measures ANOVA. This was significant for interview, $F(1,9) = 23.5$, $p < .001$, $\eta^2 = .72$, rating scale, $F(12,108) = 4.5$, $p < .001$, $\eta^2 = .33$, and the interaction between these two factors, $F(12,108) = 2.5$, $p = .007$, $\eta^2 = .22$. Target gender, $F(1,9) = 0.6$, $p = .467$, $\eta^2 = .06$, the target by interview interaction, $F(1,9) = 0.4$, $p = .544$, $\eta^2 = .04$, and the three way interaction, $F(12,108) = 1.1$, $p = .384$, $\eta^2 = .01$, were not significant. Simple-main effects of the interview by scale interaction revealed a significant increase in rating for distinctiveness, $p = .042$, age, $p < .001$, facial features, $p = .001$, hair, $p < .001$, and overall likeness, $p = .002$. Also, for H-CI composites, age ratings were significantly higher than those of brows/eyes, $p = .047$.

For the items analysis, both rating scale, $F(12,132) = 4.5$, $p < .001$, $\eta^2 = .32$, and interview, $F(1,11) = 6.6$, $p = .026$, $\eta^2 = .38$, were significant; target gender and all interactions were not, $p > .1$. None of the previous scales were significant following t-tests with a Bonferroni correction.

C*orrelational analyses*

A correlational analysis was conducted in order to explore possible factors which may influence composite quality (naming). To do this, two stepwise linear regressions were run with naming as the dependent variable, one for the internal and external feature composite sorting data, the other for the rating data. The stepwise regression used the *backward* method, which includes all variables at the start and removes them one at a time. It provides a more complete analysis than the alternative, the *forward* method, which does the reverse, by including effects of *suppressor* variables: variables that exert an influence in the presence of other variables. The following results are semi-partial or *part* correlations; each one expresses the amount of variance uniquely accounted for by that variable.

The sorting analysis revealed a significant positive partial correlation for the internal composite features, part $r = 0.71$, $p < .001$. Thus, composites that were well sorted on their internal features were also well named. The rating data revealed six low but significant partials, $p <= .05$, which were positive for individual internal features, $r = 0.27$, relational, $r = 0.28$, distinctiveness, $r = 0.28$, and kindness, $r = 0.38$; and negative for age, $r = -0.29$; and friendliness, $r = -0.30$. We believe the negative correlation for age may be explained by a difficulty in producing old-looking composites, as discussed below.

General Discussion

Constructing a good quality facial composite is a difficult task for a witness or victim: both describing the appearance of a person's face and selecting individual facial features are unnatural face perception tasks (e.g. Baddeley, 1979; Davies, 1983), yet they are key components of the traditional construction procedure. As a consequence, recognition rates from composites tend to be low and thus procedures which improve composite quality are both theoretically interesting and of practical importance. One successful approach (Frowd et al., 2005b, 2007b) employed a novel 'holistic' type of interview (HI). The procedure required laboratory-witnesses to make a number of personality-type judgements about a target face instead of providing descriptions of facial features (the norm when using a Cognitive Interview, CI). In the current work, the potential of combining these two types of interview was explored. Participant-witnesses who received a CI followed by an HI (termed an H-CI) produced composites that were correctly named more than four times as often as those who received a CI alone (and with a large effect size); they also elicited markedly fewer incorrect names. These advantages were supported by supplementary sorting and rating tasks, as discussed below. The advantage found for the H-CI is much greater than that found for the HI alone (Frowd et al., 2005c, 2007b).

Advantages of the H-CI

Past research has found general benefits for the Cognitive Interview (e.g. Finger & Pezdek, 1999; Koehnken, et al., 1994), which also apply to facial composite production (Frowd et al., 2005c; Luu & Geiselman, 1993). One reason for the effectiveness of the CI in obtaining a description of an event or a face is that recall is fragmentary in nature and more complete accounts are obtained with multiple recall attempts and subsequent prompting (e.g. Geiselman et al., 1986). With facial composites, better recall is of assistance in locating individual facial features within a composite system. However, there is an associated risk that describing a face will interfere with recognition, to produce a so-called verbal overshadowing effect (VOE) (e.g. Dodson, Johnson & Schooler, 1997; Schooler & Engstler-Schooler, 1990). While we were not specifically exploring a VOE, as discussed below, it is possible that the HI served to overcome such an effect.

While the CI assists with locating features within a composite system, the HI is believed to enhance face recognition ability, much the same as trait attribution did for Berman and Cutler (1989). In doing do so, the HI enables the composite constructor to more accurately determine when the most identifiable face has been reached. But, why should the attribution of traits improve recognition? The effect is interpretable in terms of a 'shift' in processing strategy, as suggested by Berman and Cutler (1989). The same basic notion has been used to explain the VOE (e.g. Fallshore & Schooler, 1995) and there is evidence that processing strategy can be manipulated directly (Macrae & Lewis, 2002). As mentioned above, there are two types of information which are employed when recognising a face: firstly, individual facial features; secondly, more global or *holistic* aspects of a face – sometimes called *configural* or *relational* information – which include not just the distances between features but also contextual effects, since features are perceived in the presence of other features (e.g. Bruce & Young, 1998; Davies & Christie, 1982; Cabeza & Kato, 2000; Collinshaw & Hole, 2000; Tanaka & Farah, 1993). According to the theory, describing a face shifts the processing strategy towards the featural, which is not optimal for recognition, whereas trait attribution does the opposite, prompting processing towards the configural and better recognition (Berman & Cutler, 1989; Davies & Oldman, 1999; Shapiro & Penrod, 1986; Shepherd, et al., 1978; Wells & Hryciw, 1984).

The results of the sorting task suggest that the H-CI improved the quality of both the internal and the external facial features, the regions of the face that are known to be especially important for familiar (Ellis, et al., 1979) and unfamiliar face (Young et al., 1985) perception respectively. The procedure was most notably effective in improving the internal features, likely to benefit the recognition of the composites (Frowd et al. 2007a). These effects were supported by the rating task, which suggested that both the hair, an external facial feature, and the internal features were more similar to the targets following the H-CI. Age was also found to be a better match for composites constructed with the H-CI; this attribute is selected as part of the facial texture in PRO-fit, the region of skin on the forehead and around the internal facial features, and is known to be one of the key dimensions for face perception (e.g. Shepherd, Ellis & Davies, 1977). Attaining a more accurate rendition of a target's age in a reconstructed face is therefore likely to provide a better cue for recognition. (Note that this result should not to be confused with the negative partial correlation between the rated age-match of the composites and naming, which is discussed below.)

Results from the rating data also suggest that the facial distinctiveness of the composites was a better match with the target following the H-CI. Facial distinctiveness is another key factor in face perception (e.g. Frowd et al., 2005b; Shapiro & Penrod, 1986; Valentine, 1991) and facilitates both familiar and unfamiliar recognition. Improving the distinctiveness match of a composite with its target is therefore likely to benefit identification. Indeed, in the correlational analyses, the distinctiveness match was found to be positively related to naming rate. One might ask of course whether distinctiveness was a better match as a result of participant-witnesses being prompted to think about facial distinctiveness or as a result of making a number of holistic-type judgements? The authors believe that the latter explanation is more probable since the holistic interview has been found to be beneficial in the absence of a distinctiveness judgement (Frowd et al., 2005c, 2007b). This conclusion is supported by the observation that none of the other holistic scales used in the rating task were significantly better matched in the resultant composite. The overall implication is that the prompts used in the HI are themselves unimportant, but together serve to elevate recognition proficiency.

Components of the H-CI

The HI was developed to broadly match the two components of the CI that are typically used when constructing composites: *free* and *cued* recall. Thus, our participants thought about the personality of their target face (*free* recall) and then rated for personality traits (*cued* recall). The aim was to promote a strong holistic bias. While there is evidence of bias manipulation for the latter, cued component (Frowd et al., 2007b), there is currently no direct evidence for the former. However, it is possible that freely recalling the personality of a face allows the reflection of traits not probed in the cued phase, e.g. threatening, trustworthy or swarthy, thus augmenting the holistic bias. Given the potential benefit of the HI for identifying suspects, the influence of these components would appear worthy of further research. (Berman and Cutler (1989) also suggested that trait attribution should be evaluated in a crime simulation.)

The holistic or trait attribution part of the H-CI (*cued* recall) is somewhat different to that used by Berman and Cutler (1989). They asked participants to provide up to four trait ratings for quite a large number of faces and found a recognition advantage, albeit a weak one, relative to ratings of physical features. Here, trait attribution was not only applied to a single face but was also likely to have been much stronger. Our participants were initially asked to think about the personality of their target face for a minute and then to assign seven holistic judgements. The procedure took about 5 minutes to administer, much longer than Berman and Cutler (1989). The overall benefit to composite naming was very large, and suggests that there was a large improvement in face recognition ability relative to the description only (CI) group. It also suggests that our holistic manipulation was stronger than theirs.

Experimental issues

The only planned difference between the two groups of laboratory-witnesses was the HI, which added about 5 minutes to the interview. Of course, a 5 minute filler task could have been added to the CI group, to give equal interview times. However, this was not done to maintain a realistic procedure: real witnesses do not do filler tasks! So, could the difference in interview times have influenced the results? The authors believe this is unlikely. One reason is that there is already a delay of around 5 minutes from the end of the CI to the start of the composite building process, due to the need to set up PRO-ft with the appropriate descriptors.

While going from zero delay to 5 minutes might conceivably produce a measurable effect, doubling the delay from 5 to 10, if that were the only effect of the HI component, seems much less likely to do so. Also, two previous studies (Frowd et al., 2005c, 2007b) have found that the holistic procedure is sufficient on its own to improve composite quality (i.e. without a preceding CI). Taken together with Berman and Cutler's (1989) finding that the attribution of personality traits improved recognition, the evidence favours the holistic procedure rather than a difference in delay. It is perhaps worth noting that composite quality can be very difficult to manipulate successfully, even when research elsewhere in face perception would suggest otherwise (e.g. Frowd et al., 2007b, 2007c). In this case, a 5 minute difference seems too subtle. A much longer duration may be different: Finger and Pezdek (1999) demonstrated that a 24 minute delay on average improved recognition after a descriptive phase, a so called 'release' from verbal overshadowing.

As mentioned above, might the holistic component of the H-CI itself provide a release from overshadowing (caused by the CI)? If it did, then this might be responsible for some of the large improvement in naming with H-CI construction. The design is unable to answer this directly since it is necessary to show evidence of a VOE being produced in the first place. To do that would require a third experimental condition, one where participant-witnesses do not give a verbal description and thus no VOE could be produced; and, for that condition to be superior to the CI condition. We have explored just such a situation before, in Frowd et al. (2005c), but found a slight benefit for the CI relative to no interview, which itself suggests the opposite: a verbal *facilitation* effect (e.g. Lyle & Johnson, 2004). However, this suggestion is slightly problematic as not asking a witness to generate a description causes a problem for composite construction, as discussed earlier, since facial features in PRO-fit cannot be initialised, thus limiting the effectiveness of the composite system and that experimental condition.

It turns out though that the VOE can be quite difficult to replicate (e.g. Clifford, 2003) and a recent meta-analysis by Meissner and Brigham (2001) found that its effect size is overall quite small. Given the general insensitivity of the composite process, it is somewhat unlikely that a VOE would be expressed. However, it is more likely to occur immediately after a description task and when elaborative recall instructions are given (Meissner & Brigham, 2001), a situation which parallels composite production. On-going research is exploring these issues.

Determinants of composite quality

The correlational analyses indicated that, in addition to facial distinctiveness, the individual internal features, and the match for the relational information, age, friendliness, and kindness were significant factors involved in composite naming. This finding is particularly interesting given the importance of relational information to face recognition (e.g. Leder & Carbon, 2006; O'Donnell & Bruce, 2001; Young, Hellawell & Hay, 1987) and suggests that witnesses who achieve a better configuration of features are likely to produce a more identifiable composite. Unfortunately, the rating data also suggest that relational information was not a better match with the target following the holistic interview. It is known that relational information is effortful to encode (Reinitz, Morrisey & Demb, 1994), is measurably worse after 24 hours (Hannigan & Reinitz, 2000), and would appear to be more of a recall than a recognition process. Being more recall-based, asking witnesses to explicitly describe the relational information, perhaps as part of the CI, may help to improve the constructed configuration. There is also some evidence that inaccurate configural information may be rendered less noticeable by anti-caricaturing (Frowd et al., 2007c), where both features and their configuration appear less distinctive.

The observation that the match to target of the individual internal features was positively correlated with naming is also a sensible result, given that a feature-by-feature approach was used to construct the face. However, the negative partial correlation for age (and friendliness) was unexpected, suggesting that as the age match increased, naming decreased. We interpret this negative correlation as an increase in difficulty of construction as the age of the target increases. Facial features are classified in PRO-fit by age, but the representation of features for older faces may be relatively poor, and this is likely to increase the artistic elaboration necessary for the person controlling the software (in this case, the Experimenter): the importance of such procedures is well known (e.g. Ellis, Davies & Shepherd, 1978; Frowd et al., 2005b; Gibling & Bennett, 1994). Composites of older targets may therefore look the wrong age, while still being distinctive enough to be recognised well. Ageing a face by hand is a complex procedure, requiring adjustments to several parts of the face. Fortunately, more holistically-based composite systems are now emerging, for example EvoFIT (Frowd, et al., 2004), where ageing is a dimension available for manipulation (Frowd et al., 2006).

Target and witness gender

One of the aims of the research was to explore the ability to construct both male and female targets using a modern composite system. The targets were six male and six female characters, each constructed once in each interview condition. The naming and sorting data both suggest that composites of female targets were constructed significantly better than those of male targets. However, the rating scales did not differ significantly by gender, nor were there any interactions, suggesting that differences by gender may not be a function of the composite system (for limitations in the system by gender, one would have expected differences in the rating of individual features, hair or the overall likeness, but these were not found). Instead, the observed difference in naming by gender may be a function of target properties. For example, the female targets used may simply be more distinctive in appearance than the male targets, a factor known to influence composite quality independently of the technique used to construct the face (Frowd et al., 2005b).

One of the reviewers asked whether the gender of our participant-witnesses influenced the quality of a composite? This is an interesting question to ask, as witnesses to real crime can be male or female, of course, as can be the gender of the target – although most composites are constructed of male faces; also, as there were approximately equal numbers of male and female witnesses in the study. Further, there is some evidence to support the notion of an own-gender bias (e.g. Shapiro & Penrod, 1986; Wright & Sladden, 2003): female participants are better at identifying female faces than male faces, and vice versa.

In our data, slightly more composites were correctly named when the gender of the witness matched the gender of the target (M = 27.1%, SD = 34.2) compared to when it did not (M = 22.6%, SD = 26.0). The previous participants' ANOVA was re-run with gender of witness (own / different gender to target) as a third factor. This analysis revealed that witness gender, $F(1,35) = 2.8$, $p = .101$, $\eta^2 = .08$, and all interactions involving this factor were not significant, all $p > .1$. For the incorrect naming data, levels were almost identical (M = 50.1%, SD = 37.9, own gender; M = 50.5%, SD = 33.0, different gender), and witness gender and its interactions were similarly not significant, $p > .1$.

The results thus provide no evidence that a more identifiable composite is produced if the gender of the witness is the same as the gender of the target. To the authors' knowledge, this is the first study to explore such an effect using a modern composite system (PRO-fit). Note, however, that the current design suffered from a potential confound since items were not balanced by witness gender – as this is a post-hoc type of analysis. Consequently, items did not act as their own control, unlike the case for the two other factors, and this could potentially skew the data away from the observed direction, potentially masking a significant effect; this issue also prevented items analyses from being conducted. It will be interesting to see whether an own-gender advantage might emerge from a study which deliberately set out to explore it.

Summary

In summary, the current work has replicated a previous finding that asking laboratory-witnesses to focus on the personality traits of a target face promotes a better quality composite. The new Holistic Cognitive Interview (H-CI) requires a witness or victim (a) to describe a target face as normal using a Cognitive Interview, (b) to think about the personality of the face and (c) to make seven holistic judgements. The improvement in composite quality from laboratory-witnesses was found to be impressive with the H-CI, compared to construction using the CI alone, with the number of correct names elicited increasing by a factor of four and the number of incorrect names decreasing by over 50%. The holistic part of the H-CI is believed to promote a shift in processing from individual facial features to a more global type; specifically, the interview allowed for a more accurate construction of the age, hair and facial distinctiveness in the composites. If used as part of composite construction with real witnesses and victims, the H-CI should substantially improve the identification of criminal suspects.

References

Baddeley, A. D. (1979). Applied cognitive and cognitive applied psychology: the case of face recognition. In Nilsson, L.G. (Ed.), *Perspectives on memory research*. Erlbaum, New Jersey.

Baddeley, A.D. (1990). *Human memory: theory and practice*. Hove: LEA.

Berman, G.L., & Cutler, B.L. (1998). The influence of processing instructions at encoding and retrieval on face recognition accuracy. *Psychology, Crime and Law*, 4, 89-106.

Brace, N., Pike, G., & Kemp, R. (2000). Investigating E-FIT using famous faces. In A. Czerederecka, T. Jaskiewicz-Obydzinska & J. Wojcikiewicz (Eds.). *Forensic Psychology and Law* (pp. 272-276). Krakow: Institute of Forensic Research Publishers.

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology*, 73, 105-116

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.

Bruce, V., & Young, A.W. (1986). Understanding face recognition. *British Journal of Psychology*, 77, 305-327.

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

Burton, A.M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor quality video: evidence from security surveillance. *Psychological Science*, 10, 243-248.

Cabeza, R., & Kato, T. (2000). Features are also important: Contributions of featural and configural processing to face recognition. *Psychological Science*, 11, 429-433.

Campbell, R., Coleman, M., Walker, J., Benson, P. J., Wallace, S., Michelotti, J., & Baron-Cohen, S. (1999). When does the inner-face advantage in familiar face recognition arise and why? *Visual Cognition*, 6, 197-216.

Clifford, B.R. (2003). The verbal overshadowing effect: in search of a chimera. In G. Vervaeke, M. Vanderhallen, P. Van Koppen, & J. Goethals (Eds.) *Much Ado About Crime: Chapters On Psychology And Law* (pp. 151-162). Belgium: Uitgeverij Politeia NV.

Collishaw, S.,M., & Hole, G.J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. *Perception*, 29, 893-909.
Davies, G.M., & Christie, D. (1982). Face recall: an examination of some factors limiting composite production accuracy. *Journal of Applied Psychology*, 67, 103-109.
Davies, G.M., & Oldman, H. (1999). The impact of character attribution on composite production: A real world effect? *Current Psychology: Developmental, Learning, Personality, Social*, 18, 128-139.
Davies, G.M., van der Willik P., & Morrison, L. (2000). Facial composite production: A comparison of mechanical and computer-driven systems. *Journal of Applied Psychology*, 83, 119-124.
Dodson, C.S., Johnson, M.K., & Schooler, J.W. (1997). The verbal overshadowing effect: why descriptions impair face recognition. *Memory and Cognition*, 25, 129-139.
Ellis, H.D. (1975). Recognizing faces. *British Journal of Psychology*, 66, 404-426.
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, 297-307.
Ellis, H.D., Shepard, J.W., & Davies, G.M. (1975). An investigation of the use of the photo-fit technique for recalling faces. *British Journal of Psychology*, 66, 29-37.
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.
Ellis, H.D., Shepherd, J.W., & Davies, G.M. (1980). The deterioration of verbal descriptions of faces over different delay intervals. *Journal of Police Science and Administration*, 8, 101-106.
Fallshore, M., & Schooler, J.W. (1995). Verbal vulnerability of perceptual expertise. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 6, 1608-1623.
Finger, K., & Pezdek, K. (1999). The effect of the Cognitive Interview on face identification accuracy: Release from Verbal overshadowing. *Journal of Applied Psychology*, 84, 340-348.
Fisher, G.H., & Cox, R.L. (1975). Recognizing human faces. *Applied Ergonomics*, 6, 104-109.
Frowd, C.D., Bruce, V., McIntyre, A., & Hancock, P.J.B. (2007a). The relative importance of external and internal features of facial composites. *British Journal of Psychology*, 98, 61-77.
Frowd, C.D., Bruce, V., McIntyre, A., Ross, D., Fields, S., Plenderleith, Y., & Hancock, P.J.B (2006). Implementing Holistic Dimensions for a Facial Composite System. *Journal of Multimedia,* 1, 42-51.
Frowd, C. D., Bruce, V., Ness, H., Bowie, L., Paterson, J., Thomson-Bogner, C., McIntyre, A., & Hancock, P. (2007b). Parallel approaches to composite production: interfaces that behave contrary to expectation. *Ergonomics*, 50, 562-585.
Frowd, C.D., Bruce, V., Ross, D., McIntyre, A., & Hancock, P.J.B. (2007c). An application of caricature: how to improve the recognition of facial composites. *Visual Cognition*, 15, 1-31.
Frowd, C. D., Carson, D., Ness, H., McQuiston-Surret, D., Richardson, J., Baldwin, H., & Hancock, P. (2005a). Contemporary composite techniques: The impact of a forensically-relevant target delay. *Legal and Criminological Psychology*, 10, 63-81.
Frowd, C. D., Carson, D., Ness, H., Richardson, J., Morrison, L., Mclanaghan, S., & Hancock, P. (2005b). A forensically valid comparison of facial composite systems. *Psychology, Crime and Law*, 11, 32-52.
Frowd, C.D., Hancock, P., & Carson, D. (2004). EvoFIT: a holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on Applied Perception*, 1, 19-39.

Frowd, C.D., McQuiston-Surrett, D., Kirkland, I., & Hancock, P.J.B. (2005c). The process of facial composite production. In A. Czerederecka, T. Jaskiewicz-Obydzinska, R. Roesch & J. Wojcikiewicz (Eds.). *Forensic Psychology and Law* (pp. 140-152). Krakow: Institute of Forensic Research Publishers.
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.
Gibling, F., & Bennett, P. (1994). Artistic enhancement in the production of photofit likeness: an examination of its effectiveness in leading to suspect identification. *Psychology, Crime and Law*, 1, 93-100.
Gibling, F., Ellis, H., Shepherd, J., & Shepherd, H. (1987). Processing of familiar and unfamiliar faces from their internal and external features. *Proceedings of the International Conference of Applied Psychology*, Swansea, Wales. Unpublished.
Hancock, P.J.B., Bruce, V., & Burton, A.M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, 4, 330-337.
Hannigan, S.L., & Reinitz, M.T. (2000). Influences of temporal factors on memory conjunction errors. *Applied Cognitive Psychology*, 14, 309-321.
Leder, H., & Carbon, C.C. (2006). Face-specific configural processing of relational Information. *British Journal of Psychology*, 97, 19-29.
Luu, T.N., & Geiselman, R.E. (1993). Cognitive retrieval techniques and order of feature construction in the formation of composite facial images. *Journal of Police and Criminal Psychology*, 9, 34-39.
Lyle, K., & Johnson, M.K. (2004). Effects of verbalization on line-up face recognition in an interpolated inspection paradigm. *Applied Cognitive Psychology*, 18, 393–403.
Macrae, C.N., & Lewis, H.L. (2002). Do I know you? Processing orientation and face recognition. *Psychological Science*, 13, 194-196.
Malpass, R.S. (1996). Enhancing eyewitness memory. In S.L. Sporer, R.S. Malpass & G. Koehnken (Eds.). *Psychological issues in eyewitness identification* (pp. 177-204). Hillsdale, NJ: Lawrence Erlbaum.
Mantyla, M. (1997). Recollections of faces: remembering differences and knowing similarities. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 23, 1203-1216.
Meissner, C.A., & Brigham, J.C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology*, 15, 603-616.
Navon, D. (1977). Forrest before the trees: the precedence of global features in visual perception. *Cognitive Psychology*, 9, 353-383.
O'Donnell, C., & Bruce, V. (2001). Familiarisation with faces selectively enhances sensitivity to changes made to the eyes. *Perception*, 30, 755-764.
Reinitz, M. T., Morrisey, J., & Demb, J. (1994). Role of attention in face encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 161-168.
Sadr, J., Jarudi, I., & Sinha, P. (2003). The role of eyebrows in face recognition. *Perception*, 32, 285-293.
Schooler, J.W., & Engstler-Schooler, T.Y. (1990). Verbal overshadowing of visual memories: some things are better left unsaid. *Cognitive Psychology*, 22, 36-71.
Shapiro, P.N., & Penrod, S.D. (1986). Meta-analysis of facial identification rates. *Psychological Bulletin*, 100, 139-156.
Shepherd, J.W., & Ellis, H.D. (1996). Face recall - methods and problems. In S.L. Sporer, R.S. Malpass & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 87-115). Hillsdale, NJ: Lawrence Erlbaum.

Shepherd, J.W., Ellis, H.D., & Davies, G.M. (1977). Perceiving and remembering faces. *Home Office report POL/72/1675/24/1*.
Shepherd, J.W., Ellis, H.D., McMurran, M., & Davies, G.M. (1978). Effect of character attribution on Photofit construction of a face. *European Journal of Social Psychology*, 8, 263-268.
Sporer, S.L. (1989). Verbal and visual processes in person identification. In H. Wegener, F. Loesel, & J. Haisch (Eds.), *Criminal behavior and the justice system: Psychological perspectives* (pp. 303-324). New York: Springer.
Tanaka, J.W., & Farah, M.J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology*, 46A, 225-245.
Tanaka, J.W., & Sengco, J.A. (1997). Features and their configuration in face recognition. *Memory & Cognition*, 25, 583-592.
Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. *Quarterly Journal of Experimental Psychology*, 43A, 161-204.
Wells, G.L., & Hryciw, B. (1984). Memory for faces: encoding and retrieval operations. *Memory and Cognition,* 12, 338-344.
Wells, G., Memon, A., & Penrod, S.D. (2007). Eyewitness evidence: improving its probative value. *Psychological sciences in the public interest*, 7, 45-75.
Wells, G. L., & Turtle, J.W. (1988). What is the best way to encode faces? In M.M. Gruneberg, P. Morris, & R. Sykes (Eds.). *Practical aspects of memory: current research and issues* (pp. 163-168). Chichester: Wiley.
Wright, D.B., & Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. *Acta-Psychologica*, 114, 101-114.
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 trials. *Perception*, 14, 737-746.
Young, A.W., Hellawell, D., & Hay, D.C. (1987). Configural information in face perception. *Perception*, 16, 747-759.