

Witness Interviews: Does recall of relational information improve identifiability of a facial composite?

Faye Skelton

School of Psychology, University of Central Lancashire

FCSkelton@uclan.ac.uk

Charlie Frowd

School of Psychology, University of Central Lancashire

CFrowd@uclan.ac.uk

Sally Andrews

School of Psychology University of Central Lancashire

SAndrews1@uclan.ac.uk

Abstract

Facial composites are used by Police to generate lines of enquiry; unfortunately composites made by traditional ‘feature’ systems are not often accurately named. One reason could be that these systems tend to rely on descriptions of the criminal’s facial features, when it has been shown that relationships and distances between facial features—the relational information—is of importance for face recognition. Here, we present two experiments to investigate the usefulness of probing for relational information within witness interviews. Participant-witnesses underwent a typical cognitive interview (CI), an interview in which featural information was probed for before relational information (FR), or an interview in which probing for relational information preceded probing for featural detail (RF). Composites constructed of familiar targets with no delay were recognised better in the former two interviews than the latter, suggesting that relational information interferes with subsequent recall of featural information. However, after a 24 hour delay composites constructed of unfamiliar targets did not differ significantly for naming rates by interview type. This indicates no naming benefit for recalling relational information. However, RF composites were rated as significantly better likenesses to target images after a 24-hour delay, and so future work could explore this further to assess what aspect of the image is improved by recall of relational information.

1. Introduction

Facial composites are visual likenesses of a criminal's face, often constructed by witnesses to and victims of crime in order for the Police to generate lines of enquiry by circulating the images to other police forces or via the media. Research shows that traditional feature-based composites constructed immediately after exposure to a target image are correctly named about 20% of the time [1], [2]. Unfortunately, in practice facial composites are rarely constructed within the first few hours of a crime occurring; rather, it is likely to be around 24 to 48 hours before officers are able to interview and help witnesses to construct a composite. After a 48-hour delay, correct naming drops to about 3% [1], [2]. Facial composite research is focused on ways of improving the naming rates in order to help the Police solve more crimes.

Composite construction involves two distinct cognitive processes. Firstly an interview is conducted which requires the witness to *recall* as much information as they can about the crime. Secondly, the composite software operator inputs the witness description of the face into specialised software (such as E-Fit or PROfit) to identify individual facial features (e.g. face shape, hair, eyes, nose, mouth) within the database that match the witness's description. From these selections, the witness uses *recognition* to identify the closest match for each feature to the face they have in memory—this is normally achieved by switching alternative features in and out of a complete face; the size of and distances between features can also be adjusted. Finally, artwork can be used to add wrinkles, stubble, scars, and so on.

Of these two processes, recall and recognition, it is recall (the witness description) that is critical for Police to identify possible perpetrators, and for software operators to locate features within the software databases. With a poor description, the chances of creating a good visual likeness of the perpetrator are slim. Unfortunately, it is recall that is most detrimentally affected by the passage of time [3], and so the quantity and quality of a witness description can deteriorate significantly by the time a witness is interviewed.

It is therefore essential that the main interview with the witness or victim generate as much accurate information about the criminal(s), and as few inaccuracies as possible, to make the search for those responsible more efficient and to minimise miscarriages of justice. The Cognitive Interview (CI) [4], [5] consists of a number of mnemonics to facilitate accurate recall from the witness; for example, context reinstatement (CR) requires the witness to mentally recreate the context of the crime (e.g. emotions, physical environment), and has been shown to increase the quantity and accuracy of details provided by the witness [6].

Research has also demonstrated that appreciable improvements can be made to the quality of facial composites by modifying the interview administered to witnesses. For example, a Holistic-Cognitive Interview (H-CI, [7] is known to improve the quality of composites produced using traditional feature-based systems such as PROfit. This is because feature-based systems focus on recall of individual facial features, which do not effectively discriminate between large numbers of faces. Rather, we know that faces are encoded as wholes rather than as collections of separate features [8], [9]. Holistic judgements (for example, honesty, masculinity) require a face to be encoded as a whole, thus facilitating the encoding of relationships between features and subsequently improving recognition. The implication of this interview is that participants are better able to recognise a good likeness in features presented to them during composite construction, resulting in composites that are a better likeness to the faces on which they are based [7].

As well as accurate selection of individual features, a critical part of composite construction involves setting the distances between features themselves—the relational information. It has been

demonstrated that humans are better able to recognise a familiar face if it contains an unfamiliar feature (for example a nose not belonging to that face) than if the relational information has been changed, such as if the eyes were moved further apart [9]. Furthermore, altering relational information not only impacts on recognition of that feature, but also of other features [10]. It has also been found that better quality composites have more accurate spatial relations [7]. These findings demonstrate the importance of correctly recreating the relational information in a composite. However, it would appear that witnesses are *not* normally asked to provide relational information about an offender's face—although it is entirely possible that such information may help them to position facial features correctly, thereby improving subsequent feature selection (a better context in which to select other features) and the overall identifiability of their finished composite.

Providing relational information about a target face may be valuable in another way. It has been shown that, under some circumstances, detailed recall of a face can hinder subsequent recognition, a phenomenon known as verbal overshadowing [11]. This is thought to be due to rich visual memories being overwritten by their translation into verbal descriptions, which are by nature less detailed. The effect has been found to be relatively short-lived (e.g. less than about 30 minutes, [12], but it does appear to apply to feature-based composite production [13]. Probing for relational information after featural information could plausibly focus witnesses' attention on the important spatial information directly before they construct their composite, preventing or overcoming any 'overshadowing' effect and resulting in a more identifiable composite.

The current work therefore aimed to investigate whether incorporating recall of relational information in the initial witness interview would improve recognition of resulting composites. It also investigated where in the interview such recall would be most effective. To achieve these objectives, a psychological experiment was carried out involving participants who received one of three interview types. All interviews began with rapport-building. In one condition, part of the fairly-standard police cognitive-interview procedure, constructors recalled the face in detail without interruption ('free' recall) and then were asked to provide more information about each individual feature ('cued' recall)—we refer to this condition as CI. The second condition was the same, except that constructors were subsequently asked to engage in free and then cued recall about the placement of features on the face (relational information)—feature-relational (FR); the third condition asked for free and cued recall of relational information before free and cued recall of featural information—relational-feature (RF). Interviews were followed by construction of a composite using PROfit, a widely used system in the UK, which was controlled by a suitably-trained software operator (the experimenter).

The research therefore aimed to determine whether the inclusion of 'relational' recall benefits recognition for the person constructing the face, and secondly whether any benefit is dependent upon the positioning of relational recall within the interview (i.e. before or after free/cued recall).

2. Experiment 1 – Immediate construction of familiar male targets

2.1. Composite construction

2.1.1. Design: The design was between-participants for interview type, with each participant constructing their composite after a cognitive interview (CI), featural-relational interview (FR), or relational-featural interview (RF). Participants were randomly allocated to one of these three conditions.

2.1.2. Participants: Thirty-six students and staff (12 males and 24 females; mean age 29.7 years) were recruited as participant-witnesses from the UCLan Preston campus.

2.1.3. Materials: Images of 12 famous males (Ben Affleck, Gordon Brown, David Cameron, George Clooney, Daniel Craig, Tom Cruise, Harrison Ford, Richard Hammond, Anthony Hopkins, Paul McCartney, Piers Morgan and Arnold Schwarzenegger) were sourced using the Google Images search engine on the internet. These images were printed to approximately 7cm x 8cm size and placed into an envelope (there were three envelopes, one for each condition, each containing a full set of ten images). Participants' descriptions were recorded on a standard police verbal description sheet. Composites were constructed on a PC using PROfit software.

2.1.4. Procedure: Participants were tested individually, and asked to select a photograph from one of three envelopes (for CI, FR or RF), selected randomly with equal sampling. If they were unfamiliar with the identity, then they reselected until they had a target with whom they were familiar. They were asked to study the image for one minute, after which the experimenter explained that they would undergo an interview to elicit a description of the face, and that this would be followed by a composite construction phase using PROfit software—see Figure 1 for an illustration of this system in use.

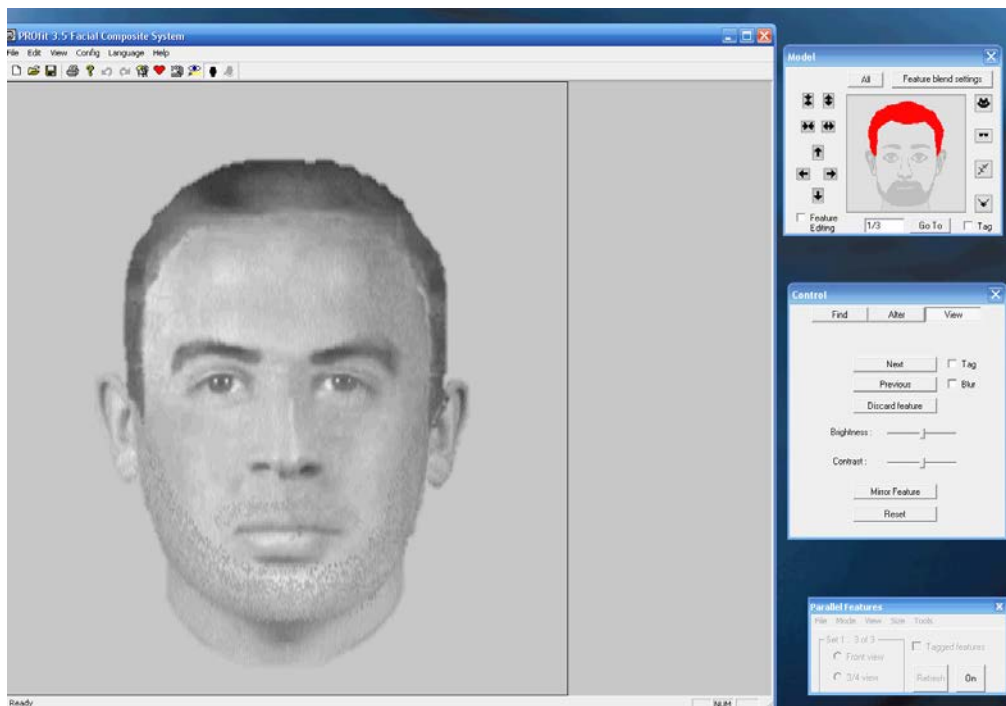


Figure 1. Screenshot of PROfit: the left panel contains the composite face; top right is the model window for selecting features to work on, and also for manipulating size and positions of features; middle right is the control panel for entering descriptions given by the constructor.

All participants underwent an initial rapport-building stage to put them at ease, followed by context reinstatement. Participants undergoing the CI then engaged in free recall whereby they described all that they could remember about the face, then had their description read back to them, with the interviewer asking after each detail whether they could recall any more about that particular feature (cued recall). Participants undergoing the FR interview were subsequently asked to focus on recalling details about the individual features (free then cued), before being asked to describe the distances between different features. Participants in the RF condition were asked instead to focus on recalling the distances between the facial features (free then cued), before moving on to describe the features

themselves. There were a total of seven ‘prompts’ for each type of information; for features these included eyes, nose, mouth, hair and for relational information they included the distance between the eyes, the position of the ears, the distance between the eyes to the tip of the nose, the nose to top lip, and so on.

Each participant’s description was entered into PROfit to narrow down to approximately 12 to 20 examples for each feature, with the witness selecting the best match for each facial feature in turn (hair, face shape, eyes, eyebrows, nose, mouth, ears). Subsequently, they were guided through resizing and positioning features and offered the option of adding artwork (for wrinkles, stubble etc.) until they were happy with the likeness (see Figure 2 for examples). Sessions took approximately an hour to complete including the time taken for the interview.



Figure 2. Example composites of UK politician Gordon Brown, constructed after a CI (left), FR (middle) and RF (right) interview. Each image was constructed by a different person. For copyright reasons, we are unable to reproduce the target photograph here.

2.2. Composite Evaluation

2.2.1. Design: Naming and rating were repeated-measures for interview type, with all participants naming or rating all 36 composites (12 from each condition). The order of presentation was randomised for each participant.

2.2.2. Participants: Nine participants (6 males; mean age 21.2 years) took part in the naming task, and 16 (11 males; mean age 27.2 years) in the likeness-rating task. Participants were students and staff recruited from the UCLan Preston campus.

2.2.3. Materials: Composites were printed twice in greyscale to approximately 7cm x 8cm. They were placed into envelopes for either the naming task or the rating task. The target images (described under composite construction) were also used for target naming and the likeness rating task.

2.2.4. Procedure: Participants carrying out the first evaluation task were asked to try and name the 36 composites of celebrities (we refer to this part as ‘spontaneous’ naming). They were informed that

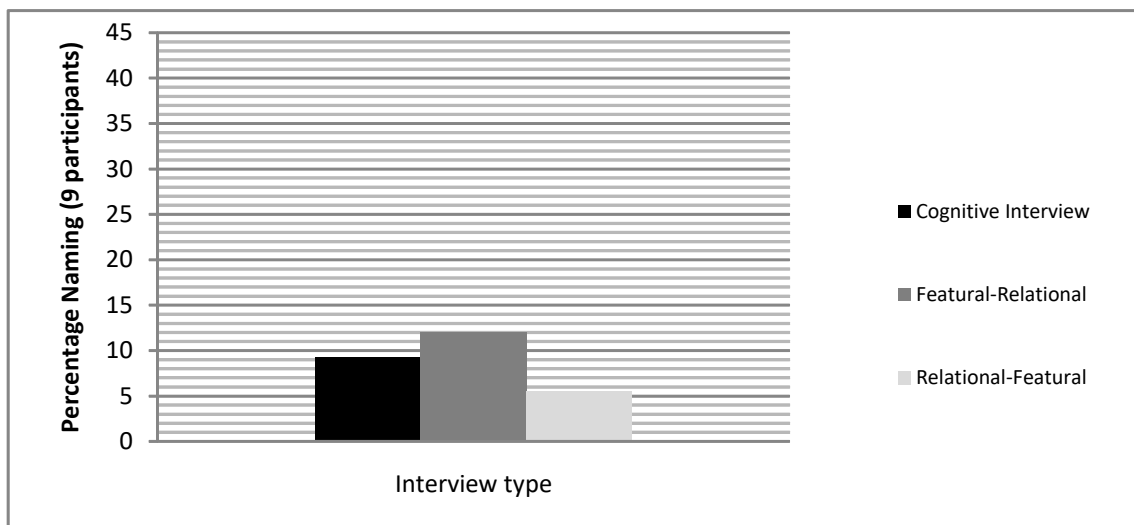
there were repeated identities. Composites were presented individually. They were then asked to name the 12 target images, to check that they were familiar with the targets. Subsequently, participants were presented with the 36 composites for a second time to name ('cued' naming).

For the rating task, other participants were presented with the composites individually. They were also shown the target images so that they could provide a judgement as to how good a likeness (1 = very poor / 10 = very good) the composite was to the target.

For both tasks, participants were tested individually. Composites and target photographs were presented in a different random order for each person.

2.3. Results

Participants' accurate spontaneous naming scores (maximum 12) were firstly subjected to a 2 (Task: spontaneous or cued naming) x3 (interview: CI, FR or RF) mixed ANOVA. Results indicated a significant effect of task, $F(1,8) = 110.22, p < .001$; as expected cued naming was significantly better than spontaneous naming. Interview was also significant, $F(2,16) = 6.42, p = .009$. Post-hoc paired t-tests confirmed that the RF composites were named significantly worse than both the CI ($t(8) = 4.00, p = .004$) and the FR composites ($t(8) = 2.92, p = .019$). Finally, the interaction between task and interview was significant, $F(2,16) = 5.09, p < .05$. To examine this interaction, one-way ANOVAs were conducted separately on spontaneous and cued naming data. The one-way repeated-measures (RM) ANOVA on spontaneous naming indicated no significant effect of interview, $F(2,16) = 2.21, p$



= .142.

Figure 3. Percentage correct spontaneous naming of famous male composites.

The RM ANOVA was significant for cued naming, $F(2,16) = 7.18, p = .006$. Post-hoc paired t-tests revealed the same pattern as for the overall ANOVA, with RF composites being accurately named significantly less than both CI ($t(8) = 3.73, p = .006$) and FR ($t(8) = 3.08, p = .008$) composites.

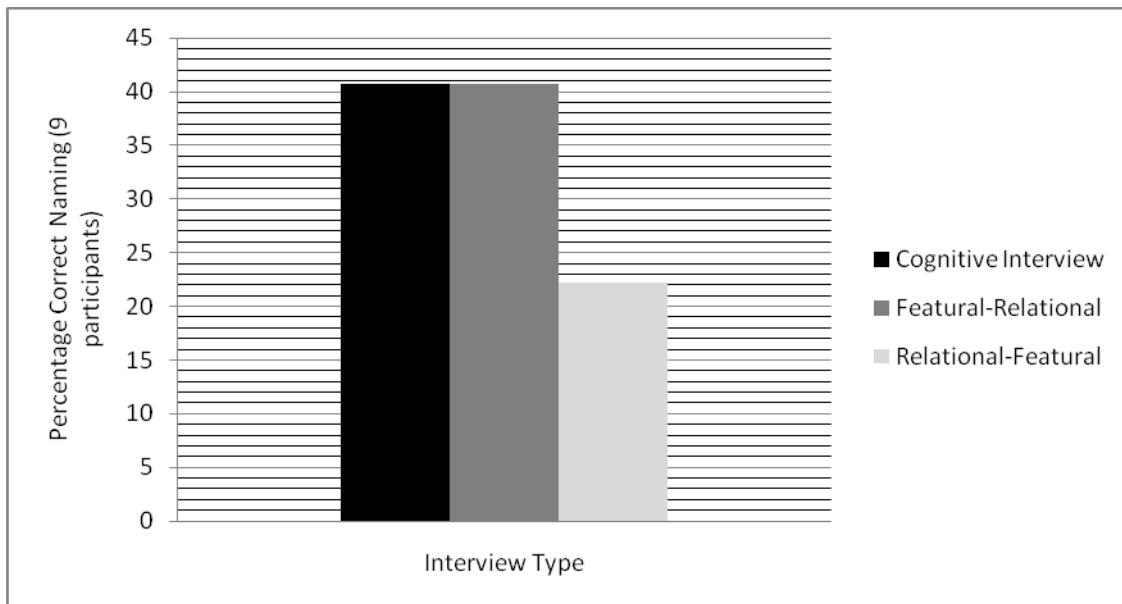


Figure 4. Percentage correct cued naming of famous male composites.

The likeness rating data were also subjected to RM ANOVA. The pattern of results was the same as for cued naming above. There was a significant effect of interview, $F(2,30) = 30.96$, $p < .001$, with post-hoc paired t-tests indicating that RF composites were rated a worse likeness to the target than both CI ($t(15) = 5.08$, $p < .001$) and FR composites ($t(15) = 8.84$, $p < .001$).

2.4. Discussion

Both naming and likeness-ratings data show that probing for relational information after featural descriptions elicits equivalent-quality composites to those produced after a CI. However, prompting for relational information first significantly worsens composite quality, indicated by correct 'cued' naming and by likeness ratings.

Experiment 2 aimed to improve the ecological validity of these findings by asking participant-witnesses to construct a composite of an unfamiliar face, and after a 24-hour delay. This not only allowed us to better extrapolate findings to the situation faced by most witnesses, but we also hoped it would provide an insight into the apparent detriment caused by recalling relational information before featural detail.

3. Experiment 2 – Delayed construction of unfamiliar male targets

Experiment 1 utilised optimal conditions, with participants constructing a composite of a familiar face with no delay. In reality, witnesses of crimes are rarely interviewed immediately a crime occurs, and so Experiment 2 aimed to investigate any differences in the three interview techniques when administered approximately 24 hours after a participant-witness had seen an unfamiliar target face.

3.1. Composite Construction

3.1.1. Design: The design was between-participants for interview type, with each participant constructing their composite after a cognitive interview, featural-relational interview, or relational-featural interview. Participants were randomly allocated to condition as before.

3.1.2. Participants: Thirty students and staff (16 males, 14 females; mean age 31.7 years) were recruited as participant-witnesses from the UCLan Preston campus.

3.1.3. Materials: Images of ten premiership level footballers (Craig Bellamy, Michael Carrick, Joe Cole, Steven Gerrard, Frank Lampard, Gary Neville, John O'Shea, Paul Scholes, Alan Smith, and John Terry) were sourced using Google Images. These were printed the same as for Experiment 1.

3.1.4. Procedure: The construction procedure was same as that of Experiment 1, except that after viewing their target photograph for one minute, participants returned between 22 and 26 hours later to be interviewed and construct a composite.



Figure 5. Example composites of UK footballer Gary Neville constructed after a 22 to 26 hour delay in the CI (left), FR (middle) or RF (right) condition.

3.2. Composite Evaluation

3.2.1. Design: Naming and rating were repeated-measures for interview type, with all participants naming or rating all 30 composites (10 from each condition). The order of presentation of composites and targets was randomised for each person.

3.2.2. Participants: Fifteen participants (13 males) took part in the naming task (mean age 24.5 years), and nine (8 males) in the likeness rating task (mean age 24.4 years). All were students or staff recruited from the UCLan Preston campus. Naming participants confirmed that they were football fans, and therefore would be familiar with the target identities.

3.2.3. Materials: Composites were printed twice in greyscale to approximately 7cm x 8cm. They were placed into envelopes for either the naming task or the rating task. The target images (described under composite construction) were also used for target naming and the likeness rating task.

3.2.4. Procedure: The procedure for evaluating composites (by naming and likeness ratings) was the same as that described in section 2.2.4 except that, for the naming task, participants were told that the composites were of premiership-level footballers.

3.3. Results

The data were initially subjected to a 2x3 mixed ANOVA, which revealed a significant effect of task, $F(1,9) = 45.00, p < .001$; again cued naming was significantly better than spontaneous naming. There was however no significant result for interview, $F(1,9) = 0.17, p = 0.85$, or for the interaction, $F(2,18) = 0.31, p = .74$. No further analyses were conducted on naming data.

However, analysis of participants' likeness ratings revealed a significant effect of interview, $F(2,16) = 3.66, p = .049$. Comparisons between RF composites and CI ($t(8) = 2.02, p = .078$) and FR composites ($t(8) = 2.10, p = .069$) were marginally significant, indicating that RF composites were rated as better likenesses.

3.4. Discussion

Unlike Experiment 1, Experiment 2 shows that after a 24-hour delay, probing for relational information does not reduce composite quality when assessed by naming tasks; rather, all three interviews lead to equivalent quality composites. Likeness-rating comparisons however reveal that RF composites were rated as better likenesses than CI and FR composites.

4. General Discussion

These experiments aimed to evaluate the usefulness of probing for relational information when interviewing victims of, and witnesses to, crime prior to face constructing using a traditional feature-based composite. Previous research had indicated that composites were named infrequently and we proposed that this may be partly due to interviews focusing on descriptions of individual features, rather than how the features are spaced in relation to each other, which is known to be important for face recognition [9], [10]. Our findings revealed that with an interview immediately following exposure to a target face, recalling relational information before recalling details of the features results in poorer quality composites, measured by both cued naming and likeness ratings. However, after 24 hours, the inclusion of a relational recall phase has no impact on resulting composite naming. Interestingly, composites produced after the RF interview were rated a better likeness to the identities they represented.

This research adds to evidence that the nature of the interview is important for composite naming, but that the effects of relational recall change over time. This is perhaps unsurprising, as it is well-known that memory decays with time [14]; however the finding that RF interviews produced worse quality images immediately but not after a 24-hour delay suggests that memory for relational information may decay at a different rate to that for featural information. Although RF composites elicited the poorest naming when constructed immediately, there was little reduction in naming after a delay, whereas both the CI and FR composites were named with much lower accuracy after a delay than immediately. Thus, recall of featural details is markedly worse after a 24 hour delay. On the contrary, recall of relational information does not appear to change. We do acknowledge, however, that participant-witnesses found it difficult to produce relational descriptions, and so perhaps recall of such information is very low whenever it is requested, leaving little room for decay.

Experiment 1 required construction of an identity with who the participant-witness was familiar, while Experiment 2 required construction of an unfamiliar face. Though face recognition research shows that familiar faces are recognised with much greater accuracy than unfamiliar faces (those seen maybe once or twice), composite research shows that familiarity does not impact upon composite

quality unless the target image is present throughout construction [1],[15]. Familiarity is therefore unlikely to explain the apparent disparities between findings of the two experiments reported here. There are, however, alternative explanations for our results.

Our participants were informed that they would need to provide a description of the face. We know that in this situation participants tend to focus on individual facial features [16]. However, it is likely that most people were not expecting to recall relational information. Therefore, this element of their interview may have acted as a “distractor” task, preventing rehearsal of feature details and subsequently reducing the amount or accuracy of detail recalled in the RF condition only. It is likely that participants’ attentional resources would be primarily directed to the more difficult task – in this case trying to remember and describe relational information- [17], leaving little possibility for rehearsal of their featural description. Lack of rehearsal would reduce the amount of featural detail available for subsequent recall. Alternatively, the relational recall element could have interfered with subsequent recall of features [18]. In either case, the long-term memory for the face would decay over 24-hours leaving little scope for the relational recall to reduce performance further, consistent with our findings from Experiment 2.

There may also be a role here for inhibition [19], in that participants in the RF condition had to suppress their featural descriptions in order to recall relational details. This could have reduced availability of featural memories when they subsequently attempted recall of the initially suppressed feature details. This would not impact upon the FR condition as they recalled feature detail first, as they most likely expected. Further research is needed to examine whether the RF decrement is a form of “overshadowing” or is simply acting as a distractor task. If the former, then the effect should be apparent where relational information is recalled directly before featural information, but not when there is a short delay between relational and featural recall. If the latter, then an unrelated task prior to feature recall should have the same effect.

Though not in keeping with the naming data, the observed advantage in likeness ratings for RF composites after a 24-hour delay does not pose a problem for these explanations. While naming is a measure of recognition and therefore of direct benefit to the Police, likeness ratings indicate the level of similarity between the composite and target image. Unlike recognition, which is holistic, this could be attributed to a particular feature or region of the face that is a particularly good likeness. Future work could therefore usefully explore what element(s) of composites are rated as a better likeness in RF as opposed to CI and FR composites. This would provide further insight into which (if any) aspects of relational recall can be useful to the composite production phase.

In conclusion, our data show that recall of relational information does not benefit facial composite recognition/naming after 24-hours, and therefore should not currently be incorporated into witness interviews. However, RF composites are rated as better likenesses when constructed 24-hours post-event, and future research is outlined which could provide further insight into the specific effects of relational recall to determine whether it can be effective at any point in the composite construction process.

References

[1] C.D. Frowd, V. Bruce, A. McIntyre, & P. Hancock. The relative importance of external and internal features of facial composites. *British Journal of Psychology* 98(1), 2007, 61-77. doi: 10.1348/000712606X104481.

- [2] C.D. Frowd, D. Carson, H. Ness, J. Richardson, L. Morrison, S. McLanaghan, & P.J.B. Hancock. A forensically valid comparison of facial composite systems. *Psychology, Crime & Law* 11(1), 2005, 1-21.
- [3] P.N. Shapiro, & S. Penrod. Meta-analysis of facial identification studies. *Psychological Bulletin* 100, 1986, 139–156.
- [4] R.E. Geiselman, R.P. Fisher, D.P. MacKinnon, & H.L. Holland. Enhancement of eyewitness memory with the cognitive interview. *The American Journal of Psychology* 98(3), 1985, 385-401
- [5] G. Wells, A. Memon, & S.D. Penrod. Eyewitness evidence: improving its probative value. *Psychological sciences in the public interest* 7, 2007, 45-75.
- [6] L. Hammond, G.F. Wagstaff, & J. Cole. Facilitating Eyewitness Memory in Adults and Children with Context Reinstatement and Focused Meditation. *Journal of Investigative Psychology & Offender Profiling* 3, 2006, 117-130. doi: 10.1002/jip.
- [7] C.D. Frowd, V. Bruce, A.J. Smith, & P.J.B. Hancock. Improving the quality of facial composites using a holistic cognitive interview. *Journal of Experimental Psychology. Applied* 14(3), 2008, 276-87. doi: 10.1037/1076-898X.14.3.276.
- [8] D. Maurer, R.L. Grand, & C.J. Mondloch. The many faces of configural processing. *Trends in Cognitive Sciences* 6(6), 2002, 255-260. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12039607>.
- [9] J. Tanaka, & M. Farah. Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology* 85, 1993, 397-405.
- [10] J.W. Tanaka, & J.A. Sengco. Features and their configuration in face recognition. *Memory & Cognition* 25(5), 1997, 583-92. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9337578>.
- [11] J.W. Schooler, & T.Y. Engstler-Schooler. Verbal overshadowing of visual memories: some things are better left unsaid. *Cognitive Psychology* 22, 1990, 36-71.
- [12] K. Finger, & K. Pezdek. The effect of the Cognitive Interview on face identification accuracy: Release from Verbal overshadowing. *Journal of Applied Psychology* 84, 1999, 340-348
- [13] C.D. Frowd, & S. Fields. Verbalisation effects in facial composite production. *Psychology, Crime & Law*. 2011, DOI: 10.1080/10683161003623264
- [14] H.D. Ellis, J.W. Shepherd, & G.M. Davies. The deterioration of verbal descriptions of faces over different delay intervals. *Journal of Police Science and Administration* 8, 1980, 101-106.
- [15] G.M. Davies, P. van der Willik, & L.J. Morrison. Facial Composite Production: A Comparison of Mechanical and Computer-Driven Systems. *Journal of Applied Psychology* 85, 2000, 119-124.
- [16] K.R. Laughery, C. Duval, & M.S. Wogalter. Dynamics of facial recall, in Ellis, H.D., Jeeves, M.A., Newcombe, F., & Young, A. (Eds.). *Aspects of face processing*, 1986, (pp. 373-387). Dordrecht: Martinus Nijhoff.
- [17] E.C. Merrill, & M. Peacock. Allocation of attention and task difficulty. *American Journal of Mental Retardation* 98(5), 1994, 588-593
- [18] M.G. Berman, J. Jonides, & R.L. Lewis. In search of decay in verbal short-term memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition* 35(2), 2009, 317-33. doi: 10.1037/a0014873.
- [19] J.T. Nigg. On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin* 126, 2000, 200–246.