Improving Composite Images of Faces Produced by Eyewitnesses

Hayley Ness, B.Sc. (hons)

Thesis submitted to the University of Stirling for the degree of Doctor of Philosophy, November 2003.

TABLE OF CONTENTS

Acknowledgements

It goes without saying that research would not be possible without the help, advice and support of numerous people. In particular, the support, encouragement and advice I have received from Prof. Vicki Bruce, my principal supervisor has been invaluable. Her intellect and humanity have not only made her an excellent supervisor, they have also made her a pleasure to know. Similarly, my second supervisor Dr. Peter Hancock has also been a star, providing me not only with much needed support and advice but also with the odd bit of cash! Thanks Peter!

I would also like to thank everyone in the Faces lab, both past (Karen Lander, Zoë Henderson, Chris O'Donnell and Anthony Cox) and present (Fraser Smith, Charlie Frowd, Helen Chang and Robbie Cooper), not only for their help and insightful comments but also for many enjoyable nights in the pub (and the odd giggle on the golf course!). The same thanks extend to Sandra Quinn, Sivakumar Anandaciva and many others. I would also like to thank all of the support staff (Bob Lavery, Stephen Stewart, Catriona Bruce, Peter Hucker and Bruce Sutherland, Kerry Fairburns) for their assistance. In particular, I would like to thank Kay Faichney and Claire Wilson not only for their invaluable administrative assistance but also for making break-times more enjoyable!

I would also like to thank Andy, Amy, my Mum and my brother Andrew. Not only for putting up with me and feeding me occasionally, but also just for being there. The last few years have been particularly difficult for all of us and I think we all deserve a pat on the back for the tremendous courage and determination that we have all displayed.

Abstract

When a witness views a crime, they are often asked to construct a facial likeness, or composite of the suspect. These composites are then used to stimulate recognition from someone who is familiar with the suspect. Facial composites are commonly used in large scale cases e.g. Jill Dando, Yorkshire Ripper, however a great deal of research has indicated that facial composites perform poorly and often do not portray an accurate likeness of the suspect. This thesis therefore examined methods of improving facial composites. In particular, it examined methods of increasing the likeness portrayed in composites, both during construction and at test.

Experiments 1 to 3 examined the effectiveness of a new three-quarter-view database in PROfit. Experiment 1 examined whether the presentation of composites in a three-quarter-view composite will aid construction. Participant-witnesses were exposed to all views of a target and the results indicated that three-quarter-view composites performed as well as full-face composites but not better. Experiments 2 and 3 then examined whether the presentation of two composites (one in a full-face view and the other in a three-quarter-view) from the same participant-witness would increase performance above the level observed for a single composite. The results revealed that two views were better than one. In addition, experiment 3 examined the issue of encoding specificity and viewpoint dependency in composite construction. All participant-witnesses were exposed to either one view of a target (full-face or three-quarter) or all views and they were asked to construct both a full-face and a three-quarter-view composite. The results indicated that performance was better when all

views of a face had been presented. When a target had been seen in a three-quarter-view, it was better to construct a three-quarter-view composite. However, when a target had been seen in a full-face view, performance for both full-face and three-quarter composites was poor.

Experiments 4 to 8 examined whether the presentation of composites from multiple witnesses would increase performance. The results revealed that morphing composites from four different witnesses (4-Morphs) resulted in an image that performed as well as or better than the best single image. Further experimentation attempted to examine why multiple composites performed well. In particular, it was asked whether multiple composites performed well because they contained varied information or whether they performed well because they just contained more information. Multiple composites from both single and multiple witnesses using the same (PROfit) and different (PROfit, E-FIT, Sketch, EvoFIT) composite techniques were compared and the results revealed that multiple composites performed well because they contained different memorial representations. This combination of different memorial representations appeared to result in an image that was closer to the ideal, or prototypical image.

Experiments 9 to 12 examined the relationship between verbal descriptions and composite quality. The results revealed that there was no clear relationship between the amount of description provided, the accuracy of the description and performance of the resulting composite. Further experimentation examined whether the presentation of a composite and a description would increase performance above the level observed for a single composite. The results revealed that the combination of a description and a composite from the same participant-witness did increase performance. This indicated that descriptions and composites might contain differing amounts and types of featural and configurational information.

Both the theoretical and practical implications of these results are discussed.

Experiments 1, 2 and 3 of this thesis have been submitted for publication. Ness, H., Hancock, P. J. B., Bowie, L. and Bruce, V. Are two views better than one? A study investigating recognition of full-face and three-quarter-view composites. *Applied Cognitive Psychology.*

Experiment 4 of this thesis appears in Bruce, V., Ness, H., Hancock, P. J. B., Newman, C. and Rarity, J. (2002). Four heads are better than one: combining face composites results yields improvements in face likeness. *Journal of Applied Psychology. 87 (5),* 894-902.

Other Publications

Frowd, C. D., Carson, D., Ness, H., Richardson, J., Morrison, L., McLanaghan, S., Hancock, P. J. B. Evaluating Facial Composite Systems. *Manuscript accepted for publication in Psychology, Crime and Law.*

Frowd, C. D., Carson, D., Ness, H., McQuiston, D., Richardson, J., Baldwin, H., Hancock, P. J. B. Contemporary Composite Techniques: The impact of a forensically-relevant target delay. *Manuscript accepted for publication in Legal and Criminological Psychology.*

1

Facial Composites

Facial composite systems have been developed in order to aid eyewitness recall of a briefly viewed unfamiliar face. In order to construct a facial composite, an eyewitness is initially asked to recall the facial features verbally. This information is then portrayed visually using a composite system. The witness is then asked to recognise whether this visual information (composite) matches his or her own internal representation of that face. Alternative features are chosen and alterations are made to the composite until the witness achieves a desirable likeness of the seen face. The facial composite is then displayed in the media in the hope that someone who is familiar with the suspect will recognise them from the composite.

Traditional methods of constructing facial composites involved choosing isolated features and physically assembling them in a frame. However, a great deal of research has indicated that this process is at odds with the way in which faces are normally processed. As a result, computerised composite systems (e.g. PROfit) have been developed in a bid to overcome many of the difficulties associated with these methods of construction. This introductory chapter will therefore begin by examining the difficulties associated with composite construction and will demonstrate that despite technological improvements in system design, facial composites still perform quite poorly. This chapter will discuss possible memorial difficulties in composite construction and will examine why recalling, describing and reconstructing an unfamiliar face are such difficult tasks. In addition, this chapter will highlight possible methods of improving the likeness of composites both during construction and at test. At the end of this chapter, the specific questions addressed in this thesis will be outlined and a brief overview of the remaining chapters will be given.

Manual composite systems

One of the earliest methods of constructing a facial likeness of a suspect involved using a manual technique. Such techniques involved physically reconstructing the face, either by assembling cut-out features, or by drawing the face. In order to sketch a likeness of a suspect, a forensic sketch artist initially spends a substantial amount of time eliciting a verbal description from a witness (an example given in Davies & Little, 1990 is 30 to 45 minutes). The artist then uses this information to sketch the face. One benefit of this approach is that the artist can add a great amount of detail to the face. Similarly, any alterations and/or fine detail can be made to the exact specification of the witness.

However, Maudlin & Laughery (1981) criticised this approach and argued that the process of constructing a composite sketch would require the witness to produce a great amount of detail, which they may not actually remember. Similarly, this approach was criticised for being time consuming and costly as specialist artists had to be employed. Therefore, there was a need to develop a system that could not only allow witnesses to quickly recognise and choose alternative features, but one that could also be used by officers who had little or no artistic ability.

Identikit and Photofit

The first such systems were Identikit and Photofit. These were mechanical systems and involved physically assembling cut-outs of facial features in a frame (Davies et al, 2000). The original Identikit, invented by Hugh McDonald, was first introduced to the U.K two years after its initial release in the United States in 1959 (Davies, 1983). Identikit consisted of individual line drawn features, which were printed onto single transparencies. A facial composite was then constructed by superimposing these feature transparencies in a frame. Further detail was then added by way of a marking pencil. This system contained a limited amount of features (130 hairlines, 102 eyes, 37 noses, 40 mouths and 52 chins) and a variety of accessories (hats, glasses, scars, moustaches, age lines and beards) (Kovera et al, 1997).

The Photofit system was first introduced in 1970. It was developed by Jacques Penry and became the main rival of Identikit (Davies, 1983). Photofit was similar to Identikit with the exception that it used black and white photographs, which were printed onto cards. A transparency was then placed on top for further modification with a marking pencil. A completed Photofit was a compilation of facial features from five different people, as only one feature (hair style with forehead, eyes, ears, chin and mouth) was extracted from each photograph during development (Bennet, 1986). There were a total of five hundred and fifty features in Photofit and three versions of the system: "male Caucasian, female Caucasian and Afro-Asian" (Ellis, Davies & Shepherd, 1978, pp297). The Identikit system was updated in 1975 (Identikit II) to include more 'lifelike' photographic elements, similar to Photofit. However, the system was still very limited as the amount of features was not increased and as such, still contained less than the Photofit system.

Photofit and Identikit composites are constructed in much the same way. An operator first elicits a verbal description from the witness. This information is then used to select individual features from arrays of alternatives, which are grouped on the basis of similarity (Ellis, Shepherd & Davies, 1975). When a set of features have been chosen, they are placed together in a frame to form a composite 'face'. The witness is then free to chose alternative features and to work with the operator to modify the image using marking pencils. Photofit differs slightly from Identikit in that a 'Visual Index' (small pictures of all features) is available for the witness to consult. In addition, a witness is also able to make extremely limited changes to the overall configuration of a Photofit by moving the forehead and chin up and down.

Evaluation of manual systems

A great deal of research has demonstrated that both Photofit and Identikit produce very poor facial representations. In particular, Ellis, Shepherd & Davies (1975) assessed the efficacy of Photofit by asking participants to reconstruct composites of two targets. Participants were presented with a Photofit of one target for ten seconds and were asked to reconstruct the image from memory. For the second target participants reconstructed the image with the target photograph available for inspection. The results revealed that participants found it extremely difficult to construct a Photofit, even when the target photograph was present during construction. Davies, Ellis & Christie (1981) also reported similar results.

In a further investigation, Ellis, Davies, & Shepherd (1978, experiment 1) manipulated target exposure time (ranging from 15 seconds to 2.5 minutes) and viewing instructions (intentional versus incidental). They found no significant effect for either length of exposure or encoding instructions. In experiment 2 the authors compared composites that had been constructed using a research operator and an experienced police operator. The results revealed no significant effect of operator, no significant differences between composites that were made with the target present or from memory and no significant benefit for artistic elaboration. A further experiment examined the usefulness of Photofit by asking participants to construct a Photofit of one target and sketch an image of a second target. All images were rated on a scale from 0-100 (0=positively misleading) and the results revealed that sketches were rated as better likenesses when they were constructed with the target present. However, there were no significant differences for Photofits; they performed equally poorly in both the target present and from-memory conditions. In addition, there was only a slight advantage for constructing a Photofit compared to allowing the participants to draw the face themselves.

These findings indicate that poor Photofit performance may be caused by limitations in the system rather than limitations in memory. Davies, Ellis & Shepherd (1978) investigated this further by examining the impact of delay on Photofit performance. In experiment 1, all participants were asked to construct composites of two targets, one immediately after exposure and one after a delay of one week. The results revealed no significant decrease in Photofit performance over time. In a second experiment, reconstruction ability was compared with recognition ability. Each participant was again asked to construct two composites, one immediately after exposure and one at a delay of three weeks. In addition, participants were required to recognise the target from a sequence of twenty faces. The results revealed that recognition performance did decrease after the three-week delay but Photofit performance did not. This led the authors to suggest that the poor performance of Photofit was indeed a result of limitations in the system, rather than limitations in memory.

Laughery & Fowler (1980) also reported similar results using Identikit. This investigation compared the efficacy of Identikit composites with sketches produced by arts graduates. Pairs of participants were exposed to a live target before constructing either an Identikit image or a sketch. In addition, each operator/artist constructed one target present image of each target. All images were rated for similarity and the results revealed that sketches performed significantly better when the target was present during construction. However, Identikit composites performed equally poorly in both conditions (target present and from memory). Similarly, McNeil, Wray, Hibler, Foster, Rhyne and Thibault (1987) reported no difference in Identi-kit performance after either a three-day or a three-week interval between target exposure and construction. This suggests that like Photofit, Identi-kit was also insensitive to factors that are known to affect recall performance.

Levels of representation

One of the major limiting factors concerns the level of representation within these systems. In a series of experiments Ellis, Davies & Shepherd (1978) examined recognition performance for Photofits. In the first experiment participants were presented with either eighteen Photofits, or eighteen photographs. At test they were required to recognise the study items from either thirty-six Photofits or photographs (type of image was held constant across study/test). The results revealed that recognition performance was higher for photographs then for Photofits. The authors then examined whether memory for Photofits was affected by the presence of feature boundary lines, inherent in all Photofit images. Using the same design as experiment one, two groups of participants were asked to recognise either a set of photographs or the same photographs with lines drawn on (similar to the feature boundary lines in Photofit). The results revealed poorer recognition performance for the lined photographs, compared to the unlined. Ellis et al then compared performance for unlined photographs, photographs with lines marking the features and photographs with random lines. The results again revealed that the unlined photographs performed better than the feature-lined photographs. Furthermore, there was no difference in performance between the two lined conditions, therefore indicating that the presence of lines interferes with face processing.

Similar difficulties in level of representation have been reported for Identikit composites (line drawn images). Davies, Ellis & Shepherd, (1978) compared performance for photographs, line drawings (outlines of the same photographs) and detailed line drawings (additional age lines, hair and eye details were added) of male celebrities. The results revealed that participants identified 90% of the photographs, 47% of the detailed line drawings and only 23% of the outline drawings. Further experimentation again revealed that even detailed line drawings were a very poor guide to facial likeness.

Leder (1999) examined why facial line drawings were difficult to process. In this investigation participants were presented with pairs of faces and were asked to decide whether the two images represented the same person. Images were presented both sequentially and simultaneously and consisted of two traced line drawings, two photographs or one of each type. Furthermore, in one block both of the images were presented at the same time, while in the second block the second image was presented 1000msec after the first. This therefore permitted an investigation of both perceptual and memorial processing of images. The results revealed that performance was better when the two images were photographs, compared to when either one or both were line drawings. In particular, sequential presentation produced poorer results when both images were line drawings, compared to simultaneous presentation, but this effect was not observed for photographs. This suggests that line drawings are particularly poor when participants are required to keep the image in memory.

Leder (1999) suggests that when faces are represented as line drawings, it is more difficult to process the face structurally because of the lack of textural information. Research by Bruce, Hanna, Dench, Healy & Burton, (1992) also indicated that line drawn faces were poorly recognised because of the absence of textural information. They examined performance for faces where both major, (e.g. eyes) and minor features (e.g. lines) were traced and line drawn faces that included shading information (information about relative areas of light and shade). The results revealed that recognition performance improved dramatically when the images contained textural information. In a later investigation, Bruce and Langton (1994) expanded on these results by demonstrating that performance improved when areas of pigmentation were preserved, rather than shading.

This research highlights major problems with the levels of representation in both Photofit and Identikit. For Photofit, the feature boundary lines have been shown to interfere with face processing. For Identikit, research indicates that the lack of textural information or pigmentation results in poor facial representations. One of the major difficulties, however, is that both systems require witnesses to recognise features in isolation.

Recognition of isolated features

Constructing an Identikit or a Photofit requires a witness to recognise isolated individual features and place them together to form a composite image. However, research has clearly demonstrated that this process may be at odds with the way in which faces are normally processed. Davies & Christie (1982, experiment 3) asked half of their participants to make similarity ratings on isolated Photofit features and half to rate the same features within the context of a whole face (a Photofit). Correlations on the similarity judgements revealed that all judgements correlated except those that required participants to make judgements on isolated features from memory, thus indicating that this procedure may result in greater errors.

Similarly, research by Farah, (1992) and Tanaka & Farah, (1993) has also found that parts of faces are easier to recognise when presented in the context of a whole face. In the study phase participants were required to learn face/name pairs. In the test phase they were presented with either two isolated features (e.g. one identical nose and one changed nose) or two faces (one where the nose had changed and one where the nose was identical). Participants were required to say which feature/face matched the target. The results revealed that performance was better when the features were presented in the context of a whole face.

Configural and featural processing

Furthermore, experiments using composite faces (Young, Hellawell & Hay, 1987; Hole, 1994) have demonstrated that when two halves of different faces are 'fused' together, it is difficult to isolate either half of the 'new' face. However, when the 'new' face is inverted, the task becomes easier. This effect has been found for both familiar and unfamiliar faces and suggests that faces are processed in terms of a facial 'gestalt' or 'schema' (Rakover, 2002).

In addition, Tanaka & Sengco (1997) asked participants to recognise features either in the same configuration, a new configuration or in isolation. The highest performance was obtained for features in the same configuration, followed by the new configuration, with the isolated feature condition achieving the lowest performance levels. Further evidence for the role of configural processing in face recognition comes from inversion studies (see Valentine, 1988 and Rakover, 2002 for reviews). In general, upright faces are thought to be encoded in terms of both featural and configurational information, while inverting a face appears to disrupt configural processing (e.g. Diamond & Carey, 1986).

In particular, Searcy & Bartlett, (1996) revealed that inversion disrupts the ability to perceive the spatial relationships between features (configuration). In their investigation they presented participants with faces that had either configurational changes (e.g. distance between nose and mouth) or local feature changes (e.g.

darkening eyebrows). Participants were asked to rate how grotesque the faces appeared and the results revealed that when the faces were inverted there was little difference in ratings for local feature changes. However, inversion did produce lower 'grotesqueness' ratings, indicating that it was much more difficult to perceive the configurational changes when the faces were inverted.

Leder and Bruce (1998) extended this work by examining this effect for memory for unfamiliar faces. Similar configurational and local feature changes were made to faces in order to increase the distinctiveness of the faces. The results indicated that faces that had been altered (either configurally or featurally) to make them more distinctive, were remembered better when presented upright. However, when the faces were inverted only the faces that had local feature changes were remembered better.

These results indicate that upright faces are processed in terms of both featural and configurational information. One hypothesis that has been put forward in an attempt to explain the interaction between configurational and featural information in face processing is the holistic hypothesis. This states that as a face is processed as a whole, attempting to break the face down into its constituent parts damages both perception and memory for the face. There are two interpretations of this holistic hypothesis. The first interpretation is that the whole face is more accessible than its constituent parts, while the second interpretation is similar in that it considers configurational information to be more important than featural information by (e.g. Farah, 1992; Tanaka & Farah, 1993; Tanaka & Sengco, 1997).

This evidence has severe implications for mechanical systems. As faces are processed holistically, not only have individual features been processed, but there is also an extensive degree of interfeature processing (Wells & Hryciw, 1984). Therefore, even when features are kept the same, the appearance of a face can be altered if the relationship between the features is altered (Bruce and Young, 1989). As Identikit and Photofit both have very limited opportunity for the manipulation of distances between features (Davies et al, 1982), it would be extremely difficult for a witness to construct a recognisable composite using such systems. This suggests that construction techniques that incorporate this information, by showing features within a whole face and allowing for the manipulation of facial configuration, may produce better facial likenesses. Indeed, Koehn & Fisher (1997) and Kovera et al (1997) both suggest that the incompatibility between configurational processing and composite construction using isolated features may explain why composites perform poorly.

Computerised composite systems

Computerised systems were developed in an attempt to overcome the difficulties inherent in the earlier manual systems. One of the earliest computerised composite systems was the Mac-A-Mug Pro system. This system contained a wide range of sketch-like facial features (184 hairlines, 117 eyes/eyebrows, 65 noses and 80 noses and accessories), displayed features within the context of a whole face, permitted changes in configuration and offered flexibility during editing with the use of the Mac-Paint programme. While the number of actual features in the system was not significantly greater than either the Photofit or Identikit systems, the ability to manipulate both features and configuration could result in an unlimited number of combinations (Kovera et al, 1997).

Effectiveness of Mac-A-Mug Pro

Early research on the effectiveness of Mac-A-Mug Pro suggested that this system had great potential. Cutler, Stocklein and Penrod (1988) examined the effectiveness of this system by asking an experienced operator to construct target-present composites of ten targets. Participants were then asked to match the composites to the target photographs. This investigation yielded high recognition rates and led the authors to conclude that this system could produce recognisable composites. Similarly, Wogalter and Marwitz (1991) reported that Mac-A-Mug Pro could produce good likenesses when the target was available for inspection during construction.

This research indicates that Mac-A-Mug Pro can produce recognisable composites and is therefore an improvement on mechanical construction. However, further research indicated that this system could only produce good likenesses when the composites had been constructed with the target in view. When the composites were constructed using a more ecologically procedure (i.e. from memory), the composites performed poorly. In particular, Kovera, Penrod, Pappas and Thill (1997, experiment 1) asked high school graduates to construct composites of familiar targets from memory (classmates and teachers from their high school). A further group of graduates were then asked to recognise the composites. These participants were asked to view fifty composites (10 targets) and give familiarity and confidence ratings as well as attempt an identification. After the identification phase, all participants were asked to rate their familiarity with the targets. The results revealed that participants rarely made an identification and where an identification was attempted (7%) only 1.7% were correct. This suggests that composites constructed from memory, even when they are of someone familiar, are of extremely poor quality. To assess this further, the authors conducted a second experiment. In this study, each composite was presented for thirty seconds, after which each participant was presented with a six-person sequential line-up (1 photograph of the target and 5 distractors). Participants were informed that the target may or may not be in the array and they were asked to rate the likelihood that each photograph was the target on a scale from 1 (definitely not the person in the composite) to 9 (definitely is the person in the composite). The results revealed that the ratings did not differ between the target and distractors, indicating that participants could not reliably 'pick' out the correct target from the array. Similar results were reported by Koehn and Fisher (1997).

This suggests that while the Mac-A-Mug Pro system was an improvement on the older mechanical systems, composites constructed from memory still performed very poorly. One of the difficulties may have been the level of representation of the images. While there was a greater number of features and a greater ability to manipulate both featural and configurational information, the features were essentially line drawings. As research has clearly demonstrated that textural information is important in face recognition (e.g. Bruce, Hanna, Dench, Healy & Burton, 1992; Bruce & Langton, 1994), computerised systems that use pigmented features (e.g. photographic features) may perform better.

E-FIT and PROfit

Two of the most commonly used computerised systems in the UK are E-Fit and PROfit. These systems differ from the Mac-A-Mug Pro system in that they contain photographic rather than sketch-like features. E-Fit and PROfit (see chapter 2 for a detailed description of PROfit) are extremely similar in design in that they both contain an extensive number of features. They display features within the context of a whole face and contain their own feature-editing tools to allow for sophisticated manipulation of both featural and configurational information. No research can be found that has compared the effectiveness of both E-Fit and PROfit to Mac-A-Mug Pro, however research has compared performance for both E-Fit and PROfit with older mechanical systems.

Davies, van der Willik and Morrison (2000) were the first researchers to compare performance for E-Fit and Photofit composites. In this investigation, participants constructed an E-Fit and a Photofit of two different targets (either unfamiliar or familiar). Each participant was initially asked to construct each composite from memory with a time limit of twenty minutes. The reference photograph was then introduced and the participant was able to make minor changes to the image. This procedure was then repeated with the second composite system and a different target. All composites (2 from memory and 2 target-present using two systems) were constructed in the course of a one-hour session. All composites were evaluated using naming and matching tasks and the results revealed that E-Fit composites performed better than Photofit composites when the target was present during construction. However, when participants constructed the composites from memory, E-Fit performed as poorly as Photofit. In addition, Davies et al state that the method of construction was extremely similar for both systems but that E-FIT was quicker. While participants were free to direct the order of construction, "participants generally followed a down-through strategy, beginning with the hair, followed by the eyes, nose, mouth and chin" (pg122). As there were no significant differences between systems when the composites were constructed from memory and method of construction was similar, it appears that despite technological advances, computerised systems do not appear to aid eyewitness recall any better than older mechanical systems.

However, other researchers (e.g. Brace, Pike & Kemp, 2000) have found that E-FIT composites perform well. While this investigation did not compare performance for E-FIT and Photofit, the results revealed that E-FIT composites can be useful investigative tools. For half of the composites in this investigation, a participant-witness worked with an operator and for the other half an operator worked alone. In addition, all composites were constructed twice, in that the initial composite was constructed from memory, then the target photograph was introduced and a second composite was made. At test, all composites were presented as pairs (1 from memory and 1 target present of the same target) to familiar participants. They were asked to try to identify the person depicted in the images and state which composite they felt best represented the person. The results revealed that the composites that had been constructed using just an operator (mean 34.72%) performed better than those that were made by a witness and an operator (mean 24.95%). In addition, for the operator alone condition, the composites that were constructed from memory were just as likely to be chosen as the 'best' composites, compared to the target-present composites. For the operator and witness condition, the target present composites were chosen as better likenesses significantly more often than the from-memory composites. Brace et al suggest that these results indicate that E-FIT composites are useful as they were identified well and in contrast to the older mechanical systems, the target-present images performed well. In addition, the authors suggest that the results indicate that the difficulty in constructing a composite appears to reflect the difficulty in translating a verbal description into a visual representation, not because the system performs poorly.

Similarly, recent research (Frowd, Carson, Ness, Richardson, McLanaghan & Hancock, 2003) has compared the effectiveness of more modern computerised systems with older mechanical systems. In this investigation, unfamiliar participants were asked to study a photograph of a famous target for one minute. After a short delay they were asked to construct one composite of the target using one of several different systems (e.g. E-FIT, PROfit, Photofit). To ensure that each system was used to its full potential, experienced operators were used and all feature-editing tools were utilised. In addition, rapport building and cognitive interview techniques were used to ensure that initial recall was as complete as possible. All composites were evaluated using naming, sorting and array tasks and the results revealed that both E-Fit and PROfit performed significantly better than Photofit, with E-Fit and PROfit achieving equivalent levels of performance. These results therefore contrast with those reported by Davies et al (2000) by indicating that E-FIT and PROfit can produce significantly better likenesses than Photofit when the images are constructed from memory.

The different findings therefore suggest potential methodological differences between both the Davies et al (2000) and Frowd et al (2003) studies. In the Davies et al study, only one hour was allowed for the construction of both composites, cognitive interview techniques were not used to elicit the initial verbal description and it is unclear whether feature editing tools were used to enhance the composites. However, in the Frowd et al (2003) study, no time limit was placed on construction, cognitive interview techniques were used and all composites were edited using the available tools within each system. While the first two factors may not have had any differential effect on composite performance, the lack of feature editing may have masked any differences between the two systems as research (Gibling & Bennet, 1994) has clearly indicated that additional editing and paintwork can significantly enhance the quality of composites. As E-FIT and PROfit both contain sophisticated editing packages, to manipulate both featural and configurational information, it is likely that editing may have significantly increased the quality of the E-FIT images and resulted in significantly better likenesses than the Photofits.

The results from these investigations therefore generally suggest that computerised systems such as PROfit and E-FIT can perform better than older systems such as Photofit. However, despite many changes to the systems (e.g. choosing features within the context of a whole face and a sophisticated ability to manipulate both featural and configurational information), when composites are constructed from memory they still do not perform particularly well. As Brace et al (2002) suggested, poor composite performance using computerised systems may reflect a difficulty in translating a verbal description into a visual representation, rather than limitations in the systems.

Methodological issues

It has been suggested that a key limiting factor in composite production is a witness's ability to verbally describe a face (Laughery and Fowler, 1980). As Ellis, Shepherd and Davies (1980, pp101) state, the verbal code "may well be too restricted a means for transmitting the rich detail contained within the human face". Indeed Paivio, (1986) indicated that visual memories are associated with multiple memory traces.

Paivio suggested that there were two distinct memory codes, a verbal memory code that relates to labels for details and a visual memory code that relates to perception.

Verbal overshadowing

Schooler, & Engstler-Schooler, (1990) suggest that verbalising a visual memory may cause interference between the verbal and visual codes, a term the authors have entitled 'recoding interference'. In the Schooler and Engstler-Schooler (1990) investigation, participants were asked to watch a video of a simulated robbery. Before a recognition test was administered, half of the participants were asked to write a description of the robber, while the other half of the participants completed a filler task. The results revealed that the participants who described the robber achieved lower levels of performance on the recognition task. This led the authors to suggest that the act of generating a verbal description may somehow impair or overshadow the visual memory of the target. This effect has been supported by other research that has also indicated that the act of verbally describing a face may impair later recognition and identification (e.g. Dodson, Johnson & Schooler, 1997; Fallshore & Schooler, 1995).

However, another possible explanation for such impairment is source confusion. That is, when two or more similar items or events are perceived, memory for the original source can become confused with memory for the additional source (Johnson, Hashtroudi & Lindsay, 1993). However, Dodson, Johnson & Schooler, (1997) examined this and demonstrated that describing an unrelated face (e.g. parent) could impair recognition performance as much as describing the target. This suggested that the verbal overshadowing effect did represent a shift in processing, rather than source confusion, as it was unlikely that an image of a parent could become confused with an unrelated image. The authors therefore suggested that the act of describing a face might impair performance. When a face is described, the face is broken down into its constituent parts and as a result there is a shift towards featural rather than holistic processing. This reliance on featural processing may therefore make a subsequent recognition task more difficult. Indeed, Fallshore and Schooler (1995) reported that any impairment due to verbalisation may be minimised when featural processing was encouraged, e.g. recognising inverted faces.

Ironically, this research suggests that verbally describing a face may be beneficial for composite construction, but only if a face has been encoded featurally. Wells & Hryciw, (1984) examined whether the 'mismatch' between encoding and retrieval operations would facilitate or hinder both recognition and composite construction. They examined holistic and featural encoding using Identikit and found that while holistic encoding instructions (character assessments) facilitated recognition performance, featural encoding instructions aided Identikit construction. Similar results have been obtained in recognition studies, where it has been found that recognition of faces is better when faces are encoded holistically (participants make judgements regarding honesty etc) rather than featurally (judging physical features) (e.g. Bower & Carlin, 1974; Patterson & Baddeley, 1977; Berman & Cutler, 1998). These findings suggest that describing a face may better facilitate feature-based construction tasks, when the face has been encoded featurally. However, evidence from face recognition indicates that faces are normally processed holistically rather than featurally, so the mismatch between processing operations may seriously impair composite production.

However, while it is clear that faces are processed holistically, many authors have failed to find any detrimental effect of describing a face on subsequent recognition performance (e.g. Clifford, Clifford & Smith, 2002; Clifford, Burke & Clifford, 2003; Memon & Rose, 2001) either for adults or children. Instead, this research supports many recognition studies that have indicated that there is no relationship between verbal descriptions and subsequent identification accuracy (e.g. Pigott & Brigham, 1985, see chapter 4 for a review). In particular, a recent meta-analysis by Meissner and Brigham (2001) concluded that while the verbal overshadowing effect was a significant effect, it was very small and only explained 1.27% of the variance. In addition, Meissner and Brigham also indicated an effect of authorship, as the majority of studies that have reported a verbal overshadowing effect have been published by Schooler and colleagues, while other researchers have difficulty in replicating the effect. This therefore suggests possible important methodological differences between the Schooler group and other researchers.

Therefore, it is unclear whether the act of describing a face may impair later performance on either a recognition task, or on composite construction. Despite the lack of clarity regarding the act of describing a face, research has indicated that verbal descriptions often contain useful identifiable information (Christie & Ellis, 1981; see Chapter 4 for a review of verbal description literature). More importantly, verbal descriptions are still, at present, a useful and necessary part of composite construction and the amount of verbal description provided by a witness is often used by police operatives as an indicator of subsequent performance (Gabbert, Dupuis, Lindsay, & Memon, 2003). Chapter 4 therefore discusses the literature on facial descriptions and examines the relationship between verbal descriptions and composite construction.

This chapter initially asks whether there is any relationship between the amount of description given by a witness, the accuracy of that description and the quality of the resulting composite. Chapter 4 also examines the type of information that is contained in a verbal description and assesses whether the combination of verbal and pictorial information can be used to increase performance at test.

Visual overshadowing

Another possible explanation for poor composite performance may be interference. When constructing a composite a witness is exposed to a varying number of similar features that are all presented within the context of a whole face. As such, Turner et al (1999) suggested that the presentation of many 'incorrect' faces (different combinations of features and configurations) might somehow 'overshadow' or interfere with memory for the original face. Turner et al investigated this by asking participants to construct an E-FIT using one of three construction methods (jigsaw, piecemeal and normal). In the jigsaw condition, participants were initially presented with a line-drawn schematic face. They were then required to choose features one at a time, building the composite much like a jigsaw. In the piecemeal condition, participants were again presented with a schematic face. However, when each individual feature had been chosen they were hidden from view until the whole face had been reconstructed. Further sets of participants were presented with all three types of composite for each target and were asked to rate them for likeness. The results revealed that the highest performance was obtained for the Jigsaw composites, followed by the normal composites, with the piecemeal composites yielding the lowest performance. Turner suggests that these results indicate that it may be perceptually harder to isolate individual features when they are presented within the context of a realistic whole face, rather than a schematic face. However, while visual overshadowing is a possibility, the mean ratings in this investigation do not seem to differ greatly (Jigsaw=4.4, Normal = 4.6) and as such it is unclear how much benefit may be gained from presenting features in the context of a less realistic face.

Interference

While there is no clear evidence for a 'visual overshadowing' effect, Turner et al's results may be explained by a potential interference effect. While participants were required to choose a feature (e.g. eyes) in the absence of other features (e.g. face shape etc), the number of presented features may have affected the results. Signal detection theory (e.g. Swets, 1964) indicates that recognition is not absolute. Instead participants make a decision based on the number of target and distractor items. If there are a large number of similar distractors, it becomes extremely difficult to distinguish the target items. Therefore, while context may have an affect on construction performance, the actual number of presented features may also affect performance. Witnesses often have to make a judgement based on a large number of similar features. Therefore, the likelihood of a more accurate decision (i.e. choosing the most similar feature) may be more difficult as the number of similar features increase.

Cognitive interview

One way to try to prevent a potential visual overshadowing effect and to restrict the number of features presented is to ensure that initial recall is as complete as possible. This can be achieved through the use of cognitive interview techniques (Geiselman, Fisher, MacKinnon & Holland, 1986). As information about an event is represented at different levels (Fisher & Cuervo, 1983), it is the role of the cognitive interviewer to maximise the amount of detailed information retrieved (e.g. facial characteristics) and minimise information from the general level (e.g. height and build etc). This is achieved through the use of four retrieval techniques: reinstating context, repeated recall in differing orders, mentally changing perspectives and emphasising to the witness that they must report everything, no matter how trivial they deem it (Finger & Pezdek, 1999).

A great deal of early research indicated that the cognitive interview increased the amount of accurate information recalled, compared to a standard interview, without increasing the amount of inaccurately recalled information (e.g. Fisher, Geiselman & Amador, 1989; Geiselman, Fisher, MacKinnon & Holland, 1985; Koehnken, Schimmossek, Aschermann & Hofer, 1995; Mantwill, Koehnken & Aschermann, 1995). However, other more recent research has reported no difference in performance between cognitive and standard interviews (e.g. Memon & Stevenage, 2002; Memon, Wark, Holley, Bull & Koehnken, 1996b).

While a full review of the cognitive interview literature is inappropriate for this thesis (see Memon, 1986 for a review), Memon (Memon, 1986; Memon, Wark, Holley, Bull & Koehnken, 1996b) has suggested that the most important effect of the cognitive interview is that it enhances communication. As an interview is an interaction between two people, memory performance will undoubtedly be affected both by the techniques used to elicit recall and by the person guiding the interview. Therefore, building rapport with a witness by using effective communication skills and eliciting recall through the use of cognitive interview techniques, may serve to ensure that recall is as complete as possible. Interestingly, research that has examined the effectiveness of both mechanical and computerised composite systems (e.g. Ellis, Davies & Shepherd, 1978; Ellis Shepherd & Davies, 1980; Koehn & Fisher, 1997; Davies, van der Willik & Morrison, 2000) has not employed any of the cognitive interview techniques described. If these techniques were employed to ensure that recall was as complete as possible, then performance for composites constructed from memory may increase.

To summarise, research has indicated that while computerised composites perform better than the older manual systems, they still do not perform well. It has been suggested that the act of describing a face may affect composite construction. However, there is conflicting evidence regarding the verbal overshadowing effect in face recognition and there is no clear evidence to indicate what effect generating a verbal description may have on subsequent construction performance. Chapter 4 therefore investigates the utility of verbal descriptions in composite construction and examines whether descriptions can be used to increase performance at test. Similarly, interference may hinder construction performance and effective use of rapport building and cognitive interview techniques may help to reduce any interference, both from the whole face and from the number of presented features. These factors may all affect composite performance, however, one of the major difficulties in composite construction is that a witness is asked to recall an unfamiliar face.

Memory for faces

Bruce and Young Model

A great deal of research has demonstrated that memory for faces can be extremely robust when faces are familiar (e.g. Klatzky and Forrest, 1984) but extremely error prone when faces are unfamiliar (see Hancock, Bruce & Burton, 2000 for a review). Models of face processing have clearly demonstrated not only how familiar faces are recognised and identified, but they have also indicated the difficulties that are inherent in unfamiliar face processing. In particular, the Bruce and Young (1986) model describes four-stages of facial identity processing. The first stage is entitled 'structural encoding' and this is where a structural representation of a seen face is formed. This representation is then 'matched' against stored representations of known faces. For a face to be recognised as familiar there should be a good match between the incoming structural description and the stored structural representation. The model describes Face Recognition Units (FRUs) which are activated when both representations 'match' and therefore enable the seen face to be classified as familiar. When the face is classified as familiar other semantic information (e.g. occupation, personality etc) about the person can be accessed via the Person Identity Nodes (PINs). The final stage is where the name of the person is accessed. The model states that names are stored separately from semantic information (PINs) but can only be accessed via the semantic store. That is, one can recall information about a person without recalling their name, but one cannot recall their name without recalling some semantic information. This model of identity processing is supported by a great deal of cognitive (e.g. Young, Hay & Ellis, 1985; Young, McWeeny, Ellis & Hay, 1986; Young, Ellis & Flude, 1988; Johnston & Bruce, 1990) and neuropsychological evidence (e.g. De Haan, Young & Newcombe, 1991; McKenna & Warrington, 1980;

Flude, Ellis & Kay, 1989). However, see Burton, Bruce and Johnston (1990) for a recent revision of this model.

Structural descriptions

The Bruce and Young (1986) model indicates that problems in recognising unfamiliar faces occur at the first stage - structural encoding. Two types of structural description are formed at this stage. The first is a view-centred description and is therefore dependent on the view that the face was seen in. The second structural description is independent of view and expression and is used to recognise familiar faces. As a witness will have only seen a face for a very brief period of time, they will be unable to form a structural description that is independent of view and expression. Therefore, memory of the face will be context or image-specific. This means that it would be much harder for a witness to recognise the face when presented with a later image of the person that differs in viewpoint, expression or lighting.

Bruce (1982) demonstrated this by presenting participants with twenty-four unfamiliar faces at study. At test, these twenty-four faces were presented with an additional set of twenty-four distractor faces. The participants were asked to recognise the faces in the same expression and viewpoint, a different expression, a different viewpoint or a different expression and viewpoint. Reaction time and accuracy were measured and the results revealed that the best performance was obtained when the images were identical, followed by a change to either viewpoint or expression, while changes in both expression and viewpoint yielded the lowest performance levels. Experiment two expanded on this by comparing performance for familiar and unfamiliar faces. The results confirmed those obtained in experiment one and revealed that decisions were faster and more accurate for familiar rather than unfamiliar faces.

Using a repetition priming task, Roberts and Bruce (1989) also demonstrated that memory for unfamiliar faces was sensitive to image transformations. In this investigation participants were presented with either personally familiar faces or unfamiliar faces. In addition, a second unfamiliar condition was used where participants were required to learn the names of the faces in the study phase. Participants were asked to respond to every face by pressing 'yes' if the face belonged in the target set and 'no' if the face did not. Spacing and view were manipulated, such that the prime and target faces were presented in either the same or different view and with either no intervening items or one intervening item. The results revealed a robust repetition priming effect for familiar faces despite changes in viewpoint and spacing. However, while repetition priming was observed for unfamiliar faces at both spacing intervals, the effect was only found when view was identical. In addition, performance for unfamiliar faces was the same regardless of whether the name had been learned.

These results suggest that memory for unfamiliar and familiar faces differ. Whereas, memory for unfamiliar faces appears to rely on a pictorial or image-specific code, memory for familiar faces may rely primarily on a more robust structural code. However, differences in performance for unfamiliar and familiar faces have also been observed in perceptual tasks that require no memory load.

Kemp, Towell and Pike (1997) revealed that matching a photograph on a credit card to a live target was extremely error prone. Only half of supermarket cashiers in this study performed accurately, either accepting or correctly rejecting the cards when the photograph was a high quality image. When the photograph resembled someone of similar appearance, only thirty-six percent of cashiers correctly rejected the card.

Similarly, in the Bruce, Henderson, Greenwood, Hancock, Burton & Miller, (1999) investigation, participants were presented with a high quality video still of an unfamiliar male together with either a target absent or target present array. Images were either matched or unmatched for view and expression and participants were required to either 'pick out' the target from the array, or correctly reject the array. The results revealed high error rates even when the images were matched for view and expression. When view differed between target and array, performance decreased further. Henderson, Bruce and Burton (2001) extended this work to show that even when only two images were presented with a video still of the target and participants were required to state 'which one is the target?' accuracy was still low.

In contrast, identification of familiar faces from CCTV footage has been shown to be extremely good when the face is in view, even when the footage is of extremely poor quality (Burton, Wilson, Cowan & Bruce, 1999). Bruce, Henderson, Newman and Burton (2001) extended this work by asking groups of unfamiliar and familiar participants to match target images (high quality video clips or stills) to high quality photographs. While viewing the target footage, participants were presented with a high quality photograph and were required to state whether the two images were the same person. The results revealed a large effect of familiarity. That is, when the participants were familiar with one or both of the images (video image and/or photograph), they were able to either correctly match the images or correctly reject them with a high degree of accuracy (90%).

These results suggest that familiar and unfamiliar faces are processed differently. For the matching tasks it appears that matching a viewed face to a stored memorial representation of that face (familiar faces), is more successful than matching to another image of the same person (unfamiliar faces). For recognition tasks familiar faces are matched faster and more accurately than unfamiliar faces, despite changes in viewpoint and expression. Similarly, repetition priming occurs across different views for familiar faces but not for unfamiliar faces and for perceptual matching tasks where no memory load is involved, familiar faces are matched more accurately than unfamiliar faces.

This suggests that memory for an unfamiliar face relies primarily on a pictorial or image-specific code, as performance is sensitive to image changes. In contrast, memory for familiar faces may rely primarily on structural rather than pictorial codes, as performance did not decrease despite changes in view and expression. However, memory for familiar faces does not rely solely on structural codes. In an earlier repetition priming investigation Bruce & Valentine, (1985) revealed that performance for familiar faces increased when the same picture was used as prime and target compared to when different pictures were used. This indicates that familiar face processing is facilitated by both pictorial and structural codes, whereas unfamiliar face processing is facilitated by pictorial codes.

These findings may be explained by differing amounts of exposure to faces. For familiar faces, the face has been seen on multiple occasions and in different contexts. Therefore a more robust representation of the face will have been formed, with a greater number of memory traces (Hintzman, 1986) that may greater facilitate changes in viewpoint and expression. In contrast, the memorial representation of an unfamiliar face contains very few traces, as exposure is limited to a single instance in recognition tasks. Similarly, because of the very brief exposure period in an eyewitness situation, it is unlikely that a structural description will be formed and recognition and reconstruction ability will therefore rely heavily on image-specific or pictorial details i.e. how well the presented image 'matches' the witnesses stored internal representation. Therefore, as face recognition evidence indicates that unfamiliar face memory is image-specific, would composite performance increase if composite systems utilised this information and displayed images in a more image-specific manner?

Recall versus recognition

When a witness constructs a composite they are asked to both recall and recognise whether the presented image 'matches' their internal representation of the face. Single-process, or generation-recognition models of memory (e.g. Anderson & Bower, 1972; Kintsch, 1970) assume that recall and recognition involve the same underlying process and indicate that it is the level of activation that determines whether an item will be remembered. That is, if an item is weakly represented in memory, the item will be less likely to be recalled, but effective retrieval cues contained in a recognition test may lead to successful recognition. As recall involves both a generation and a recognition stage, whereas recognition only involves the latter, these models assume that recognition will always be better than recall.

Encoding Specificity

However, research has also indicated that participants can recall words that they fail to recognise (Flexser & Tulving, 1978; Tulving & Flexser, 1992), indicating that recall and recognition may in fact be two distinct aspects of the same process. Flexser and Tulving demonstrated that the relationship between recognition and recall is a function of the amount of information that is available. In particular, the information that is available in a recognition task is uncorrelated with the information that is available in a recall task and reflects retrieval independence. This instead provides support for the encoding specificity principle (Thomson & Tulving, 1970; Tulving & Thomson, 1973), rather than single-process models of memory. The encoding specificity principle states that recall and recognition are distinct aspects of the same retrieval system, and that successful retrieval is dependent on the degree of overlap between the information at encoding and the information that is available at retrieval. As such, more accurately matching the cues at encoding and retrieval may increase recall performance. Indeed, Penrod et al (1992a) suggested that composite performance might increase when the properties available in the composite more accurately matched the properties at encoding. Penrod et al suggested that presenting features within the context of a whole face might encourage more holistic processing and thus improve construction performance. However, as described earlier, all modern composite systems display features within the context of a whole face and performance is still poor.

This suggests that while the processes at encoding and retrieval may be more accurately matched (i.e. displays composites in a more holistic way), other properties of the composite may also need to be matched more accurately (e.g. viewpoint). The image that is displayed on a composite system will never accurately 'match' the viewed face, as it will not contain an exact copy of the individual's facial features. However, as unfamiliar face memory is image or context-specific, it may be easier for a witness to build a composite that is matched for viewpoint and expression. The more modern composite systems (e.g. PROfit and E-FIT) allow for the manipulation of expression through the use of their sophisticated editing tools. However, one of the difficulties with all composite systems is that they do not allow viewpoint to be matched to the view seen at encoding. Until recently, all composite systems have only permitted a witness to construct an image in a full-face view, regardless of the view at encoding. Therefore, chapter 2 will investigate the role of encoding specificity in composite construction and examine whether composite performance increases when the view at encoding matches the view at retrieval.

Nature of stored representations

As composite construction is a retrieval task, more accurately matching encoding and retrieval cues may serve to increase recall and subsequent construction performance. However, the way in which facial information has been encoded and stored will ultimately affect performance. Research (e.g. Valentine, 1991a) has indicated that faces may be represented as deviations from the norm, or deviations from the prototypical face. In particular, Valentine suggests that faces are normally distributed across different featural and configurational dimensions, where the norm or prototypical face occupies the centre of 'face space' and all other faces are distributed around this. Therefore, faces are represented in terms of 'face space' according to their featural and/or configurational deviations from the average or norm face. As a result, typical faces that share similar properties (e.g. average face shape, average length nose) may be clustered together in face space. This would therefore make it much harder to distinguish one typical face from another. Similarly, a face that has either distinctive configurational (eyes very close together) or featural information (very large nose) will be located on the edges of 'face space', and would therefore be located much quicker and with much less confusion than a typical face (e.g. Valentine & Bruce, 1986). This suggests that despite more accurately matching encoding and retrieval cues, composite construction may be more error prone for typical rather than unusual faces, as the assumption of 'face space' indicates that it would be much harder to distinguish the identity of one typical target face from another.

Accessing the stored representation of a typical face can therefore lead to confusion between different faces/features. In addition, the act of constructing a composite can also lead to confusion and error. After a witness has described a face, they are asked to search through a series of similar features and choose the feature that is most similar to the target. If it is more difficult to identify the target face it may be more difficult to distinguish between these similar features and construct an accurate likeness of the target.

Likewise, research suggests that memory for the original target may become integrated with either memory of another face, or memory for the presented features. In particular, research on the effect of prototype faces has indicated a potential memory integration effect. For example, Solso & McCarthy (1981) reported that 35 out of 36 participants reported seeing an unstudied prototype face. Thus suggesting that memory for faces/features is somehow integrated with memory for similar items, either through trace blending (e.g. Metcalfe, 1991), by storing integrated traces (e.g. Hintzman, 1986) or as a result of competing patterns of interaction (e.g. McClelland & Rumelhart, 1985).

Memory integration

Metcalfe (1990) examined the integration of similar items in memory and developed a model entitled CHARM (Composite Holographic Associative Recall Model). The central tenet of this model was that memory traces were superimposed or blended in memory and that retrieval was dependent on the items that were associated in storage. While Metcalfe examined the integration of words, specifically testing the work of Loftus and colleagues (e.g. Loftus, 1975, 1979; Loftus & Loftus, 1980, Loftus, Miller & Burns, 1981) on misinformation effects, she also suggests that as facial features are in essence continuous variables, it is likely that similar features may become associated in memory and result in retrieval or recall of an alternative. In fact, Metcalfe states that "It is possible that a lure that was never viewed before may be 'better' recognised than even the actual target face" (pg158).

The CHARM model is similar in design to other distributed network models (e.g. McClelland & Rumelhart, 1985), as it assumes that an encoded item is represented as a set of semi-independent features in a distributed pattern. Whereas, McClelland and Rumelhart indicated that false recognition could occur through competing patterns of interaction, in the CHARM model items are associated or blended through convolution. When two items are similar, these items will be positively correlated.

When the items are associated, the ordered sets of features that make up these items are combined to create a new memory trace. Retrieval then occurs through a process of correlation. More specifically, a retrieval cue (e.g. an array of faces or a set of features) is 'matched' or correlated to the composite trace and the best 'match' is the retrieved item. Metcalfe suggests that the model can explain the prototype effect, as the model indicates that prototypes are more likely to form between similar items. In the case of composite construction, this model suggests that the 'best match', or chosen feature, could be any similar feature that is presented (e.g. a similar nose). In particular, if facial features share several similar properties or are positively correlated (e.g. are all long and pointy), the model suggests that the feature that is most highly correlated will be chosen.

This model of memory is similar to other multiple memory theories proposed by other researchers (e.g. Bower, 1967; Hintzman, 1986, 1988; Nosofsky, 1991). In particular, Hintzman suggests that traces or parts of original events/items are stored in memory and that it is these combined traces that are retrieved. This output is therefore much like a schema of the original event or item. In particular, Hintzman found that when participants studied words that were perceptually or conceptually similar to a new set of words at test, high rates of false recognition occurred. Similarly, Nosofsky (1991) revealed that both classification and recognition of items were based on similarity comparisons with exemplars that were stored in memory. These models therefore indicate that retrieval is based on similarity, suggesting that a witness will retrieve or choose a feature that shares similar properties to the encoded feature.

Schema theory

Hintzman (1986, 1988) suggests that a retrieved item or event is a schema of the original item/event. While Hintzman did not specifically examine memory for faces, other researchers have also suggested that faces are processed according to existing schemas (e.g. Rakover, 1999, 2002). In particular, it has been demonstrated that an important schema is the spatial layout of the face e.g. eyes above nose, nose above mouth etc. Similarly, there are schemas for individual facial features, which describe the individual properties of each feature e.g. a nose has two nostrils etc. However, Rakover suggests that there is a hierarchy of schemas, with the most important schema representing the spatial layout or configuration of the face. In addition, researchers have suggested that this general processing occurs before the individual features are processed (e.g. Bruce, 1988).

Other researchers have also highlighted the reconstructive nature of recall (e.g. Hasher & Griffin, 1978; Alba & Hasher, 1983), noting that the information that is encoded in memory is guided by a schema. The information is then guided into hierarchical slots and missing information, or missing slots can be filled in with information that is accordant with the existing schema. This guiding schema can actively select and modify information, in order to arrive at a coherent, organised framework of information. Therefore, when memory is searched during retrieval, the schema is activated and the information that is recalled may either reflect accurate information (i.e. information that was actually encoded), or inaccurate information (i.e. information that was used to fill in the gaps).

To summarise, the conceptualisation of a 'face space' indicates that it is much harder to distinguish typical faces than it is to distinguish distinctive faces. This is because faces are clustered in 'face space' according to the degree of deviation from the norm or prototype face. This suggests that as it may be more difficult to identify a typical target face it may be more difficult to distinguish between similar features and construct an accurate likeness of the target. Similarly, research has indicated that memory for similar items/faces may become integrated or blended, as evidenced by the prototype effect in face recognition. Therefore, rather than recalling accurate details of a target, a witness may confuse the target face with another face in 'face space. Memory for the target may become confused or blended with memory for other faces or features during construction, or a witness may recall facial/featural schemas but have difficulty distinguishing specific characteristics of the target.

External versus internal features

One of the added difficulties with unfamiliar face recall is that as there is restricted opportunity to build a structural representation of the face (see pages 26-29), facial features may be differentially represented in memory. Evidence for this comes from investigations that have examined performance of familiar and unfamiliar faces where both the internal (eyebrows, eyes, nose and mouth) and external features (hair and face shape) have been isolated. For example, Ellis, Shepherd & Davies, (1979, experiment 1) asked participants to identify famous faces from either their internal or external features. The results revealed an overall accuracy rate of 80%, where 50% of the faces were identified from the internal features and 30% were identified from the external features. Experiment 2 examined this effect for unfamiliar faces using a standard recognition task and the results revealed no significant difference between internal and external features. Experiment 3 replicated experiment one (famous faces) using a standard recognition task and the results again revealed an advantage for internal features. Similar results were reported by Young, Hay, McWeeny, Flude & Ellis, (1985) and Clutterbuck and Johnson, (2002) using matching tasks.

These results suggest that while there is an internal feature advantage for familiar faces, there is no such advantage for unfamiliar faces. However, other research has not only supported the internal feature advantage for familiar faces, it has also revealed an external feature advantage for unfamiliar faces (Bruce, Henderson, Greenwood, Hancock, Burton & Miller, 1999; Bonner, Burton & Bruce, 2001). In particular, Bonner et al examined performance for unfamiliar faces over a time period of three days, to examine the effects of familiarisation. The results revealed that performance was better when the external features of unfamiliar faces were matched compared to the internal features. In addition, over the three-day period, performance for the internal features improved and was equivalent to external feature performance.

This suggests that as a face becomes familiar there is a shift from external to internal features. O'Donnell and Bruce, (2001) examined this process of familiarisation in more detail. This investigation manipulated the internal features (eyes, nose and mouth) and the external features (hair and chin) both configurally (distances) and featurally (features were swapped). Half the faces were unfamiliar and half had been learned in the study phase. Participants were presented with two faces (one unchanged and one changed) and they were required to state whether the two faces were physically the same or different. The results revealed that changes to the eyes were detected better for familiar faces than for unfamiliar faces.

To summarise, these results suggest that familiar and unfamiliar faces are processed differently. In both memory tasks (Ellis, Shepherd & Davies, 1979) and perceptual tasks (Young et al 1985; Clutterbuck and Johnson, 2002) there appears to be an internal feature advantage for familiar faces. While these studies report no advantage for unfamiliar faces, other research (Bruce et al, 1999; Bonner et al, 2001) has reported an external feature advantage for unfamiliar faces. In addition, familiarisation research suggests that there is a shift from external to internal features as a face becomes familiar (e.g. Bonner et al, 2001) and this starts with a differential shift to the eye region (O'Donnell and Bruce, 2001).

One possible reason for these findings may be that external features such as hair and face shape occupy such a large part of the facial image. As a result, when a face has only been viewed for a very short period of time, the larger more dominant external features may be perceived better than smaller less dominant internal features. Consequently, the external features of unfamiliar faces may be better represented in memory. In contrast, for familiar faces, external features such as hair are more likely to change over time and are less stable than the internal features. As a structural representation of the face is built up over multiple presentations of the face, in different contexts, views, expressions and with perhaps different hairstyles, it is likely that the internal features would be a better guide to identity.

This indicates a paradox for composite construction in that an eyewitness may have perceived, recalled and reconstructed the external features of the face better than the internal features. However, when the composite is released via the media in the hope that someone who is familiar with the target may make an identification, this familiar person may utilise internal feature information to identify the facial composite. As the internal features of a composite may be less accurate than the external features, the likelihood of a correct identification may be diminished.

While matching encoding and retrieval cues more accurately may aid construction of composites, it is likely that external features will still more be accurately represented than internal features. One way to improve the performance of facial composites however, may be to examine methods of improving likeness at test, rather than at construction. Each composite is a representation of an individual witness's memory of a seen face and is likely to contain information that is both similar and dissimilar to the original target face. So, how would one witness's representation of a seen face compare with another witness's representation of the same face? Research has examined the effect of presenting composites from more than one witness and has revealed that presenting composites from multiple witnesses increases performance significantly (Bennet, Brace, Pike, & Kemp, 1999; McNeil, Wray, Hibler, Foster, Rhyne & Thibault, 1987; Bruce, Ness, Hancock, Newman, & Rarity, 2002, experiment 1). This suggests that each composite contains differing types and amounts of similar and dissimilar information. Interestingly, research on the effect of prototypes in face recognition (e.g. Solso & McCarthy, 1981; Bruce, Doyle, Dench & Burton, 1991; Homa et al, 2001), indicates that each composite may be a deviation from the ideal or prototypical image. Therefore, when composites from multiple witnesses are presented, this information may serve to 'create' an image that is closer to the 'ideal' or prototypical image. Chapter 3 therefore investigates whether combining composites from different witnesses aids performance at test. In particular, chapter 3 examines performance for multiple composites from both multiple and single witnesses using both PROfit and multiple composite systems, to assess what type of information serves to increase performance. If composites from multiple witnesses perform well because they just contain more information, then combining other types of information (e.g. adding textural information) may also serve to increase performance.

Thesis questions and review of remaining chapters

This introductory chapter has highlighted many general issues concerning facial composite construction. Firstly, this chapter has provided a detailed evaluation of both traditional and computerised methods of composite construction, in particular, demonstrating that despite advances in system design, facial composites still do not perform particularly well. Secondly, possible methodological reasons for poor composite performance were considered and issues such as verbal overshadowing and interference effects were discussed. Thirdly, this chapter considered literature on unfamiliar face processing, focusing in particular on the Bruce and Young model to highlight the difference between familiar and unfamiliar face processing. Finally, this chapter considered relevant memory research, which has highlighted potential difficulties with both the storage and retrieval of facial information.

This introductory chapter has described and evaluated the general difficulties and issues concerned with composite construction and as demonstrated, most composite research to date has focused on difficulties associated with construction. As such, much is known about the memorial difficulties associated with facial composite construction. However, very little research has examined methods of improving facial composite likeness. This thesis will therefore build on the research presented in this introductory chapter by examining specific methods of improvement. In particular, this thesis will examine both methods of improving both the construction of facial composites and the subsequent verification and identification of composite images.

Will the development of a three-quarter-view database improve composite performance?

The starting point for this investigation was the finding that despite technological advances in system design, facial composite systems still do not produce particularly good likenesses. This indicated that poor composite likeness might be a result of memorial difficulties. However, as constructing a facial composite is a retrieval task, memorial difficulties may in fact be a result of system design. For example, a witness may have viewed a three-dimensional face, yet they are asked to construct a two-dimensional full-face composite. As unfamiliar face memory is image or context-specific, the disparity between the view at encoding and the view at retrieval may result in poorer quality composites. Therefore, the first question addressed whether composite systems may be more effective if they displayed faces in a more three-dimensional manner. Chapter 2 therefore examines the effectiveness of a new three-quarter-view database in PROfit. It begins by discussing why a three-quarter view composite system may be helpful and provides a review of the three-quarter-view literature with regard to unfamiliar face recognition. This chapter addresses three specific questions.

The first question focuses on the role of composite systems as effective retrieval aids and asks whether performance will increase when composites are constructed in a three-quarter view. At present, a witness is only permitted to construct a composite in a full-face view, even though they may have viewed a three-dimensional face. Therefore, experiment 1 examines whether the presentation of features in a three-quarter view will be a more effective retrieval aid than the presentation of features in a two-dimensional full-face view.

In addition, experiments 2 and 3 ask whether the presentation of multiple views from the same witness will increase composite performance. In particular, these experiments examine whether the presentation of both a full-face and a three-quarter-view composite from the same witness, serve to increase performance above the level observed for a single full-face composite. Furthermore, experiment 3 examines the issue of encoding specificity and asks whether composite performance will increase when the view at encoding matches the view at retrieval?

Will the presentation of composites from multiple witnesses increase performance?

This question is addressed in chapter 3, which begins by examining the literature on multiple witness composites and prototype effects in face recognition. In particular, the experiments described in this chapter (experiments 4 – 8) initially extend the work of Bruce, Ness, Hancock, Newman & Rarity (2002) by examining whether the combination of composites from four different witnesses serves to increase performance above the level observed for a single composite.

Furthermore, this chapter asks whether the combination of different memorial representations results in an image that is closer to the ideal or prototype. In order to assess why combined composites perform well, the type of information that is contained within the composites is manipulated. Therefore, comparisons are made between morphed composites from multiple and single witnesses using both a single system (PROfit) and multiple systems (PROfit, E-FIT, EvoFIT and Sketch). If the combination of composites from multiple witnesses results in an image that is closer to the ideal or prototype, then performance should only increase when different memorial representations are combined.

What is the relationship between verbal descriptions and composite performance?

This question is addressed in chapter 4, which begins by evaluating the literature on verbal descriptions. The experiments in this chapter (experiments 9 – 12) begin by assessing whether there is any relationship between the amount of description a witness provides, the accuracy of the description and the quality of the resulting composite. As discussed in the introductory chapter, it is unclear whether the act of describing a face impairs subsequent recognition performance. However, what is clear is that verbal descriptions are an important and necessary part of composite construction and are in fact used by many police operatives as an indicator of composite performance. Experiments 9 and 10 therefore examine whether there is any relationship between verbal descriptions and composite performance. Experiments 11 and 12 then consider whether the combination of a verbal description and a composite from the same witness will serve to increase performance above the level observed for a single composite.

2

Multiple Views from the Same Witness

A general overview of composite systems and the difficulties associated with constructing a composite was given in the previous chapter. However, one potential difficulty that has not been discussed is that of viewpoint. This chapter will therefore begin by discussing why a three-quarter view composite system may be helpful and will provide a review of the relevant literature with regard to unfamiliar face recognition. It has not previously been possible to investigate any effect of view on composite construction, as all composite systems have only contained full-face databases. However, the introduction of a new three-quarter-view female database in PROfit has enabled a careful consideration of view effects. This new database not only allows a witness to construct a three-quarter-view composite, it also has the facility to automatically generate three-quarter view images from full-face composites. There are therefore three aims to this investigation. The first aim is to examine performance for constructed full-face and three-quarter view composites under different exposure conditions (all views in experiment 1 and full-face view, three-quarter view and all views in experiment 3). The second aim is to examine performance for three-quarter view composites that have been automatically generated from the full-face composites. Finally, the third aim is to examine whether presenting a full-face and three-quarter composite together (constructed *and* automatically generated) serves to increase performance above the level achieved for a single full-face composite.

Why construct a three-quarter-view composite?

In a real-life situation a witness will have viewed a previously unfamiliar three-dimensional moving face. However, when a witness is invited to build a composite likeness of the face, they are asked to construct a two-dimensional full-face image. Evidence on the role of movement in unfamiliar face recognition has indicated that movement may help to build a robust three-dimensional representation of the face (e.g. Schiff, Banka & De Bordes Galdi, 1986; Bruce & Valentine, 1988; Pike, Kemp, Towell & Philips, 1997) (see Bonner et al, 2003, for null effects). As this research suggests that a witness may have encoded and stored a three-dimensional representation of the face, will it be easier for a witness to construct a three-quarter view composite, which reveals more about three-dimensional structure, rather than a full-face composite?

Is a three-quarter-view better than a full-face?

Several researchers have investigated whether there is a one particular view that is preferred in face recognition (e.g. Bruce, Valentine and Baddeley, 1987; Schyns & Bülthoff, 1994; Hill, Schyns & Akamatsu, 1996; Newell, Chorizo & Valentine, 1999). This research stems partly from research on object recognition that has suggested not only that object recognition may be viewpoint dependent (e.g. Edelman & Bülthoff, 1992; Tarr & Pinker, 1990), but also that certain views of an object are often preferred (e.g. Palmer et al, 1981).

For faces, it has been speculated that an angled view (or three-quarter view) may represent a canonical view. As this view is centred between both the full-face and profile views it is possible that this view may contain information that is available in both the full-face and profile views. In particular, research by Wells (1985) indicates that three-quarter-views may contain more information than full-face views. In this investigation participants were asked to describe faces that were presented in either a full-face or a three-quarter-view and the results revealed that more descriptors were provided after presentation of a three-quarter-view image.

In order to examine whether a three-quarter-view is preferred view, researchers have presented all views of a face at study and examined recognition performance at test. Any difference in performance would therefore be a result of the testing view, therefore indicating whether one particular view was preferred or canonical. Schyns & Bülthoff (1994, Experiment 1) examined performance for 3D laser-scanned heads with shaded surface models. Participants were presented with all views (-36, -18, 0, 19, and 36 degrees) at study and two faces at test (one of the target and one distractor, both in the same view). The results revealed that no one view was preferred. Similar results were also obtained by Hill, Schyns & Akamatsu (1996) who presented five views at test (-90, -45, 0, 45, and 90) and found no effect of test view.

These experiments indicate that a side view (18 – 45 degrees) is not canonical. Using a different methodology, Bruce, Valentine and Baddeley (1987) also found that the three-quarter view was not canonical. They examined performance for three poses (0, 45, 90 degrees) and participants were asked to decide if two sequentially presented pairs of faces, in two differing views, represented the same person. Recognition accuracy and latency were measured and the results revealed a three-quarter-view advantage for unfamiliar faces, but not for familiar faces. Thus, the three-quarter view does not make it easier to access representations about known faces.

Similarly, Newell, Chorizo & Valentine (1999, Experiment 3) presented all three views (0, 45, 90) in quick succession (1 second per view) to give the appearance of movement at study. They found that recognition was impaired for the profile view, while there were no significant differences between the full-face and three-quarter views. Thus, it appears that while the profile view achieves the poorest performance, there appears to be little difference in performance for the full-face and three-quarter views. Furthermore, Logie, Baddeley & Woodhead (1987, Experiment 4) examined the effect of view for 'live' targets. They found that while performance for the three-quarter view was slightly higher than for the full-face, there were no significant differences, leading the authors to suggest that view is not particularly important in real-life situations.

These results suggest that while there is some evidence that certain views (side views) are preferred in object recognition (e.g. Palmer at al, 1981) neither the full-face or three-quarter view seem to be preferred in face recognition. However, performance does appear to be lower for the profile view. This may be because faces contain a great deal of information (structural, featural, configural) and this information may differentially support viewpoint generalisation. For example, the information that is contained in a full-face view looks very different when that face is presented in a profile view (e.g. eyes and eyebrows). Therefore, some information may be viewpoint dependent (featural) whereas other information may be viewpoint invariant (structural). So, what are the patterns of viewpoint dependence and what do they tell us about the information contained in the different views of a face?

Seeing a full-face view at study

Patterson & Baddeley, (1977) and Woodhead, Baddeley & Simmonds, (1979) both presented faces in a full-face view at study and examined recognition performance at test by either presenting three-quarter and profile views (Patterson & Baddeley, 1977) or full-face, three-quarter and profile views at test (Woodhead, Baddeley & Simmonds, 1979). In addition Patterson & Baddeley (1977) also combined pose with disguise (with beard, wig, glasses etc). Despite these differences in methodology, both studies indicated that a three-quarter-view performed better than a profile view.

These results indicate that when there is a change in view between study and test, the three-quarter view performs better than the profile view, while there appears to be no difference in performance between the three-quarter view and full-face views. In a more recent investigation, Newell, Chorizo & Valentine (1999, Experiment 1) participants were also presented with a full-face view at study. After an interval of 500 milliseconds, they were presented with a second face in one of five views (-90, -45, 0, 45, and 90 degrees) and their task was to decide whether the two faces represented the same person. For the matched trials the results revealed that the profile views performed significantly poorer than either the full-face or three-quarter views. For the mismatch trials there was a significant effect of view for typical faces, with the three-quarter view conditions obtaining significantly more correct responses than the profile conditions. These results support those of Patterson & Baddeley, (1977) and Woodhead, Baddeley & Simmonds, (1979) by revealing poorer performance for the profile views and comparable performance for both the full-face and three-quarter views.

These results indicate that while the three-quarter view performs well, it does not perform better than the full-face view. However, when only one view of a face is presented at study, any differences between test views are likely to be a result of the view that was presented at study. These studies all presented a full-face view at study and the results indicate that there is little difference in performance for the three-quarter and full-face views at test. One possible reason for this concerns the degree of angular depth rotation. Lui & Chaudhuri (2002) suggest that performance is better when the angle of rotation between study and test is 45-degrees (i.e. full-3/4 etc) and that performance is poorer for profiles when there is a 90-degree angle of rotation (full-profile). Likewise, similar performance has been observed for novel objects, where generalisation from a single view decreases with increasing angles of rotation (e.g. Edelman & Bülthoff, 1992).

However, another possible explanation concerns symmetry (Schyns & Bülthoff, 1994; Troje & Bülthoff, 1996; Hill, Schyns & Akamatsu, 1997; Troje, 1998). They note that a face is essentially bilaterally symmetrical (albeit not perfectly). As such, side-views of a face can be thought of as non-singular, as a symmetrical view can be generated from them and full-face views are singular, as a symmetrical view cannot be generated. This argument is based on evidence from object recognition (Poggio & Vetter, 1992) that has demonstrated an interaction between the view at study and the view at test. Poggio & Vetter, (1992) suggested that 'virtual views' could be generated from one non-singular view of a bilaterally symmetrical object and that if only one non-singular view was presented at study, recognition of novel views could be achieved at test.

This may explain why a three-quarter-view advantage was not obtained when only a full-face view was presented at study. As the full-face view may be non-singular it would be extremely difficult to generalise to a novel view at test. Similarly, the symmetry argument suggests that when a three-quarter view is presented at study, a 'virtual view' could be generated and this may result in the successful recognition of the face in a novel view i.e. the full-face. If the symmetry argument is correct, then some viewpoints should also be better generalised than others after seeing a three-quarter view at study.

Seeing a three-quarter view at study

The angle of rotation hypothesis suggests that there should be little difference in performance between the full-face and three-quarter views. However, Krouse (1981) did find a difference between these two views. She presented participants with previously unfamiliar faces in both a three-quarter and full-face view. She then tested for recognition by presenting faces in the same (matched) or different (unmatched) pose, either immediately or after a two day delay. The results indicated a three-quarter view advantage which was unaffected by both time delay and change of pose. Further studies also seem to provide support for a three-quarter-view advantage at learning. Baddeley & Woodhead, (1983) presented faces in full-face, three-quarter and profile views at study and compared performance for all three views at test. The results revealed that the three-quarter view performed better than either the full-face or profile. Similar, results were also obtained by Logie, Baddeley & Woodhead (1987).

This suggests that there may be a three-quarter-view advantage when more than one view is presented at study. However, Lui & Chaudhuri (2002) failed to find such an advantage. They presented each face in only one view at study (8 faces in a full-face view, 8 in ¾ view and 8 in profile view). At test the twenty-four targets were presented in the same direction as the study phase and were shown with an additional twenty-four distractors. Participants were required to perform a yes/no recognition task and the results revealed no main effect of view. In their review of the three-quarter literature Lui & Chaudhuri (2002) suggest that there is little evidence for a three-quarter-view advantage and that most differences in performance can be explained by their 'angle of rotation' hypothesis.

Other studies have also failed to find a three-quarter-view advantage (e.g. Laughery, Alexander & Lane, (1971, Experiment 2; Davies, Ellis & Shepherd (1978, Experiment 2). In particular, Davies, Ellis & Shepherd (1978) combined view (full-face, three-quarter) and mode of presentation (line drawings, photographs) as between-subjects factors at study. At test, the participants were asked to recognise the targets in both full-face and three-quarter view, in the same mode of presentation as the study phase (photographs or lines drawings). The results revealed no difference in performance between the full-face and three-quarter view. These results appear to provide support for the 'angle of rotation' hypothesis by showing no difference in performance for views that differ by a 45-degree angle.

Thus it appears that there are inconsistencies in the literature. Some researchers have found a three-quarter advantage (Krouse, 1981; Baddeley & Woodhead, 1983; Logie, Baddeley & Woodhead, 1987) whereas others have not (Lui & Chaudhuri, 2002;

Laughery, Alexander & Lane, 1971, Experiment 2; Davies, Ellis & Shepherd (1978, Experiment 2). Furthermore, while the angle of rotation hypothesis may explain some of the inconsistencies, it cannot explain the three-quarter-view advantage found by Krouse (1981).

The symmetry argument suggests that when a three-quarter view is presented at study, a 'virtual view' could be generated and this may result in the successful recognition of the face in a novel view i.e. the full-face. This may perhaps explain why several authors have failed to find a three-quarter-view advantage when a three-quarter view was presented at study.

Schyns & Bülthoff (1994, Experiment 2) investigated the effect of symmetry by examining performance for five views of laser scanned 3D heads (-36, -18, 0, 19, 36) at study (between-subjects) and at test (within-subjects factor). The results revealed a main effect of study view, no main effect of test view and an interaction between the two. In particular, there was a significant difference in performance between 0 and all other views. When the full-face view had been learnt (0) sharp decreases in performance were observed for increased angle of rotation (an inverted U shape). The results also indicated that performance for the 36-degree study condition was highest when the test faces were either 36 or –36 degrees. Therefore, the strongest generalisation performance for the 36-degree (three-quarter-view) condition was to its *symmetrical* view, with slightly lower generalisation performance to the 18-degree condition and lowest generalisation performance to the 0 condition (full-face).

Similarly, Hill, Schyns & Akamatsu (1997, Experiment 2) also examined generalisation performance from a single view. They presented three views (0, 45 & 90 degrees) as a between-subjects factor at study and five views (-90, -45, 0, 45, 90) as a within-subjects factor at test. The results revealed no main effect of study view but there was a significant interaction between study and test views. The results supported those obtained by Schyns & Bülthoff (1994) by again revealing an inverted U shape performance for the full-face (0) images. In all conditions, performance decreased with increasing angle of rotation, however, for the three-quarter-view (45°) a peak in performance for the opposite three-quarter was also observed - the *symmetrical* view.

These results appear to suggest support for both the 'angle of rotation' hypothesis proposed by Lui & Chaudhuri (2002) and the symmetry hypothesis. It appears that when a face is rotated in depth by 90-degrees performance decreases sharply. Hill, Schyns & Akamatsu (1997, Experiment 3) used more naturalistic stimuli (3D shape and texture models) to examine this further. They found that the full-face and profile views did not generalise well because the information contained in these views is very different. For example, eyes and eyebrows look very different in a full-face view compared to a profile view. Similarly, the results also revealed that generalisation from a three-quarter-view did not depend on the test view. This again supports the suggestion that the three-quarter-view is non-singular.

These results all suggest that unfamiliar face recognition is viewpoint dependent and that generalisation to novel views from only one view is dependent on the learning view and *not* the testing view. More importantly, different patterns of viewpoint dependence are observed for different learned views. In particular, performance for the full-face view appears to reflect an inverted U shape function. Similarly, while generalisation performance for side views also decreases slightly, there is often a peak in performance for the opposite view – the symmetrical view. This suggests that learning a side view of a face may result in better generalisation performance than learning a full-face view.

Further research has attempted to determine the best angle for learning. It has been suggested (Troje & Bülthoff, 1996) that generalisation performance is greatest somewhere between 20 and 70-degrees, but the best angle for learning view is dependent on the type of stimulus used. When Troje & Bülthoff (1996) used more naturalistic textured faces, they found optimal performance in the range of 25 to 40 degrees.

Movement

The results from the previous investigations suggest that when all views of a face are presented at study, there is no difference in subsequent recognition performance between a full-face and three-quarter view. Some of these investigations presented different views in quick succession to give the appearance of movement (e.g. Newell, Chorizo & Valentine, 1999), or examined performance for a 'live' target (Logie, Baddeley & Woodhead (1987). While these studies did not specifically evaluate the effects of movement on unfamiliar face recognition, the results from other studies (e.g. Schiff, Banka & De Bordes Galdi, 1986; Bruce & Valentine, 1988; Pike et al, 1997) all suggest that movement may help to build a more robust three-dimensional representation of the face. This is particularly important, as in a real-life situation a witness will have viewed a previously unfamiliar moving face.

Encoding specificity

As viewing a moving face may lead to the encoding of three-dimensional, structural information about the face, then composite construction may be more successful if three-dimensional cues were available at test. While there was no difference in performance for recognition tasks when all views were presented, composite construction is primarily a recall task. When new features are presented, witnesses *search* their memory, extract information and decide whether the presented feature 'matches' the feature stored in memory. The process of recalling a face/feature, is inherently more difficult than *recognising* a face, which may be facilitated by *familiarity* rather than conscious recollection. Indeed, authors have noted the reconstructive nature of recall (e.g. Bartlett, 1932; Davies, Ellis & Shepherd, 1978). In particular, Davies et al (pg. 22) state that "Photofit making...becomes an act not of *reproduction* but of *reconstruction...*". Furthermore, Bartlett (1932) argued that stored items could become combined at retrieval, resulting in the recollection of incorrect information, thus suggesting that successful retrieval may be dependent on the cues available at test. Indeed, the encoding specificity principle (e.g. Thomson & Tulving, 1970; Tulving & Thomson, 1973) states that retrieval will be more successful when retrieval cues more accurately match those in the original encoded experience. However, while the retrieval cues (i.e. features) in a composite will never accurately match those in the original face, retrieval may be more successful when the cues (features) are represented in a more three-dimensional way i.e. in a three-quarter-view.

Presenting multiple views

As well as examining whether the *type* of information available at construction would improve composite performance, this investigation also examined whether increasing both the *type* and *amount* of information at test would serve to enhance performance levels. Recent research (e.g. Bennet et al, 1999; Bruce et al, 2002) and chapter 3 of this thesis has found that presenting varied information at test increases identification performance. In particular, combining composites from four *different* witnesses increased identification performance above the level observed for a single composite (Bennet et al, 1999; Bruce et al, 2002; chapter 3). The results from these investigations suggest that the combination of information from different witnesses results in an image that is closer to the ideal image or prototype (e.g. Solso & McCarthy, 1981; Bruce, Doyle, Dench & Burton, 1991; Homa et al, 2001). However, one of the aims of this investigation was to examine performance for the presentation of *different* kinds of composite from the *same* witness. While the increase in performance of the combined composites in the Bruce et al (2002) study, appears to reflect the combination of different *memories*, it is unclear whether performance would still increase when different types of information are presented at test. For example, two different views of the same person can look very different and in fact can look more different than images of different people presented in the same view (Hill et al, 1997). Therefore would the presentation of these different *types* of information (i.e. a three-quarter view and full-face view composite) also result in higher performance levels?

PROfit

In this investigation all composites were constructed using the PROfit composite program (formerly known as CD-fit). PROfit is used by police forces around the world (including the U.K) and is similar to its other computerised competitors (e.g. E-Fit). Features are displayed within the context of a whole face and can be resized and shaped. This program also contains its own feature editing tool kit, a PROwarp tool that allows for the manipulation of large areas of the image as well as individual features (e.g. to create expressions), eleven different databases and a wide choice of features. For example the two female databases (full-face and three-quarter view[1]) that were used in this investigation both contain 343 hairstyles, 281 faces shapes, 214 eyes, 316 noses, 317 lips, 76 eyebrows and 51 ears. In order to create the databases two photographs were taken of each volunteer – one at full-face and one at three-quarter view. Four features were then taken from each of these photographs. In order to create the 'generation' procedure, every feature was given 'anchor points' in order to determine its location within the face and a three-digit identity code. When the program is asked to generate a composite, it uses an index table to correctly identify the full-face features and the corresponding three-quarter view features (ensuring that matching features are used).

Aims

The main aim of this investigation was to examine the effectiveness of this new three-quarter-view database in PROfit. More specifically, Experiment 1 examined whether 'participant witnesses' could construct a more identifiable composite in a three-quarter view compared to the standard full-face view, after presentation of *all* views of a face. All target faces were presented on video and displayed equal amounts of all views, in an attempt to emulate everyday interaction. As such, this experiment was investigating whether composite performance would increase when the retrieval cues were more similar to encoding (i.e. more three-dimensional). As stated earlier, while face *recognition* studies have often failed to find a three-quarter advantage when all views have been presented at study, composite construction is essentially a *recall* task. Therefore, more accurately 'matching' encoding and retrieval cues may facilitate conscious remembering (recall) rather than familiarity based judgements (recognition). This experiment also examined performance for three-quarter-view composites that had been automatically generated from the full-face composites using PROfit. While there are no ACPO (Association of Chief Police Officers) guidelines prohibiting the construction of more than one composite, it is still current practice to only invite a witness to construct one image. Examining automatically generated composites should further enhance our understanding of any three-quarter-view effect, because if a three-quarter-view acts as a more efficient retrieval cue then the constructed composites should perform better than the automatically generated ones. However, if automatically generated composites perform well, either alone or with their corresponding full-face composite, then it would still be possible for a witness to only construct one composite. Experiment 2 examined whether the presentation of both three-quarter and full-face view composites would increase performance above the level achieved for a single composite. Experiment 3 examined the encoding specificity principle in more detail. View at study was a between-subjects factor (full-face, three-quarter and all views) while construction view (full-face and three-quarter-view) was a within-subjects factor. That is, participants were allocated to *one* viewing

[1] This database displays composites at a 30° angle, which is consistent with research (e.g. Troje & Bülthoff, 1996) which has found optimal performance for recognition between 25° and 40°

condition and were required to construct *two* composites of the target (one in a full-face view and one in a three-quarter-view). At test, full-face and three-quarter-view composites were presented alone and in pairs (i.e. one full-face and one three-quarter-view that had been constructed by the same participant).

Experiment 1: Construction of full-face and three-quarter-view composites

The main aim of this experiment was to investigate whether participants could construct a more identifiable composite in a three-quarter-view compared to a full-face view, after presentation of all views of *one* target face. The targets were female members of staff from the psychology department at the University of Stirling. In the first stage of the experiment participants viewed a 30-second video clip of an unfamiliar female target. They were then asked to construct two composites of her face from memory (one at full-face and one at three-quarter-view). In stage 2 a further set of three-quarter-view composites was automatically generated from the full-face composites. In stage 3, participants who were unfamiliar with the targets rated the composites for likeness. In stage 4, participants who were familiar with the targets attempted to identify the composites.

Stage 1: Construction of Composites

Materials

In this experiment target faces were taken from the same video and photographic database used to create the female database in PROfit. In order to create the new database two photographs were taken of each volunteer – one at full-face and one at three-quarter view. As a maximum of four features were taken from each of these photographs (the same features from each photograph), it was impossible to recreate the faces perfectly. Each feature was given 'anchor points' in order to determine its location within the face and a three-digit identity code. When the program is asked to generate a composite, it uses an index table to correctly identify the full-face features and the corresponding three-quarter view features (ensuring that matching features are used).

Video frames were extracted and digitised without sound, using the Media 100 video-editing package. A thirty-second video clip was created for each target. Each clip consisted of fifteen seconds of movement (rotating in chair from left to right: shaking head from side to side, nodding up and down) and fifteen seconds of full-face view. Composites were constructed using PROfit (Windows version 3.0) on an ASUS Hi-Grade UltiNote AS8400 laptop computer.

Participants

Sixteen adults aged between eighteen and forty years were recruited from the psychology department of Queen Margaret University College, Edinburgh. All participants were unfamiliar with the targets. Each participant received a £10 payment.

Design

A 4 (target) by 2 (construction view) mixed design was adopted, with target as a between-subjects factor and construction view (full-face and three-quarter view) as a within-subjects factor. Each participant viewed a thirty-second video clip of *one* target and constructed two composites of the *same* target (one in a full-face view and one in a three-quarter view) from memory. There were four targets and sixteen participants, creating a total of thirty-two composites (8 per target). The order of construction was counterbalanced so that eight subjects constructed a three-quarter-view composite first and eight constructed a full-face view composite first. The result was that sixteen full-face (4 per target) and sixteen three-quarter-view composites (4 per target) were constructed.

The sixteen full-face composites were then used to generate a further set of three-quarter view composites using PROfit. This created another sixteen composites and resulted in a total of forty-eight composites. See Figure 1 for an example and Stage 2 for a description of the procedure.

Procedure

Each participant was asked to view a thirty-second video clip. The participant was not initially told that they would have to remember this person. After the participant had viewed the clip they were informed of the true nature of the experiment. The procedure for the cognitive interview and construction of the composites was then explained. As rapport building is an important aspect of the cognitive interview procedure (prior to eliciting a description), the experimenter then chatted to the participant about their interest/work etc in order for them to feel as relaxed and familiar with their surroundings as possible. The total average (mean) time spent on explanations and rapport building was 12 minutes. The participant was then encouraged to close their eyes and visualise the face. For the first recall attempt (free recall) they were asked to describe the features in any order and were encouraged to describe everything they could see, even if they thought it was irrelevant. The second recall attempt was more structured in that the participant was asked to focus on each feature separately, starting at the top of the head and working their way down the face slowly. If a third recall attempt was needed the order was varied (e.g. starting at the bottom of the face and working upwards). If the participant had omitted any information, questions were then directed at these areas (e.g. Can you recall/describe the shape of the mouth?). No questions were directed at features or aspects of features that the participant had said that they could not recall. This description was then entered into either the full-face or three-quarter view database in PROfit.

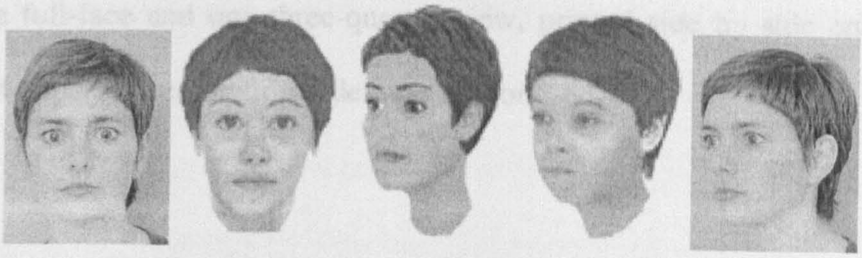
PROfit is very similar to other computerised composite systems as it displays a small facial shaped icon. A drop-down menu that provides a breakdown of each part of the feature accompanies every feature in this icon. For example, when you click on the face, the drop-down menu displays 'face shape, chin shape, length, width, age, fleshiness, forehead' etc. Within each of these categories there are a range of options. For example, for 'face shape' the options are 'oval, round, triangular, square and angular'. If a descriptor did not match the word(s) the participant had used to describe that feature, then the participant chose the descriptor that they felt was the closest alternative. The experimenter offered no advice. If a participant did not recall a feature or aspect of a feature e.g. size of eyes, then the 'average' option was entered. Where this was not possible, no descriptor was chosen.

When the *full* CI elicited description had been entered into PROfit, the participant and experimenter worked together to produce a facial likeness, by viewing chosen features, selecting alternative features and editing both features (e.g. changing size, shape, shade etc) and configuration. All features were edited using the tools available in PROfit. If further alterations were needed (e.g. highlights, shadows, laughter lines) the composite was exported into Adobe Photoshop 7. Construction of the composite ceased when the participant was either confident that the image represented a good likeness of the target, or they could not make any further changes.

On completion of the first composite, the description that was elicited from the cognitive interview was then used to construct the second composite (i.e. the same description that was used to construct the first composite). The description was entered into the second PROfit database (either full-face or three-quarter view). Both databases contain the same features but they are not in the same order, so this ensures that the participant cannot simply choose the same features, thereby replicating the first composite in a different view. The participant and experimenter then worked together to construct the second composite. No suggestions were offered during construction of this second image. No time limit was placed on this procedure, however the total average time to conduct the cognitive interview and construct both composites was 90 minutes.

Stage 2: Automatic generation of three-quarter view composites

A further set of three-quarter view composites was automatically generated from the full-face composites. In order to generate the image the program uses an index table to ensure that matching features are used. However, any alterations that are made to the full-face composite by the witness are not 'transferred' to the automatically generated image. As a result, a detailed list of all alterations was kept by the operator and each generated composite was then altered in exactly the same way as the original full-face composite, for example if the fringe had been removed on the full-face image it was removed on the generated image. This procedure commenced when all of the composites had been constructed and was repeated for all sixteen full-face composites, see figure 1 for an example. The participants were not present during this process.



Target Full-Face Constructed ¾ Generated ¾ Target

Figure 1: Composite example

Stage 3: Likeness Ratings

Materials

Full-face and three-quarter view composites were presented alone. Each composite was presented with monochrome photographs depicting the target in both views (one in full-face and one in three-quarter view). This ensured that as much information as possible was available for the task. All images measured 13cm in height. The photographs were edited using Microsoft Photo Editor to ensure that brightness and contrast were constant.

Participants

Forty unpaid participants were recruited from Queen Margaret University College and local Tesco supermarkets. All were unfamiliar with the targets.

Design

Unfamiliar participants rated the composites for likeness on a scale from one (low) to ten (high). The composites were divided into two books each containing twenty-four composites (8 full-face, 8 constructed ¾ view and 8 generated composites). Each participant saw only one book, with twenty participants rating the composites in book one and twenty rating the composites in book two. Each composite was printed on a single sheet of A4 paper and displayed with two monochrome photographs of the target (one full-face and one three-quarter view, printed side by side on a separate sheet of A4 paper). Presentation order was randomised.

Procedure

Each participant was told that the composites were constructed after a 'participant witness' had only seen the target face for 30 seconds. It was stressed that the composites were constructed from memory and that they represented a likeness of the original target. Each participant was then informed that his or her task was to rate how good the likenesses were. They were asked to study each set of images (composite and photographs) and rate the composites for likeness on a scale of 1 (low) to 10 (high). This was repeated for all twenty-four composites. No time limit was placed on this procedure.

Results

The highest likeness ratings were achieved for the three-quarter view composites (mean = 4.1, s.d.=0.95) followed by the full-face composites (mean=3.8, s.d.=1.27) then the generated three-quarter view composites (mean=2.51, s.d.=0.89). A 4 (target) by 3 (type of composite) repeated measure analysis of variance revealed a significant effect of type of composite ($F(2,78) = 36.413$ $p<0.01$), a significant effect of target ($F(3,117) = 33.939$ $p<0.01$) and a significant interaction ($F(6,234) = 9.139$ $p<0.01$). Further analyses on type (paired-samples t-tests) revealed that there were no significant differences between the full-face and three-quarter view composites (t (39)

=1.595 p>0.05). However there were significant differences between the full-face and generated composites (t (39) = 6.555 p<0.01) and between the three-quarter view and generated composites (t (39) = 8.058 p<0.01). Further analyses on each target (one-way repeated measure anovas and paired-sample t-tests) revealed that for three out of the four targets, the generated composites performed significantly poorer than both the full-face and three-quarter view composites (p<0.01 for all). Interestingly, for one target (target 1), the three-quarter-view composites performed significantly better than the full-face composites (t (39) = -5.027 p<0.01). However, one target performed at floor level across all conditions. It is unclear why this target performed poorly in all conditions. The pattern of verbal descriptions for these targets is examined in chapter 4 and this reveals that the amount and type of descriptors recalled for target 2 did not differ significantly from the other three targets, suggesting that participants could describe this face well. Therefore, the poor quality of the composites may have reflected difficulty in translating the description into a pictorial representation.

To examine any possible effects of construction order, the mean ratings were calculated for each participant's composites. A paired samples t-test revealed no effect of order of construction (t (15) = 0.844, p = 0.4). In addition, as this experiment used targets that were represented in the PROfit database, the frequency of 'correct' feature choices was examined (i.e. features originally taken from the targets). As the frequencies were very low, they were not subjected to formal analyses. The number of features that were correctly chosen overall were ears (2), nose (2), eyes (3), hair (5). No single participant chose more than one correct feature and the frequency of 'correct' choices did not differ between the different views. In addition, choosing a 'correct' feature did *not* increase performance of the composite.

Stage 4: Identification

Participants

Thirty-two members of staff from the department of psychology at the University of Stirling. All participants were familiar with the target faces.

Design

Participants who were familiar with the targets were asked to identify the composites. To avoid priming effects, each participant was presented with only one composite for each target. Twelve books were constructed, each containing one type of composite for each of the four targets. Each participant saw only one book (i.e. four composites).

Procedure

Each participant was informed that the composites were constructed after a 'participant witness' had only seen the target face for 30 seconds. It was stressed that the composites were constructed from memory and that they represented a likeness of the original target. Pilot work with non-specific instructions led participants to assume that the targets were famous rather than familiar. Therefore it was necessary to change the instructions and participants were told that the composite represented someone from the psychology department. A better method may have been to say that the targets were personally familiar to the participants. However, they may still have assumed that the targets were members of the department. While these instructions decreased the number of possible targets, the total number of targets (i.e. women working in the department) was still at least thirty-six. Participants were encouraged to provide a name or some identifiable semantic information about the person. On completion, participants were told who the targets were.

Results

Equivalent rates (% of participants correctly identifying composites) were found for the full-face composites (23% correct with 9% false positives), and the three-quarter view composites (22% correctly identified with 9% false positives), with the generated composites performing more poorly (13% correctly identified and 28% false positives). The data was collapsed across targets and a Friedman test was conducted on the hit rate. This revealed that there were no significant differences (X^2=1.55; df=2 p>0.05) between the different types of composite although the trend is clearly in line with the rating scores.

Discussion

The results from this experiment did not show a three-quarter view advantage, but instead revealed that the three-quarter view composites performed as well as the full-face composites. Interestingly, these results are similar to those obtained for recognition tasks (e.g. Hill, Schyns & Akamatsu, 1997; Schyns & Bülthoff, 1994). These studies also demonstrated that when all views were presented at study, no one view was preferred at test. The results from this investigation suggest that participants may have been exposed to enough ‘instances’ of each view at study, to ensure successful generalisation to both views at construction. Therefore, a three-quarter-view composite is as good as a full-face composite when all views have been presented, but not better. Encoding specificity and viewpoint dependency will be explored in more detail in experiment 3.

Whereas experiment 1 asked whether constructing an image in a three-quarter-view could increase composite performance, experiment 2 examines whether *adding* an additional image (three-quarter-view) at test would serve to increase performance above the level observed for a single full-face image. Stage 1 examined performance for full-face and three-quarter-view composites that had both been constructed by 'participant witnesses'. Stage 2 examined performance for full-face and automatically generated three-quarter-view composites (i.e. generated from the full-face composites using PROfit).

Experiment 2: Presenting multiple views

Stage 1: Presenting full-face and three-quarter-view composites

Participants

Thirty-two participants were recruited from the psychology department at the University of Stirling. They consisted of third and fourth year psychology students and three members of staff. All participants were familiar with the targets.

Design

The aim of this experiment was to examine whether *adding* an additional image (a three-quarter-view) would increase performance. From the ratings given in experiment 1 it was possible to determine the highest rated (best) and intermediate rated (average) full-face composites for each target face. The full-face composites were shown alone and with their corresponding three-quarter view (i.e. the one that was constructed by the same participant). This created a total of sixteen composites (4 best and 4 intermediate at full-face alone and 4 best and 4 intermediate at full-face and three-quarter view together). Participants were shown one type of composite of each of the four targets.

Procedure

Participants were approached and asked to attempt to identify the person depicted in each of the four composites. No participant had taken part in experiment 1. The procedure was identical to the identification procedure in stage 3 of experiment 1.

Results

The mean percent correct identifications are summarised in figure 2.

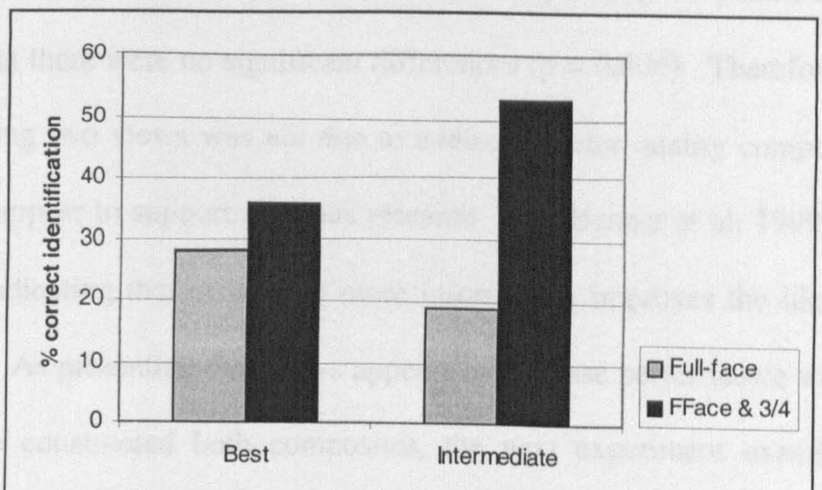


Figure 2: Percentage correct identifications per condition

The data was collapsed across target and a Cochran's Q test revealed that the full-face and three-quarter composites shown together were identified significantly more often than the single full-face composites ($Q = 8.43$, $df = 3$ $p<0.05$). Further analysis using Mcnemar tests revealed that there were significant differences between the intermediate rated full-face and full-face & three-quarter composites ($p< 0.05$), but not between the best composites ($p> 0.05$). In particular, the intermediate composites displayed a marked increase jumping from 19% correct identifications for the single full-face to 53% when both views were presented. The best composites increased from 28% for the single full-face to 37% when both views were presented.

These results indicate that there is a benefit for presenting two views rather than one. However, as the presentation of both views was determined by the quality of the *full-face* composites (best & intermediate) is this effect due to just 'adding in' a better quality three-quarter-view composite? To examine this, the overall mean likeness ratings for the best and intermediate full-face composites were compared with the overall mean ratings for the corresponding three-quarter-view composites (i.e. the ones that were presented with the full-face composites). A paired-samples t-test revealed that there were no significant differences ($p = 0.606$). Therefore, the benefit for presenting two views was *not* due to adding a better quality composite. Instead, the results appear to support previous research (e.g. Bennet et al, 1999; Bruce et al, 2002) by indicating that presenting more information improves the identification of composites. As presenting two views appears to increase performance when the same witness has constructed both composites, the next experiment examined whether performance would also increase when full-face composites were presented with automatically generated three-quarter-views (i.e. generated from the full-face composites using PROfit).

Stage 2: Presenting full-face and 'automatically generated' three-quarter views

Materials

An identification task was not undertaken due to the limited number of participants who were familiar with the targets. Instead, a six alternative forced-choice task was undertaken. To create the arrays, five distractors were chosen for each of the four target faces. These were chosen on the basis of the verbal descriptions given in stage 1 of experiment 1. These descriptions were often extensive and varied, so distractors were matched only on information concerning hair, face shape and age. These were presented with the target as black and white photographs on a single sheet of A4 paper. Microsoft Photo Editor was used to ensure that brightness and contrast was consistent. See figure 3 for an example of arrays.

Figure 3: Example of arrays

Participants

Forty-eight participants aged between 17 and 50 years were recruited from local businesses in Edinburgh. All participants were unfamiliar with the targets. No participant had taken part in any earlier experiments.

Design

The same best and intermediate rated composites were used. The full-face composites were shown alone and with both the corresponding three-quarter view and generated composites. There were twenty-four composite types in total (8 full-face, 8 full-face and constructed three-quarter and 8 full-face and automatically generated three-quarter). Each participant was shown one type of composite for each of the four targets. Each composite type was presented with an array of six black and white photographs (one of the target and five distractors). This method was used as a way of assessing the quality of the composites and was not designed as a formal 'line-up'.

Procedure

Participants were shown a line-up of six photographs together with a set of composite images. They were told that the composites represented a likeness of the target and that both composites represented two views of the same person. Participants were told that the target may or may not be in the array. They were asked to examine all of the images closely and to indicate whether or not they thought the target was in the array. If they thought the target was present, participants were asked to point to the appropriate photograph. This was repeated for all four targets.

Results

The overall percentage correct matches were 52% for the single full-face composites, 73% for the full-face with ¾ view composites and 42% for the full-face with generated composites. A Friedman test revealed that the observed differences were significant, (X^2 (2, 48) = 7.078 $p<0.05$). Further analyses using Wilcoxon Signed Rank tests revealed that there were significantly more correct matches for the full face & three-quarter view composites, compared to the full-face and automatically generated composites ($p<0.05$). The difference between the full-face and three-quarter and the single full-face composites did not quite reach significance ($p= 0.068$), although the trend is clearly in line with the results obtained in stage 1. In addition, there were no significant differences between the single full-face and the full-face and generated composites (p=0.369). Furthermore, there was an almost significant reduction in performance for the best (highest rated) full-face composite when an automatically generated three-quarter-view composite was presented alongside it (p=0.059). Similarly, the best full-face and three-quarter-view composites (presented together) performed almost significantly better than the full-face and generated composites (presented together) (p=0.090).

These results therefore suggest that while there appears to be a benefit for presenting two views, this benefit is only apparent when the composites have actually been constructed. The automatically generated composites performed poorly when presented alone *and* when presented with the full-face composites. These results suggest that a three-quarter-view composite *does* act as an efficient retrieval cue, as performance is significantly better for the constructed three-quarter composites compared to the automatically generated images. The results also suggest that just presenting *more* information does *not* facilitate increased performance at test. Instead, these results provide supporting evidence for the presentation of different *types* of information, as reported by Bennet et al (1999), Bruce et al (2002) and chapter 3 of this thesis.

The results from Experiment 1 revealed that a three-quarter-view performed as well as a full-face view when all views were presented at study. This is line with face recognition research (e.g. Hill, Schyns & Akamatsu, 1997; Schyns & Bülthoff, 1994). The following experiment examined the effect of encoding specificity in more detail.

Participants were allocated to one of three encoding conditions (full-face, three-quarter-view or all views). They were then asked to construct both a full-face and a three-quarter-view composite of the same target.

Experiment 3: Encoding Specificity

Stage 1: Composite construction

Materials

Four females from a different university (Queen Margaret University College, Edinburgh) agreed to act as targets in this experiment. Each target was videotaped individually using a Sony Hi8 camcorder for approximately three minutes. They were asked to sit in a chair and converse with an experimenter while both looking straight ahead and moving (rotating in chair from left to right: shaking head from side to side, nodding up and down). Three thirty-second video clips were then created for each target. The first clip displayed the target looking straight-ahead (full-face condition), the second clip displayed the target at a thirty degree angle (three-quarter view condition) and the third clip displayed equal amounts of the previous two conditions (15 seconds looking straight ahead and 15 seconds of movement: the all view condition). Frames were extracted and digitised without sound, using the Media 100 video-editing package. Targets were also photographed using a Digital Olympus C-900 camera in two different positions (full-face and three-quarter view).

Participants

Twenty-four adults aged between eighteen and forty years were recruited from Stirling University. All participants were unfamiliar with the targets. Each participant received a £10 payment.

Design

A 4 (target) by 3 (encoding view) by 2 (construction view) mixed factorial design was adopted, with target and encoding view (full-face, three-quarter view and all views) as between-subject factors and construction view (full-face and three-quarter view) as a within-subjects factor. Each participant saw *one* viewing condition (full-face view, three-quarter view, all views) of *one* unfamiliar target. They were then asked to construct *two* composites of that target, one in a full-face view and one in a three-quarter view from *memory*. There were six participants for each of the four targets, ensuring that for every target two participants saw the target in a full-face view, two saw the target in a three-quarter view and two saw all views of the target. This created a total of 48 composites; 12 per target. Target order was randomised and construction order was counterbalanced.

Procedure

The procedure for composite construction was identical to the procedure in stage 1 of experiment 1.

Stage 2: Likeness Ratings

Materials

Each full-face and three-quarter view composite was presented with monochrome photographs depicting the target in both views (one in full-face and one in three-quarter view). All images measured 13cm in height. The photographs were edited using Microsoft Photo Editor to ensure that brightness and contrast were constant. Each composite was printed on a single sheet of A4 paper and displayed with two monochrome photographs of the target (one full-face and one three-quarter view, printed side by side on a separate sheet of A4 paper).

Participants

Twenty-two participants aged between 18 and 45 years were recruited from the University of Stirling. Participants had not taken part in any of the previous experiments and all were unfamiliar with the targets.

Design

Unfamiliar participants rated the composites for likeness on a scale from one (low) to ten (high). All forty-eight composites were randomly ordered in one presentation book. Presentation order was randomised.

Procedure

This procedure was identical to the likeness rating procedure used in stage 3 of experiment 1.

Results

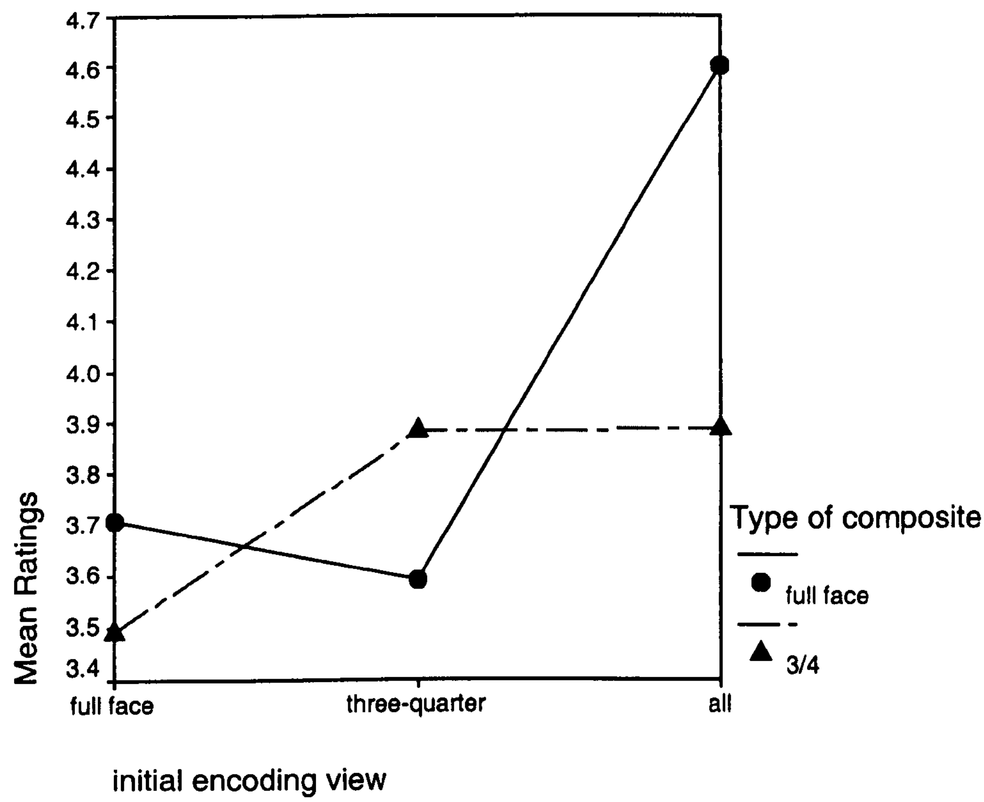


Figure 4: Mean ratings for full-face and three-quarter-view composites

The mean ratings for the full-face and three-quarter view composites are summarised in figure 4. A 2 (type of composite - full-face, ¾) by 3 (encoding view - full-face, ¾, all views) repeated measure analysis of variance was conducted. This revealed a significant effect of type ($F(1,21) = 5.356$ $p<0.05$), an effect of view ($F(2,42) = 19.242$ $p<0.01$) and a significant interaction between type and view ($F(2,42) = 16.187$ $p<0.01$). Further repeated measures anovas revealed a significant effect of view for both the three-quarter view composites ($F(2,42) = 5.024$ $p<0.05$) and the full-face composites ($F(2,42) = 32.208$ $p<0.01$). Pairwise comparisons using paired samples t-tests revealed that for the three-quarter view composites there were significant differences between the all view condition and the full-face condition ($p< 0.05$) and between the three-quarter view and full-face condition ($p < 0.05$). For the full-face composites there were significant differences between the all condition and both the three-quarter view ($p< 0.01$) and the full-face condition ($p<0.01$).

More importantly figure 4 does reveal a moderate encoding specificity effect. The three-quarter view composites performed significantly better than the full-face composites when the target was seen in a three-quarter view ($p< 0.05$). However, while the full-face composites appeared to perform better when constructed after viewing the target in full-face (mean = 3.7), compared to viewing the target in a three-quarter-view (mean = 3.59) this difference was not significant. Similarly, the full-face composites performed significantly better than the three-quarter view composites ($p<0.01$) in the all view condition.

These results provide initial support for the encoding specificity principle by suggesting that presenting the same view at study and test is better than introducing a different view at test. However, they also suggest that presenting as much information as possible is even better. Taken together, the results from experiments 1,2 and 3 indicate that then you have seen all views of a face and only construct *one* composite, it doesn't matter whether you construct a full-face composite or a three-quarter-view composite, as both perform equally well (although experiment 3 all view condition appears to be an exception here). The results from experiment 2 suggest that constructing two views of the same person may be preferable, as there is a significant increase in performance when both views are presented at test. The results from experiment 3 also show that performance is good for both the full-face and three-quarter-view composites when all views of the face have been seen. However, the results also indicate that if you have only seen a face in a three-quarter-view, it is *better* to construct a three-quarter-view composite.

Stage 2: Array Task

Materials

Target absent and target present arrays were constructed for each of the four targets. The target absent arrays contained monochrome photographs of six similar looking females. The target present arrays contained one monochrome photograph of the target and five distractor photographs. The same distractors were used in both arrays. Due to inconsistencies in the verbal descriptions given by participants the faces were matched visually for hairstyle/colour, face shape and approximate age. All images were standardised for height (7cm) and were presented on a single sheet of A4 paper. Microsoft Photo Editor was used to ensure that brightness and contrast were consistent. Four different sets of arrays were constructed (target present full-face view, target present ¾ view, target absent full-face view and target absent ¾ view).

Participants

Two hundred and eighty eight participants aged between 18 and 55 years were recruited from cafeterias and student unions at both the University of Glasgow and Queen Margaret University College, Edinburgh. Participants had not taken part in any of the previous experiments and all were unfamiliar with the targets.

Design

As each participant had constructed two composites of *one* target (one composite in a full-face view and one composite in a three-quarter view), each composite was presented alone and with its corresponding partner as a 'pair' (i.e. the full-face and three-quarter-view composite that had been constructed by the same participant were presented together). There were 48 single composites and 24 'pairs'. All composites were presented with both target present and target absent arrays. View was held constant i.e. three-quarter view composites were presented with three-quarter view arrays and full-face composites were presented with full-face arrays. This created a total of 144 presentations (48 single composites and 24 'pairs' of composites presented with both target present and target absent arrays). Careful consideration was given to array view for the pairs. As each pair contained one full-face composite and one three-quarter-view composite, the optimum array would contain both views. However, this was not possible in this experiment and as an advantage for presenting both views had previously been found in experiment one using full-face arrays, this procedure was adopted here. To ensure that each participant only saw one composite for each of the four targets, thirty-six separate presentation books were constructed. Each book was balanced for type of composite, initial encoding view and array type.

Procedure

Participants were told that the composites were constructed from memory and that they represented a likeness of the original target. They were told that when they saw two composites, these represented two views of the same person. Participants were asked to examine all of the images closely and were told that the target may or may not be in the array. They were asked to indicate whether or not they thought the target was in the array. If they thought the target was present, participants were asked to point to the appropriate photograph.

Results

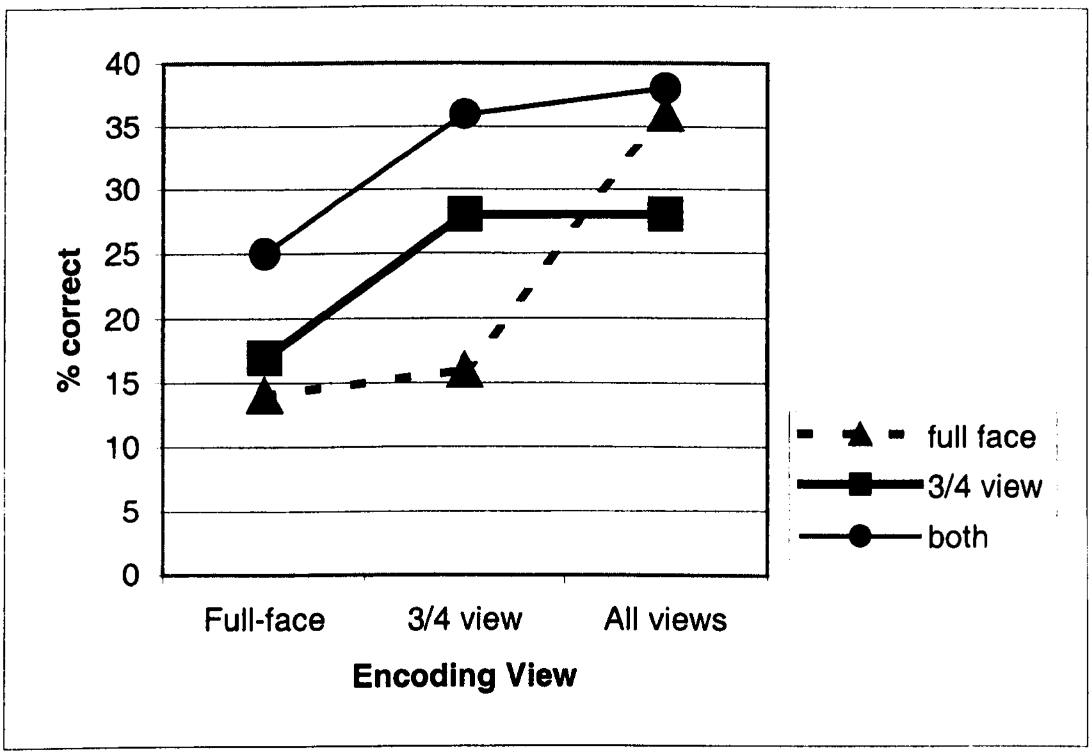


Figure 5: Mean no of correct matches (hit rate and correct rejections) for full-face, three-quarter-view and both composites presented together

Figure 5 displays the percentage of correct matches for both target present arrays (hits) and target absent arrays (correct rejections). As can be seen, the same pattern of performance was obtained in both experiments 2 and 3. Presenting both views appears to be better than presenting a single view (experiment 2) and it is better to construct a three-quarter-view composite when you have only seen a face in a three-quarter-view (experiment 3 likeness ratings). Although the design of this experiment lent itself to parametric analyses, as the data was categorical (0=incorrect match, 1=correct match) non-parametric tests were used. The data was initially collapsed across target and encoding view and a Friedman test was conducted on the overall hit rate. This revealed significant differences (X^2=6.049; df=2 p>0.05). Further analyses using Wilcoxon Signed Rank tests revealed that when both composites were presented performance was significantly better than the single full-face composites ($p < 0.05$). The presentation of both composites appeared to perform better than the presentation of a single three-quarter-view composite, however this difference did not quite reach significance ($p=0.083$).

A Cochran's Q test on the no of correct matches by type and view revealed significant overall differences ($Q = 22.688$, $df = 8$ $p<0.05$). Further tests by individual type of composite revealed significant differences for the full-face composites ($Q = 11.806$, $df = 2$ $p<0.05$). Pairwise comparisons using Mcnemar tests revealed that the all view condition was significantly higher than both the three-quarter-view condition ($p<0.05$) and the full-face condition ($p< 0.05$). No significant differences were observed for the three-quarter-view composites, however the difference between the full-face condition and the three-quarter view condition almost reached significance ($p= 0.065$). Similarly, no significant differences were observed for both composites.

There were no significant differences in false positives for both the full-face and three-quarter view composites ($X^2=2.339$; $df=5$ $p>0.05$). For the full-face condition there were 33% false positives for the full-face composites, 37% for the three-quarter-view composites and 37% when both composites were presented. For the three-quarter-view encoding condition, there were 36% false positives for the full-face composites, 38% for the three-quarter-view composites and 34% when both views were presented. For the all view encoding condition, there were 44% false positives for the full-face composites, 30% for the three-quarter-view composites and 36% when both composites were presented. Presenting both composites resulted in significantly more false positives when a full-face had been encoded ($p<0.05$ for comparisons with both single full-face and single three-quarter-view composites). However, when a three-quarter view had been encoded presenting both composites resulted in fewer false positives, although this difference did not quite reach significance ($p = 0.056$ for both comparisons).

Figure 5 displays a similar pattern to the likeness rating data in figure 4. In the three-quarter encoding condition the three-quarter view composites again perform better than the full-face composites, although this time the difference does not quite reach significance ($p = 0.059$). However, the moderate benefit for the full-face composites in the full-face encoding condition has disappeared ($p > 0.05$). In the all view condition there are no significant differences between the full-face and three-quarter-view composites.

The results from both the rating and array tasks indicate that performance will be better when a 'participant witness' has encoded all views of a target face. While there is no increase in performance for the three-quarter view composites, performance is still high and there is a marked increase in performance for the full-face composites. The results also suggest that when a 'participant witness' has seen a face in a three-quarter-view, performance will be better when a three-quarter-view composite is constructed. Interestingly, the results suggest that when a full-face view has been encoded, performance will be low when a full-face composite is constructed.

The performance of the three-quarter-view composites in the three-quarter encoding condition provides initial support for the encoding specificity principle. However, the performance of the full-face composites does not. Furthermore, the similar performance of the three-quarter-view composites in both the three-quarter and all view encoding conditions indicates that similar information was encoded from both presentations. This may provide support for the symmetry argument proposed by Poggio & Vetter (1992) who state that learning one view of a bilaterally symmetrical object can be sufficient to generalise to other views. As a face is generally bilaterally symmetrical, then a side view (the symmetrical view), which is non-singular, may contain enough information to generalise to other views (Hill, Schyns & Akamatsu, 1997; Schyns & Bülthoff, 1994).

The results from experiment 1 are also supported by this research by indicating that when all views of a face are presented, no one view is preferred (similar results were obtained by Hill, Schyns & Akamatsu, 1997 & Schyns & Bülthoff, 1994). However, if the symmetry argument was correct, then performance for the full-face composites should have been higher when a three-quarter-view had been encoded. However, the results for the full-face composites are generally supported by the symmetry argument. While the results for the ratings task provided initial support for the encoding specificity principle by indicating that full-face composites were better when a full-face had been encoded, this difference was not significant. Furthermore, the results from the array task in experiment 3 revealed that performance for the full-face and three-quarter-view composites was low when a full-face view had been encoded.

Therefore, several patterns of viewpoint dependence have emerged when generalising from a single view. When a face has only been seen in a three-quarter-view, it is *better* to construct a three-quarter-view composite. When a face has been seen in a full-face view performance is low for both full-face *and* three-quarter-view composites. However, when all views of a face have been encoded, composite performance is equally good in both views.

In addition, experiments 2 and 3 both suggest that when more information is provided at test, performance increases. This increased performance for presenting two views of a face is only observed when both composites have been constructed, as performance decreases when one of the composites has been automatically generated. This supports previous research (e.g. Bennet et al, 1999; Bruce et al, 2002; chapter 3) by suggesting that the presentation of *varied* information increases identification. The experiments in chapter 4 examine the presentation of differing types and amounts of information in more detail. In particular, these experiments have found that simply presenting *more* information does not serve to increase identification (i.e. the presentation of more than one composite by the *same* participant in the *same* view). This may explain the poor performance for presenting a full-face composite with a very similar automatically generated three-quarter-view. While the generated composite was presented in another view, it contained the *same* information as the full-face composite.

To conclude, when a witness has seen a side view of a suspect the results indicate that a three-quarter-view composite should be constructed. However, while standard full-face composites perform well when all views of the face have been encoded, care should be taken when a person has only seen a face in a full-face view, as composites in both views achieved low levels of performance when a full-face view had been encoded. The results also indicate that it would be beneficial for a witness to construct two composites of a suspect, one in full-face view and one in a three-quarter-view. This seems particularly important when a witness has only seen a full-face view. As the main aim of this investigation was to examine the effectiveness of a new three-quarter-view database in PROfit, it can be concluded that this database is a useful and beneficial aid, in increasing the recognition and identification of composite images.

PAGE MISSING IN ORIGINAL

3

Multiple Composites from Multiple Witnesses

Chapter 2 examined whether composite performance would increase when two different composites from the same witness were presented together. In contrast, this chapter examined whether the use of composites from multiple witnesses would serve to increase performance. Composites are quite often used in large scale cases e.g. the Jill Dando and Yorkshire Ripper investigations, and as a result there may be several eyewitnesses. However, until very recently the ACPO (Association of Chief Police Officers, 1999, pg11) guidelines have stated that "Where there is more than one witness to a single incident, each witness should be assessed individually on their level of recall to produce a composite image. One image should then be produced using a witness with good recall. The remainder of the witnesses can then be kept for other formal means of identification".

This statement limits the number of composite constructions to one, and also implies that it is possible to identify the witness with the most accurate recall. Limiting the amount of composites may not necessarily be problematic *if* the 'best' witness constructs the image. However, how do you know which witness is the *best* witness

The above statement suggests that 'good' recall is an indicator of eyewitness accuracy and further statements expand on this (ACPO guidelines pg.'s 9-10) "It is essential that the witness has seen the suspect's face and is able to visualise or describe the facial features. The witness must have a clear mental image of the person and more importantly can visualise or describe, however simply, the facial features to be reproduced...". This suggests that the witness's ability to construct a 'good' composite is dependent on their ability to recall the facial features. However, an investigation into the possible relationship between verbal descriptions and composite construction is reported in chapter 4. No reliable relationship between amount of verbal description, accuracy of descriptions and composite quality could be found. Therefore the quality of the verbal description seems to be an unreliable indicator of eyewitness performance.

Implicit in this guideline is the reliance on eyewitness confidence. It is stated that a witness should be able to "visualise" *OR* "describe" the face. Therefore, there is a reliance on a witness's self reported measure of accuracy and confidence. The guidelines state that "A witness being confident that they could identify the suspect again is recognition. This is NOT the same as being able to confidently recall the facial features, without which a composite cannot be constructed" (pg10). Therefore, a composite will not be constructed unless a witness is *confident* that they can recall the facial features. However, numerous studies have indicated that confidence is *not* a reliable indicator of eyewitness accuracy (e.g. Sporer, Penrod, Read & Cutler, 1995). Other factors such as personality and intelligence have also failed to predict eyewitness accuracy (e.g. Loftus, Levidow & Duensing, 1992). Consequently, there appears to be no formal method, at present, of identifying the *best* witness. As a result there is no way of knowing which witness will construct the most identifiable composite.

As there is no way of determining whether a witness will produce a 'good' quality composite, one solution may be to invite *more* than one witness to construct a facial likeness. However, the recent ACPO guidelines state that "The production of a composite image with multiple witnesses must not be attempted, as this will amount to cross contamination of each witness's primary memory" (1999, pg11). Nevertheless, while this statement appears to state that witnesses should not construct a composite 'together', it does not appear to preclude the construction of individual composites by multiple witnesses.

Presenting Multiple Composites

Bennet, Brace, Pike & Kemp (1999) were the first researchers to examine whether the presentation of multiple composites would increase identification performance. They asked 8 'participant witnesses' to view a 'live' simulated crime (person stealing a camera from a car). The witnesses were exposed to the target for 1 minute and saw his face in all angles. After a short interval (10 minutes) the participants gave a verbal description (elicited by Cognitive Interview techniques) and constructed a composite of the target with an E-FIT operator. At test, both quality *and* number of composites were manipulated. Firstly, all composites were presented to the original witnesses (the ones that had *constructed* the composites) and they were asked to rank all composites in order of likeness. This permitted the identification of the 'best, average and worst' composites. Secondly, a further group of participants were asked to examine the eight composites and choose the four that most resembled each other. The images were then presented to a group of participants who were familiar with the targets. These participants saw all *eight* composites, *four* similar composites and a *single* composite. Participants were asked to try to identify the person depicted in the composites. The results indicated that performance increased when composites from multiple witnesses were presented. In particular, presenting four similar composites yielded the highest performance (42% correct identifications) followed by the presentation of eight composites (29% correct identifications) with the single composite yielding the lowest performance (13% correct identifications). Further investigation manipulated the quality of the four composite condition. This revealed that the presentation of the four *worst* composites still performed better than either the single or the eight composite condition.

This investigation indicates that performance is best when four composites are presented but that performance decreases when this number is increased. The authors suggested that performance for the eight composite condition may have been lower because this condition contained the *best* and *worst* composites. While the presentation of four *poor* quality composites performed better than the other conditions, it may be that when *different* quality composites are presented together identification becomes more problematic. If the quality varies greatly the composites may essentially resemble different people.

Combining Composites

Composites from multiple witnesses will invariably range in quality. Some witnesses may construct a good likeness of the target, while others may construct a very poor image. As presenting different quality images at test may affect performance, one way to counter this would be to combine the composites to produce a 'new' image. This was investigated by McNeil et al (1987). They asked two groups of participants to construct Identi-Kit composites both immediately after viewing a 'live' target and at a three-week delay. New 'modal' composites were then constructed by selecting the most frequently chosen features. All composites were rated for likeness against a black and white photograph of the target, on a scale from 1 (poor likeness) to 8 (very good likeness). The results revealed that the new 'combined' composites performed better than the single Identi-Kit images in all conditions. While these results indicate that performance increases when information from multiple witnesses is combined, the 'modal' method of combining information would be extremely difficult to adapt to a new computerised system such as PROfit. Composite construction with a system such as PROfit permits greater ability to manipulate both featural and configurational information, as well as enhancing images with paintwork. Therefore, *if* more than one witness chose the same initial feature (e.g. nose), it is unlikely that the features would look exactly the same in the final composite. Consequently, combining information in this simple way would be impossible.

A more effective way of combining information from multiple witnesses may be to just 'merge' or 'morph' the composites together. Bruce et al, (2002 Experiment 1) examined performance for morphed composites. In this investigation each participant was asked to construct a composite of two *different* targets. The first composite was constructed with the target image present and served to familiarise participants with the construction procedure. Participants then constructed a second composite of a different target from memory (after a 30-second exposure to target photograph). Half of the targets were famous and half were unfamiliar. Construction order was counterbalanced such that half of the participants constructed a composite of a famous target first and half constructed a composite of an unfamiliar target first. Combined composites were then created by morphing composites from two different witnesses (2-Morphs) and four different witnesses (4-Morphs). Performance for both individual and morphed composites was initially assessed by likeness ratings (1-low to 10-high). The overall results revealed that the highest performance was obtained for the 4-Morphs, followed by the 2-Morphs, with the individual composites yielding the lowest performance rate. The best and worst composites were identified from the likeness ratings and these were compared with the 4-Morph and the set of four (all 4 composites presented together) for each target. A six alternative forced-choice array task was employed and the overall results revealed that the 4-Morph performed better than the set of 4 condition. However, for the more ecologically valid condition, where participants constructed composites of an unfamiliar target from memory, performance was low and did not differ significantly between conditions.

While this research provides further support for the presentation of composites from multiple witnesses, it is unclear whether the use of morphed composites would be particularly advantageous in a more forensically valid condition. Therefore, experiment 1 of this chapter extended this research by examining performance for morphed composites with a more ecologically valid method.

Composites as Prototypes?

The investigations by Bennet et al (1999), McNeil et al (1987) and Bruce et al (2002, experiment 1) all suggest that performance increases when information from multiple witnesses is presented. This evidence together with evidence from research on the effect of prototypical faces suggests that each composite may be a deviation from the ideal image or prototype. Therefore, when the composites are combined, they may produce an image that is closer to the ideal, thus serving to increase performance. Research on prototypes has found that participants falsely recognise unstudied prototype faces with greater confidence than previously seen faces (e.g. Solso & McCarthy, 1981; Bruce, Doyle, Dench & Burton, 1991; Homa et al, 2001). In these investigations participants were required to examine a set of faces for a few seconds each. They were then presented with the same faces, together with a set of new prototypical faces and were asked to recognise the faces that were presented in the study phase (old/new judgement).

In particular, in the Solso & McCarthy (1981) study, these new faces contained differential amounts of prototypical information (100% prototypes, 75%, 50%, 25%, 0%). The results revealed that 35 out of 36 participants reported that they had seen the 100% prototype in the study phase. This suggests that memory for faces/features is somehow integrated with memory for similar items, either through trace blending (e.g. Metcalfe, 1991), by storing integrated traces (e.g. Hintzman, 1986) or as a result of competing patterns of interaction (e.g. McClelland & Rumelhart, 1996) (see chapter one for a review of memory models). Therefore, information is combined in memory and results in the false recognition of unstudied faces.

However, instead of combining information in memory, both the McNeil et al (1987) and Bruce et al (2002, experiment 1) investigations combined information 'artificially' at test. In particular, the Bruce et al study used a morphing procedure to combine composites. This procedure 'averages' information, which may result in the reinforcement of similar or 'correct' information while inconsistent or 'incorrect' information is 'averaged out'. As such, this procedure may be particularly effective when the composites contain different kinds of information. For example, if four different composites all portray noses of different sizes and shapes, the combination of this information may result in a nose that is closer to the 'ideal' nose.

Evidence from both the Bennet et al (1999) and Bruce et al (2002) investigations suggest that composites *do* differ greatly in quality. The results from the likeness ratings from both studies revealed that the composites ranged from very poor or 'worst' to very good or 'best'. However, whereas information was combined in the 'modal' (McNeil et al, 1987) and 'morphed' (Bruce et al, 2002, experiment 1) composites, performance was also good when four composites were presented together (Bennet et al 1999; Bruce et al, 2002).

Research on the effect of figural after-affects on the perception of faces (e.g. Webster & MacLin, 1999; Leopold et al, 2001) has examined how the presentation of one face, affects the presentation of another face. Webster & MacLin (1999) asked participants to either match or rate faces before or after they were presented with a distorted face. They found that the presentation of a distorted face affected both the perception and recognition of the original face. More specifically, after presentation of a distorted face, the original (undistorted face) appeared distorted in the *opposite* direction. Extending this work, Leopold et al (2001) examined performance for faces that were distorted along a continuum (anti-face or anti-caricature, average, caricature and original face). In particular, this investigation asked whether the presentation of a distorted face would affect the identity judgement of a second face. The results revealed that exposure to a single face *did* produce a significant bias in the perception of the second face. More specifically, it was found that the perceptual bias moved in the opposite direction to the studied face, passing through the face space continuum. This led the authors to state that "The encoding of faces…draws upon…mechanisms that reference the central tendency of the stimulus category" (pg89). Much like the prototype effect discussed earlier.

While composites are not distorted per se, they can be thought of as *distortions* or *deviations* from the ideal, or prototypical image. As such, they would also range along a perhaps less well-defined face space continuum. This research suggests that when more than one composite is presented, each composite (or deviation) may have an effect on the perception of subsequent composites. In particular, the investigation by Leopold et al (2001) suggests that this perceptual bias may result in the perception of an image that is closer to the ideal or prototype. Therefore, while it may be possible to 'create' a composite that is closer to the prototype by combining information artificially e.g. morphing, presenting multiple composites may also result in the *perception* of an image that is closer to the ideal or prototype.

Aims

In summary, research has indicated that performance increases when composites from multiple witnesses are presented and that this information may serve to 'create' an image that is closer to the 'ideal' or prototypical image. The main aim of this investigation was therefore to examine performance for morphed composites using a more ecologically valid procedure. For experiment 4, target faces were taken from a Stirling University database and *unfamiliar* participants at a different university were asked to construct the composites. A further group of *familiar* participants at Stirling University were then asked to identify the composites. Experiment 5 examines the prototypical effect in more detail by combining composites from multiple and single witnesses. If morphed composites represent an image that is closer to the ideal or prototype, then morphed composites from multiple witnesses should perform significantly better than morphed composites from single witnesses (due to the combination of memorial representations). Similarly, experiment 6 examines the prototypical effect when multiple composites are presented (sets of 4) using the same design. In addition, participants are asked to provide self-reported identification strategies to examine whether participants do reference information from all four composites when making an identification.

While research has indicated that performance increases when information from multiple witnesses are presented, this has only been observed for composites that have been constructed using the same system and operator. As composites are used as investigative tools in large scale cases e.g. Jill Dando inquiry, it is likely that there may be different witnesses in different areas of the U.K. As different police forces in the U.K employ different methods of composite construction, it is highly probable that composites from multiple witnesses may be obtained using these different systems. The effect of combining information from multiple witnesses, operators and systems is therefore unclear.

A recent investigation by Frowd et al, (in press) has indicated that composites from different techniques or systems contain different types of information. This investigation evaluated performance for several commonly used composite techniques in the U.K. including E-FIT, PROfit, Forensic Sketch Artist and EvoFIT, a new holistic composite system (Hancock & Frowd, 2001). In general, it was found that the two most common systems in the UK (PROfit & E-FIT) performed equally well, while performance on identification tasks was lower for Sketches and Evo-FIT. The equivalent performance rates for E-FIT and PROfit was not surprising as the systems are extremely similar. Both systems display features in the context of a whole face and they allow for the manipulation of features and configuration. However, the type of information that is contained in the images may explain the differing levels of performance for the other techniques. While the sketches contained very detailed shape and featural information, they contained very little textural information, which has been shown to be important in face recognition (e.g. Bruce, 1991; Bruce & Langton, 1994). In contrast, Evo-FIT images contained not only holistic shape information they also contained more realistic textural information, almost of a photographic quality. However, there is very limited opportunity to manipulate individual features. Therefore, while PROfit and E-FIT contain very similar types of information, the Sketch and EvoFIT images contain differing amounts of shape and/or textural information (See figure 14, page 138 for an example).

While the aims of experiment 4, 5 and 6 are to examine the effect of combining different <u>memorial representations</u>, experiment 7 extends this by examining what impact these different <u>types</u> of information may have on the resulting morph. This experiment will therefore investigate the effect of morphing composites from different witnesses <u>and</u> different systems. More specifically, this investigation will examine whether the combination of information from both witnesses and systems will serve to increase performance for a single composite i.e. will performance for a single PROfit increase when we combine it with a different type of composite (e.g. sketch)? Experiment 8 will examine performance for 2-Morphs in more detail by asking the same witness to construct a PROfit and an EvoFIT. If an increase in performance is observed for the 2-Morphs in this condition, this will help to identify whether morphed composites perform well because they contain different memorial representations, or whether they perform well because they just contain more information.

Experiment 4: Multiple composites from multiple witnesses using PROfit

The first aim of this investigation was to examine performance for morphed composites with a more ecologically method, than that used by Bruce et al (2002, experiment 1). In Stage 1 participants were asked to view a thirty-second video clip of an unfamiliar female. They were then asked to construct a composite of this target. In Stage 2 unfamiliar participants rated all composites for likeness. In Stage 3 participants who were familiar with the targets attempted to identify the composite images. As the pool of participants who were equally familiar with the targets was quite limited (i.e. departmental staff and final year honours students), in Stage 4 a further group of unfamiliar participants were asked to undertake a six alternative forced-choice array task.

Stage 1: Composite construction and morphing procedure

Materials

Composites were constructed using PROfit (version 3.0.2) and morphed using Sierra Morph 2.5. Target faces were members of staff at the University of Stirling and were taken from existing photographic and video databases.

Participants

Sixteen adults aged between 18 and 40 years were recruited from Queen Margaret University College in Edinburgh. All participants were unfamiliar with the targets and were paid £10.

Procedure

Stage 1: This experiment used the full-face composites that were constructed during the three-quarter-view investigation in chapter 2. The construction procedure was therefore identical. Participants who were *unfamiliar* with the targets were asked to view a thirty-second video clip of one target. The operator then used cognitive interview techniques to elicit a verbal description. Each participant was then asked to construct two composites of the same target (one in a full-face view and one in a three-quarter view). Please see chapter 2, pg. 64 for a full description of procedure.

Morphing Procedure

For each target there were four full-face composites. However, the Sierra Morph software only permits pairs of images to be morphed at any one time. Therefore, the first two composites were morphed together by placing points (markers) around all of the key features on the first face (e.g. eyes, eyebrows, facial outline, hairline, nose etc). For every marker that was placed on the first image, a corresponding marker appeared on the second image. The markers on the second image were then moved so that they corresponded exactly to the position in the first image. See figure 6 for an example.

Figure 6: Example of placement of feature markers during morphing procedure

An average of 100 points was placed on each composite. Particular care was taken when the composites had long hair (as above) and extra markers were often placed around the bottom of the hairstyle. When different hairstyles are morphed together, the result can often be very blurry. However, the placement of extra markers ensured that this was kept to a minimum. Occasionally slight blurring did occur due to different hairstyles. If this occurred outside of the composite, might hinder later identification attempts (i.e. off-putting for participants), and was superfluous it was removed in Photoshop. See figure 7 for an example. Blurring was only removed if it occurred outside of the image; <u>no</u> element of the composite itself was 'retouched'.

a) With hair shadow b) Hair shadow removed

Figure 7: Morph of composites in figure 6

Each pair of composites were morphed together to produce an image that contained fifty percent of each face (e.g. figure 7). This procedure was repeated with the next two composites of the same target. The pair of 2-Morphs were then morphed together to create a final 4-Morph that contained 25 percent of each composite. See figure 8.

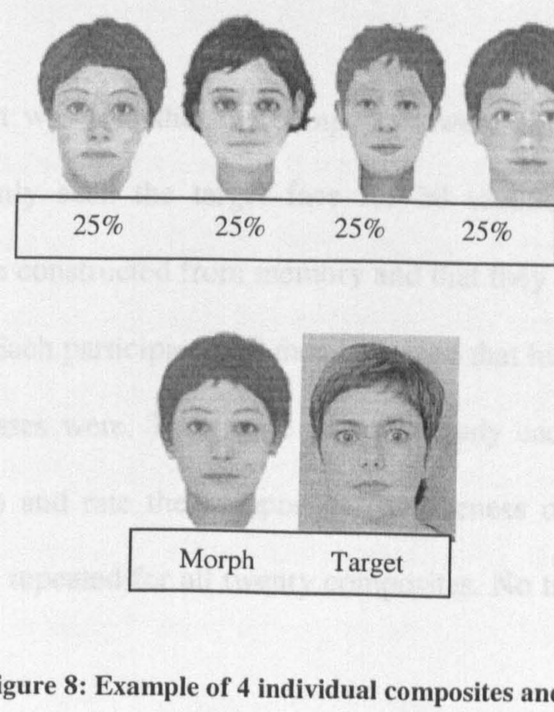


Figure 8: Example of 4 individual composites and 4-Morph

Stage 2: Likeness ratings

Participants

Twenty unpaid participants were recruited from staff members at Tesco stores in Edinburgh. All were aged between 17 and 60 years and were unfamiliar with the targets.

Design

A within-subjects design was adopted with each participant viewing all images. One presentation book was constructed which contained all twenty composites (4 individual composites and one 4-Morph for each of the targets). Composites were standardised for height (13cm) and printed on a single sheet of A4 paper (landscape). Each composite was presented alongside a full-face monochrome photograph of the target. Presentation order was randomised.

Procedure

Each participant was told that the composites were constructed after a 'participant witness' had only seen the target face for 30 seconds. It was stressed that the composites were constructed from memory and that they represented a likeness of the original target. Each participant was then informed that his or her task was to rate how good the likenesses were. They were asked to study each set of images (composite and photograph) and rate the composites for likeness on a scale of 1 (low) to 10 (high). This was repeated for all twenty composites. No time limit was placed on this procedure.

Results

From the likeness ratings it was possible to identify the worst (lowest rated) and best (highest rated) composites. The results indicated that the 4 morphs (mean likeness rating 5.4, SD = 2.0) were rated as well as the best individual composites (mean likeness rating 5.5, SD = 1.1). This supports Bruce et al (2002, Experiment 1) by indicating that morphed composites perform either as well as or better than the best single image. The overall mean likeness ratings for all of the individual composites was then calculated and a comparison was made between the morphed composites and the average of the individual composites. A 4 (target) by 2 (individual and morphed composites) analysis of variance revealed significant main effects of target (F (3,57) = 36.7, p<0.001), type of composite (F (1,19 = 38.06, p<0.001) and an interaction (F (3,57) = 7.22, p<0.01. Simple main effects analysis (Bonferroni) revealed that the mean likeness rating was higher for the morphed composites for all four targets. An analysis by items revealed that performance for the morphed composites was significantly higher for two of the targets (p<0.05).

The results from the likeness ratings therefore support previous research (e.g. McNeil et al, 1987; Bruce et al, 2002) by indicating that a single 'combined' or morphed composite performs as well as or better than the best individual composite. However, previous research (e.g. Bennet et al, 1999) has also indicated that when composites from four different 'participant witnesses' are presented together (i.e. all 4 are presented) performance is significantly better than the presentation of a single composite. The 4-Morph does appear to perform better than the 'best' single composite but how will it compare to the presentation of all 4 composites? In order to examine this, in Stage 3 participants who were familiar with the targets were asked to try to identify the 'best' single composite, the 'worst' single composite, all 4 composites and the 4-Morph.

Stage 3: Identification

Participants

Thirty final year (4th year honours) psychology students and two members of staff attempted to identify the composites. All participants were recruited from the psychology department at the University of Stirling.

Design

A 4 (target) by 4 (type of composite) within-subjects design was adopted. There were 16 composites in total (one best, one worst, one all 4 and one 4-Morph for each of the 4 targets). These sixteen composites were divided into four presentation books using a Latin square design. This ensured that each participant only saw one type of composite (best, worst, all 4 and 4-Morph) for each of the four targets. Within each book the order was rotated. Books were presented to participants who were familiar with the targets. There were a total of eight identification attempts per composite.

Procedure

Each participant was informed that the composites were constructed after a 'participant witness' had only seen the target face for 30 seconds. It was stressed that the composites were constructed from memory and that they represented a likeness of the original target. All participants were told that the composite represented someone from the psychology department. While this decreased the pool of potential targets, the total number of potential targets was still at least thirty-six. This procedure was identical to the identification stage in chapter 2 (please see pg. 62). Participants were asked to examine the composites carefully and to attempt to identify the person depicted in the composites. They were encouraged to provide a name or some identifiable semantic information. To ensure that all participants were familiar with the targets they were debriefed. If any participant was unfamiliar with any one of the targets their results were not used.

Results

The identification rates were relatively low and a Cochran's Q test revealed that performance did not differ significantly between the four conditions (X^2=3.750; df. =3 p=0.290). The highest rate was found for the condition where all four composites were presented (38% and no false positives). This was followed by the 4 morph (28% correct, 6% false positives) then the best individual composites (22% correct and 6% false positives) with the worst composite yielding the lowest identification performance (16% correct and 6% false positives). While there were no significant differences between conditions, the trend is clearly in line with the likeness ratings obtained in stage 2 and previous research (Bruce et al, 2002 Experiment 1) by indicating that the 4-morph again performs as well as the best individual composite. Performance was highest for the 'all 4' condition which supports the findings by Bennet et al (1999). However, as overall identification performance was relatively low and did not differ significantly between conditions, a further experiment was undertaken.

Stage 4: Six alternative forced-choice array task

The experiments in both stage 3 of this investigation and those in chapter 2 (experiments 1 and 2) had exhausted the number of participants who were familiar with the targets. Therefore, this experiment adopted a method that was suitable for use by unfamiliar participants - a six-alternative forced-choice task.

Materials

Each type of composite presentation (all 4, 4-Morph, best, worst) was presented with an array of six photographs. Two types of array were constructed for each target– target present and target absent. For the target present arrays, one photograph of the target was presented with photographs of five distractors. For the target absent arrays, six distractors were presented. All distractors were chosen on the basis of the verbal descriptions. As the descriptions varied greatly between participants, the distractors were matched for general hairstyle/colour, face shape and approximate age. The photographs were presented in two rows of three on a single sheet of A4 paper (landscape). All photographs were monochrome and were standardised for height (7cms), brightness and contrast. The position of distractors (target absent arrays) and target *and* distractors (target present arrays) was varied. See figure 9 for an example.

Figure 9: Example of arrays

Participants

Sixty-four participants aged between 17 and 60 years were recruited from an Open University summer school held at the University of Stirling. All participants were unfamiliar with the targets and the distractors in the arrays. All had normal or corrected vision.

Design

The same four conditions were compared (4-Morph, best, worst and all 4) for each of the four targets. The sixteen composites were presented with both a target present array and a target absent array, creating a total of thirty-two separate trials. These thirty-two composites were divided into eight different presentation books, using a Latin square design. Each book contained one type of composite (4-Morph, best, worst, all 4) for each of the four targets. In addition, half of the trials in each book contained target-present arrays and half contained target-absent arrays. Therefore, there were a total of eight different conditions (four types of composite with both target present or target absent arrays for each). Unfamiliar participants saw only one book and there were a total of eight participants per book.

Procedure

Each participant was informed that the composites were constructed from memory and represented a likeness of the target. They were asked to look carefully at the composite(s) and the accompanying array. All participants were told that the target may *or* may not be in the array. If they thought that the composite *did not* resemble anyone in the array they were instructed to say that the target was not there. Similarly, if they thought that the composite *did* resemble someone in the array, they were instructed to point to the appropriate photograph. This was repeated for all four arrays. No time limit was placed on this procedure.

Results

Composite	**4-Morph**		**All Four**		**Best**		**Worst**	
Array Type	Present	Absent	Present	Absent	Present	Absent	Present	Absent
Correct response	41	41	28	13	34	19	28	28
Incorrect rejection of array	22		9		15		9	
False positive		59		87		81		72
Miss (incorrect choice)	37		63		50		63	

Table 1: Percentage responses for each condition

Table 1 summarises the performance for the six alternative forced-choice array task. As can be seen, the 4 Morph condition produced the highest number of correct responses from both target present and absent arrays (mean 41% correct for both) and generated the lowest number of false positive choices for either type of array. The poorest performance was obtained when all four composites were presented (overall rates =21% correct, 75% false choices), while similar performance was observed for both the worst individual (28% correct, 68% false choices) and best individual (27% correct, 66% false choices) composites. A Cochran's Q test revealed that the number of correct responses for all four conditions did not differ significantly ($X^2 = 4.914$, df = 3, $p = 0.178$). While there were no overall significant differences, a further pairwise comparison was conducted in order to examine performance for the two multiple conditions (4-Morph and all 4). A McNemar test revealed that observed difference between the 4-Morph and the all-4 condition almost reached significance ($X^2 = 3.704$, $p = 0.054$). These results therefore indicate that the 4-Morph performs as well as the best individual composites and almost significantly better than the sets of 4. The incorrect responses were then condensed (misses, false positives, incorrect rejections) and the overall number of incorrect responses for each condition was compared. A Cochran Q test revealed that there were no significant differences in overall incorrect responses ($X^2 = 5.122$, df=3, $p > 0.05$).

Discussion

The results from this investigation indicate that the presentation of composites from more than one witness increases performance. In particular, the results from all stages of this investigation indicate that the 4-Morph performs as well as or better than the best single composite. Of particular interest are the results obtained in stage 3 (identification) and stage 4 (6Alt forced-choice task). While the trend in stage 3 indicated that performance was higher when all four composites were presented, the opposite trend was observed in stage 4, with all 4 composites yielding the lowest performance. This change in performance level may have reflected task demands. In stage 4, participants were presented with four different composites (set of 4) together with an array of six photographs. While the participants were told that all four composites depicted the same person, the composites often looked like four different people. Therefore it was extremely difficult for participants to both correctly match the composites to the target photograph and to correctly reject an array. This is reflected in the level of incorrect matches illustrated in table 1. However, for the identification task, which is a more ecologically valid procedure, sets of 4 composites appear to perform well.

These results indicate that performance increases when information from multiple witnesses is presented. In particular, the consistent performance of the 4-Morphs supports previous research by Bruce et al (2002, experiment 1) and indicates that the combination of different memorial representations may result in an image that is closer to the ideal image or prototype. Similarly, the results from the identification task also suggest that it may be possible to 'extract' or perceive an image that is closer to the ideal image when multiple composites are presented. Bennet et al (1999) had previously reported an increase in performance when four composites of similar quality were presented. However, this investigation has extended this research by revealing that four composites of varied quality (from worst to best) can also perform well. Importantly, as there is currently no formal method of determining which witness is likely to construct the 'best' composite, the results from this and previous investigations indicate that it may be beneficial to invite more than one witness to construct a composite.

However, while the results suggest that the presentation of information from different witnesses increases composite performance, it is unclear exactly why multiple composites perform well. While it is suggested that 4-Morphs perform well because they result in an image that may be closer to the ideal, it is unclear whether this actually occurs. When composites are morphed together the differences are 'averaged out' and the similarities are reinforced. Therefore, as the differences are 'averaged out' it is unclear how important this varied information is. Thus, it is unclear whether morphed composites perform well because they contain varied information (different memories) or whether they perform well because they just contain more information.

If morphed composites perform well because they result in an image that is closer to the ideal or prototype, then performance should be higher when memorial representations are combined. However, if morphed composites perform well because they just contain more information, then performance may be as good when composites from the same witness are combined. Experiment 5 will therefore examine performance for morphs that have been created using composites from both single and multiple witnesses.

Experiment 5: Why do morphed composites perform well?

This experiment examined performance for morphs that have been created using composites from both single and multiple witnesses. The same female targets were used. However, to ensure that the operator was blind to target identity a different operator worked with the participants to construct the composites. In the first stage of the experiment participants viewed a 30-second video clip of an unfamiliar female target. They were then asked to construct one full-face composite from memory.

Multiple images were obtained from each participant by capturing the screen at regular intervals. In the second stage, composites (sets of 4) from both single (4 from each participant) and multiple (1 from each of the 4 participants) were morphed together to create one 4-Morph for each target (from different participants) and one 4-Morph for each participant. In stage 3, a group of unfamiliar participants rated the composites for likeness. In stage 4, the composite performance was evaluated using a six-alternative forced choice array task.

Stage 1: Composite construction

Materials

Composites were constructed using PROfit (version 3.0.2) and morphed using Sierra Morph (version 2.5). Target faces and video clips were identical to those used in experiment 1 (please see chapter 2, pg. 62). In order to obtain multiple images from each participant, the 'screen capture' function in Paint Shop Pro (version 7.00) was used.

Participants

Sixteen adults aged between 18 and 45 years were recruited. They consisted of members of the public and students from other departments within Stirling University. No participant had previously constructed a composite and all participants were unfamiliar with the targets. Each participant was paid £10.

Design

Each participant was asked to view a thirty-second video clip of *one* target and then construct *one* full-face composite of the seen target. There were four targets and sixteen participants (4 per target), creating a total of sixteen full-face composites. Construction order was randomised and the experimenter worked blind to the identity of the targets. In order to obtain multiple images from individual participants, Paint Shop Pro was used to 'capture' the screen during composite construction. As construction times can vary greatly (e.g. between 20 minutes to 1 hour), the screen capture function was set at sixty seconds to ensure a sufficient number of composites from each participant. After each composite had been constructed the experimenter chose the very first composite image that had been captured and the final image. The total number of screen captures were then divided by four and the second and third images were chosen. This created a total of sixty-four composites (4 from each of the 16 participants). This procedure was identical for all sixteen participants. Participants were not present during this procedure.

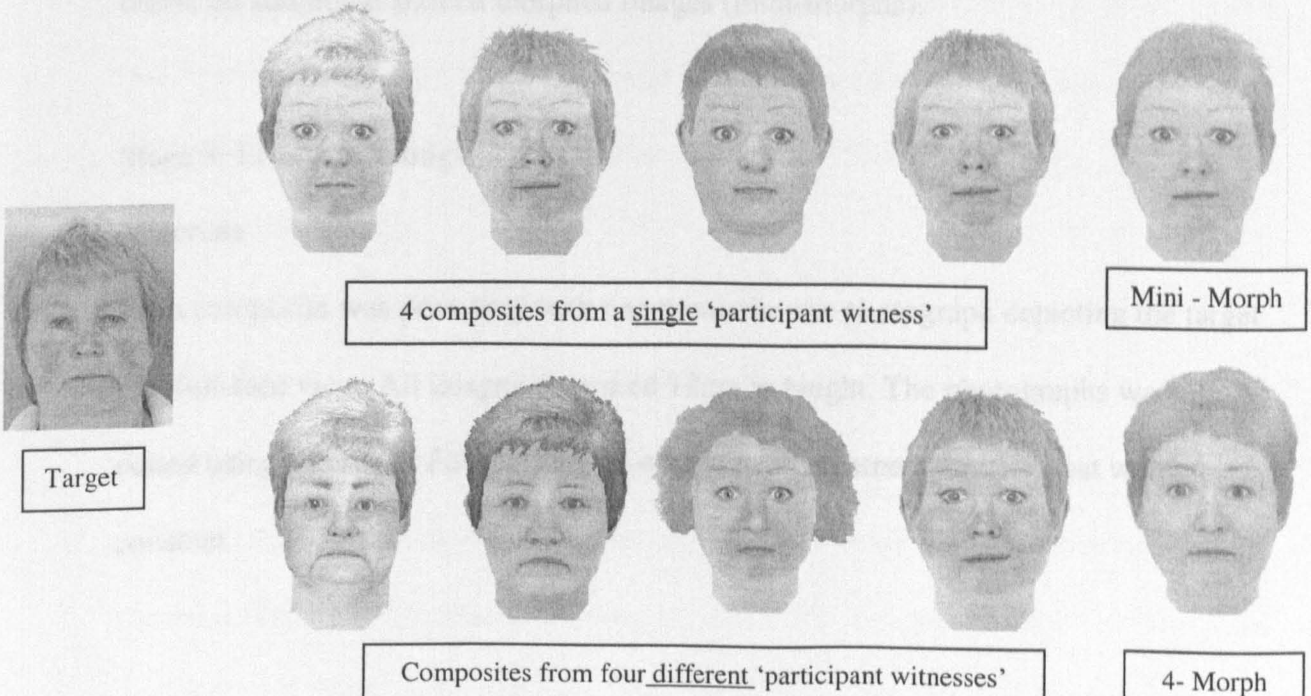


Figure 10: Example of 4-Morph and Mini-Morph

Construction procedure

The construction procedure was similar to that described in Chapter 2. Participants who were unfamiliar with the targets were asked to view a thirty-second video clip of a female target. Immediately after target exposure, the operator used cognitive interview techniques to elicit a verbal description. Each participant was then asked to construct one composite of the seen target. Please see chapter 2, pg. 64 for a full description of the interview and construction procedures.

Stage 2: Morphing procedure

The morphing procedure was identical to that used in experiment 1. For the multiple or 4-Morph condition, the four composites (i.e. those constructed by 4 different participants) were morphed together to create a 4-Morph for each target. For the single condition, the four composites from each participant were morphed together to create an additional sixteen morphed images (mini-morphs).

Stage 3: Likeness Ratings

Materials

Each composite was presented with one monochrome photograph depicting the target in a full-face view. All images measured 13cm in height. The photographs were edited using Microsoft Photo Editor to ensure that brightness and contrast were constant.

Participants

Twenty-four unpaid participants aged between 25 and 55 years were recruited from an Open University summer school based at the University of Stirling. All participants were unfamiliar with the targets.

Design

A 4 (target) by 3 (type of composite) repeated measures design was adopted. The types of composite were 4-Morphs (morphs from different participants), mini-morphs (morphs from the same participant) and single full-face composites (the final composite from each participant). There were four 4-Morphs (1 per target), sixteen mini-morphs (1 per participant and 4 per target) and sixteen single full-face composites (1 per participant and 4 per target). All thirty-six composites were randomly ordered in a single presentation booklet. Each composite was printed on a single sheet of A4 paper and displayed with one monochrome photograph of the target (printed side by side on the same sheet of A4 paper). Unfamiliar participants rated the composites for likeness on a scale from one (low) to ten (high). Presentation order was randomised.

Procedure

Each participant was told that the composites were constructed after a 'participant witness' had only seen the target face for 30 seconds. It was stressed that the composites were constructed from memory and that they represented a likeness of the original target. Each participant was then informed that his or her task was to rate how good the likenesses were. They were asked to study each set of images (composite and photograph) and rate the composites for likeness on a scale from 1 (low) to 10 (high). This was repeated for all thirty-six composites. No time limit was placed on this procedure.

Results

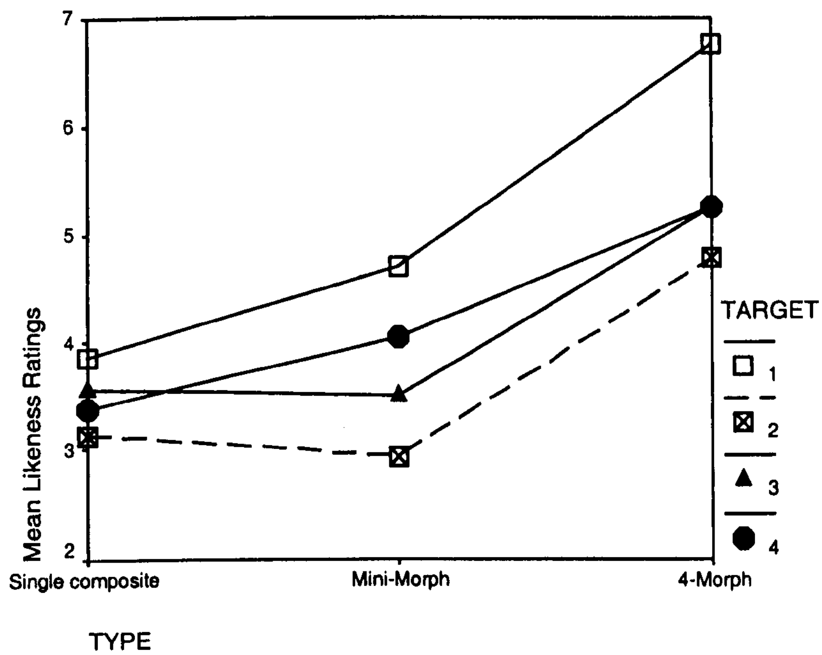


Figure 11: Mean likeness rating for each target and type of composite

The mean likeness ratings obtained in each condition of the experiment for each target are displayed in figure 11. Overall, the highest likeness ratings were obtained by the 4-Morphs (mean 5.5, S.D. 1.4), followed by the mini-morphs (mean 3.8, S.D. 1.1) while the lowest performance was observed for the single composites (mean 3.5, S.D. 1.1). A 4 (target) by 3 (type) repeated measures analysis of variance revealed a significant effect of type of composite ($F(2,46) = 22.099$ $p<0.01$), a significant effect of target ($F(3,69)=56.596$, $p< 0.01$) and an interaction ($F(6,138)=2.863$, $p < 0.05$). Further analyses using one-way repeated measures anovas and paired t-tests revealed that for all four targets the 4-Morphs performed significantly better than both the mini-morphs ($p <0.001$) and the single composites ($p <0.001$). In addition, for two of the targets the mini-morphs performed significantly better than the single composites ($p<0.001$).

These results therefore provide additional support for experiment 1 by suggesting that when information from different witnesses is combined to produce a new, single image, this image performs significantly better than a single composite. Similarly, as the 4-Morphs performed significantly better than the mini-morphs, these results also suggest that morphed composites perform well because they contain varied rather than just more information. However, the mini-morphs did perform significantly better than the single composites for two of the targets. Therefore, in order to examine performance further, the composites were evaluated with a six-alternative forced-choice array task.

Stage 4: Six-alternative forced choice array task

Materials

An identification task was not undertaken due to the limited number of participants who were familiar with the targets. Instead, a six-alternative forced choice task was undertaken. The arrays were identical to those used in experiment 1 (see figure 3 pg. 75).

Participants

Thirty-two participants aged between 25 and 55 years were recruited from an Open University summer school based at the University of Stirling. All participants were unfamiliar with the targets.

Design

From the likeness ratings it was possible to determine the highest rated (best) and lowest rated (worst) mini-morphs (i.e. morphs from a single participant) for each target. These were compared with the 4-Morphs and single composites for each target. This created a total of sixteen composite types (4 worst mini-morphs, 4 best mini morphs, 4 4-Morphs and single composites). The composites were divided into 4 presentation booklets using a Latin square design. Each booklet contained one type of composite (4-Morph, best, worst and single composite) for each of the four targets. As there were four single composites per target, presentation order was rotated within each booklet so that all four composites were presented equally. Performance for the other three conditions was then compared with the <u>average</u> of the single composites. Unfamiliar participants saw only one booklet and there were a total of eight participants per book.

Procedure

Participants were told that the composites were constructed from memory and that they represented a likeness of the target. They were asked to examine all of the images closely and were told that the target may or may not be in the array. Participants were asked to indicate whether or not they thought the target was in the array. If they thought that the composite did not resemble anyone in the array, they were asked to reject the array. If they thought the target was present, participants were asked to point to the appropriate photograph. This was repeated for all four targets. No time limit was placed on this procedure.

Results

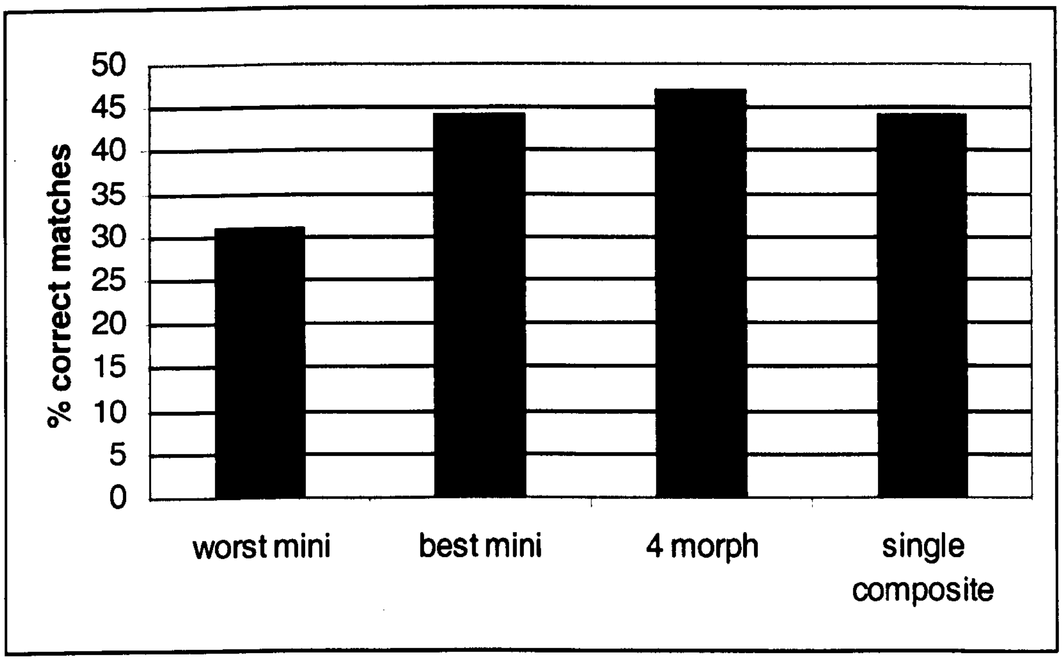


Figure 12: Mean no. of correct matches for each type of composite

Figure 12 displays the percentage correct matches for each condition. The overall percentage correct matches were 47% for the 4-Morphs, 44% for the best mini-morphs, 44% for the single composites and 31% for the worst mini-morphs. These results were low and did not differ significantly between conditions (X^2=1.227, DF=3, $p > 0.05$), however the trend clearly supports previous experiments by indicating that the 4-Morphs perform better than the single composites. Further analyses revealed there were no significant differences both within each target ($p> 0.05$) and between targets ($p<0.05$).

Therefore, while the results from the likeness ratings indicated that combining composites from a single witness may be beneficial, the results from the array task do not support this. The trend supports previous experiments by indicating that the 4-Morphs perform well. Similarly, the results indicate that combining four poor quality composites yields the lowest performance. Interestingly, the trend also suggests that combining composites from a 'good' witness does not increase performance. Therefore, even if it were possible to predict witness performance, combining four composites from the 'best' witness would not increase performance above the level achieved for an average single composite.

These results suggest that morphed composites do not perform well because we are just combining more information, i.e. more than one composite from the same witness. Even when we combine more good quality information (four composites from a good witness) performance does not increase. Instead, it seems that morphed composites perform well because of the combination of varied information (composites from different witnesses). Therefore, it appears that morphed composites perform well because they result in an image that is closer to the ideal or prototype.

In summary, experiment 5 indicated that performance increased when composites from four different witnesses were combined to create a new morphed composite. Previous research on the effect of prototypes in face recognition (e.g. Solso & McCarthy, 1981; Bruce, Doyle, Dench & Burton, 1991; Homa et al, 2001) suggested that morphed composites may perform well because they result in an image that is closer to the ideal, or prototype. Experiment 6 examined this in more detail by manipulating the information contained in the morphs. A comparison was made between morphs from single witnesses (more information) and morphs from multiple witnesses (varied information). The results indicated that overall the morphs from multiple witnesses performed better than the morph from single witnesses. This indicates that morphed composites perform well because they contain different memorial representations. This combination of memorial representations therefore appears to have resulted in an image that was closer to the ideal or prototype.

Experiment 4 also reported an increase in performance when multiple composites were presented (sets of 4). Research (e.g. Webster & MacLin, 1999; Leopold et al, 2001) has indicated that the presentation of one 'distorted' face can alter the perception of a second face. This suggests that when more than one composite is presented, each composite (or deviation) may have an effect on the perception of subsequent composites. In particular, the investigation by Leopold et al (2001) suggests that this perceptual bias may result in the perception of an image that is closer to the ideal or prototype. While the results from experiment 5 indicate that morphed composites perform well because they contain varied information, it is unclear why multiple composites perform well. Experiment 6 will therefore examine this by replicating experiment 2 with multiple rather than morphed composites. Sets of 4 composites will be presented from both single (more information) and multiple (varied information) witnesses. If it is possible to 'extract' an image that is closer to the ideal or prototype, then performance should be higher when composites from multiple witnesses are presented. In addition, even if performance is higher for the multiple condition, it is extremely difficult to know whether participants are utilising information from all composites and what kind of information they are using to make an identification. As each set of 4 will invariably contain both good and poor quality composites, participants may only focus on one or two 'good' quality composites, rather than using information from all four. Indeed, as each set of 4 will contain one or two poor quality composites, utilising information from all four may hinder rather than facilitate identification performance. Therefore, after participants have made an identification attempt they will be asked to report how they reached their decision. A comparison will then be made between the number of composites participants used and identification accuracy.

Experiment 6: Why do sets of 4 composites perform well?

This experiment replicated experiment 5 with multiple rather than morphed composites. Sets of 4 composites from both single (more information) and multiple (varied information) witnesses were compared with single composites. In stage 1, each 'participant witness' was asked to view a thirty-second video clip of one male target. They were then asked to construct one full-face composite of the target. Paint Shop Pro was used to 'capture' multiple images from each 'participant witness' during construction. All individual (final) composites were initially rated for likeness in order to determine which 'participant witness' had constructed the 'best' (highest rated) composite and which 'participant witness' had constructed the 'worst' composite (lowest rated) for each of the four targets. The sets of 4 composites from the 'best' and 'worst' participant witnesses were then compared with the sets of 4 composites from four different participants and the single full-face composites. In stage 2, participants who were familiar with the targets attempted to identify the composites. Where an identification attempt was made, participants were asked to report how they reached their decision i.e. how many composites did they base their decision on?

Stage 1: Construction of composites

Materials

Targets were male members of staff from the Department of Psychology, University of Stirling. To ensure that the experimenter was blind to target identity, images were chosen by an independent researcher and were selected from existing photographic and video databases. Video frames were extracted and digitised without sound, using the Media 100 video-editing package. A thirty-second video clip was created for each target. Each clip consisted of fifteen seconds of movement (rotating in chair from left to right: shaking head from side to side, nodding up and down) and fifteen seconds of full-face view.

Composites were constructed using PROfit (Windows version 3.0) on an ASUS Hi-Grade UltiNote AS8400 laptop computer. Paint Shop Pro (version 7) was used to 'capture' the screen during composite construction.

Participants

Sixteen adults aged between eighteen and forty years were recruited from Queen Margaret University College, Edinburgh. All participants were unfamiliar with the targets. Each participant received a £10 payment.

Design

Each participant was asked to view a thirty-second video clip of *one* target and then construct *one* full-face composite of the seen target. There were four targets and sixteen participants (4 per target), creating a total of sixteen full-face composites. Construction order was randomised and the experimenter worked blind to the identity of the targets. In order to obtain multiple images from individual participants, Paint Shop Pro was used to 'capture' the screen during composite construction. As construction times can vary greatly (e.g. between 20 minutes to 1 hour), the screen capture function was set at sixty seconds to ensure a sufficient number of composites from each participant. After each composite had been constructed the experimenter chose the very first composite image that had been captured and the final image. The total number of screen captures were then divided by four and the second and third images were chosen. This created a total of sixty-four composites (4 from each of the 16 participants). This procedure was identical for all sixteen participants. Participants were not present during this procedure.

Construction procedure

The procedure for composite construction was identical to that used in both experiment 1 and chapter 2 (please see pg. 64 for a full description).

Stage 2: Identification

All individual composites (i.e. the final composites from each 'participant witness') were initially rated for likeness. This permitted a comparison between sets of 4 composites from the 'best' witness, 'worst' witness and all four witnesses for each target.

Participants

Thirty staff members and two senior students attempted to identify the composites. All participants were recruited from the psychology department at the University of Stirling and were familiar with the targets.

Design

A 4 (target) by 4 (type of composite) within-subjects design was adopted. The types of composite were sets of 4 from the best and worst participants, sets of 4 from all four participants and single full-face composites (i.e. the final composite from each participant) for each of the four targets. This created a total of sixteen different types of composite (four 'best' sets of 4, four 'worst' sets of 4, four sets of 4 from different participants and 4 single full-face composites). The composites were divided into four presentation books using a Latin square design. This ensured that each participant only saw *one* type of composite presentation (best, worst, multiple and single full-face) for each of the four targets. There were a total of eight identification attempts per book. In addition, as there were four single composites for each target (i.e. one from each of the four participants per target), presentation order was rotated such that there were two identification attempts for each of the four single composites. This permitted a comparison between the average of the single composites with the three multiple conditions. Within each book the order of presentation was rotated and all books were presented to participants who were familiar with the targets.

Procedure

Each participant was informed that the composites were constructed after a 'participant witness' had only seen the target face for 30 seconds. It was stressed that the composites were constructed from memory and that they represented a likeness of the original target. It was also stressed that all four composites represented the same person. Participants were informed that each composite represented someone who was personally familiar to them. Participants were asked to examine the composites carefully and to attempt to identify the person depicted in the composites. They were encouraged to provide a name or some identifiable semantic information. In addition, after each identification attempt participants were encouraged to explain how they reached their decision i.e. they were asked 'what information did you use to reach your decision'. No suggestions were offered by the experimenter. To ensure that all participants were familiar with the targets they were debriefed. If any participant was unfamiliar with any *one* of the targets their results were not used.

Results

Identification rates were high with the sets of 4 from four different participants achieving the highest performance (63% correct identifications) with the lowest number of false positives (15%). This was followed by the sets of 4 from the 'best' witnesses (53% correct identifications and 25% false positives). Equivalent rates were observed for the sets of 4 from the 'worst' participants (38% correct identifications and 28% false positives) and the single full-face condition (38% correct identifications and 25% false positives). The data was initially collapsed across target and a Cochran Q test was conducted on the hit rate. This revealed that there were no significant differences ($X^2=5.243$, $DF=3$, $p >0.05$). There were also no significant differences in false positives ($X^2=1.543$, $DF=3$, $p >0.05$). However, further analyses on the hit rate for both type and target revealed significant effects for the same targets ($X^2=84.860$, $DF=15$, $p <0.01$). Target 4 was equally well identified in all conditions (ranging from 75% to 100%), whereas target 3 did not obtain any correct identifications in any condition. Performance for target 3 (a lecturer within the psychology department) was unfortunate and may have been affected by a period of absence from the department during this investigation. Target 4 achieved extremely high identification rates across all conditions and may have been the result of a particularly distinctive hairstyle. For the remaining two targets the sets of 4 composites from different participants performed significantly better than the other three conditions ($p<0.05$). Therefore, for three out of four targets the sets of 4 from different participants performed as well as or better than both the single composites and the single sets of 4 (i.e. those from both the 'best' and 'worst' witnesses).

The results therefore suggest that the presentation of multiple composites from four different witnesses yield higher performance levels than both multiple composites from single witnesses and single (individual) composites. This provides additional support for experiment 4 by indicating that composites from four different witnesses perform well. The results are also similar to those obtained in experiment 5 and may indicate that multiple composites perform well because they contain different memorial representations, not just more information. The results from experiment 5 indicated that morphed composites perform well because they resulted in an image that was closer to the ideal or prototype. However, in order to determine whether participants do in fact reference information from multiple (all four) composites and perhaps 'extract' an image that is closer to the ideal or prototype, an examination of self-reported identification strategies was undertaken.

After a participant had made an identification they were asked to report (if they could) how they reached their decision (i.e. how many composites did they base their decision on?). See table 2 for the percentage number of composites used for each of the three multiple conditions (best set of 4, worst set of 4 and 4 from four different witnesses).

%	'Best' sets of 4	'Worst' sets of 4	Multiple sets of 4
1 face	**55**	**35**	**31**
2 faces	7	17	21
3 faces	4	0	3
All 4	3	17	21
None	31	31	24
Total	100%	100%	100%

Table 2 : No of composites participants used to make an identification

Table 2 illustrates the percentage number of composites used by participants to make an identification. As can be seen, the largest percentage of participants in all conditions reported that they based their decision on one composite, rather than using information from all four composites. The number of composites used for each set of 4 was analysed (Cochran's Q) and the results revealed significant differences for the best sets of 4 (X^2=32.400 DF=3, $p<0.01$). Further pairwise comparisons (Mcnemar) revealed that significantly more participants based their identification on just one composite, compared to either two, three or all four composites ($p<0.01$ for all). For the worst sets of 4, significantly more participants reported using one composite compared to three composites ($p<0.05$). For the multiple sets of 4 (from 4 different witnesses) there were no significant differences although the trend is in the same direction with more participants using one rather than multiple conditions to make an identification. However, multiple images seem to be used more often when different witnesses have constructed all four.

These results suggest that when participants are presented with four composites of the same target, they tend to prefer to make an identification decision using just one composite. However, these results are based on self-reported strategies and in order to assess the effectiveness of multiple composites, it is essential to examine the number of faces used to make an identification attempt with the accuracy of these identification attempts. To examine this the data was initially collapsed across type (sets of 4) and the number of correct identifications for each reported strategy (e.g. the no. of composites participants used) was compared, see figure 13.

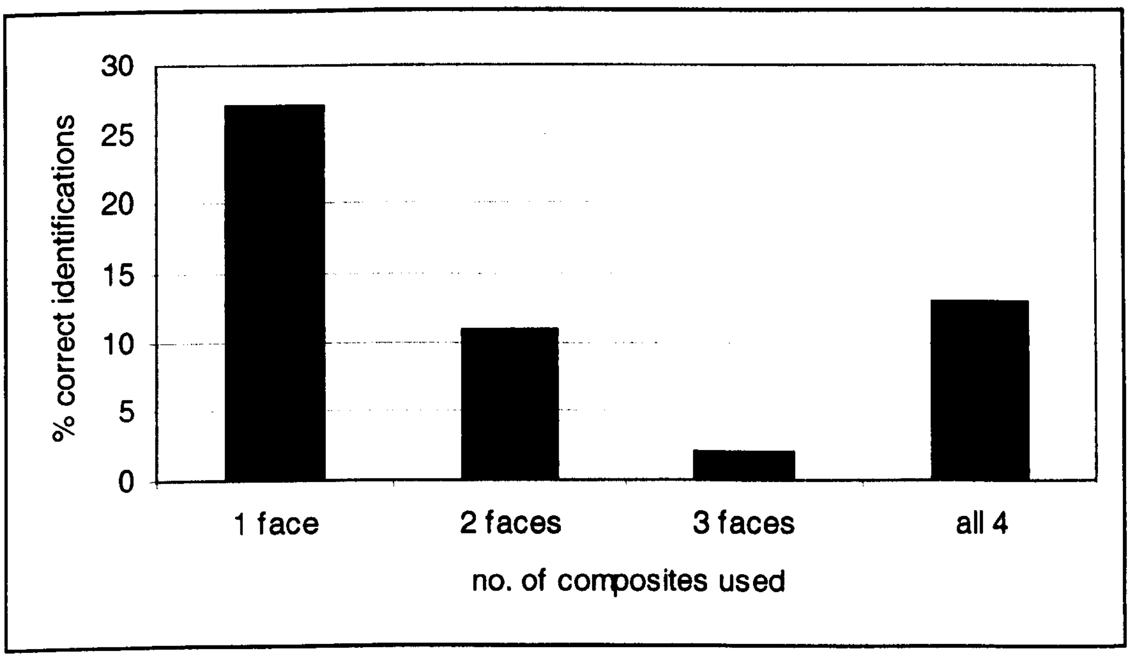


Figure 13: No of correct identifications for each number of composites used to make an identification

A Friedman test revealed significant overall differences (X^2=18.113, DF=3 P<0.01). Further pairwise comparisons using Wilcoxon signed rank tests revealed that there were significantly more correct identifications when only one face was used compared to both 2 faces (p<0.05), 3 faces (p<0.01) and all 4 faces (p<0.05). Further analyses on each set of 4 using Cochran's Q tests revealed significant differences for the best sets of 4 (X^2=21.429, DF=3 P<0.01). There were significantly more correct identifications when one face was used compared to both 2 faces ($p<0.01$), 3 faces ($p<0.01$) and all 4 faces ($p<0.01$). There were no significant differences in accuracy for the worst sets of 4. For the multiple sets of 4 (from 4 different witnesses) there were more correct identifications when one face was used compared to three faces ($p<0.05$).

These results indicate that when participants are presented with four composites from the 'best' witness, they are more accurate when they base their decision on one composite (the final good quality image). This supports the results from the identification stage and suggests that adding more information from one 'good' quality witness does not serve to increase performance. Similarly, adding more information from a 'poor' witness (worst) does not increase performance. While there was no relationship between the number of composites used to make an identification and identification accuracy, a large percentage of participants preferred to base their identification on one composite rather than all four. These results are similar to those obtained in experiment 5 and suggest that either morphing or presenting multiple composites from a single witness does not serve to increase performance.

The results from experiment 5 indicated that the combination of memorial representations in the 4-Morphs resulted in an image that was closer to the ideal, or prototype. However, the results from this experiment provide little evidence for this effect when multiple composites are presented. Thirty-one percent of participants reported basing their identification on one composite, twenty-one percent used two composites and twenty-one percent used all four composites. Similarly, when the number of faces used were compared with identification accuracy there were no significant differences. If participants were referencing information from all four composites and perhaps perceiving an image that was closer to the ideal or prototype, then identification performance should have been higher when participants used all four composites to make an identification. However, there were no differences in accuracy for participants who reported using two composites compared with four. Therefore, while sets of 4 composites from four different witnesses perform well, there is little evidence for any prototype effect. More importantly, participants seem to prefer to use only <u>one</u> composite to make an identification.

In summary, the results from experiments 4, 5 and 6 have indicated that identification performance increases when information from multiple witnesses is presented. The results from experiment 4 and previous research (Bruce et al, 2002, experiment 1) indicated that morphed composites performed as well as or better than the best single composites. Experiment 5 demonstrated that morphed composites from different witnesses performed better than morphed composites from single witnesses, thus suggesting that 4-Morphs result in an image that it closer to the ideal or prototype. The same trend was observed in experiment 6 for multiple composites, however there was little evidence for any prototype effect and self-reported identification strategies revealed that participants preferred to make an identification attempt based on a <u>single</u> composite. These results suggest that while multiple composites perform well, the information that is obtained from multiple witnesses (i.e. different memorial representations) may be more effective when they are combined to form a single morphed image.

These results therefore indicate that morphed composites may be a more effective investigative tool. However, one potential difficulty with the previous experiments is that they have only examined performance for morphs when the composites have been constructed using the same composite system (PROfit). As composites are generally used in large-scale cases e.g. murder inquiries, several police forces may be involved. As different police forces in the U.K employ different methods of composite construction it is highly probable that composites from multiple witnesses may be obtained using these different systems. Experiments 4 and 5 have indicated that morphed composites perform well because they contain varied information in the form of different memorial representations. However, it is unclear whether other types of varied information may increase performance. Experiment 7 will therefore examine performance for morphed composites that have been constructed using techniques that produce very different images (e.g. Sketch artist, EvoFIT, PROfit). The first aim will be to replicate the 4-Morph effect using composites from different systems and different witnesses. The second aim is to examine whether the combination of information from both witnesses and systems will serve to increase performance of a single composite i.e. will performance for a single PROfit increase when we combine it with a different type of composite (Sketch or EvoFIT)?

Experiment 7: Composites from multiple witnesses and systems

This experiment examined performance for morphed composites that had been constructed using different techniques. The results from research in our own laboratory (Frowd et al, submitted) have indicated that composites from different techniques contain different types of information. Frowd et al (submitted) compared performance for several commonly used composite techniques i.e. E-FIT, PROfit and

Forensic Sketch Artist, together with the older manual Photo-Fit system and EvoFIT, a new holistic composite system (Hancock & Frowd, date). In general, it was found that the two most common systems in the UK (PROfit & E-FIT) performed equally well, while performance on identification tasks was lower for Photo-Fit, Sketches and Evo-FIT. The equivalent performance rates for E-FIT and PROfit was not surprising as the systems are extremely similar. Both systems display features in the context of a whole face, they allow for the manipulation of features and configuration and they both contain similar numbers of features. More importantly, both contain similar shape and texture information.

Similarly, while the sketches contained very detailed shape and featural information, they contained very little textural information, which has been shown to be important in face recognition (e.g. Bruce, 1991; Bruce & Langton, 1994). In contrast, Evo-FIT images contained not only holistic shape information they also contained more realistic textural information, almost of a photographic quality. However, there is very limited opportunity to manipulate individual features.

Therefore, while PROfit and E-FIT contain very similar types of information, the Sketch and EvoFIT images contain differing amounts of shape and/or texture. See figure 14 for an example.

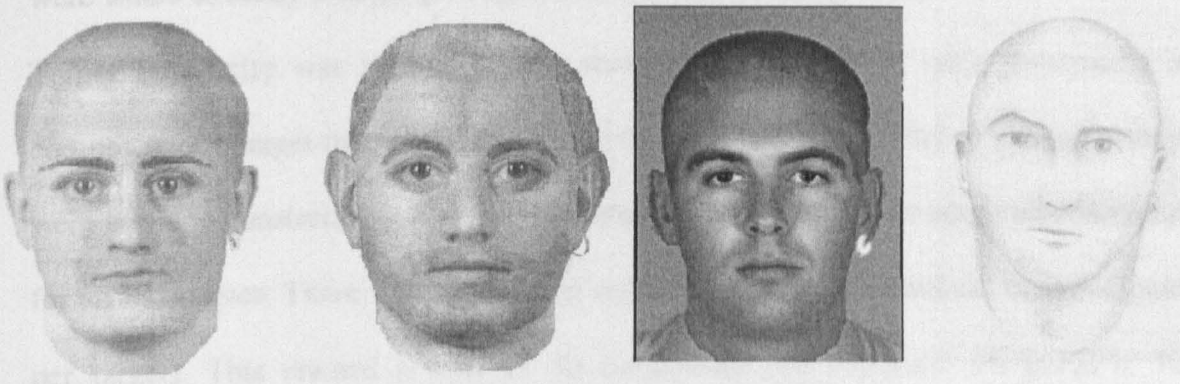


Figure 14: Composites of Andre Agassi from 4 different construction techniques

The composites in figure 14 contain both different memorial representations (as they are constructed by different witnesses) and different types of information. Previous experiments have indicated that morphed composites perform well because they contain varied information. So far, this varied information has consisted of different memorial representations, however it is unclear whether other types of varied information may also facilitate performance for morphed composites. This experiment will therefore investigate the effect of morphing composites from different witnesses and different systems. More specifically, this investigation will attempt to replicate the 4-Morph effect. It will also examine whether the combination of information from both witnesses and systems will serve to increase performance for a single composite i.e. will performance for a single PROfit increase when we combine it with a different type of composite (Sketch or EvoFIT)?

Identification

Materials

This experiment used the PROfits, E-FITs, EvoFITs and Sketches that had been constructed for the Frowd et al (submitted) study. In this investigation participants were asked to study a target photograph of a young famous white male for 1 minute. Target familiarity was pre-checked to ensure that participants only constructed a composite of a target that was unfamiliar to them. Following a delay of 3-4 hours they were asked to construct one composite of the seen target. This procedure was identical for all techniques. There were ten targets and ten 'participant witnesses' (1 participant per target). This created a total of 40 composites (10 PROfit's, 10 E-FIT's, 10 Sketches and 10 EvoFIT's all constructed by different 'participant witnesses'). Target photographs were obtained from various famous face databases available on the Internet. All composites were morphed using the Sierra Morph software. The morphing procedure was identical to that used in Experiment 4.

Participants

Forty participants aged 18 – 50 years were recruited. Participants were students from Stirling University and staff members from Tesco stores in Edinburgh. All participants were familiar with the targets.

Design

PROfit and E-FIT contain very similar types of information and as Bruce et al (2002, experiment 1) had previously found an advantage for morphing two PROfits these composites were not combined with each other. As the main aim of this investigation was to examine the effect of combining different types of images, two separate experiments were conducted. The first experiment compared performance for a single PROfit, a morph containing the PROfit and Sketch, a morph containing the PROfit and Evo-FIT and a morph containing all 4 composites (4-Morph) for each of the 10 targets. This created a total of 40 composites (4 types of composite for each target).

These 40 composites were divided into five presentation booklets using a Latin square design. Each booklet was balanced for condition and rotated around target. Therefore, each booklet contained 2 single PROfits, 2 PROfit/Sketch morphs, 2 PROfit/EvoFIT morphs and 2 4-Morphs. Each participant was presented with one booklet. There were a total of 8 participants per booklet.

Figure 15: Example of morphed composites of Andre Agassi (originals can be seen in Figure 14)

Procedure

Each participant was told that they would be asked to look at eight individual composites. They were informed that the composites were constructed after a ‘participant witness’ had only seen the target face for 1 minute. It was also stressed that the composites were constructed from memory and that they represented a likeness of the original target. The participants were not informed that the composites were constructed using different systems *or* that some of the composites had been morphed together. However, as the texture of the EvoFIT images often resulted in a considerably darker final morph, the participants were told that the composites depicted white males. All participants were told that the composites depicted famous men and that their task was to examine the composites carefully and to attempt to identify the person depicted in the composites. They were encouraged to provide a name or some identifiable semantic information. At the end of the procedure each participant was asked to identify the targets from the original photographs. If any participant was unfamiliar with any *one* of the targets their results were not used.

Results

The highest performance was obtained by the 4-Morphs (31% correct identifications, 17% false positives), followed by the PROfit/EvoFIT morph (22% correct identifications, 10% false positives), with almost identical performance for both the single PROfit condition (15% correct identifications, 17% false positives) and the PROfit/Sketch morph (15% correct identifications, 11% false positives). A Friedman test was conducted on the hit rate. This revealed that the overall differences in performance were almost significant (X^2=7.686, DF=3, p = 0.053). Further pairwise comparison using Wilcoxon tests revealed that the 4-Morph performed significantly better than both the single PROfit and the PROfit/Sketch morph ($p< 0.05$). Similarly, while the PROfit/EvoFIT morph appeared to perform better than both the single PROfit and PROfit/Sketch conditions, this difference did not quite reach significance ($p = 0.08$ and 0.09 respectively). False positives did not differ significantly between conditions (X^2=2.599, DF=3, p = 0.458).

These results suggest that when composites from 4 different 'participant witnesses' are combined, performance is significantly better than presenting a single composite. However, the results also provide support for combining different <u>types</u> of information. While there was no benefit for adding a sketch (adding shape information), there was a slight benefit for adding an EvoFIT (textural information).

In order to ensure that the benefit for the 4-Morph was not due to the presence of the E-FIT, this experiment was repeated with E-FIT as the 'base' composite. The design and procedure were identical to the previous experiment. Performance was compared for a single E-FIT, an E-FIT/Sketch morph, an E-FIT/EvoFIT morph and the same 4-Morph for each target. Forty participants were recruited from Stirling University and were asked to identify the composites. All participants were familiar with the targets.

Results

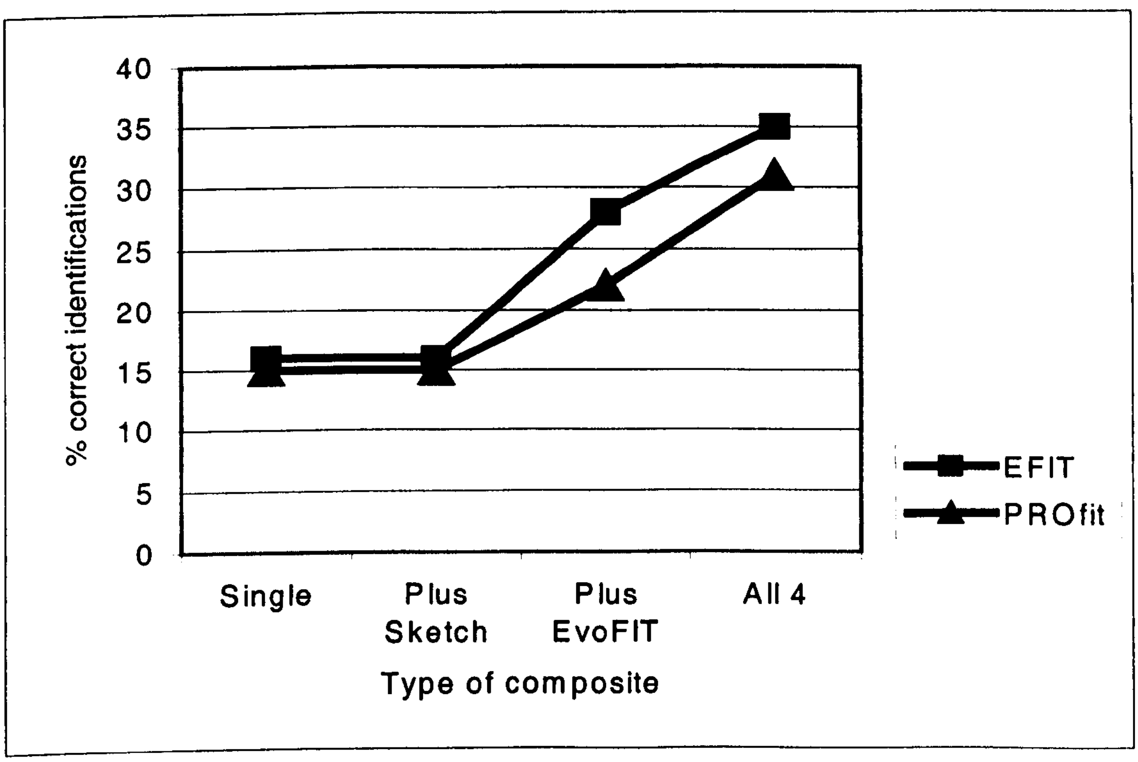


Figure 16: Performance for morphed composites for both E-FIT and PROfit

As can be seen from figure 16 the same effect was found. The highest performance was obtained for the 4-Morph condition (35% correct identifications and 22% false positives), followed by the E-FIT/EvoFIT morph (28% correct identifications and 15% false positives) while similar performance was again obtained for the single E-FIT (16% correct identifications and 38% false positives) and E-FIT/Sketch (16% correct identifications and 25% false positives) conditions. A Friedman test was conducted on the hit rate and this revealed significant differences (X=7.971, DF=3, $p < 0.05$). Further pairwise comparisons using Wilcoxon signed ranks tests revealed that again the 4-Morph performed significantly better than both the single ($p < 0.05$) and E-FIT/Sketch ($p < 0.05$) conditions. The E-FIT/EvoFIT morphs also performed significantly better than the E-FIT/Sketch morphs ($p < 0.05$) and almost significantly better than the single E-FIT's ($p = 0.07$). False positives did not differ significantly between conditions (X=5.105, DF=3, $p = 0.164$). An analysis by items revealed that for half (5) of the targets the 4-Morphs performed as well as the other conditions and for the other half the 4-Morphs performed significantly better than the other conditions ($p<0.05$).

These results therefore indicate that combining different <u>types</u> of information from both multiple witnesses and multiple systems increases performance. In particular, these results provide further support for experiment 4 and previous research (Bruce et al, 2002, experiment 1) by revealing that composites from four different witnesses increases performance above the level achieved for a single composite.

For the 2-Morphs, (or pairs of composites) there was no benefit for adding a sketch. As these composites were constructed by two different witnesses, this finding is contrary to the reported findings by Bruce e al (2002, experiment 1) who reported an increase in performance for 2-Morphs. This may reflect the type of information that is contained in a sketch and suggests that adding featural/shape information does not serve to increase performance. However, the results suggest that performance does increase when an Evo-FIT image is added. Evo-Fits obviously contain shape information, but it is likely, due to the poor performance of the sketch morphs that the benefit for adding an Evo-FIT reflects a benefit for adding textural information, which has been shown to be important in face recognition (e.g. Bruce, 1991; Bruce & Langton, 1994).

The results so far indicate that when combining multiple composites from the same system (i.e. PROfit) there is a significant increase in performance when different witnesses have constructed the composites. This suggests that the combination of different memorial representations in the 4-Morph results in an image that is closer to the ideal or prototype. Additionally, the results from experiment 6 indicate that this 'prototypical' effect only occurs when composites are morphed together. However, the results from experiment 7 indicate that performance may increase when other types of information are combined (e.g. textural information). The results suggest an increase in performance when a PROfit is combined with an EvoFIT. However, as this experiment did not specifically separate composite information (e.g. texture) from memorial information, it is unclear whether the increase in performance was due to the combination of different memories or the combination of different types of information (e.g. shape from PROfit and texture from Evo-FIT). Experiment 8 will investigate this by asking the same witness to construct both an EvoFIT and a PROfit. Any increase in performance here would suggest a benefit for combining different types of information. However if there is no increase, then the slight benefit observed in experiment 7 for combining a PROfit and an EvoFIT may be due to the combination of memorial representations rather than different types of information.

Experiment 8: Multiple systems with the same witness

The aim of this experiment was to examine performance for morphed composites that had been constructed by the same witness using different systems. This investigation was conducted in conjunction with a larger study (Frowd et al, submitted), which evaluated several different composites techniques (e.g. PROfit, E-FIT, Faces, EvoFIT, Sketch Artist) with a more forensically relevant delay of two days between target exposure and composite construction. Unfamiliar participants were asked to study a photograph of a famous male for one minute. Two days later they were asked to describe the face (cognitive interview) and construct two composites of the target, one PROfit and one EvoFIT (Stage 1). In stage 2, unfamiliar participants rated the composites for likeness. In stage 3, a further group of unfamiliar participants attempted to 'pick out' the target from an array (6 alternative forced-choice task).

Stage 1: Composite construction and morphing procedure

Materials

Full-face target photographs were obtained from various famous face databases available on the Internet. All composites were constructed using PROfit version 3 and EvoFIT. Composites were morphed using PROmorph, a new morphing programme, developed by Frowd (2002). PROmorph is similar to Sierra Morph (see pages 103-105 for a description) with the exception that PROmorph can morph several images at any one time, whereas Sierra Morph can only morph pairs of images.

Participants

Ten participants were recruited from Stirling University. All participants were screened to ensure that they were unfamiliar with at least one target. Each participant was paid £15.

Design

A 10 (target) by 2 (composite technique) mixed design was adopted, with target as a between-subjects factor and technique (PROfit and EvoFIT) as a within-subjects factor. Each participant viewed a full-face photograph of one target for one minute. Two days later they were asked to construct two composites of the same target (one PROfit and one EvoFIT) from memory. Order of construction was counterbalanced such that five participants constructed a PROfit first and five constructed an EvoFIT first. There were ten targets and ten participants (1 per target) creating a total of twenty composites (1 PROfit and 1 EvoFIT per target). See figure 17 for an example.

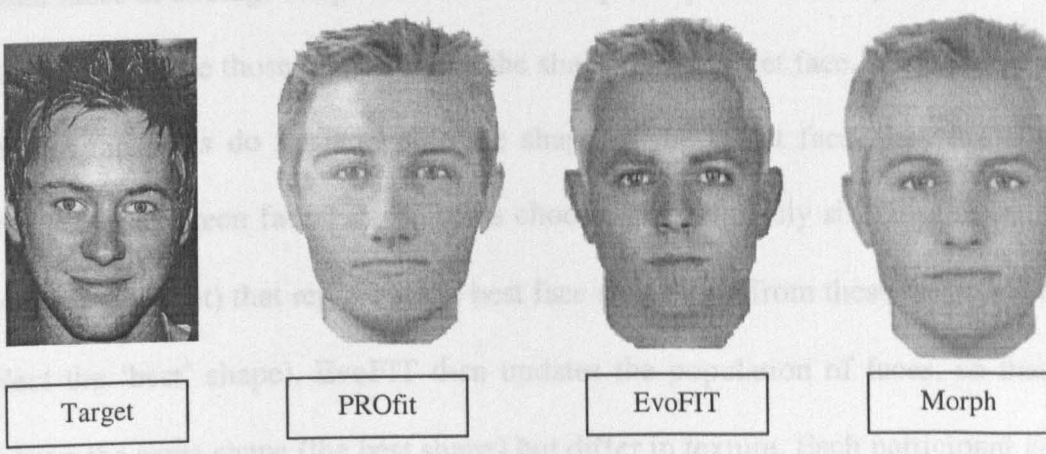


Figure 17: Example of PROfit/EvoFIT morph from a single 'participant witness'

Construction Procedure

A verbal description was initially elicited from each participant (using cognitive interview techniques) prior to construction. The cognitive interview procedure was identical to that used in previous experiments (Please see chapter 2, pg. 64 for a full description). Each participant was then asked to construct two composites (one PROfit and one EvoFIT). The procedure for constructing a PROfit composite was identical to that described in chapter 2 (please see pg. 64 for a description).

EvoFIT Construction Procedure

In order to construct an EvoFIT image, participants initially select a hairstyle from the PROfit database. This hairstyle is then applied to one EvoFIT image using editing tools available in Photoshop. The operator then asks EvoFIT to automatically update the whole database, so that each face is displayed with the same hairstyle. To construct an EvoFIT image, each participant is presented with a screen displaying eighteen faces of average shape and texture. The participant is initially asked to study the faces and choose those that resemble the shape of the target face. If the participant states that the faces do not resemble the shape of the target face, they are shown another set of eighteen faces. Participants choose approximately six faces (often one or two from each set) that represent the best face shape (and from these they are asked to select the 'best' shape). EvoFIT then updates the population of faces, so that all faces have the same shape (the best shape) but differ in texture. Each participant again selects six faces with the 'best' texture and from these are asked to pick the best one. The best shapes and textures are then presented and the participant is asked to pick the 'best' face (closest representation of target). It is then possible to manipulate the face and the size and shape of individual features using an integral morphing programme.

EvoFIT then 'breeds' together the information from the shape and texture choices to create a new set of faces. This process can continue for up to six generations (average of 4 generations in this investigation) and can take between 2 to 3 hours. Due to the length of this procedure, participants were given a short break (5 to 10 minutes) before the construction of the second composite.

Each pair of composites (i.e. the PROfit and EvoFIT composites that had been constructed by the same participant) were morphed together using PROmorph. Participants were not present during this procedure.

Stage 2: Likeness Ratings

Materials

All composites were presented with a monochrome full-face photograph of the target. All images measured 13cm in height. The photographs were edited using Microsoft Photo Editor to ensure that brightness and contrast were constant.

Participants

Twenty participants aged 25 to 55 years were recruited from an Open University summer school based at the University of Stirling. All participants were unfamiliar with the targets.

Design

There were three composites per target (1 PROfit, 1 EvoFIT and 1 morph). As there were 10 targets, this resulted in a total of thirty composites (10 PROfits, 10 EvoFITs and 10 morphs). Each composite was printed on a single sheet of A4 paper and displayed with one monochrome full-face photograph of the target (on the same sheet of A4 paper). All thirty composites were randomly ordered in one presentation booklet. Unfamiliar participants rated all composites for likeness on a scale from one (low) to ten (high). Presentation order was randomised.

Results

Equivalent rates were observed for both the EvoFIT composites (mean 3.7 S.D. 1.2) and the morphed composites (mean 3.7, S.D 1.3), while the PROfit composites yielded the lowest performance (mean 3.1 S.D 1). A 3 (type) by 10 (target) repeated measures analysis of variance revealed a significant effect of type of composite ($F (2,38) = 9.805$, $p < 0.01$) an effect of target ($F (9,171) = 10.196$, $p < 0.01$) and an interaction ($F (18,342) = 10.329$, $p < 0.01$). Further paired samples t-tests revealed that performance for the PROfits was significantly lower than both the EvoFIT ($p<0.05$) and morphed composites ($p<0.05$). Further analyses using one-way repeated measures anovas and paired samples t-tests revealed that for seven targets the morphed composites performed better than the PROfits (significantly so for 5 targets, $p<0.05$). Similarly, for five of the targets the EvoFITs performed significantly better than the PROfits ($p<0.05$). Therefore, the lowest performance was observed for the single PROfits, while equivalent performance levels were obtained for the EvoFITs and morphs. This indicates that there is no benefit for combining a PROfit and an EvoFIT when the <u>same</u> witness has constructed both composites and suggests that the benefit observed in experiment 4 was primarily a result of combined memorial representations, rather than combined shape and textural information. In order to examine this further the composites were evaluated using a six-alternative forced-choice array task.

Stage 3: Six-alternative forced-choice array task

Materials

To create the arrays, five distractors were chosen for each of the four target faces. These were chosen on the basis of the verbal descriptions. These descriptions were often extensive and varied, so distractors were matched only on information concerning hair, face shape and age. These were presented with the target as black and white photographs on a single sheet of A4 paper. Microsoft Photo Editor was used to ensure that brightness and contrast was consistent.

Participants

Forty participants were recruited from an Open University summer school based at the University of Stirling. All participants were aged between 32 and 56 years of age and were unfamiliar with the targets.

Design

There were thirty composites in total (1 PROfit, 1 EvoFIT and 1 Morph for each of the ten targets). To ensure that participants saw only one type of composite for each target, all composites were divided into five separate presentation booklets using a Latin square design. The booklets were balanced for condition (i.e. they contained 2 PROfits, 2 EvoFITs and 2 Morphs) and rotated around target. Each composite was presented with a target present array. Unfamiliar participants were asked to try to 'pick out' the target from the array. There were a total of eight attempts per booklet.

Procedure

Participants were told that the composites were constructed from memory and that they represented a likeness of the original target. Participants were asked to examine the images closely and were told that the target may or may not be in the array. They were asked to indicate whether or not they thought the target was in the array. If they thought the target was present, participants were asked to point to the appropriate photograph. This procedure was repeated for all six targets.

Results

The morphed composites achieved 27% correct matches, followed by the EvoFIT composites (25% correct matches), while the PROfits yielded the lowest performance (23% correct matches). Performance was low and did not differ significantly between conditions (X^2 =. 146, DF=2, p >0.05). An analysis by-item revealed no significant differences (X^2 = 39.371, DF=29, p >0.095). While the differences were not significant, the trend is clearly in line with the likeness ratings, by indicating that the combination of two different composites from the same witness does not serve to increase performance.

These results therefore suggest that the observed benefit for combining a PROfit and EvoFIT in experiment 7 seems to be primarily due to the combination of memorial representations rather than the combination of textural and shape information. In addition, while experiment 4 indicated that morphs perform well when they contain composites from four different witnesses, the results from experiment 7 indicate that performance can also increase when composites from two different witnesses are combined. This supports previous research by Bruce et al (2002, experiment 1) who reported a smaller increase in performance 2-Morphs (2 PROfits) compared to 4-Morphs.

However, the results from experiment 7 also indicate that 2-Morph performance is dependent on the type of information that is combined. While the results indicated a benefit for adding an EvoFIT to either a PROfit or an E-FIT, no such benefit was observed when a sketch was added. This suggests that performance for morphed composites may not be entirely due to just the combination of memorial representations, but that the nature of these memorial representations may also be an important factor. This finding may be specific to sketches, or other line drawn composite systems such as Mac-A-Mug Pro, as benefits for combining PROfit, E-FIT and EvoFIT composites have been observed. However, as forensic sketches are still used as investigative tools in the U.K., further investigations should be undertaken to examine the effect of combining these images with other composite images.

General Discussion

The results from experiment 4 indicated that combining composites from four different witnesses increased performance above the level achieved for a single composite. Experiment 1 reported consistently high performance for composites that had been morphed from different witnesses using the same system. This supports previous research (McNeil et al, 1987; Bruce et al, 2002, experiment 1) and initially suggested that morphed composites performed well because they contained a combination of different memorial representations. Experiment 5 examined this in more detail by morphing composites from single and multiple witnesses. The results indicated significantly higher performance for the morphs from multiple witnesses compared to the morphs from single witnesses. This suggests that morphed composites perform well because the combination of memorial representations results in an image that may be closer to the ideal image or prototype (e.g. Solso & McCarthy, 1981; Bruce, Doyle, Dench & Burton, 1991; Homa et al, 2001).

Similarly, the results from the identification task in experiment 4 also suggested that it may be possible to perceive an image that is closer to the ideal image (e.g. Leopold et al, 2001) when multiple composites were presented. Bennet et al (1999) had previously reported an increase in performance when four composites of *similar* quality were presented. However, this investigation has extended this research by revealing that four composites of varied quality (from worst to best) could also perform well. This initially suggested that the differing quality of each composite may have affected the perception of subsequent composites, as reported by Webster & MacLin (1999) and Leopold et al (2001) and resulted in the *perception* of an image that is closer to the ideal or prototype. However, experiment 6 investigated this in more detail by asking participants to provide self-reported identification strategies. The results indicated that overall participants preferred using only one composite to base their identification decision and identifications based on a single composite were generally more accurate. This suggested that participants were not referencing information from all composites and provided further support for the presentation of a single morphed composite.

Similarly, while morphed composites performed well using PROfit, it was unclear whether this effect would still be obtained when composites from different composite systems/techniques were combined. As these techniques result in images that contain very different types of information it was unclear what effect these types of information may have on the resulting morph. Therefore, experiment 7 attempted to replicate the 4-Morph effect with composites from different witnesses and systems and investigated whether adding a different type of image could increase performance for a single composite. The results revealed that combining composites from four different witnesses and systems resulted in an image that performed significantly better than a single composite. This initially suggested that performance for a single composite could be increased when textural information was added (EvoFIT) but not featural/shape information (Sketch). Experiment 8 investigated whether the benefit for adding an EvoFIT was due to the combination of memorial or composite (texture/shape) information. PROfits and EvoFITs were morphed using composites that had been constructed by the same witness and the results indicated no increase in performance when both images were morphed. This suggests that the benefit observed in experiment 7 was primarily due to the combination of memorial rather than composite (texture) information.

The results from all of the experiments in this chapter indicate that performance for morphed composites is dependent on the type of information combined. The combination of any four composites results in a significantly better image, when different witnesses have constructed the composites. Similarly, a smaller increase in performance may be observed when composites from two witnesses are combined (Experiment 4 & Bruce et al, 2002, experiment 1). However, this 2-Morph effect is dependent on the type of composites that are combined. Performance increases when two PROfits are combined (Bruce et al, 2002, experiment 1) and when a PROfit/E-

FIT and EvoFIT are combined (experiment 4) but not when a PROfit/E-FIT and Sketch are combined.

These results are particularly important with regard to current construction procedures, as they indicate that the best performance levels are obtained when composites from four different witnesses are combined. At present only one witness is invited to construct a composite, while other witnesses are kept for other identification procedures (e.g. line-ups). As such, the recent ACPO guidelines stated that "The production of a composite image with multiple witnesses must not be attempted, as this will amount to cross contamination of each witness's primary memory" (pg11). While this guideline does not specifically prohibit the construction and combination of composites from more than one witness, it does not actually state that composites from more than one witness *can* be constructed, nor does it state that these composites can be used in combination.

As there is no way at present, of determining which witness will construct the 'best' quality composite, the results from this investigation and previous research (e.g. McNeil et al, 1997; Bennet et al, 1999; Bruce et al, 2002, experiment 1) all indicate that it would be beneficial to obtain composites from multiple witnesses. As the experiments in this chapter have revealed, the best and most consistent performance is observed when composites from four different witnesses are combined, while smaller increases in performance can be obtained when composites from two witnesses are combined.

To conclude, chapter 2 revealed that the presentation of composites from the same witness in two different views increased performance above the level achieved for a single full-face composite. This chapter did not find an increase in performance when composites from the same witness were presented (either morphed or multiple) but instead revealed that performance can increase when composites from two witnesses are combined. However, the highest and most consistent performance levels were obtained when composites from four different witnesses were combined. Importantly, this chapter has revealed that composite performance is dependent on the type of information presented. Chapter 4 will also investigate the combination of different types of information. This chapter focuses primarily on verbal descriptions and will ask whether the presentation of a composite and a verbal description from the same witness will serve to increase performance, above the level obtained for a single composite. The nature of the relationship between verbal descriptions and composite performance will also be examined.

4

Composites and Verbal Descriptions from the Same Witness

The previous chapters have investigated different procedures for improving the verification of composite images. Chapter 2 investigated whether composite performance would increase when the composites were constructed in a three-quarter view. The results indicated that performance for three-quarter view composites was better when the target had been seen in a three-quarter view. Similarly, the results also indicated that when two composites from the same participant witness were presented, performance increased above the level observed for a single full-face composite. Furthermore, chapter 3 investigated whether the combined use of memorial information from more than one witness would also serve to increase composite performance at test. The results indicated that when composites from four different witnesses were combined to form a new morphed image, performance was as good as or better than the best single composite.

Therefore, the results from both chapters 1 and 2 indicate that performance can increase when different types of information are combined at test. While this information has so far been confined to information contained within a composite image (either during construction or at test), another important source of information is the verbal description that is used to construct the composite. While composite systems have been developed in order to aid eyewitness recall, verbal descriptions are still the most important means of communicating information about a face.

Verbal descriptions are obviously an important source of information in composite construction, as the description provided by a witness is used as the 'starting point' for construction. However, one reason why verbal descriptions are particularly important in a forensic context is that there is no formal method at present, of determining which witness will construct the 'best' composite likeness. Factors such as confidence (e.g. Sporer, Penrod, Read & Cutler, 1995), perceived memory ability (Woodhead, Baddeley & Simmonds, 1979), intelligence (e.g. Brown, Deffenbacher & Sturgill, 1977), occupation (Ainsworth, 1981) viewing distance and sex of subject (Christiaansen, Sweeney & Ochalek, 1983) have all been shown to be very poor indicators of eyewitness accuracy. One recent investigation has indicated a relationship between cognitive style (field dependency) and recall performance (Emmet, Clifford and Gwyer, in press), however it is unclear whether cognitive style would be a positive indicator of composite performance. Therefore, as there is no accurate indicator of performance, the most readily available indicator (to police operatives) is the amount of verbal description supplied by a witness.

The use of descriptions as indicators is highlighted in the current ACPO (Association of Chief Police Officers) guidelines. The guidelines state that "...each witness should be assessed individually on their level of recall to produce a composite image. One image should then be produced using a witness with good recall..."(pg11). This statement clearly states that witnesses should be assessed on their 'level' of recall but there is no indication of what is meant by 'good' recall. Similarly, further statements fail to expand on this, "The witness must have a clear mental image of the person and more importantly can visualise or describe however simply, the facial features..." (pg. 11). As illustrated, these guidelines are very ambiguous and as a result, different composite operators use different criteria (personal communications). Some operators will only construct an image with a witness who can fully describe all of the facial features, while others will construct a composite with a witness who cannot recall one or two features. However, there is much ambiguity as this often depends on the feature. A witness is more likely to be asked to construct a composite if they cannot recall the ears and nose, rather than key features such as eyes and hairstyle. Most operators however, will not ask a witness to construct a composite if that witness cannot recall two or more key features.

Interestingly, Gabbert, Dupuis, Lindsay and Memon (2003) recently questioned sixteen composite operators in Scotland and asked them to state how they determined the quality of a verbal description. The data revealed that the most important factors were attention to detail, exposure to perpetrator, amount of detail, followed by corroboration of evidence, reliability of witness and confidence of witness. This again reveals that two of the most important factors that field operators use to determine the quality of a verbal description are the type and amount of detail.

This raises two important issues. The first is that it is unclear how much information a witness should recall. While it is clear that a witness should describe the facial features, what does 'however simply' mean? Does it matter whether a witness can only recall one descriptor for each feature (e.g. black hair, blue eyes, long nose, wide mouth etc), or should they provide a fuller more detailed description? Second and more importantly is the underlying assumption that the reconstruction ability of a witness can be determined by the amount of information they can recall and describe.

Amount of information recalled

Shepherd, Davies and Ellis (1988) conducted three experiments where participants were required to describe unfamiliar and familiar faces, from memory and with the target photograph present. Further groups of participants were then asked to identify the person depicted in the description either from memory or from an array of four photographs (1 target, 3 distractors). They reported that participants only generated an average of 7.5 descriptors when asked to describe a face with the target photograph present and 5.7 when asked to describe a familiar colleague from memory. More importantly, longer descriptions did not result in an increase in performance, leading the authors to suggest that shorter, more precise descriptions may be more effective than longer descriptions. These results indicate that only a very limited description is needed in order to make a successful identification. However, a more important question to consider is whether the amount of verbal description is indicative of performance on a subsequent task.

Do good describers make good recognisers?

Much research has examined the relationship between verbal descriptions and performance on subsequent recognition tasks. Goldstein, Johnson & Chance (1979) examined whether people who were good at describing faces were also good at recognising them. More specifically, they asked whether the way in which a witness described a face would have any relation to the accuracy of that description. For example, if a witness described a face clearly and fluently, would they be more accurate than a witness who was more hesitant? In their investigation 'gold standard' descriptions of ten targets were compiled by three judges using the FFAL (Facial Features Checklist). In the study phase participants were presented with each photograph for 2.5 seconds. Immediately after presentation, each participant was asked to use the FFAL to compile a description of the target. The participant descriptions were compared with the 'gold standard' descriptions and accuracy scores (%) for each participant's description were calculated. The results revealed that descriptions from memory were poorer than the 'gold standard' descriptions. In addition, the results indicated a fairly narrow range of accuracy scores (29%-63%), with the 'best' describer correct on only 14 out of 23 items. This narrow range indicated that while participants appeared to differ in their ability to describe a face, not all faces were equally describable.

The recognition task took place one week later and was unrelated to the description phase (i.e. different targets were used). In the study phase, ten new faces were presented for 2.5 seconds each, with an inter-stimulus interval of 0.5 seconds. At test, all thirty-nine faces were presented and participants were required to make a standard old/new judgement. Participants were reminded that only ten faces had been presented at study and were asked to try and make only ten old judgements. The results revealed no correlation between accuracy of description and recognition performance. The recognition data was split into three groups (best, middle, worst describers) and the results suggested that better describers were slightly better recognisers. However, there was little evidence for a predictive relationship between verbal ability and recognition performance.

These results indicate that better describers are slightly better recognisers. However, this investigation did not test whether a description related to the identification of the same face. Pigott & Brigham (1985) were the first to test this in a more forensically relevant experiment. A 2 (depth of processing) by 2 (array type) by 2 (target) between-subjects design was adopted. A live target entered the testing room for fifteen seconds and participants were asked to rate either how honest he looked (deep encoding) or how tall he was (shallow encoding). Participants were then required to provide a description of the target by completing a description checklist. This checklist required participants to describe both facial (e.g. hair colour/length/style, skin colour, eye colour/shape/brow) and physical characteristics (e.g. weight, build, sex, age, height etc). Participants then individually viewed one line-up (target present or absent) and were told that the target may or may not be in the array. All participants were then asked to rate how confident they were in their decision. Accuracy was measured both by comparing participant descriptions to those compiled by judges and by comparing participant descriptions to the person chosen from the line-up (target or distractor), termed description congruence. The results revealed no relationship between description accuracy and recognition performance. Similarly, there was no relationship between description congruence and recognition accuracy. Depth of processing also had no effect.

Therefore, Pigott & Brigham (1985) failed to support the finding by Goldstein et al (1979) that better describers were slightly better recognisers. However, Goldstein et al did report an effect of target. Similarly, both studies only examined the relationship between description accuracy and recognition performance and did not measure the amount of description provided.

Wells (1985) extended this research by investigating whether faces that were easy to describe were also easy to identify, measuring both amount and accuracy of description. In this investigation participants were presented with a target face and were asked to make a trait judgement. They were then required to describe and attempt to recognise the target from either a target present or target absent array. Two judges scored the descriptions. Each judge initially counted the number of facial features described to give a score for completeness. Then each judge scored the descriptions for accuracy, assigning values of –1, 0 and 1 for each feature according to whether the description was inaccurate, ambiguous or accurate. Descriptions were then coded for congruence by assigning the same values (-1, 0 and 1) except that the descriptions were now compared with the chosen photograph (face chosen in the array task) rather than the actual target photograph. The results revealed that description completeness did not correlate with either description accuracy or congruence. However, there was a low (0.27) correlation between description accuracy and recognition performance. Wells suggested that as this correlation was not observed in the Pigott & Brigham (1985) investigation, that the effect may be due to target characteristics. As two participants had described and attempted to recognise each target, it was therefore possible to examine this further, by comparing the first participants' description with the second participants' recognition attempt. This between-subject correlation was the same as the within-subject correlation. Further analyses examined whether one participants description of a face could predict another participants recognition performance of the same face. The between-subject correlation was again similar to the within-subject correlation, leading to the conclusion that the relationship between description and recognition was not due to the ability of individual participants, but could instead be explained by the finding that faces that were easy to describe were also easy to recognise.

The results from these investigations therefore indicate that there is no relationship between the amount of descriptors recalled and subsequent recognition performance. More specifically, participants who provide full, detailed descriptions are not more likely to perform better on subsequent recognition tasks. Similarly, the amount of description (number of descriptors) does not appear to correlate with accuracy of the description.

Why can't descriptions predict recognition performance?

The results from the previous experiments indicate that recall and recognition yield different results. Similarly, research has indicated that we are much better at recognising faces than we are at recalling them (e.g. Shepherd & Ellis, 1973; Shepherd, Ellis & Davies, 1978). When we recognise a face, the target information (face) is provided and it is necessary to retrieve some contextual information (e.g. is this the person I saw earlier?). However, when we recall a face, aspects of the context are often provided (e.g. through context reinstatement) but it is necessary to retrieve the target (facial) information.

A great deal of research has indicated that the process of retrieving facial information may be contrary to the way in which faces are normally processed. Many studies have clearly demonstrated that both unfamiliar and familiar faces are perceived and recognised holistically (e.g. Carey & Diamond, 1977; Haig, 1984; Sergent, 1984; Young, Hellawell & Hay, 1987; Tanaka & Farah, 1993; Hole, 1994; Tanaka & Sengco, 1997; see chapter 1 for a review of literature). However, in order to retrieve and describe facial information, it is necessary to 'break down' the face into its constituent parts, a process that seems to be at odds with the way in which faces are processed.

Similarly, research on word recognition has highlighted the independent nature of recall and recognition (e.g. Tulving & Watkins, 1973; reviews by Flexser & Tulving 1978, 1982). The results from these experiments have indicated that participants can recall a word then fail to recognise it on a subsequent recognition test. This finding had been termed retrieval independence, where cues that are present on a recognition task are uncorrelated with those on a recall task. This finding together with the investigations discussed earlier indicate that verbal descriptions are unlikely to predict recognition performance. However, while verbal descriptions appear to be uncorrelated with recognition tasks, the relationship between composite construction, recall and recognition is less clear.

Composites, descriptions and recognition

Davies, Ellis & Shepherd (1978) examined the effect of constructing a composite (Photofit) on subsequent recognition performance. Groups of participants were required to either construct a Photofit and perform a subsequent recognition task, or just perform the recognition task. The results revealed no significant difference in recognition performance, although performance after composite construction was slightly lower. Similar results have recently been reported by Brace and colleagues (2002) who found no decrement in recognition following E-FIT construction. While these investigations suggest that reconstruction and recognition tasks may be uncorrelated, no comparison was made with the verbal descriptions.

Maudlin & Laughery (1981) examined the relationship between composite construction (Identikit) and verbal descriptions on subsequent recognition performance. After presentation of a target face participants were asked to either construct an Identikit (group 1), complete an adjective check list and write a description of the target (group 2) or complete a personality inventory (control group). The test phase consisted of a standard old/new recognition task at varying time delays (30 minutes and 2 days). The results revealed that the participants who had constructed the Identikit (mean 5.5) performed better on the recognition task than either the description (mean 4.9) or the control groups (mean 4.2). There was also no effect of delay, which led the authors to suggest that the benefit for constructing a composite may have been a result of the development of more effective retrieval mechanisms during construction.

These results indicate that composite construction may be uncorrelated with recognition performance. However, in order to examine the relationship between verbal descriptions and composite construction in more detail, it is necessary to examine whether verbal descriptions can predict composite performance.

Composite construction and verbal descriptions

While it has been suggested that a witness's ability to verbally describe a face is a key limiting factor in composite construction (e.g. Laughery & Fowler, 1980), Christie & Ellis (1981) argued that descriptions may contain useful and accurate information. However, the difficulty may lie in transferring these descriptions into an accurate visual likeness.

In their investigation, (Christie & Ellis, 1981) participants were asked to both verbally describe and construct a Photofit likeness of a target face from memory. There were six target faces and six participants per target. In addition, as a control condition all participants constructed a composite of another target with the reference photograph available for inspection. Order of construction was counterbalanced such that half of the participants constructed the target-present composite first and half constructed the from-memory composite first. Participants were asked to describe the face immediately after a one-minute exposure to the target photograph.

All composites and descriptions were evaluated using array and sorting tasks. For the array tasks, the composites and descriptions were divided into separate sets of six (i.e. 6 descriptions, 6 composites from memory etc). Each set of six was presented with an array of twenty-four photographs (6 targets and 18 distractors) and participants were asked to try and 'pick out' each of the six targets. For the sorting task, a further set of participants were given the target photographs together with either the from-memory composites, the target-present composites or the descriptions. They were asked to match each image to its corresponding target.

The results revealed that the descriptions performed significantly better than the composites (both from-memory and target-present) in both tasks, suggesting that the descriptions contained more identifiable information than the Photofits. Interestingly, the authors reported that there was an increased likelihood of an accurate Photofit, when the description was more detailed (contained more descriptors). When the descriptions and composites were presented together, performance did not increase significantly above the level achieved for the descriptions alone. Interestingly, in this condition, thirty-five out of sixty participants reported finding the descriptions more useful than the Photofits, while only ten participants found the Photofits useful.

These results clearly indicate that verbal descriptions can contain useful information. The authors however suggest that the difference in performance between the descriptions and Photofits may have been a result of the construction process, in that the act of constructing the composite may have interfered with the internal representation of the face. Therefore while initial recall may be good, the presentation of similar features during construction may have caused the image to 'disappear', thus resulting in a poorer composite image.

However, as stated in chapter one, the older mechanical composite systems such as Photofit produced very poor facial likenesses (e.g. Ellis, Shepherd and Davies, 1975).

Photofit was limited in the number of features it contained, especially key features such as hair and face shape (Ellis, 1986). Similarly, the shading boundaries that occurred when Photofit features were placed together was found to impair recognition (Ellis, Davies and Shepherd, 1978). This led Ellis et al (1978, pg305) to state that "...Photofit is too crude an instrument for the generally successful recall of faces..."

Given these findings, it is unclear whether the benefit for verbal descriptions in the Christie and Ellis (1981) investigation was influenced by the superiority of descriptions or the inferiority of the Photofit system. As research has indicated that mechanical systems produce very poor likenesses and indeed produce poorer likenesses than those constructed by sketch artists (Laughery & Fowler, 1980), it is perhaps unsurprising that verbal descriptions were found to contain more useful information than Photofit images. Therefore, it is unclear how useful descriptions are and how they compare to composites that have been constructed using more modern computerised systems such as PROfit.

Recent findings by Gabbert, Dupuis, Lindsay and Memon (2003) indicate that perhaps the results reported by Christie and Ellis (1981) were influenced by the inferiority of the Photofit system. The Gabbert et al (2003) investigation examined the relationship between verbal descriptions and E-FIT performance. In their investigation, each description was presented with a six-person, target present photographic array. Participants were then required to rate the degree of similarity between the description and each of the six photographs on a scale from 1 (not at all similar) to 9. This determined a mean rating for each description. In addition, the E-FIT composites were also rated for likeness against one photograph of the target. This allowed a comparison to be made between the description ratings and the composite ratings. The results revealed no association between the two measures, leading the authors to suggest that the ability to accurately recall a face does not relate to the ability to construct a good composite.

Therefore, there are conflicting results regarding the predictive utility of verbal descriptions in composite construction. The results by Gabbert et al (2003) suggest that positive findings reported by Christie and Ellis (1981) may have been influenced by the inferiority of the Photofit system. However, as the Gabbert et al findings are based on similarity ratings and not on identification, the differing results may in fact reflect task demands. Similarly, Gabbert et al do not report any findings concerning the amount of description reported, nor do they state how well the descriptions performed. Therefore, the exact nature of the relationship between verbal descriptions and composites is still unclear.

Aims

This investigation will therefore initially examine the relationship between verbal descriptions and composites constructed using PROfit. Experiment 1 will examine the relationship between the amount of verbal description and performance of the resulting composites. Experiment 2 will expand on this by comparing the amount <u>and</u> accuracy of descriptions with composite quality using a different set of targets.

In addition, as chapters 2 and 3 have indicated that composite performance increases when more information is presented, this investigation will also examine whether performance will increase when composites and descriptions from the same participant are presented together. Christie and Ellis (1981) reported no increase in performance for combined composites and descriptions, however, the composites were of poor quality. Experiment 3 will therefore examine performance for descriptions and PROfit composites presented alone and together. The results from chapters 2 and 3 have indicated that performance increases when memorial information is combined at test, rather than just more information. Experiment 4 will expand on this by examining performance for composites, participant generated descriptions and 'perfect' descriptions (generated by two independent judges). Composites and descriptions will be presented alone and together. That is, the participant generated composite will be presented with both its corresponding description (from the same participant) and with a 'perfect' description. It is expected that 'perfect' descriptions will perform better than participant generated descriptions both when presented alone and when presented with a composite.

Experiment 9: Is amount of description correlated with composite quality?

Christie and Ellis (1981) had reported an increased likelihood of an accurate Photofit, when the description was more detailed (i.e. contained more descriptors). This investigation will therefore provide an initial examination of the relationship between the amount of description provided and the quality of composites using a more modern computerised composite system (PROfit). This experiment used the full-face composites that were constructed for experiment 1 (chapter 2). The construction procedure was therefore identical. Participants who were unfamiliar with the targets were asked to view a thirty-second video clip of one target. The operator then used cognitive interview techniques to elicit a verbal description. Each participant was then asked to construct two composites of the same target (one in a full-face view and one in a three-quarter view). Please see chapter 2, pg. 64 for a full description of procedure. All full-face composites had been previously rated for likeness against a photograph of the target, on a scale from 1 (low) to 10 (high). This experiment will therefore examine each description and mark the amount of descriptors provided. Correlations will then be performed on the number of descriptors and the quality of the composite for each participant.

Stage 1: Coding verbal descriptions

Materials

This experiment used composites and descriptions of targets (female staff members from the psychology department) that had been constructed for experiment 1 (chapter 2). During composite construction the experimenter had written each verbal description on a standard composite description sheet. This sheet was adapted from those used at Grampian Police force (please see appendix for an example). This sheet contains a list of facial features (face shape, hair, eyes, eyebrows, nose, mouth, ears) with space between each feature to write the corresponding description (please see chapter 2, page 64 for a description of interview procedure).

Coding descriptions

For every description each unit of information or descriptor was marked and verified by a second independent judge. Minor inconsistencies in marking between the judges were identified after the first description had been assessed. This was mainly caused by one judge marking 'fleshy round cheeks' as one unit of information, whereas the second judge felt that they should be treated as two units. A consensus was reached and every descriptor that provided information about a particular feature was marked (e.g. fleshy and round were marked as two descriptors). As well as marking all adjectives (repetitions were not marked), all attributional comments were also marked e.g. 'friendly face'. Shepherd, Ellis & Davies (1988) previously reported no difference in performance for descriptions that contained both physical and attributional statements, however attributional statements alone led to very poor identification performance.

When each descriptor had been marked, the total number of descriptors was added together to give a total recall score for each participant. In addition to marking the total number of descriptors, it was noted that many descriptions contained 'negative' descriptors. That is, some participants could not actually describe the size or shape of a feature (e.g. round) but they could recall that it was not a different shape (e.g. it's not rectangular). It is important to note that a distinction was made between a 'negative' descriptor and a negative statement (e.g. she isn't wearing earrings). If a unit of information was clearly stating that the target did not have certain characteristics (e.g. she doesn't have wrinkles) then this was marked as a positive descriptor. In addition, some participants appeared to reach a description through a process of elimination. For example, one participant stated that the 'face is not round, not long, not thin, not fat or chubby'. However, they did not actually state what shape the face was. Dunning & Perretta (2002) had reported that participants who made identifications automatically were more accurate than those participants who consciously thought about their decision. Therefore, it is possible that these 'negative' descriptors either reflect a more conscious search process, or reflect general uncertainty, perhaps resulting in a less accurate description and/or composite. These 'negative' descriptors were therefore also coded separately, resulting in three scores for each participant, one total recall score, one 'positive' descriptor score and one 'negative' descriptor score. Figure 18 provides an example of the coding procedure using the description given by participant number 3. The light blue colour denotes a marked descriptor and the yellow colour denotes a marked negative descriptor. The total recall score for this participant was 26, with 22 'positive' descriptors and 4 'negative' descriptors.

Middle aged – 40's. Friendly face. Isn't wrinkly. Not a double chin but a bit jowly. Round, quite plump face with fleshy round cheeks. A lot of colouring in the cheeks. Not a really strong jawline – not angular. Deep lines around the mouth – laughter lines. Hair is short and blonde – dyed. Brown eyebrows. Blue eyes that are normally spaced apart. Lines underneath the eyes but not much. Nose isn't too big – not prominent. In the middle and in proportion. A bit pointy from the side. A small triangular mouth with a big bow in the middle.

Figure 18: Example of coding procedure

Composite evaluation

In experiment 1 all full-face composites were rated for likeness by twenty unfamiliar participants on a scale of 1 (low) to 10 (high). A comparison was then made between the mean likeness rating and the amount of description recalled for each participant's composite.

Results: Verbal descriptions

The amount of information recalled and the composite likeness ratings were initially subjected to separate analyses. Each participant recalled between 14 and 28 descriptors (mean 21.5, SD 4.4). The data was split into eight descriptive categories, which included one category for general information and one category for each of the seven main physical features (i.e. 1-general information including age, build, attributions, 2 - facial shape, 3 - hair, 4 – eyebrows, 5 – eyes, 6-nose, 7-mouth, 8-ears). A 4 (target) by 8 (feature) repeated measures analysis of variance revealed no effect of target ($F (3,9 = 0.247, p=0.861$), a significant effect of feature ($F (7,21 = 29.624, p<0.001$) and no interaction ($F (21,63 = 1.093, p=0.379$). Further analyses (one-way repeated measure anovas and paired samples t-tests) on each target, revealed that while the pattern of description differed slightly for each target, in general more descriptors were recalled for external features compared to internal features, with the exception of ears, see figure 19.

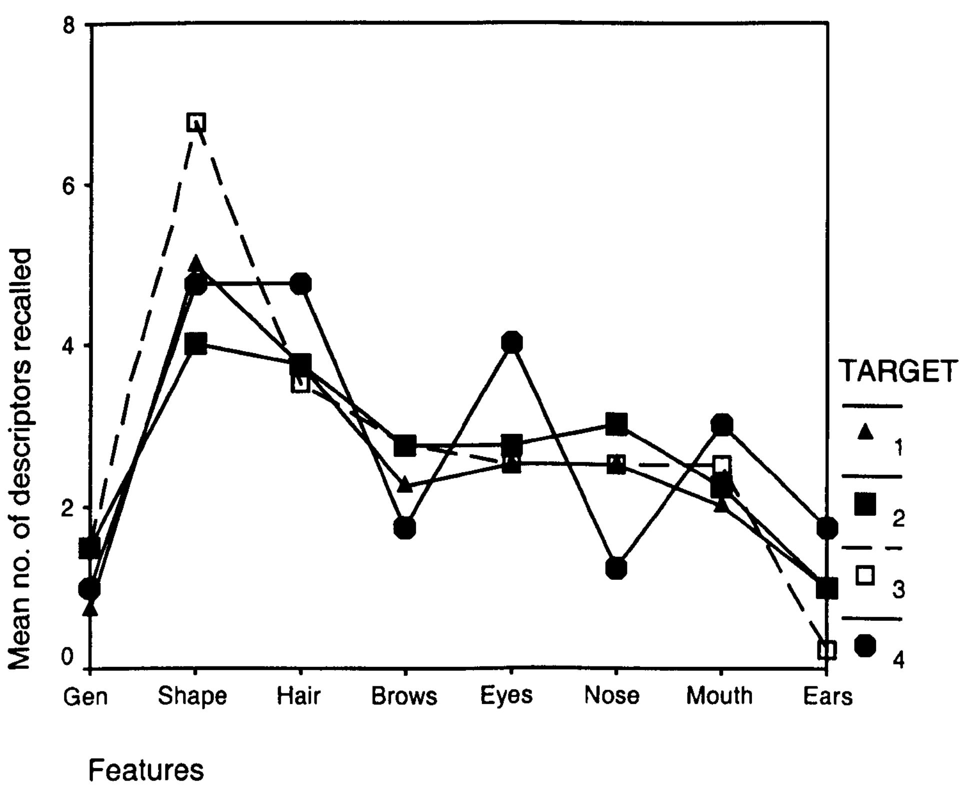


Figure 19: Mean no of descriptors recalled for each feature and target

Composites

The mean likeness ratings for composites ranged from 1.65 to 6.45 (mean 4.2, SD 1.6). A one-way repeated measures analysis of variance revealed a significant effect of target (F (3,57 = 24.011, p<0.001). Further analyses using paired samples t-tests revealed that the composites for target 2 (mean 2.9, SD 1.5), obtained significantly lower likeness ratings than either target 1(mean 4.5, SD 1.3), target 3 (mean 4.5, SD 1.2), or target 4 (mean 4.9, SD 1.1), (p<0.001 for all). This indicates that while participants could recall information about target 2 and describe the features well (see figure 2) there was difficulty in translating this description into a 'good' quality pictorial representation. As the pattern of descriptors recalled was similar to the other targets, this indicates a possible difficulty with the representation of features within PROfit, rather than a memorial difficulty.

Relationship between verbal descriptions and composites

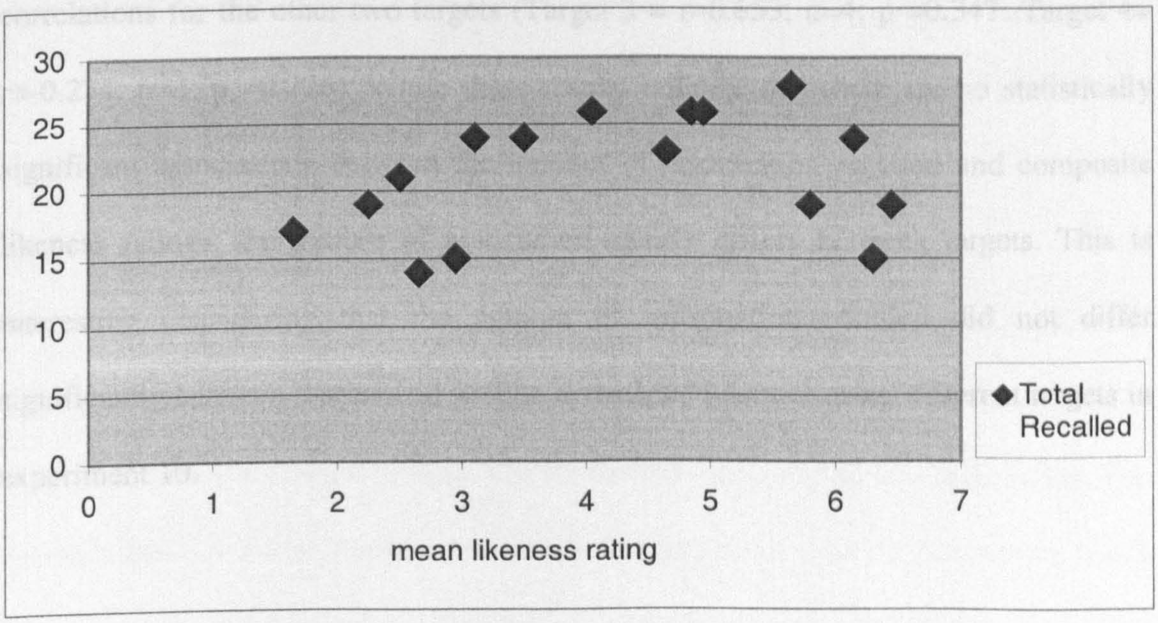


Figure 20: Graph illustrating the total amount of information recalled and mean likeness ratings for each participant

The total amount of information recalled and the mean likeness ratings for each participant are displayed in figure 20. As can be seen, there is no clear relationship between the amount of descriptors recalled and the quality of the resulting composites and a correlation (Pearson) revealed no significant association ($r = 0.278$; $n=16$; $p=0.297$). The number of negative descriptors was then compared with the mean likeness ratings and again no significant association was observed ($r=0.234$; $n=16$; $p=0.384$). The only significant associations obtained were between the number of positive and negative descriptors and the total amount of information recalled ($r=0.876$; $n=16$; $p<0.001$ for positives and $r=0.586$; $n=16$; $p<0.05$ for negatives).

Correlations on each individual target revealed that for two of the targets there was in fact a low negative correlation between the amount of information recalled and performance of the resulting composites (Target 1 = $r=-0.149$; $n=4$; $p = 0.851$: Target 2= $r=-0.149$; $n=4$; $p = 0.851$). However, there were low to moderate positive correlations for the other two targets (Target 3 = $r=0.653$; $n=4$; $p = 0.347$: Target 4= $r=-0.274$; $n=4$; $p = 0.726$). While these results indicate that there are no statistically significant associations between the amount of information recalled and composite likeness ratings, the pattern of association clearly differs between targets. This is interesting considering that the amount of information recalled did not differ significantly between targets and will be investigated further using different targets in experiment 10.

These results therefore do not offer any clear evidence to support the suggestion by Christie and Ellis (1981) that participants who provide larger descriptions (i.e. contained more descriptors) are more likely to produce better quality composites.

Instead, the results are similar to those reported in recognition studies (e.g. Pigott & Brigham, 1985; Wells, 1985; Shepherd, Davies and Ellis, 1988), and support Gabbert et al (2003) by clearly indicating that there is no relationship between the amount of information recalled and the quality of the resulting composite.

Furthermore, Shepherd, Davies and Ellis, (1988) suggested that as longer descriptions did not result in increased performance, shorter, more precise descriptions may be more effective. The results from this investigation provide little evidence that shorter descriptions result in a better quality composite, however Shepherd at al examined performance for descriptions rather than composites.

To assess this further, a further experiment was undertaken using a different set of targets. Whereas experiment 9 only examined the amount of verbal description, experiment 10 will compare the amount of verbal description with accuracy of description. Accuracy will be assessed using a six-alternative forced choice array task.

Experiment 10: Length and accuracy of description

This experiment used the full-face composites that were constructed for experiment 6. The construction procedure was therefore identical. Participants who were *unfamiliar* with the targets were asked to view a thirty-second video clip of one male target. The operator then used cognitive interview techniques to elicit a verbal description. Each participant was then asked to construct one full-face composite of the target. All composites had been previously rated for likeness against a target photograph on a scale from 1 (low) to 10 (high). In stage 1, all verbal descriptions were marked using the same procedure as the previous experiment (experiment 9) to assess the amount of information recalled for each participant. In stage 2, all descriptions were presented with target present arrays to assess accuracy. Correlations were then performed on the amount of description, the accuracy of the description and the quality of the composite (likeness ratings).

Stage 1: Coding descriptions

All descriptions were marked using a similar procedure as experiment 9 (see page 172 for procedure). As experiment 9 had indicated that 'negative' descriptors did not contain any predictive element, the total number of descriptors was only marked in this experiment. Total recalled was then compared with the accuracy of descriptions.

Stage 2: Assessing accuracy of descriptions

Materials

This experiment used the composites and descriptions that had been obtained in experiment 6 (please see chapter 3, pg. 120 for details). The targets were four male members of staff at the University of Stirling. All descriptions were presented with a target present array. See figure 21 for an example.

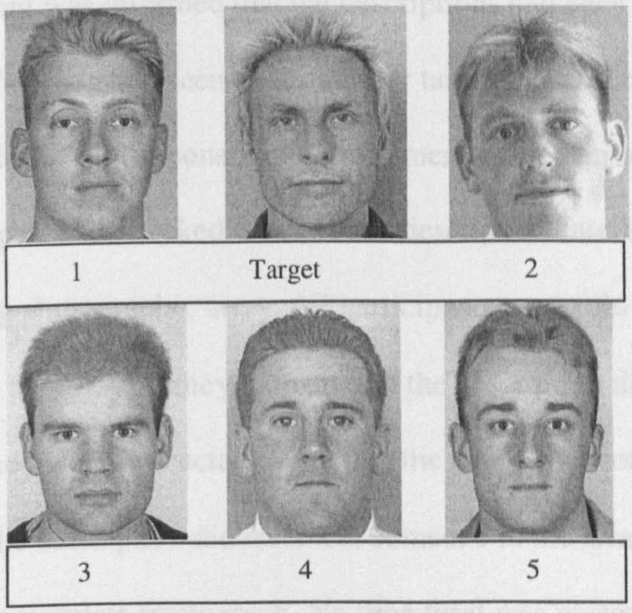


Figure 21: Example of male arrays

Participants

Thirty-two participants were recruited from student unions in Glasgow University. All participants were unfamiliar with the targets.

Design

There were 16 descriptions in total (4 per target). These were divided into four presentation books using a Latin square design. This ensured that each participant only saw <u>one</u> description for each of the four targets. Within each book the order was rotated. Books were presented to participants who were unfamiliar with the targets. Participants were asked to try to identify the target from an array (1 target and 5 distractors). There were a total of eight attempts per description.

Procedure

Each participant was informed that the descriptions had been provided by mock 'witnesses' after they had seen an unfamiliar target for 30 seconds. It was emphasised that the descriptions were constructed from memory and represented a likeness of the target. Participants were asked to read each description carefully and examine the accompanying photographic array. All participants were told that the target may or may not be in the array. If they thought that the description did not resemble anyone in the array they were instructed to say that the target was not there. Similarly, if they thought that the description did resemble someone in the array, they were instructed to point to the appropriate photograph. No time limit was placed on this procedure.

Results: Amount of description

The amount of information recalled and the composite likeness ratings were again initially subjected to separate analyses. Participants recalled between 22 and 42 descriptors (mean 33.3, SD 6.1). The data was split into the same eight descriptive categories as experiment 9 (i.e. 1-general information including age, build and attributions, 2 - facial shape, 3 - hair, 4 – eyebrows, 5 – eyes, 6-nose, 7-mouth, 8-ears). A 4 (target) by 8 (feature) repeated measures analysis of variance revealed no effect of target (F (3,9 = 0.553, p=0.659), a significant effect of feature (F (7,21 = 21.661, p<0.001) and no interaction (F (21,63 =1.086, p=0.386). Further analyses (one-way repeated measure anovas and paired samples t-tests) on each target, revealed a similar pattern of description to experiment 9, with more descriptors for external features compared to internal features, see figure 22.

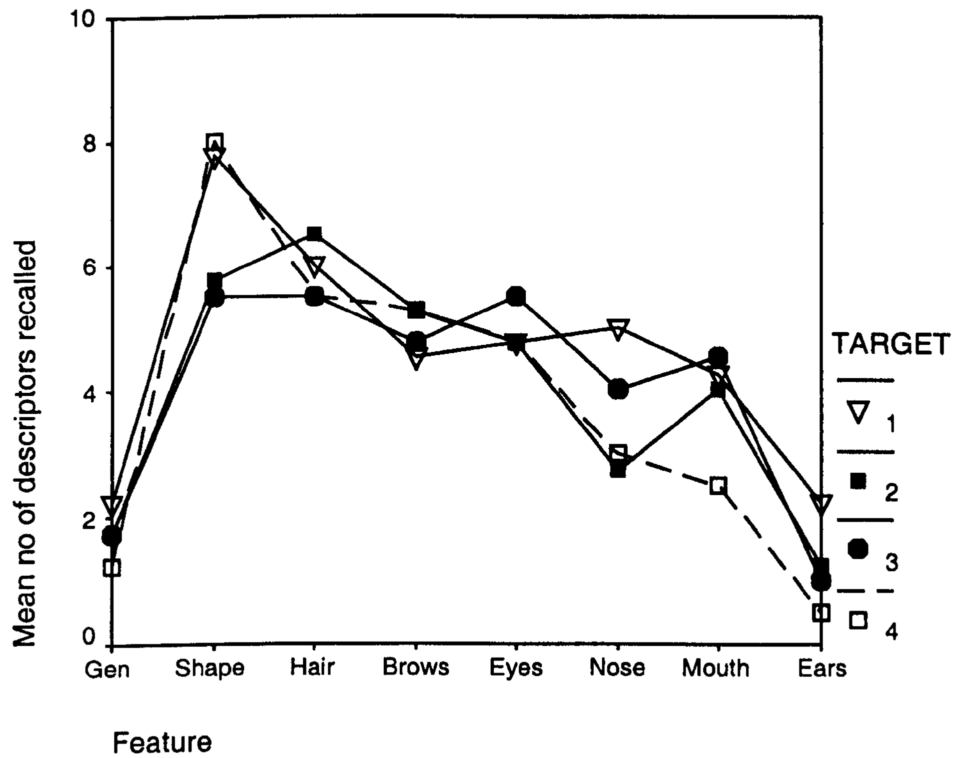


Figure 22: Mean number of descriptors recalled for each feature and target

Accuracy of description

The number of correct matches (hits) for all sixteen descriptions ranged from 25% - 100% (mean 67%, SD 22.4%). A Friedman test revealed that the number of correct matches differed significantly across targets (X^2 =7.944, DF=3, $p<0.05$). Further analyses using Wilcoxon Signed Ranks tests revealed that there were significantly more correct matches for target 2, compared to target 3 ($p<0.05$) and almost significantly more correct matches compared to targets 1 and 4 ($p=0.054$; $p=0.072$ respectively). This indicates that while the pattern of information recalled was similar for all targets, the description for target 2 was slightly more effective in an identification task.

Composites

The mean likeness ratings for composites ranged from 2.3 to 6.15 (mean 4.5, SD 1.08). A one-way repeated measures analysis of variance revealed a significant effect of target (F (3,57 = 15.983, $p<0.001$). Further analyses using paired samples t-tests revealed that the composites for target 3 (mean 3.6, SD 1.7) obtained significantly lower likeness ratings than either target 1(mean 5, SD 1.4), target 2 (mean 4.5, SD 1.4), or target 4 (mean 4.8, SD 1.3), ($p<0.05$ for all). A comparison between the mean number of correct matches for each target's descriptions and the mean likeness ratings for each target's composites, revealed that both the descriptions and the composites for target 3 performed poorer than the other three targets.

This initially suggests that there were difficulties in both describing and constructing a composite of this target. However, as description performance for target 3 did not differ significantly from targets 1 and 4, it appears that there was difficulty in translating the description into a pictorial representation. While these results illustrate the pattern of performance for composites and descriptions for each target, the following analyses will examine associations between composites and descriptions for each participant.

Accuracy and amount

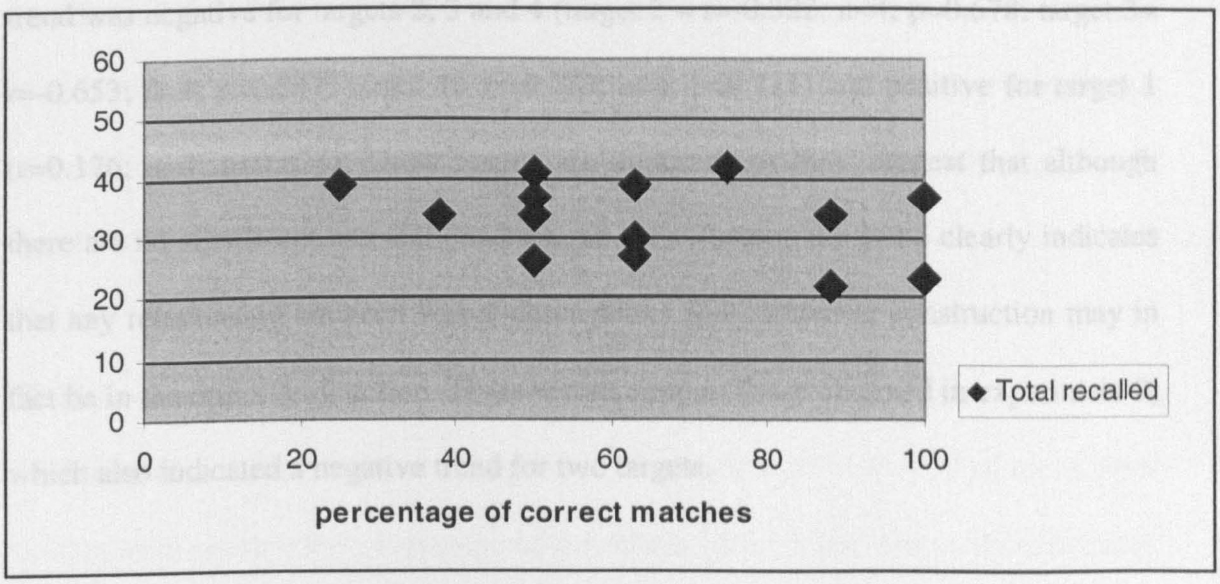


Figure 23: Graph showing relationship between amount and accuracy of description

Figure 23 displays the relation between the total amount of information recalled and the percentage number of correct matches for each participant's description. As can be seen, there is an indication of a slight negative association between the two factors and while a Pearson's correlation revealed no significant association (r=-0.349; n=16; p=0.186), the association is <u>negative</u>. Similarly, while there was no significant association between the amount of information recalled and the quality of the resulting composite (r=-0.123; n=16; p=0.651), this association was also negative. However, the association between accuracy of description and composite quality was positive (r=0.314; n=16; p=0.237).

Further correlations were performed on the amount of information recalled, accuracy of description and composite performance (likeness ratings) for each target. For the amount of information recalled and description accuracy, the trend was negative for targets 1, 3 and 4 (target 1 = r=-0.686; n=4; p=0.314; target 3= r=-0.357; n=4; p=0.643; target 4= r=-0.592; n=4; p=0.408) and positive for target 2 (r=0.585; n=4;

p=0.415). For the amount of information recalled and composite performance, the trend was negative for targets 2, 3 and 4 (target 2 = r=-0.322; n=4; p=0.678; target 3= r=-0.653; n=4; p=0.347; target 4= r=-0.279; n=4; p=0.721) and positive for target 1 (r=0.176; n=4; p=0.824). These results are important as they suggest that although there are no significant associations between these factors, the trend clearly indicates that any relationship between verbal descriptions and composite construction may in fact be in the opposite direction. These results support those obtained in experiment 9, which also indicated a negative trend for two targets.

Discussion

The results from both experiments 9 and 10 clearly highlight individual differences in recall ability. Even under ideal learning conditions (i.e. video clearly displaying all views of the face), where descriptions were provided immediately after target presentation, the amount of recall differed significantly between participants.

In addition, the results from both experiments also indicate that there is no relationship between the amount of information recalled and composite quality. These experiments therefore provide no support for the suggestion that participants who provide longer descriptions (i.e. those containing more descriptors) are more likely to produce better quality composites (Christie and Ellis, 1981). Instead, the results are similar to those reported in recognition studies (e.g. Pigott & Brigham, 1985; Wells, 1985; Shepherd, Davies and Ellis, 1988), by clearly indicating that there is no relationship between the amount of information recalled and the quality of the resulting composite.

The results may provide some support for the suggestion that shorter, more precise descriptions may be more effective than longer descriptions (Shepherd, Davies and Ellis, 1988), as figure 23 illustrates a slight negative trend between the amount and accuracy of descriptions. However, due to small sample sizes and lack of statistical significance future research should investigate this further.

These findings are particularly important for police operatives, as the recent ACPO guidelines clearly state that a witness should be assessed on their level of recall prior to composite construction. However, as the results indicate that length of description is not correlated with either accuracy of description or composite quality, any assessment based on the amount of information recalled may be ineffectual.

However, one of the difficulties of assessing the relationship between descriptions and composites under ideal learning conditions, is that all participants recalled information about most features. As stated earlier, most composite operators will not invite a witness to construct a composite if they cannot recall more than two features. However, it was impossible to assess the importance of this in these experiments. In experiment 9 only six participants could not recall any information about the ears and eyebrows and in experiment 10 only five participants could not recall information about either the ears, nose or eyebrows. Future experiments should therefore adopt a more ecologically valid procedure using both a live target and a substantial delay between target exposure and the cognitive interview.

Despite this, the results do indicate that composite performance is not related to either the amount or accuracy of description. While this suggests that it impossible to identify which witness will construct the best quality composite, by assessing the amount of verbal description provided, the descriptions did contain useful identifiable information. The percentage number of correct matches for all descriptions ranged from 25% and 100%, indicating that it was possible to identify the target from even the poorest description.

As both chapters 2 and 3 have reported an increase in performance when different types of information have been combined at test, it is possible that the combination of verbal descriptions and composites from the same participant may also serve to increase performance. Experiment 11 will therefore examine this using the descriptions and composites from experiment 9. This experiment will adopt a within-subjects design where unfamiliar participants will be presented will both composites and descriptions, together with target present arrays (6alt forced choice task). All composites and corresponding descriptions (i.e. the ones generated by the same participant) will be presented alone and together. Participants will be asked to attempt to 'pick out' the target from the array.

Experiment 11: Combining descriptions and composites

The aim of this experiment was to investigate whether the combination of information from both a composite and a description (from the same participant) would serve to increase performance above the level observed for a single composite. This investigation used the same composites and descriptions as experiment 9 (female staff members at the University of Stirling). All composites and corresponding descriptions (i.e. the ones generated by the same participant) were presented alone and together.

Unfamiliar participants were asked to attempt to 'pick out' the target from an array (6alt forced choice task).

Materials

This investigation used the same composites and descriptions as experiment 9. The targets were four female staff members from the department of psychology, University of Stirling. All composites and descriptions were presented on a single sheet of A4 paper (landscape) together with a target present array, containing one photograph of the target and five distractors, see figure 24 for an example of arrays.

Figure 24: Example of female arrays

Participants

One hundred and twenty-eight participants were recruited from the University of Glasgow and Queen Margaret University College, Edinburgh. All participants were unfamiliar with the targets.

Design

Sixteen participants (4 per target) had constructed composites in the original experiment. This resulted in a total of sixteen composites and sixteen verbal descriptions (i.e. those generated by the same participants). Composites and descriptions (from the same participant) were presented alone and together. In order to examine the best method of combining verbal and pictorial information, two different presentation formats were used. A 4 (target) by 4 (presentation type) within-subjects design was adopted. The presentation types were composites presented alone, descriptions presented alone, composites and descriptions presented side by side, and composites and descriptions together (i.e. each feature descriptor was placed next to the corresponding feature). This resulted in a total of 64 presentations (16 composites, 16 descriptions, 16 composites and descriptions presented side by side and 16 composites and descriptions together). The 64 images were divided into sixteen presentation books using a Latin square design. This ensured that each participant only saw *one* type of presentation (single composite, single description, composite and description side by side and composite and description together) for each of the four targets. Within each book the order was rotated. Books were presented to participants who were unfamiliar with the targets. There were a total of eight attempts per composite. See figure 25 for an example of presentation types.

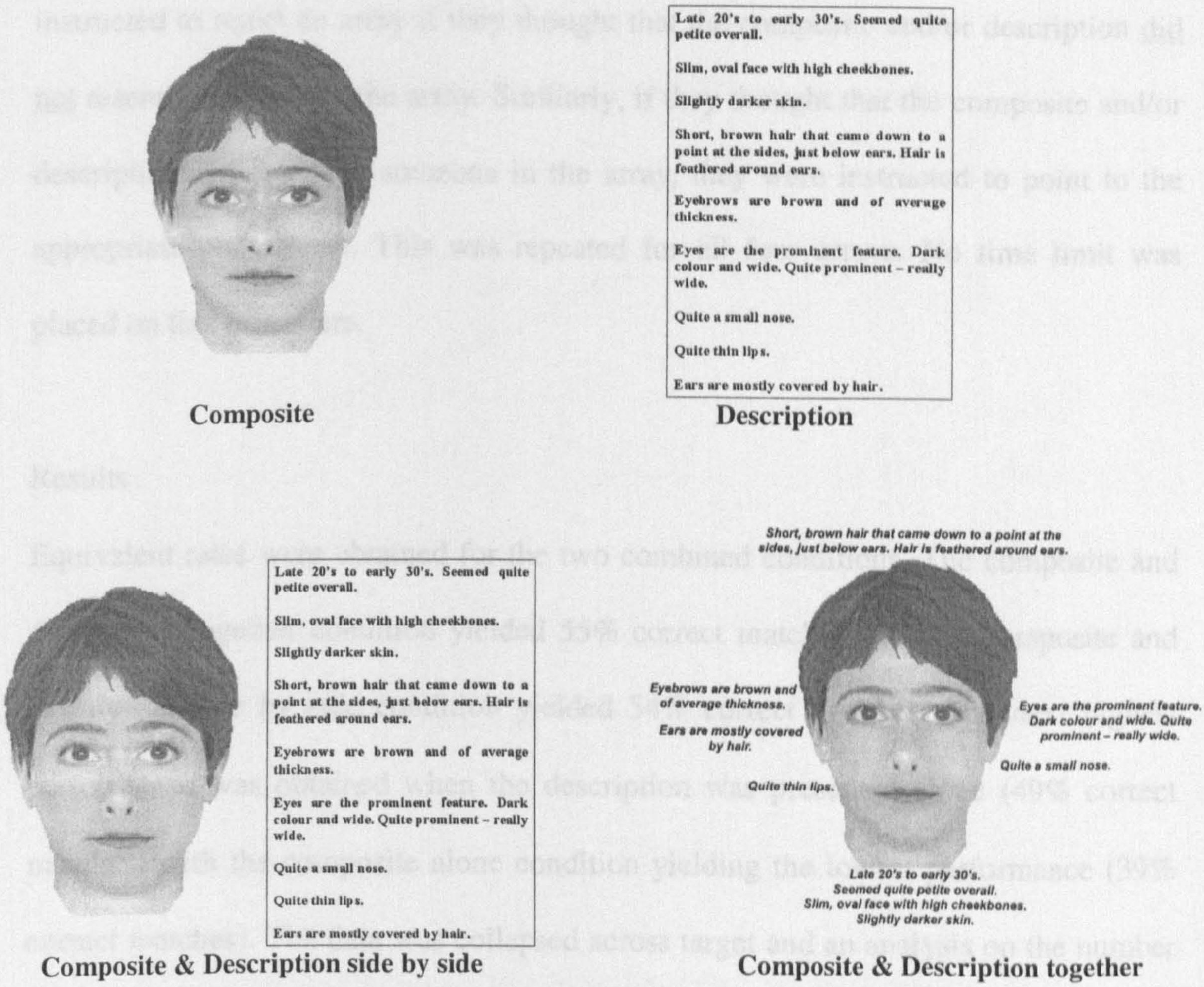


Figure 25: Example of stimuli (composites and descriptions were generated by the same witness)

Procedure

Each participant was informed that they would be asked to look at composites and descriptions of faces. They were told that the stimuli had been constructed in an earlier experiment, where participants had been asked to describe a face and construct a composite likeness of the person, after only seeing the face for 30-seconds. Participants were therefore informed that when they saw a composite and a description together, these had been generated by the same 'participant witness'. Participants were asked to examine all stimuli carefully together with the accompanying array. The arrays were target-present, however to avoid bias all participants were told that the target may or may not be in the array. Participants were instructed to reject an array if they thought that the composite and/or description did not resemble anyone in the array. Similarly, if they thought that the composite and/or description did resemble someone in the array, they were instructed to point to the appropriate photograph. This was repeated for all four arrays. No time limit was placed on this procedure.

Results

Equivalent rates were obtained for the two combined conditions. The composite and description together condition yielded 55% correct matches, and the composite and description side by side condition yielded 54% correct matches. The next highest performance was obtained when the description was presented alone (49% correct matches) with the composite alone condition yielding the lowest performance (39% correct matches). The data was collapsed across target and an analysis on the number of correct matches (Cochran Q) revealed significant differences across conditions ($X^2=7.938$, DF=3, $p < 0.05$). Further pairwise comparisons using Mcnemar tests revealed that the composite alone condition yielded significantly fewer correct matches than both the composite/description together condition ($p < 0.05$) and the composite/description side by side condition.

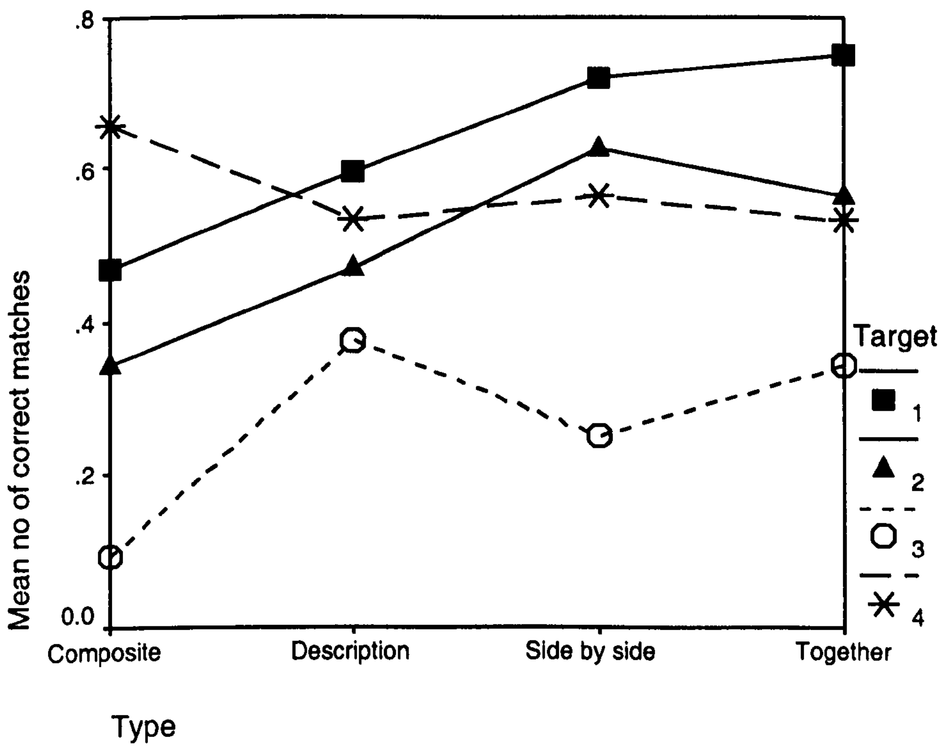


Figure 26: Mean no of correct matches for each target and type

Figure 26 displays the mean no of correct matches for each target and type. This illustrates that for targets 1, 2 and 3, the combined conditions performed better than the single composite condition. An analysis by target revealed significant overall differences (X^2= 58.636, DF=3, $p < 0.01$). Further analyses (Cochran Q and Mcnemar tests) revealed that for target 1, the composite and description together condition performed significantly better than the single composite ($p<0.05$), with the side by side condition performing almost significantly better than the single composite ($p=0.057$). For target 2, the combined (side by side) condition performed significantly better than the single composite ($p<0.05$). Similarly, for target 3 the combined (composite and description together) condition performed significantly better than the single composite ($p<0.05$). In addition, the description for target 3 achieved significantly more correct matches then the composite ($p<0.05$). For target 4, there were no significant differences.

The results therefore indicate that for three out of four targets, presenting a composite with a description (either side by side or together) increased performance significantly above the level observed for a single composite ($p < 0.05$). Importantly, while there were no significant differences for target 4, the combined presentation of composites and descriptions did not decrease performance levels. The lack of any beneficial effect of combining information for this target may have reflected a possible ceiling effect. As figure 26 illustrates, the composite performed very well when presented alone and suggests that it was a very good composite. As a result, adding additional information may not have produced a beneficial effect. However, there is no detrimental effect on performance when composites and descriptions are combined for this target and there is a beneficial effect for combining information for the other three targets.

In addition, the results also provide support for experiment 10 by suggesting that descriptions do contain useful identifiable information. In this experiment the descriptions performed as well as or better than composites when presented alone.

Similarly, these results also support the findings reported in both chapters 2 and 3 by indicating that performance increases when more information is presented. In particular, the results are similar to those in chapter 2, by indicating that when two different types of information from the same participant witness are presented together, performance increases above the level observed for a single composite.

Importantly, chapters 2 and 3 revealed that increased performance levels were dependent on the type of information that was combined. More specifically, the observed increase in performance in both chapters was found to be a result of the combination of different memorial representations, rather than the combination of more information. Experiment 12 will therefore examine performance for the combined presentation of descriptions and composites in more detail. This experiment will firstly attempt to replicate the results from the previous experiment using a different set of targets. In addition, this experiment will also expand on the results obtained in experiment 10, by examining performance for descriptions that were obtained 2 days after target exposure. As the descriptions were obtained after a significant delay it is hoped that some participants may fail to recall information about key features. Therefore an examination of the effect of this on description and composite accuracy may be permitted.

In addition to examining performance for the combined presentation of composites and descriptions, this experiment will also examine performance for 'perfect' descriptions (i.e. those generated by independent judges). Chapter 2 reported an increase in performance only when memorial information was combined, (i.e. when two constructed composites were presented together). Performance did not increase significantly when one constructed composite (memorial information) was presented with an automatically generated image (more information). Therefore, while it is expected that the 'perfect' descriptions will perform better than the participant generated descriptions when presented alone, it is unclear how they will perform when combined with a composite. The results from chapters 2 and 3 indicate that there may be little increase when they are combined with a composite. However, while the previous chapters have examined the combination of more versus varied information, they have not examined performance for combined memorial and 'optimal' information. Combining a composite with accurate descriptive information should increase performance above the level observed for the composite alone. However, it is unclear whether performance will increase above the level observed when the 'perfect' description is presented alone i.e. it is unclear what information participants will use to make an identification decision. If participants only use the descriptive information, then performance for the combined condition should be equivalent to the description alone condition. However, if the 'perfect' descriptions differ greatly from the composites and participants use both types of information to make a decision, performance may be lower.

Experiment 12: Why do composites and descriptions perform well?

This experiment examined performance for composites and descriptions in more detail. The first aim was to attempt to replicate the results from experiment 11 with a different set of targets. In addition to examining performance for the combined presentation of composites and descriptions, this experiment will also compare performance for descriptions from memory with those generated by a group of judges with the target present ('perfect' descriptions). As the results from both chapters 2 and 3 indicated that performance increased only when memorial information was combined at test, the 'perfect' descriptions will also be presented with the 'participant witness' generated composites. While it is expected that the 'perfect' descriptions will perform better than the participant generated descriptions when presented alone, the results from chapter 2 indicate that there may be little increase when they are combined with a composite.

Materials

This investigation used the PROfit composites that had been constructed for Frowd et al (submitted), and experiment 7 (see experiment 7, pg. 138 for an example). In experiment 7, unfamiliar participants were asked to study a photograph of a famous male for one minute. Two days later they were asked to describe the face (cognitive interview) and construct two composites of the target, one PROfit and one EvoFIT (Stage 1). However, in order to ensure that performance for the PROfit composites was not affected by construction of an EvoFIT another five participants were recruited. These additional participants followed the same procedure (e.g. constructing both a PROfit and an Evo-FIT of the same target), however the order of construction was such that all five participants constructed the PROfit first. This ensured that for all ten targets in this investigation, the PROfit was constructed first and was not affected by construction of an Evo-FIT. The five additional EvoFIT's were not used but instead served to keep the procedure the same. All composites and descriptions were presented on a single sheet of A4 paper (landscape) together with a target present array (one photograph of the target and five distractors).

'Perfect' descriptions

Five undergraduate psychology students were asked to compile target present descriptions of all ten targets, as part of a larger cognitive psychology project. Targets were allocated such that each 'perfect' or prototypical description contained information from two students. Each student was given a photograph of a target, together with a standard description sheet (i.e. identical to those used during cognitive interview). The description sheets contained a list of the seven main physical features (e.g. face shape, hair, eyebrows, eyes, nose, mouth, ears) with a space between each feature to write a description (see appendix). Students were instructed to describe all features.

Participants

Forty participants were recruited from Stirling University. They consisted of postgraduate students, secretarial and technical staff.

Design

There were ten 'participant witnesses' and ten targets (1 participant per target). This resulted in a total of ten descriptions and ten composites (generated by the same participant witnesses). This resulted in a total of fifty presentations. A 5 (type of presentation) by 10 (target) within-subjects design was adopted. The types of presentation were composite, description, 'perfect' description, composite with corresponding description presented together (i.e. the composite and description from the same participant) and composite and corresponding 'perfect' description presented together (i.e. composite and target present description of the same target). To ensure that participants only saw one type of presentation for each of the ten targets, the images were divided into 5 presentation booklets using a Latin square design. Each booklet contained ten presentations, one of each target and two of each type. All images were presented with a target present array. Order of presentation was rotated within each booklet. There were eight attempts per booklet.

Procedure

Participants were told that the composites and descriptions were constructed from memory and that they represented a likeness of the target. Participants were informed that when a composite and description were presented together, these had been generated by the same 'participant witness' and therefore represented the same person. They were asked to examine the images closely and were told that the target may or may not be in the array. If participants thought that the composite and/or description did not resemble anyone in the array, they were asked to reject the array. If they thought that the target was present, participants were asked to point to the appropriate photograph. This was repeated for all ten targets. No time limit was placed on this procedure.

Results

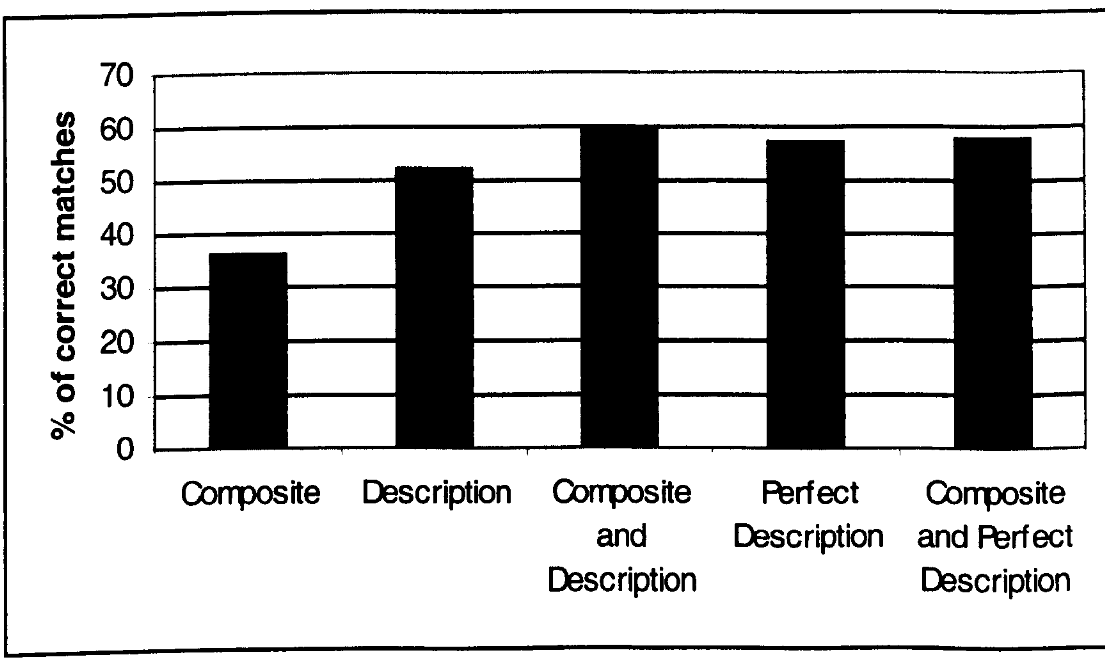


Figure 27: Percentage no of correct matches for each condition

The percentage of correct matches are displayed in figure 27. The data was initially collapsed across target and a Friedman test revealed that the overall differences did not quite reach significance (X^2=8.387, DF=4, p=0.078). However, as the primary aim of this investigation was to examine whether performance would increase when composites and descriptions were presented together, further analyses was undertaken using Wilcoxon signed ranks tests. These revealed that the combined composite and description condition (mean 1.2, s.d. 0.7) performed significantly better than the composite alone (mean 0.7, s.d.0.8) condition ($p<0.05$). This therefore provides further support for experiment 11 by indicating that the combination of memorial information from both a composite and a description serves to increase performance above the level observed for a single composite. Analyses for each type and target (Cochran Q and Mcnemar) revealed that the combined composite/description performed better than the single composite for eight of the targets and significantly higher for four targets.

The descriptions performed well, again providing support for both experiments 10 and 11 by indicating that descriptions do contain useful identifiable information. However, while they performed better than composites when presented alone, this difference did not reach significance ($p=0.095$). Interestingly, the 'perfect' descriptions did not perform significantly better than the participant generated composites ($p=0.528$). This initially suggests inadequacies with the 'perfect' descriptions, however performance levels for both the generated and 'perfect' descriptions were fairly high (52% and 57% respectively) and the mean no of descriptors was similar (21.5 from memory and 22 with target present). However the information contained within the descriptions may have had an effect on performance, as the target present descriptions did not contain any attributional information. While Shepherd et al (1988) reported that attributional descriptors alone performed poorly in identification tasks, the combination of attributional and physical feature descriptions in the participant generated descriptions (from memory) might have been a useful identification aid, thereby increasing performance to the level obtained by the 'perfect' descriptions.

Similarly, the combined composite/perfect description did not perform significantly better than the combined composite/participant generated description ($p=0.909$). This again indicates the similarity between the two types of description but also indicates that participants may have relied more on the descriptive information when making a decision. Furthermore, adding a 'perfect' description to a composite did increase performance significantly above the level observed for a single composite ($p<0.05$), however the combined composite and 'perfect' description did not perform significantly better than the perfect description alone ($p=0.876$). This again indicates that participants may have relied more on the descriptive information. Indeed, an examination by target revealed that for all ten targets there was no significant difference in performance between the generated description and combined/generated description condition, indicating that for both combined conditions, participants relied heavily on the descriptive information.

This is contrary to the results reported in previous chapters, as both chapters 2 and 3 indicated that the combination of information increased performance, rather than the reliance on one particular composite. However, the previous chapters only combined pictorial information whereas this investigation combined both verbal and pictorial information. During the experiment, it was noted that participants preferred to read the description first, comparing each descriptor to the photographs in the array. Therefore, participants had begun to make a decision before they had examined the information contained within the composite. In contrast, when participants were presented with two composites (chapter 2) they seemed to find it easier to extract information from both images at the same time. This difference in strategy may explain the difference in results.

However, while the combined conditions did not perform significantly better than the descriptions, combining a composite with a description did perform significantly better than a composite. This is particularly important as it indicates that descriptions do contain useful information. Of particular interest is the finding that all participants gave full and detailed descriptions after a two-day delay. The mean no of descriptors reported in this investigation was 48.5 (SD 10.3), compared to 21.5 (experiment 9) and 33.3 (experiment 10). The higher number of descriptors in this experiment may therefore reflect rehearsal and consolidation processes. Ellis, Shepherd & Davies (1980) previously reported no difference performance for descriptions that had been provided either one hour or one day after target exposure. Therefore, an examination of the relationship between the failure to recall key features on composite construction may require a more substantial delay of at least one week.

General Discussion

The results from these experiments are particularly important for police operatives, who use the amount of information recalled to assess eyewitnesses. Experiments 9 and 10 revealed that there is no relationship between the amount of information recalled and the quality of the resulting composite. In addition, experiment 10 revealed that there is no relationship between the amount of description recalled and the accuracy of that description. While there appears to be no clear relationship between descriptions and composite performance, experiments 11 and 12 indicated that descriptions are useful and can improve performance when combined with a composite at test. This suggests that the emphasis on descriptions should be changed from one of trying to identify the 'best' witness, to trying to improve composite performance.

The results from experiments 9 and 10 therefore provide no support for the suggestion that participants who provided larger descriptions (i.e. those containing more descriptors) were more likely to produce better quality composites (Christie and Ellis, 1981). Instead, the results are similar to those obtained by Gabbert et al (2003) who also found no relationship between descriptions and composite performance. Similarly, the results also support recognition studies (e.g. Pigott & Brigham, 1985; Wells, 1985; Shepherd, Davies and Ellis, 1988), by clearly indicating that there is no relationship between the amount of information recalled and the quality of the resulting composite. Similarly, the results provide little support for the suggestion that shorter, more precise descriptions may be more effective than longer descriptions (Shepherd, Davies and Ellis, 1988). Interestingly, all of the descriptions used in this investigation contained substantially more descriptors than those obtained by Shepherd et al (5.7 compared to 21.5, 33.3 and 48. 5 in this investigation), which may reflect the use of cognitive interview techniques. Similarly, descriptions still performed well after a substantial delay between exposure and interview (2 days).

These results indicate that descriptions contain useful information but the absence of any relationship between descriptions and composite performance suggests that the amount of verbal description is a poor indicator of composite performance. However, verbal descriptions are an important source of information, therefore experiments 11 and 12 attempted to build on the results from chapters 2 and 3, by examining whether performance would increase when composites and descriptions from the same participant witness were combined at test. Both experiments reported that performance did increase significantly above the level observed for a single composite.

In addition, experiment 12 attempted to examine why performance increased when a composite was presented with a description. The results indicated that participants tended to rely more on the information contained in the description rather than the information contained within the composite. This is contrary to the results from chapters 2 and 3 and may reflect different strategies adopted for examining verbal and pictorial information. However, the finding that participants relied more on the description together with the results indicating that the combined presentation of composites and descriptions increased performance, indicates that the descriptions contained useful identifiable information. Future work could examine the effect of presenting verbal and pictorial information further, by randomly assigning a composite to a description. Indeed, in experiment 12 the descriptions from memory performed as well the 'perfect' descriptions. While this initially indicated inadequacies in the 'perfect' descriptions, performance was still relatively high. While it is impossible to compare witness descriptions with target present descriptions in a real-life situation, as the target is unknown, these results indicate that descriptions are useful investigative tools.

The results from experiment 12 indicate that perhaps descriptions should be presented rather than composites. However, all participants in these experiments reported a preference for composites and descriptions together rather than descriptions alone. Similarly, in a real-life situation it may perhaps be more useful to display a composite with a description in a newspaper, rather than just a description.

One important aspect of these experiments concerns the issue of ecological validity. One of the difficulties of assessing the relationship between descriptions and composites under ideal learning conditions is that all participants recalled information about most features. As stated earlier, most composite operators will not invite a witness to construct a composite if they cannot recall more than two features. However, it was impossible to assess the importance of this in these experiments. In all experiments participants recalled information about most features, with only a small number of participants failing to recall information about either the eyebrows, ears or nose. Future experiments should therefore adopt a more ecologically valid procedure using both a live target and a substantial delay (at least one week) between target exposure and the cognitive interview, in order to assess the impact of the failure to recall key features on composite quality.

However, despite this, the results clearly indicate that there is no relationship between the amount of information recalled and the accuracy of both the description and resulting composite. Similarly, the results also indicate that descriptions do contain useful information and can improve performance when combined with a composite at test. This suggests that verbal descriptions should not be used to assess which witness will construct a composite. Instead the results indicate that the emphasis should be moved away from using descriptions as 'identifiers', to using descriptions to try to improve composite performance.

5

Review, future work and conclusions

This final chapter will begin with a recap of the original questions posed in this thesis. It will then review the work in the preceding chapters and discuss the findings from all of the reported experiments, drawing conclusions and offering recommendations for future work.

Thesis questions

As stated in chapter 1, the aim of this thesis was to improve performance of facial composites. The starting point for these investigations was that while a great deal of research has evaluated the effectiveness of composite systems, this research has tended to focus on the difficulties associated with construction such as inadequacies in system design and memorial difficulties. As a result, much is known about the issues that affect construction of an unfamiliar face from memory. However, very little work has examined whether composite performance can be improved at test. This thesis therefore attempted to examine methods of improving facial composites both during construction and at test. One of the major findings of composite research to date is that despite technological advances in system design, composites still do not perform particularly well. This indicated that poor composite likeness might be a result of memorial difficulties. As constructing a facial composite is a retrieval task, memorial difficulties may in fact be a result of system design. For example, a witness will have viewed a three-dimensional face, yet they are asked to construct a two-dimensional full-face composite. As unfamiliar face memory is image or context-specific, the disparity between the view at encoding and the view at retrieval may result in poorer quality composites. Therefore, the first question addressed whether composite systems may be more effective if they displayed faces in a way revealing more three-dimensional information. The second question followed on from this, but instead of examining whether three-quarter view composites would aid construction, it asked whether the presentation of a full-face and three-quarter-view composite from the same witness would increase performance at test. The third question expanded on the role of composites as effective retrieval cues and examined the issue of encoding specificity. In particular, it asked whether composite performance would increase when the view at encoding more accurately matched the view at retrieval.

Therefore the questions that were posed in chapter 2 examined whether the amount and/or type of information that was presented both during construction and at test could increase performance. Chapter 3 expanded on these experiments by examining whether multiple types of information could increase performance at test. In particular, it asked whether the presentation of composites from multiple witnesses could increase performance. This work initially expanded on the work of Bruce et al (2002) by examining performance for composites from four different witnesses using a more ecologically valid procedure, than in Bruce et al's initial work (Bruce et al, 2002, experiment 1). This chapter then went on to ask whether combining composites from different witnesses resulted in an image that was closer to the ideal, or prototypical image.

The previous chapters therefore examined whether the presentation of different types of information could increase composite performance. One important type of information that is elicited from a witness is a verbal description. Due to the importance of verbal descriptions chapter 4 therefore initially asked whether there was any relationship between the amount and accuracy of description and performance of the resulting composite. This chapter then expanded on the preceding chapters by asking whether the presentation of different types of information from a single witness (a description and a composite) would increase performance at test.

The following sections of this chapter will review the experiments reported in the preceding chapters. Each thesis question will be addressed separately, new questions will be posed and further research will be suggested.

Will the development of a three-quarter-view composite database improve composite performance?

Is a three-quarter-view better than a full-face view?

Chapter 2 initially asked whether a three-quarter-view database would aid construction of composites. More specifically, would the presentation of composites in a way revealing more three-dimensional information aid construction? The results suggest that three-quarter-view composites do act as an efficient retrieval cue, as performance was significantly better for the constructed three-quarter view composites compared to the automatically generated images. That is, when the images were automatically generated and participant-witnesses had no opportunity to interact with the composites, they performed poorly. However, the results also suggested that three-quarter-view composites were not more effective than full-face composites. This result was surprising as some evidence suggests that a witness may have encoded and stored a three-dimensional representation of the face (e.g. Schiff, Banka & De Bordes Galdi, 1986; Bruce & Valentine, 1988; Pike, Kemp, Towell & Philips, 1997). The principle of encoding specificity (e.g. Thomson & Tulving, 1970; Tulving & Thomson, 1973) states that recall will be more effective when the cues available at retrieval more accurately match those at encoding. As such, using more three-dimensional retrieval cues may have aided both recall and reconstruction of the seen face. However, one difficulty with experiment 1 is that participant-witnesses were exposed to all views of a target face. As a result, participants may have been exposed to enough 'instances' of each view at study, to ensure successful generalisation to both views at construction. Therefore, instead of providing support for encoding specificity, these results are similar to those obtained in recognition tasks (e.g. Hill, Schyns & Akamatsu, 1997; Schyns & Bülthoff, 1994), by indicating that when all views are presented at study, no one view is preferred at test.

Are two views better than one?

The second question asked in chapter two was whether the development of a three-quarter-view database would aid performance at test. As full-face and three-quarter-view composites performed equally well, experiments 2 and 3 examined whether the presentation of both a full-face and a three-quarter-view composite from the same participant would increase performance at test. Experiment 2 examined whether 'adding' a three-quarter-view to a full-face composite would increase performance, and experiment 3 examined performance for two views versus one under differing encoding conditions. The results from these experiments suggest that performance does increase when more information is presented. However, this effect was only found when both composites had been constructed. When one composite had been automatically generated performance decreased. The automatically generated composite was essentially the same composite in a different view i.e. the same features were presented. Therefore, the benefit for presenting two composites in differing views cannot be solely attributed to the fact that more information was presented. Instead, it appears that it may be the presentation of varied information that serves to increase performance. Often two views of the same person can look more different than two different faces in the same view. This may be because faces contain a great deal of information (structural, featural, configural) and this information can look very different in differing views. For example, a three-quarter view may display more structural information (e.g. face shape and outline of the nose), whereas a full-face view may display more featural information (e.g. mouth and eye regions).

However, if it were just the case that performance increased because composites in differing views display different types of information, then the automatically generated composites should have performed better. Instead, what may have happened is that the information that was retrieved during construction reflected the view that the composite was constructed in. Therefore, the three-quarter-view composites may have contained more accurate (i.e. more similar to the target) structural information (e.g. face shape and outline of the nose), while the full-face composites may have contained more accurate featural information (e.g. eye and mouth regions). The differing amounts of more accurate information contained in the composites may explain why the full-face and three-quarter-view composites achieved equivalent levels of performance when presented alone. There may not have been enough accurate information in either type of composite to generate an increase in performance when presented alone. However, when both types of composite were presented together, it may have been possible to extract information from both views and therefore reach a more accurate identification decision. Therefore, the combination of information from both types of composite may have resulted in the perception of an image that was closer to the ideal image. The apparent benefit for presenting varied rather than just more information supports previous research that has examined the effect of presenting composites from multiple witnesses (e.g. Bennet et al, 1999; Bruce et al, 2002) and is consistent with the findings reported in chapters 3 and 4.

Therefore, the results from both experiments 1 and 2 indicate that performance for full-face and three-quarter-view composites is similar when all views of a face have been encoded. However, the information that is contained within the composites may differ and the combination of both types of information may serve to increase performance.

What happens when a witness has only seen one view of a face?

Experiment 3 expanded on these results and examined the pattern of viewpoint dependence in composite construction. In particular, this experiment explored the issue of encoding specificity and examined whether composite performance would increase when the view at encoding more accurately matched the view at retrieval.

From an applied perspective, this would tell us how composite performance differs when a witness has only seen a suspect in a single view. The results from this experiment initially provided moderate support for the encoding specificity hypothesis, by revealing that three-quarter-view composites achieved a higher level of performance when a three-quarter-view had been presented at study, compared to a full-face view. Similarly, full-face composites performed slightly better when a full-face view had been encoded, compared to a three-quarter view. However, when the composites were evaluated using a six alternative forced choice array task, the effect for the full-face composites disappeared. Furthermore, while there was a significant increase in performance when all views of a face had been presented, compared to a single view for the full-face composites, the three-quarter-view composites achieved equivalent levels of performance in both the three-quarter and all view encoding condition.

Unfamiliar face memory has been shown to be image or context specific (e.g. Bruce, 1982; Bruce and Young, 1986) due to a reduced ability to form a robust structural representation of the face. Therefore, it is perhaps surprising that performance for the full-face composites was not better when the images were matched for view. Similarly, the encoding specificity principle suggests that recall will be better when the view at encoding more accurately matches the view at retrieval.

These results may suggest that the encoding specificity effect was found for three-quarter-view composites but not for full-face composites. However, another possible explanation for these results may be the symmetry argument proposed by Poggio & Vetter (1992). This states that learning one view of a bilaterally symmetrical object can be sufficient to generalise to other views. As a face is generally bilaterally symmetrical, then a side view (the symmetrical view), which is non-singular, may contain enough information to generalise to other views. However, as a full-face view is singular, generalising from this view to other views is extremely difficult (Hill, Schyns & Akamatsu, 1997; Schyns & Bülthoff, 1994). This hypothesis indicates that generalisation from a side-view is greatest to the symmetrical side view and decreases with angle of rotation much like an inverted U-shape. This may explain the results of the three-quarter-view composites, however it doesn't explain why performance for the full-face composites was poor in the full-face condition, neither does it explain why the three-quarter-view composites achieved similar levels of performance both when a three-quarter-view and when all views were presented at study.

One possible explanation for the poor performance of full-face composites comes from unfamiliar face memory research. As there is a restricted opportunity to build a robust structural representation of a face, the information that will be dominant in memory will be the information that occupies most of the image – the hair and face shape (e.g. Bruce et al, 1999; Bonner et al, 2003). As performance for the full-face composites was generally poor when a full-face view had been presented at study compared to all views, this suggests that a full-face view does not display enough information about the structural properties of the face.

Furthermore, the finding that three-quarter-view composites performed similarly when both a three-quarter-view and when all views of a face had been presented, suggests that similar types of information had been encoded and stored. Therefore, it is possible that three-dimensional structural information had been encoded from both presentations and resulted in similar levels of performance. However, it could be argued that if the encoding of three-dimensional structural information was responsible for these performance levels, then perhaps the full-face composites should have performed better when a three-quarter-view had been encoded. However, as the symmetry hypothesis suggests that performance levels decrease with increasing angle of rotation, then performance would be lower for full-face composites compared to three-quarter-view composites. Similarly, full-face composites performed significantly better when all views had been encoded, again suggesting that the encoding of three-dimensional information is important in composite construction.

While there is a suggestion that movement may help to build a more robust three-dimensional representation of a face (e.g. Schiff, Banka & De Bordes Galdi, 1986; Bruce & Valentine, 1988; Pike, Kemp, Towell & Philips, 1997), the results from this investigation do not support this. All of the videos used in this investigation depicted the target moving. In both of the single view conditions the target was seen talking (played without sound) nodding and expressing. Therefore, if movement per se was responsible for building a more robust three-dimensional representation of the face, then performance levels for the full-face and three-quarter-view composites may have been expected to be more similar. These results therefore provide tentative support for other research which has reported no beneficial effects of movement in face recognition (e.g. Christie & Bruce 1998; Bonner, Burton & Bruce, 2003). However, it is difficult to pinpoint the precise role of movement and perhaps a further investigation could examine the effect of presenting static with moving images.

However, the main reason for using moving video images was to attempt to emulate everyday interaction. Presenting a moving, three-dimensional target is more ecologically valid than presenting a static two-dimensional photograph of a target. However, one of the difficulties of conducting applied experimental eyewitness research is the issue of ecological validity. While the experiments in chapter 2 used an ecologically valid experimental procedure (i.e. unfamiliar witnesses, familiar identifiers), it is unclear whether the same patterns of viewpoint dependence would emerge using a 'live' target in a mock crime scenario. Therefore further research should be conducted in order to assess the impact of the effects of view on composite construction.

What are the practical implications?

The development of a three-quarter-view database has clearly aided composite performance and furthered our understanding of view effects in composite construction. In particular, experiments 2 and 3 indicated that presenting two views from the same witness increased performance above the level observed for a single composite. Similarly, important patterns of viewpoint dependence were observed in experiment 3.

When a face has only been seen in a three-quarter-view, the results suggest that it may be *better* to construct a three-quarter-view composite. When a face has been seen in a full-face view performance is low for both full-face *and* three-quarter-view composites. However, when all views of a face have been encoded, composite performance is equally good in both views. Therefore when a witness has seen a side view of a suspect the results indicate that a three-quarter-view composite should be constructed. However, while standard full-face composites perform well when all views of the face have been encoded, care should be taken when a person has only seen a face in a full-face view, as composites in both views achieved low levels of performance when a full-face view had been encoded.

At present, it is unclear whether view effects are taken into consideration before constructing a composite with a witness. The only ACPO guideline (Association of Chief Police Officers) that makes any reference to how the witness viewed the face states that "It is essential that the witness has seen the suspect's face and is able to visualise or describe the facial features. The witness must have a clear mental image of the person and more importantly can visualise or describe, however simply, the facial features to be reproduced..." (pg.'s 9-10). While this statement makes it clear that a witness must be able to visualise the face, it makes no reference to the view that the suspect was seen in. In this investigation, all participant-witnesses could clearly describe the face despite seeing the targets in differing views.

Due to the pattern of viewpoint dependence observed in this investigation, it perhaps seems important that composite operators should pay close attention to view effects. However, perhaps the reason that view has not been considered, is that construction of a composite in a view other than full-face has not been previously possible. However, even for full-face composites the results indicate a significant difference in performance when a single view has been encoded compared to all views. Similarly, given that the three-quarter-view database is now an integral part of the PROfit composite system and is now available to field operators, the effects of view both at the time of encoding and at construction should be considered carefully.

Will the use of composites from multiple witnesses increase performance?

Chapter 3 initially asked whether the use of composites from multiple witness would serve to increase performance. The results from experiment 4 indicated that combining composites from four different witnesses resulted in an image that performed as well as or better than the best single composite. This supported previous research (McNeil et al, 1987; Bruce et al, 2002, experiment 1) and initially suggested that morphed composites performed well because they contained a combination of different memorial representations. Experiment 5 examined this in more detail and asked whether morphed composites performed well because they just contained more information, or whether they performed well because they contain varied information (memorial representations). In order to examine this, composites from both single (more information) and multiple witnesses (varied information) were morphed. The results indicated that the morphs from multiple witnesses (4-Morphs) performed significantly better than the morphs from single witnesses (mini-morphs). These results therefore indicate that morphed composites perform well because the combination of memorial representations results in an image that may be closer to the ideal image or prototype (e.g. Solso & McCarthy, 1981; Bruce, Doyle, Dench & Burton, 1991; Homa et al, 2001).

Similarly, the results from the identification task in experiment 1 also suggested that it may be possible to perceive an image that is closer to the ideal image (e.g. Leopold et al, 2001) when multiple composites were presented. Bennet et al (1999) had previously reported an increase in performance when four composites of *similar* quality were presented. However, this investigation has extended this research by revealing that four composites of varied quality (from worst to best) could also perform well. This initially suggested that the differing quality of each composite may have affected the perception of subsequent composites, as reported by Webster & MacLin (1999) and Leopold et al (2001) and resulted in the *perception* of an image that is closer to the ideal or prototype. However, experiment 3 investigated this in more detail by asking participants to provide self-reported identification strategies. The results indicated that overall participants preferred using only one composite to base their identification decision and identifications based on a single composite were generally more accurate. This suggested that participants were not referencing information from all composites and provided further support for the presentation of a single morphed composite.

However, while the self-reported identification strategies revealed that participants preferred to use only one composite to make their identification decision, presenting composites from four different witnesses can perform well (experiment 6) and often achieve levels of performance that are better than a single composite (experiment 4 identification task). This suggests that it may in fact be possible to reference information from multiple composites and reach a more accurate identification decision.

Why do multiple composites perform well?

The finding that multiple composites perform well, either when presented together or when presented as a new morphed image, may highlight something about the nature of the information contained within the composites. Due to the restricted ability to form a robust structural representation of an unfamiliar face, the information that will be dominant in memory will be the hair and face shape. Therefore, a witness will be more likely to accurately recall the hair and face shape of a seen face, than they will the internal features. However, would presenting four composites that all contain fairly accurate hair and face shape and inaccurate internal features increase performance? As research has demonstrated that familiar faces are easier to identify and match using internal rather than external features (Bruce et al, 1999; Ellis et al, 1979; Young et al, 1985) due to a more robust structural code (Bruce & Young, 1986), it is unlikely that presenting composites that only contained similar hair and face shape information would increase performance.

However, memory research (e.g. Hintzman, 1986, 1988; Metcalfe, 1990) indicates that witnesses may recall different types of information. In particular, Hintzman suggests that traces or parts of original events/items are stored in memory and that it is these combined traces that are retrieved. This output is therefore much like a schema of the original event or item. As encoded information is stored and retrieved according to an existing schema, the information that it will be combined with will be particular to the individual. Therefore, when this information is retrieved during composite construction, a witness will choose a feature that is similar to their existing schema of the face/feature. Consequently, each witness may choose a different 'similar' feature. While Hintzman did not specifically examine memory for faces, other researchers have also suggested that faces are processed according to existing schemas (e.g. Rakover, 1999, 2002). This research indicates that each composite will contain differing amounts of similar and dissimilar information and importantly, these differences are unlikely to be correlated.

Similarly, Metcalfe (1990) suggests that memory for the original target may become integrated with memory of another face/features, stating that "It is possible that a lure that was never viewed before may be 'better' recognised than even the actual target face" (pg158). In particular, research on the effect of prototype faces has indicated a potential memory integration effect. For example, Solso & McCarthy (1981) reported that 35 out of 36 participants reported seeing an unstudied prototype face. Thus suggesting that memory for faces/features is somehow integrated with memory for similar items, either through trace blending (e.g. Metcalfe, 1990), by storing integrated traces (e.g. Hintzman, 1986) or as a result of competing patterns of interaction (e.g. McClelland & Rumelhart, 1985, 1996).

While this research indicates that information can become combined in memory and result in recognition of an unstudied prototype, it appears that combining memorial representations by morphing them together at test also results in an image that is closer to the ideal or prototypical image. Similarly, presenting multiple composites may also result in an image that is closer to the prototype. While the self-reported identification strategies in experiment 6 revealed that participants generally preferred to use one composite to make an identification attempt, sets of four composites still perform well. Therefore, a similar prototype effect may be occurring that is perhaps automatic in nature. In particular, Leopold et al (2001, pg89) state that "The encoding of faces...draws upon...mechanisms that reference the central tendency of the stimulus category" (pg89), which is much like the prototype effect.

Is this effect specific to a particular composite system?

While it as evident that morphed composites performed well using PROfit, it was unclear whether this effect would still be obtained when composites from different composite systems/techniques were combined. As these techniques result in images that contain very different types of information it was uncertain what effect these types of information may have on the resulting morph. Therefore, experiment 7 attempted to replicate the 4-Morph effect with composites from different witnesses and systems and investigated whether adding a different type of image could increase performance for a single composite. The results revealed that combining composites from four different witnesses and systems resulted in an image that performed significantly better than a single composite. This initially suggested that performance for a single composite could be increased when textural information was added (EvoFIT) but not featural/shape information (Sketch). Experiment 8 investigated whether the benefit for adding an EvoFIT was due to the combination of memorial or composite (texture/shape) information. PROfits and EvoFITs were morphed using composites that had been constructed by the <u>same</u> witness and the results indicated no increase in performance when both images were morphed. This suggests that the benefit observed in experiment 7 was primarily due to the combination of memorial rather than composite (texture) information.

The results from all of the experiments in this chapter indicate that performance for morphed composites is dependent on the <u>type</u> of information combined. The combination of <u>any</u> four composites results in a significantly better image, when <u>different</u> witnesses have constructed the composites. Similarly, a smaller increase in performance may be observed when composites from two witnesses are combined

(Experiment 4 & Bruce et al, 2002, experiment 1). However, this 2-Morph effect is dependent on the type of composites that are combined. Performance increases when two PROfits are combined (Bruce et al, 2002, experiment 1) and when a PROfit/E-FIT and EvoFIT are combined (experiment 4) but not when a PROfit/E-FIT and Sketch are combined.

What are the practical implications

These results are particularly important with regard to current construction procedures, as they indicate that the best performance levels are obtained when composites from four different witnesses are combined. At present only one witness is invited to construct a composite, while other witnesses are kept for other identification procedures (e.g. line-ups). As such, the current ACPO guidelines state that "The production of a composite image with multiple witnesses must not be attempted, as this will amount to cross contamination of each witness's primary memory" (pg11). While this guideline does not specifically prohibit the construction and combination of composites from more than one witness, it does not actually state that composites from more than one witness *can* be constructed, nor does it state that these composites can be used in combination.

As there is no way at present, of determining which witness will construct the 'best' quality composite, the results from this investigation and previous research (e.g. McNeil et al, 1997; Bennet et al, 1999; Bruce et al, 2002, experiment 1) all indicate that it would be beneficial to obtain composites from multiple witnesses. As the experiments in chapter 3 have revealed, the best and most consistent performance is observed when composites from four different witnesses are combined, while smaller increases in performance can be obtained when composites from two witnesses are combined.

What is the relationship between verbal descriptions and composite construction?

Chapter 4 initially asked whether there was any relationship between the verbal description that is initially provided by a witness and performance of the resulting composite. Experiment 9 examined whether the amount of verbal description provided was related to composite quality. The results revealed that there was no relationship between these factors. In addition, the results from the verbal description data revealed that more external feature information was recalled than internal feature information. This supports face recognition research that has demonstrated that unfamiliar face memory is dominated by external features (e.g. Bruce et al, 1999; Bruce & Young, 1986; Ellis et al, 1979; Young et al, 1985). However, the amount of verbal description provided did not support previous research (e.g. Shepherd et al, 1988). The larger amount of description provided in this investigation perhaps reflects the use of cognitive interview techniques to elicit the verbal descriptions.

Experiment 10 then extended this by examining the relationship between the amount and accuracy of descriptions for a different set of targets. The results supported those obtained in experiment 9 by revealing that significantly more external feature information was recalled than internal feature information. However, there was no clear relationship between the amount and accuracy of descriptions.

The results from experiments 9 and 10 therefore provide no support for the suggestion that participants who provided larger descriptions (i.e. those containing more descriptors) were more likely to produce better quality composites (Christie and Ellis, 1981). Instead, the results are similar to those reported in recognition studies (e.g. Pigott & Brigham, 1985; Wells, 1985; Shepherd, Davies and Ellis, 1988), by clearly indicating that there is no relationship between the amount of information recalled and the quality of the resulting composite. Similarly, these results provide no support for verbal overshadowing; the suggestion that providing a verbal description may impair later performance (e.g. Dodson, Johnson & Schooler, 1997; Fallshore & Schooler, 1995; Schooler & Engstler-Schooler, 1990). Instead, they provide support for research that has demonstrated no detrimental effect on performance (e.g. Clifford, Clifford & Smith, 2002; Clifford, Burke & Clifford, 2003; Memon & Rose, 2001).

However, as the dual code hypothesis (Paivio, 1986) indicates that both the verbal and visual code may differentially support different types of facial information (verbal = featural, visual = configurational), is it possible that these different types of information are represented in the different tasks? I.e. do verbal descriptions contain more featural information, while the composites contain more configurational information? The assessments of the verbal descriptions in experiments 9 and 10 indicate that a large proportion of information relates to featural information, in particular, focusing on external feature information due to the nature of unfamiliar face memory. However, assessing the amount of configurational information contained within a composite is extremely difficult. However, one method of assessing whether the descriptions and composites contain different types of information is to present them together. Chapters 2 and 3 both demonstrated that the presentation of different types of information increased performance. However, the presentation of similar types of information (more than one composite from the same participant-witness) did not.

Will the combination of a description and a composite from the same witness increase performance

Experiments 11 and 12 therefore examined whether the combination of a description and a composite from the same participant-witness would increase performance. The results from both experiments revealed that combining a composite and a description from the same-participant witness did increase performance above the level observed for a single composite. Interestingly, descriptions performed either as well as or better than composites when presented alone. This not only suggests that participant-witnesses can recall useful identifiable information. It also suggests that composites and descriptions may in fact contain different types of information. As performance for descriptions was often as good as or better than composites when presented alone, this suggests that both the visual and the verbal presentations contain useful identifiable information.

In addition, experiment 12 attempted to examine this further by comparing performance for participant-generated descriptions/composites and 'perfect' descriptions/composites. The results indicated that participants tended to rely more on the information contained in the description rather than the information contained within the composite when making an identification decision. As a comparison of the information contained within the generated and perfect descriptions was not undertaken, it is therefore difficult to ascertain whether the perfect descriptions contained more configurational information than the generated descriptions. However, as both types of description achieved similar levels of performance both when presented alone and when presented together, it can perhaps be assumed that both contained similar amounts of featural information. However, further research will need to assess this formally.

It is extremely interesting that the descriptions performed so well and that participants appeared to rely greatly on the information contained within the descriptions. One possible reason for this may be explained by the nature of unfamiliar face memory. It has been demonstrated that external features dominate unfamiliar face, because they occupy such a large part of the image (Bruce et al, 1999; Bruce & Young, 1986; Ellis et al, 1979; Young et al, 1985). Therefore, it is likely that participants constructed a composite that contained more accurate external feature information compared to internal feature information. However, experiments 9 and 10 also revealed that participants recalled more external feature information, so why would this result in a composite and a description that contained differing types of information?

Furthermore, any difficulties in accessing accurate memorial information due to the location of the face in 'face space' (Valentine & Bruce, 1986; Valentine, 1991a), or because of assimilation or integration effects (Bower, 1967; Hintzman, 1986, 1988; Metcalfe, 1990; Nosofsky, 1999; Rakover, 1999, 2002) may similarly effect both recall tasks. Therefore, it is difficult to ascertain why composites and descriptions might contain differing types of information.

One possible reason however may concern procedural or methodological differences between the two tasks. Composite construction using modern computerised systems such as PROfit attempts to encourage holistic processing. For example, features are presented within the context of a whole face and witnesses are encouraged to correct external featural information prior to internal features, thereby providing an 'accurate' facial context that may promote automatic recognition of features. It is therefore possible that a witness may accurately recognise and reconstruct configurational information. As Rakover (1999, 2002) has suggested that there is a hierarchy of facial schemas, it is possible that a witness may recognise and reconstruct both the overall schema of the face and individual featural schemas, without being able to recognise and reconstruct individual featural components (e.g. colour of eyes).

In contrast, when a verbal description is elicited, recall of featural information is encouraged. In particular, witnesses are encouraged to provide detailed information regarding the shape of the face, hairstyle, eyes, eyebrows, nose, mouth and ears. While witnesses are encouraged to visualise the whole face during the initial recall attempt, further recall cycles ask a witness to focus on specific features, thereby promoting recall of featural rather than configurational information.

These differing tasks therefore appear to result in the 'extraction' of different types of information. This therefore helps to explain why the presentation of a description with its corresponding composite (i.e. the one generated by the same participant-witness) serves to increase performance, as the composite may contain more accurate configurational information, while the description may contain more accurate featural information.

What are the practical implications?

The research described in chapter 4 together with recent research (Gabbert et al, 2003) suggests that there is no relationship between the quality of description and the quality of the resulting composite. This research therefore clearly indicates that verbal descriptions lack any predictive utility. As the amount of verbal description provided is often used by police operatives to assess an eyewitness (Gabbert et al, 2003) this research clearly indicates that descriptions should not be used to assess witness performance.

However, one of the difficulties of assessing the relationship between descriptions and composites under ideal learning conditions, is that all participants recalled information about most features. It was therefore impossible to assess the importance of failing to recall information about key features on composite construction. Future experiments should therefore either use a more substantial delay between target exposure and the cognitive interview, or adopt a more ecologically valid procedure using both a live target and a substantial delay.

Despite this, the results from experiments 11 and 12 clearly indicate that descriptions do contain useful information and can improve performance when combined with a composite at test. Additionally, these results indicate that descriptions and composites contain different types of information and it is the combination of this varied information that serves to increase performance. This research therefore suggests that the emphasis on verbal descriptions should be changed from one of trying to identify a 'good' witness, to trying to improve composite performance.

Concluding comments

A great deal of research has examined facial composite performance and has highlighted many of the factors that make facial recall and composite construction such difficult tasks. However, while this research has increased our understanding of both of the process of recalling an unfamiliar face and the process of reconstructing a likeness of a briefly viewed face, very few improvements in composite performance have been attained. However, the experiments contained within this thesis have clearly demonstrated that composite performance can be improved, using information from both single witnesses and multiple witnesses.

However, while this thesis has demonstrated that facial composites can be improved, the process of conducting this research has raised many more questions about composite construction and has highlighted several areas for future research. The first area of future research concerns the improvement of composites at construction. Chapter 2 asked whether more accurately matching encoding and retrieval cues (view) might serve to increase performance. This was based not only on the encoding specificity principle (e.g. Tulving & Thomson, 1973) but also on research that has revealed that memory for unfamiliar faces is context or image specific (e.g. Bruce, 1982). Chapter 2 did provide some support for more accurately matching view at encoding and retrieval, however this was limited primarily to three-quarter-view composites. As research has highlighted the importance of context in both visual and verbal memory (see Memon & Bruce, 1985-6 for a review), more accurately matching context at encoding and retrieval may yield higher performance levels. Future research should therefore examine the role of context and encoding specificity in more detail.

Potential experiments might examine the impact of the background image during construction, perhaps placing an image of the seen of the crime behind the facial image. Additionally, examining whether composite performance increases when the composite is made at or near the scene of the crime, rather than in a police station may not only provide useful practical information, but would also highlight the role of context effects in composite construction in more detail. Similarly, if context is important during composite construction, it may also be important at the identification stage. Therefore, presenting the final composite against a background depicting the scene of crime may also aid identification.

Another method of more accurately matching the cues available at encoding at retrieval may be to utilise CCTV footage during construction. Often when a crime is committed, the event has been captured on CCTV. If this footage displays the suspect's face then it can be shown in the media, in the hope that someone who is familiar with the person will recognise them (e.g. Bruce et al, 2001). However, if the suspect's face is not visible in the CCTV footage due to the angle/direction of the camera, but has been seen by a witness, this witness may then be asked to construct a facial composite. It would then be interesting to investigate whether this CCTV footage would increase composite performance, by allowing the witness to view the footage during construction.

The second area of future research concerns what makes a good witness. Chapter 3 examined performance for sets of four composites from both multiple and single witnesses. The results revealed a 4-Morph effect for composites from multiple witnesses but not from single witnesses. This suggested that the 4-Morph effect was a result of the combination of different memorial representations, rather than just more information. Similarly, the results also suggested that each individual composite was a deviation from the ideal or prototypical image and that these deviations were unlikely to be correlated. However, what is also interesting is the degree to which each composite differed and why. Why do some witnesses construct 'good' likenesses of an unfamiliar face, while other witnesses produce 'poor' likenesses? Understanding why construction ability differs between witnesses may help lead to a method of identifying a 'good' witness.

One interesting experiment may be to use an eye tracker to examine what individual witnesses look at when they are exposed to an unfamiliar target. Patterns of eye movements could be analysed to assess whether some witnesses focus more and for longer on individual facial features, while others may make more general, global eye movements. This may tell us whether there are individual differences in encoding strategy. The patterns of eye movements could then be correlated with composite performance to assess whether there is any relationship i.e. is performance better when the face has been encoded featurally? Similarly, the pattern of featural encoding could be examined in more detail. Do witnesses focus more on external rather than internal features, as indicated by unfamiliar face research (e.g. Bonner et al, 2003). While there is a danger in assuming that eye movements correspond to attention and encoding (see Christianson, Loftus, Hoffman & Loftus, 1991), examining the type and amount of eye movements may help to highlight potential differences in construction ability between witnesses.

Another method of potentially examining differences between witnesses may be to utilise the remember/know paradigm used by many memory researchers (e.g. Inoue & Bellezza, 1998; Lindsay & Kelley, 1996; Rajaram, 1993). In typical remember/know experiments, participants are required to perform a standard recognition task. In the test phase, they are required to state whether an item is old or new, whether they actually remember the item from the study phase, or whether they are just familiar with it and how confident they are in their judgement. This is based on Tulving's (e.g. 1973) experiential approach that indicates that recognition memory reflects at least two distinct states of awareness. The first is conscious recollection, or remembering, where an item is recognised because it has actually been *remembered* from the original encoded experience (a remember judgement). The second is where an item is recognised because it seems familiar (a know judgement). While research has examined remember/know judgement in recognition tasks, no research has examined these judgements in the context of composite construction.

While a recognition task is very different to a reconstruction task, these judgements could be adapted quite easily to composite construction. When a witness chooses a feature they could be asked whether they remember that the chosen feature is similar (i.e. the witness has a clear image of the original feature and can recognise that this feature is similar), or whether it just seems familiar (i.e. the witness cannot actually remember the original eyes but they just look right). As unfamiliar face research has indicated that external features may be more dominant in memory as they occupy such a large part of the image (e.g. Bruce et al, 1999; Bonner et al, 2003). It would be interesting to examine whether witnesses report more remember judgements when constructing external features and more familiarity (know) judgements when constructing internal features. In addition, these judgements could be compared with eye movements. An analysis of the amount of time spent on each feature or each group of features (internal versus external) could be compared with reported remember/know judgements and performance of the resulting composites.

However, in order to understand why composite performance differs between witnesses, it is important to understand the exact nature of the task. Composite construction is essentially a reconstruction task and as such, it has been noted that this task may have very little in common with standard recognition tasks (e.g. Davies et al, 1978). In order to reconstruct a face, a witness is asked to initially recall the facial features, then when this information is presented in the form of a composite image, the task involves both recognising and recalling whether the presented feature/face 'matches' matches their own internal representation. This task is very different to a standard recognition test, where the task involves recognising whether a presented face or feature had previously been encountered.

As we are much better at recognising faces than recalling them, it has been assumed that composite systems that ask witnesses to recognise whole faces (e.g. EvoFIT), rather than reconstructing the face featurally may be more successful. However, recent research (Frowd et al, 2003a, 2003b) indicates that recognition based systems such as EvoFIT do not perform better than feature based systems such as E-FIT and PROfit. Furthermore, recent research in our own laboratory has revealed a similar pattern of performance for both PROfit and EvoFIT, in that both systems produce better composites when the face has been encoded featurally rather than holistically. This is surprising and perhaps indicates that the assumption that EvoFIT was a holistic, recognition-based system was incorrect.

As remember/know judgements are thought to reflect different states of awareness in recognition memory, they may help to highlight the processes involved in constructing a composite for both standard feature based systems such as PROfit and more holistic systems such as EvoFIT. If composite construction using PROfit is essentially a recall task then there may be no relationship between remember/know judgements and composite construction. However, as it has been assumed that systems such as EvoFIT are more recognition-based, then it would be interesting to examine the pattern of remember/know judgements when constructing an EvoFIT and compare these with the judgements reported when constructing a PROfit.

However, there is a danger in examining memorial processes based on self-reported judgements, as witnesses may not be explicitly aware of the processes they are using to identify a face or a feature. Perhaps more invasive methods such as ERP's and techniques such as functional MRI scans may highlight the exact nature of facial reconstruction. Composite systems such as PROfit and EvoFIT encourage witnesses to construct a face in a certain way, by either asking them to select individual features within the context of a face (PROfit), or by asking them to select whole faces (EvoFIT). However, it is unclear whether individual witnesses are actually constructing the composites in this manner. Are they purely recognising whether faces are similar or dissimilar in EvoFIT or does this comparison process require active searching and recall of individual features? Similarly for PROfit, it is unclear exactly what memorial processes witnesses are using when they construct a composite. As memory research (e.g. Flexser & Tulving, 1978; Tulving & Flexser, 1992) has indicated that recall and recognition are in fact two distinct aspects of the same process, it is likely that witnesses are both recognising and recalling facial information during construction. Similarly, multiple memory theories (e.g. Metcalfe, 1990; Hintzman, 1986) indicate that memory output can be a combination or schema of several similar items. However, while this indicates the difficulties associated with construction, this does not tell us why witnesses differ in their ability to reconstruct a face. While the future research described earlier may not pinpoint the exact reasons why witnesses differ in construction ability, they are directing research in the right direction. A great deal of composite research to date has examined the general problems with construction and as a result much is known about the difficulties associated with constructing an unfamiliar face from memory. However, this thesis has helped to move this research forward by examining methods of improving composite construction. Future research should therefore expand on this by not only examining methods of improving composite construction, but also by gaining an understanding of how and why individual witnesses differ in their ability to construct a facial composite.

Similarly, another very important area of research concerns how composites should be evaluated. Throughout this thesis several different methods have been used to evaluate facial composites. These include identification, likeness ratings and six-alternative forced-choice array tasks. Identification is the most ecologically valid task as it is most similar to a real-life situation, where composites are displayed in the media, or in police stations, in the hope that someone who is familiar with the suspect will recognise them from the image. However, in psychological experiments, where there is often a shortage of participants who are familiar with the target, an assumption has been made that other tasks such as array tasks are equivalent to identification. However, identification is a memorial task whereas matching a composite to an array is a perceptual matching task, therefore, different strategies may be used. As research has demonstrated that face recognition is holistic (see introduction), whereas matching is often feature-based (matching composite and photograph on individual features), matching a composite to an array may not be the best proxy for an identification task. However, recent research in our lab has started to investigate different methods of composite evaluation. This research has indicated that while tasks such as matching are feature-based (participants match individual features rather than the whole face), recognition of composites may also be feature-based. The results from these investigations suggest that as composites are often poor representations of faces, it is necessary to examine individual features in order to reach a recognition judgement. This is contrary to the way in which good representations (e.g. photographs) of faces are normally processed and future work will need to examine this in more detail.

To conclude, this thesis has clearly demonstrated that facial composites can be improved. In particular, it has demonstrated that the presentation of multiple types of information from both single and multiple witnesses can increase performance at test. This thesis has also highlighted many important areas of future research. Of particular interest is gaining a fuller understanding of how and why individual witnesses differ in their ability to construct a composite. Increasing our understanding of individual differences may not only lead to a method identifying a 'good' witness, it may also increase our understanding of the exact processes involved in composite construction. While this thesis has clearly demonstrated that facial composites can be improved, increasing our understanding of individual differences may help to further improve both the construction and identification of facial composites.

References

Alba, J. W. & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93 (2),* 203-231.

Anderson, J. R. & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79,* 97-123.

Association of Chief Police Officers. (1999). *[National working practices in facial imaging].* Unpublished document.

Baddeley, A. (1990). *Human memory: theory and practice.* Lawrence Erlbaum Associates. UK.

Baddeley, A. & Woodhead, M. (1983). Improving face recognition ability. In S. Lloyd-Bostock & B. Clifford, *Evaluating witness evidence,* Chichester, Wiley.

Bennet, P. J. (1986). Face recall: A police perspective. *Human Learning, 5,* 197-202.

Bennet, P., Brace, N., Pike, G. & Kemp, R. (1999). Making the most of E-Fits: Are multiple witnesses more accurate than a single witness? *Paper presented at the 1st joint conference of the American Psychology-Law Society and the European Association of Psychology and Law,* Dublin, Ireland.

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

Bonner, L., Burton, A. M. & Bruce, V. (2003). Getting to know you: How we learn new faces. *Visual Cognition, 10(5),* 527-536.

Bower, G. H. (1967). A multicomponent theory of the memory trace. In K. W. Spence and T. J. Spence (Eds). *The Psychology of Learning and Motivation: Advances in Research and Theory (Vol 1).* Academic Press. New York.

Bower, G. H. & Karlin, M. B. (1974). Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology, 103,* 751-757.

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.* Institute of Forensic Research Publishers. Krakow.

Brown, E., Deffenbacher, K. & Sturgill, W. (1977). Memory for faces and the circumstances of encounter. *Journal of Applied Psychology, 62,* 311-318.

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology, 73,* 105-116. Bruce, V., Doyle, T., Dench, N. & Burton, M. (1991). Remembering facial configurations. *Cognition, 38,* 108-144.

Bruce, V., Hanna, E., Dench, N., Healey, P. & Burton, M. (1992). The importance of 'mass' in line drawings of faces. *Applied Cognitive Psychology, 6,* 619-628.

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

Bruce, V., Henderson, Z., Newman, C. & Burton, A.M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied, 7(3),* 207-218.

Bruce, V. & Langton, S. (1994). The use of pigmentation and shading information in recognising sex and identities of faces. *Perception, 23,* 803-822.

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 (5),* 894-902.

Bruce, V. & Valentine, T. (1985). Identity priming in the recognition of familiar faces. *British Journal of Psychology, 76,* 373-383.

Bruce, V. & Valentine, T. (1988). When a nod's as good as a wink. The role of dynamic information in facial recognition. In M. M. Gruneberg & E. Morris (Eds), *Practical Aspects of Memory: Current Research and Issues.* Lawrence Erlbaum Associates.

Bruce, V., Valentine, T. & Baddeley, A. D. (1987). The basis of the ¾ View advantage in face recognition. *Applied Cognitive Psychology, 1,* 109-120.

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

Bruce, V., & Young, A. (1989). *In the eye of the beholder.* New York. Oxford University Press.

Burton, A. M., Bruce, V. & Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology, 81,* 361-380.

Burton, A.M. & Vokey, J.R. (1998). The face-space typicality paradox: understanding the face-space metaphor. *The Quarterly Journal of Experimental Psychology, 51A(3),*475-483.

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.

Carey, S. & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science, 195,* 312-313.

Christie, F. & Bruce, V. (1988). The role of dynamic information in the recognition of unfamiliar faces. *Memory and Cognition, 26 (4),* 780-790.

Christie, D. F. M. & Ellis, H. D. (1981). Photofit constructions versus verbal descriptions of faces. *Journal of Applied Psychology, 66 (3)*, 358-363.

Clifford, B., Burke, K. & Clifford, C. (2003). Do children or adults exhibit a verbal overshadowing effect? *Paper presented at the Second Joint Meeting of the American Psychology-Law Society, the European Association of Psychology and Law and Australian and New Zealand Association of Psychiatry, Psychology and Law.* Edinburgh, July 9th – 12th.

Clifford, B. & Burke, K. J. (2002). The verbal overshadowing effect in adults and children. *Paper presented at the 12th meeting of the European Association of Psychology and Law*, Leuven, Belgium, Sept 16-17th.

Clutterbuck, R. & Johnson, R. A. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception, 31*, 985-994.

Cutler, B., Stocklein, C. J. & Penrod, S. D. (1988). Empirical examination of a computerized facial composite production system. *Forensic Reports, 1*, 207-218.

Davies, G. M. (1983). Forensic face recall: the role of visual and verbal information. In S. M. A. Lloyd-Bostock and B. R. Clifford (Eds), *Evaluating witness evidence.* John Wiley & Sons.

Davies, G. & Christie, D. (1982). Face recall: An examination of some factors limiting composite production accuracy. *Journal of Applied Psychology, 67(1)*, 103-109.

Davies, G., Ellis, H. & Christie, D. (1981). Exploring new strategies for face recall. *Medicine, Science and Law, 21(2)*, 137-145.

Davies, G., Ellis, H. & Shepherd, J. (1978). Face identification: The influence of delay upon accuracy of Photofit construction. *Journal of Police Science and Administration, 6 (1)*, 35-42.

Davies, G. M., Ellis, H. D. & Shepherd, J. W. (1978) Face recognition as a function of mode of representation. *Journal of Applied Psychology, 63*, 180-187.

Davies, G. & Little, M. (1990). Drawing on memory: exploring the expertise of the police artist. *Medicine, Science and Law, 30 (4)*, 345-353.

Davies, G., van der Willik, P. & Morrison, L. J. (2000). Facial composite production: A comparison of mechanical and computer-driven systems. *Journal of Applied Psychology, 66*, 29-37.

De Haan, E. H., Young, A. W. & Newcombe, F. (1991). A dissociation between the sense of familiarity and access to semantic information concerning familiar people. *European Journal of Cognitive Psychology, 3(1)*, 51-67.

Diamond, R. & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115,* 107-117.

Dodson, C. S., Johnson, M. K. & Schooler, J. W. (1997). The verbal overshadowing effect: Source confusion or strategy shift? *Memory & Cognition, 25,* 129-139.

Dunning, D. & Perretta, S. (2002). Automaticity and eyewitness accuracy: A 10-to 12-second rule for distinguishing accurate from inaccurate positive identifications. *Journal of Applied Psychology, 87 (5),* 951-962.

Edelman, S. & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research, 32,* 2385-2400.

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

Ellis, H. D., Shepherd, J. W. & Davies, G. M. (1975). An Investigation of the use of the Photofit system 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 (1),* 101-106.

Emmet, D., Clifford, B. & Gwyer, P. (2003). The influence of field dependency on eyewitness accuracy in free recall and cued recall. In M. Vanderhallen, G. Vervaeke, P. J. Van Koppen & J. Goethals (Eds). *Much Ado about Crime: Chapter on Psychology and Law.* Politeia. Belgium.

Fallshore, M. & Schooler, J. W. (1985). The verbal vulnerability of perceptual expertise. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 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 (3),* 340-348.

Fisher, R. P. & Cuervo, A. (1983). Memory for physical features of discourse as a function of their relevance. *Journal of Experimental Psychology: Learning, Memory & Cognition, 9,* 130-138

Fisher, R. P., Geiselman, R. E. & Amador, M. (1989). Field test of the cognitive interview: Enhancing the recollection of actual victims and witnesses of crime. *Journal of Applied Psychology, 74,* 722-727.

Flexser, A. J. & Tulving, E. (1978). Retrieval independence in recognition and recall. *Psychological Review, 85 (3),* 153-171.

Flexser, A. J. & Tulving, E. (1982). Priming and recognition failure. *Journal of Verbal Learning and Verbal Behavior, 21,* 237-248.

Flude, B. M., Ellis, A. W. & Kay, J. (1989). Face processing and name retrieval in an anomic aphasia: Names are stored separately from semantic information about familiar people. *Brain and Cognition, 11,* 60-72.

Frowd, C., Carson, D., Ness, H., Richardson, J., McLanaghan, S. & Hancock, P.J.B. (2003). Evaluating Facial Composite Systems. *Psychology, Crime & Law.* Manuscript submitted for publication

Gabbert, F., Dupuis, P., Lindsay, R. C. L. & Memon, A. (2003). Investigating the forensic value of the E-FIT composite program. *Poster presented at the Fifth Biennial Conference of the Society for Applied Research in Memory and Cognition.* Aberdeen.

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

Gibling, F. & Bennet, P. (1994). Artistic enhancement in the production of Photofit likenesses. *Psychology, Crime & Law, 1,* 93-100.

Goldstein, A. G., Johnson, K. S. & Chance, J. (1979). Does fluency of face description imply superior face recognition? *Bulletin of the Psychonomic Society, 13(1),* 15-18.

Haig, N. D. (1984). The effect of feature displacement on face recognition. *Perception, 3,* 505-512.

Hancock, P.J.B., Bruce, V. & Burton, A.M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Science, 4(9),* 330-337.

Hasher, L. & Griffin, M. (1978). Reconstructive and reproductive processes in memory. *Journal of Experimental Psychology: Human Learning and Memory, 4 (4),* 318-330.

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

Hill, H., Schyns, P. G. & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition,* 201-222.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple trace memory model. *Psychological Review, 93,* 411-428.

Hintzman, D. L. (1988). Judgements of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528-551.

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

Homa, D., Smith, C., Macak, C., Johovich, J. & Osorio, D. (2001). Recognition of facial prototypes: The importance of categorical structure and degree of learning. *Journal of Memory and Language, 44*, 443-474.

Inoue, C. & Bellezza, F. S. (1998). The detection model of recognition using know and remember judgements. *Memory & Cognition, 26(2)*, 299-308.

Johnson, M. K., Hashtroudi, S. & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114 (1)*, 3-28.

Johnston, R. A. & Bruce, V. (1990). Lost properties? Retrieval differences between name codes and semantic codes for famous people. *Psychological Research, 52*, 62-67.

Klatzky, R. L., & Forrest, F.H. (1984). Recognizing familiar and unfamiliar faces. *Memory and Cognition, 12(1)*, 60-70.

Kintsch, W. (1970). Models for free recall and recognition. In D. A. Norman (Ed). *Models of Human Memory*, Academic Press. New York.

Koehn, C. E., & Fisher, R. P. (1997). Constructing facial composites with the Mac-A-Mug Pro system. *Psychology, Crime and Law, 3*, 209-218.

Koehnken, G., Schimmossek, E., Aschermann, E. & Hofer, E. (1995). The cognitive interview and the assessment of the credibility of adults' statements. *Journal of Applied Psychology, 80,* 671-684.

Kovera, M. B., Penrod, S., Pappas, C. & Thill, D. (1997). Identification of computer-generated facial composites. *Journal of Applied Psychology, 82,* 235-246.

Krouse, F. L. (1981). Effects of pose, pose change and delay on face recognition performance *Journal of Applied Psychology, 66,* 651-654.

Laughery, K. R., Alexander, J. F. & Lane, A. B. (1971). Recognition of human faces: Effects of target exposure time, target position, pose position and type of photograph. *Journal of Experimental Psychology: Human Perception & Performance, 22(2),* 395-416.

Laughery, K.R. & Fowler, R.F. (1980). Sketch art and Identi-Kit procedures for recalling faces. *Journal of Applied Psychology, 65,* 307-316.

Leder, H. (1999). Matching line drawings of faces. *Perception, 28,* 1171-1175.

Leder, H. & Bruce, V. (1988a). Local and relational aspects of face distinctiveness. *The Quarterly Journal of Experimental Psychology, 51A,* 449-473.

Leopold, D. A., O'Toole, A. J., Vetter, T. & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience, 4 (1),* 89-94.

Lindsay, S. D. & Kelley, C. M. (1996). Creating illusions of familiarity in a cued recall remember/know paradigm. *Journal of Memory and Language, 35,* 197-211.

Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology, 7,* 560-572.

Loftus, E. F. (1979). Reactions to blatantly contradictory information. *Memory and Cognition, 7(5),* 368-374.

Loftus E.F. & Loftus, G. (1980). On the permanence of stored information in the human brain. *American Psychologist, 35,* 409-420.

Loftus, E. F., Miller, D. & Burns, H. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 19-31.

Logie, R. H., Baddeley, A. D. & Woodhead, M. M. (1987). Face Recognition, pose and ecological validity. *Applied Cognitive Psychology, 1,* 53-69.

Lui, C. H. & Chaudhuri, A. (2002). Reassessing the ¾ view effect in face recognition. *Cognition, 83,* 31-48.

Mantwill, M., Koehnken, G. & Aschermann, E. (1995). Effects of the cognitive interview on the recall of familiar and unfamiliar events. *Journal of Applied Psychology, 80,* 68-78.

Maudlin, M. A. & Laughery, K. R. (1981). Composite production effects on subsequent facial recognition. *Journal of Applied Psychology, 66 (3),* 351-357.

McClelland, J. L. & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General, 114 (2),* 159-188.

McKenna, P. & Warrington, E. K. (1980). Testing for normal dysphasia. *Journal of Neurology, Neurosurgery and Psychiatry, 42,* 781-788.

McNeil, J.E. Jr., Wray, J.L., Hibler, N. S., Foster, W. D., Rhyne, C. E. & Thibault, R. (1987). Hypnosis and Identi-Kit: A study to determine the effect of using hypnosis in conjunction with the making of Identi-Kit composites. *Journal of Police Science and Administration. 15 (1),* 63-67.

Meissner, C. A. & Brigham, J. C. (2001). A Meta-Analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology, 15,* 603-616.

Memon, A. (1986). Telling it all: the Cognitive Interview. In Memon, A., Vrij, A. & Bull, R. (Eds). *Psychology and law: truthfulness, accuracy and credibility.* McGraw-Hill. London.

Memon, A. & Bruce, V. (1985-6). Context effects in episodic studies of verbal and facial memory. *Current Psychological Research & Reviews, Winter,* 349-369.

Memon, A. & Rose, R. (2001). Identification abilities of children: Does a verbal description hurt face recognition? *Psychology, Crime & Law, 0,* 1-14.

Memon, A. & Stevenage, S. (1996a). Interviewing witnesses: What works and what doesn't? *Psycholoquy,7 (6), psych.96.7.06.witness memory.1.memon.*

Memon, A., Wark, L., Holley, A., Bull, R. & Koehnken, G. (1996b). Reducing suggestibility in child witness interviews. *Applied Cognitive Psychology, 10,* 503-518.

Metcalfe, J. (1990). Composite holographic associative recall model (CHARM) and blended memories in eyewitness testimony. *Journal of Experimental Psychology, General, 119,* 145-160.

Newell, F. N., Chirozo, P. & Valentine, T. (1999). Recognizing unfamiliar faces: The effects of distinctiveness and view. *The Quarterly Journal of Experimental Psychology, 52A (2),* 509-534.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance, 17 (1),* 3-27.

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

Paivio, A. (1986). *Mental representations: A dual coding approach.* Oxford University Press. London.

Palmer, S., Rosch, E. & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds), *Attention and Performances XI.* Lawrence Erlbaum.

Patterson, K. E. & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory, 3 (4),* 406-417.

Pigott, M. & Brigham, J. C. (1985). Relationship between accuracy of prior description and facial recognition. *Journal of Applied Psychology, 70 (3),* 547-555.

Pike, G. E., Kemp, R. I., Towell, N. A. & Phillips, K. C. (1997). Recognising moving faces: The relative contribution of motion and perspective view information. *Visual Cognition, 4 (4),* 409-437.

Poggio, T. & Vetter, T. (1992). Recognition and structure from one 2D model view: Observations on prototypes, object classes and symmetries. *AI Memo 1347.* http://www.ai.mit.edu/projects/cbcl/web-pis/poggio/memos.htm.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21(1),* 89-102.

Rhodes, G., Brennan, S. & Carey, S. (1987). Identification and rating of caricatures: Implications for mental representations of faces. *Cognitive Psychology, 19,* 473-497.

Rakover, S. S. (1999). Thompson's Margaret Thatcher illusion: When inversion fails. *Perception, 28,* 1127-1230.

Rakover, S. S. (2002). Featural vs. configural information in faces: A conceptual and empirical analysis. *British Journal of Psychology, 93,* 1-30.

Roberts, T. & Bruce, V. (1989). Repetition priming of face recognition in a serial choice reaction-time task. *British Journal of Psychology, 80,* 201-211.

Schiff, W., Banka, L & De Bordes Galdi, G. (1986). Recognising people seen in events via dynamic 'mugshots'. *American Journal of Psychology, 99,* 219-231.

Schooler, J. W. & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22,* 36-71.

Schyns, P. G. & Bülthoff, H. H. (1994). Viewpoint dependence and face recognition. *Proceedings of the XVI Meeting of the Cognitive Science Society,* 789-793.

Searcy, J. H., & Bartlett, J. C. (1996). Inversion and processing of component and spatial-relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 904-915.

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology, 75,* 221-242.

Shepherd, J.W., Davies, G. M. & Ellis, H. D. (1988). How best shall a face be described? In Gruneberg, M. M., Morris, P. E. & Sykes, R. N. (Eds). *Practical Aspects of Memory: Current Research and Issues.*

Solso, R. L. & McCarthy, J. E. (1981) Prototype formation of faces: A case of pseudo-memory. *British Journal of Psychology, 72,* 499-503.

Sporer, S. Penrod, S., Read, D. & Cutler, B. L. (1995). Choosing, confidence and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118,* 315-327.

Swets, J. A. (1964). *Signal Detection and Recognition by Human Observers: Contemporary Readings.* John Wiley & Sons, London.

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

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

Tarr, M. J. & Pinker, S. (1990). When does human object recognition use a viewer-centred reference frame? *Psychological Science, 1,* 253-256.

Thomson, D. M. & Tulving, E. (1970). Associative encoding and retrieval: weak and strong cues. *Journal of Experimental Psychology, 77,* 593-601.

Troje, N. (1998). Generalization to novel views of faces: Psychophysics and models concerning the role of bilateral symmetry. In C. Taddei-Ferretti & C. Musio (Eds), *Downward Processes in the Perception Representation Mechanisms.* World Scientific.

Tversky, A. & Krantz, D. H. (1969). Similarity of schematic faces: A test of interdimensional additivity. *Perception and Psychophysics, 5,* 124-128.

Troje, N. F. & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research, 36 (12),* 1761-1771.

Tulving, E. & Flexser, A. J. (1992). On the nature of the Tulving-Wiseman function. *Psychological Review, 99,* 543-546.

Tulving, E. & Thomson, D. (1973) Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80 (5),* 352-373.

Tulving, E. & Watkins, M. J. (1973). Continuity between recall and recognition. *American Journal of Psychology, 86,* 352-373.

Turner, J., Pike, G., Kemp, R. & Brace, N. (1999). Applying perceptual research to E-FIT construction: A minimal face experiment. *Paper presented at the 10th Annual Conference of the European Association of Psychology and Law, April, Cyprus.*

Valentine, T. (1988). Upside - Down faces: A review of the effect of inversion on face recognition. *British Journal of Psychology, 79,* 471-491.

Valentine, T. (1991a). A unified account of the effects of distinctiveness, inversion and race in face recognition. *Quarterly Journal of Experimental Psychology, 43A.* 161-204.

Valentine, T. & Bruce, V. (1986). Recognising familiar faces: The role of distinctiveness and familiarity. *Canadian Journal of Psychology, 40(3),* 300-305.

Vetter, T., Poggio, T. & Bülthoff, H. H. (1994). The importance of symmetry and virtual views in three-dimensional object recognition. *Current Biology, 4,* 18-23.

Webster, M. A. & MacLin O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomic Bulletin & Review. 6 (4),* 647-653.

Wells, G. L. (1985). Verbal descriptions of faces from memory: Are they diagnostic of identification accuracy? *Journal of Applied Psychology, 70 (4),* 619-626.

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

Wogalter, M. S & Marwitz, D. B. (1991). Face composite construction: In-view and from-memory quality and improvement with practice. *Ergonomics, 34(4),* 459-468.

Woodhead, M. M., Baddeley, A. D. & Simmonds, D. C. V. (1979). On training people to recognise faces. *Ergonomics, 22 (3),* 333-343.

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., Hay, D. C. & Ellis, A. W. (1985). The face that launched a thousand slips: Everyday errors in recognising people. *British Journal of Psychology, 76(4),* 495-523.

Young, R. K., Hellawell, D. & Hay, D. C. (1987). Configural information in face perception. *Perception, 16,* 747-759.

Young, A.W., McWeeny, K. H., Ellis, A. W. & Hay, D. C. (1986). Naming and categorizing faces and written names. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 38A(2)*, 297-318.

Appendix

Experiment 1: Likeness Ratings
Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
FULLFACE	40	1.63	6.50	3.7594	1.2768
THREEQUA	40	2.13	5.63	4.0719	.9557
GENERAT	40	1.25	4.50	2.5156	.8975
Valid N (listwise)	40				

Repeated measure anova on likeness ratings - 4 (target) by 3 (type of composite)

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TARGET	Sphericity Assumed	164.418	3	54.806	33.939	.000
	Greenhouse-Geisser	164.418	2.919	56.334	33.939	.000
	Huynh-Feldt	164.418	3.000	54.806	33.939	.000
	Lower-bound	164.418	1.000	164.418	33.939	.000
Error(TARGET)	Sphericity Assumed	188.936	117	1.615		
	Greenhouse-Geisser	188.936	113.827	1.660		
	Huynh-Feldt	188.936	117.000	1.615		
	Lower-bound	188.936	39.000	4.845		
TYPE	Sphericity Assumed	216.879	2	108.440	36.413	.000
	Greenhouse-Geisser	216.879	1.997	108.588	36.413	.000
	Huynh-Feldt	216.879	2.000	108.440	36.413	.000
	Lower-bound	216.879	1.000	216.879	36.413	.000
Error(TYPE)	Sphericity Assumed	232.288	78	2.978		
	Greenhouse-Geisser	232.288	77.893	2.982		
	Huynh-Feldt	232.288	78.000	2.978		
	Lower-bound	232.288	39.000	5.956		
TARGET * TYPE	Sphericity Assumed	103.146	6	17.191	9.139	.000
	Greenhouse-Geisser	103.146	4.896	21.068	9.139	.000
	Huynh-Feldt	103.146	5.684	18.148	9.139	.000
	Lower-bound	103.146	1.000	103.146	9.139	.004
Error(TARGET*TYPE)	Sphericity Assumed	440.188	234	1.881		
	Greenhouse-Geisser	440.188	190.941	2.305		
	Huynh-Feldt	440.188	221.663	1.986		
	Lower-bound	440.188	39.000	11.287		

T-Tests on type

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	FULLFACE - THREEQUA	-.3125	1.2391	.1959	-.7088	8.377E-02	-1.595	39	.119
Pair 2	FULLFACE - GENERAT	1.2438	1.1999	.1897	.8600	1.6275	6.555	39	.000
Pair 3	THREEQUA - GENERAT	1.5563	1.2215	.1931	1.1656	1.9469	8.058	39	.000

Anova for target 1

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	159.879	2	79.940	39.392	.000
	Greenhouse-Geisser	159.879	1.796	89.016	39.392	.000
	Huynh-Feldt	159.879	1.877	85.164	39.392	.000
	Lower-bound	159.879	1.000	159.879	39.392	.000
Error(TYPE)	Sphericity Assumed	158.288	78	2.029		
	Greenhouse-Geisser	158.288	70.047	2.260		
	Huynh-Feldt	158.288	73.215	2.162		
	Lower-bound	158.288	39.000	4.059		

T-Tests for target 1

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	T1FF - T134	-1.3125	1.6513	.2611	-1.8406	-.7844	-5.027	39	.000
Pair 2	T1FF - T1GEN	1.5125	2.2346	.3533	.7978	2.2272	4.281	39	.000
Pair 3	T134 - T1GEN	2.8250	2.1109	.3338	2.1499	3.5001	8.464	39	.000

Anova for target 2

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	.929	2	.465	.317	.729
	Greenhouse-Geisser	.929	1.902	.488	.317	.719
	Huynh-Feldt	.929	1.997	.465	.317	.729
	Lower-bound	.929	1.000	.929	.317	.577
Error(TYPE)	Sphericity Assumed	114.404	78	1.467		
	Greenhouse-Geisser	114.404	74.192	1.542		
	Huynh-Feldt	114.404	77.894	1.469		
	Lower-bound	114.404	39.000	2.933		

Anova for target 3

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	57.950	2	28.975	11.335	.000
	Greenhouse-Geisser	57.950	1.824	31.767	11.335	.000
	Huynh-Feldt	57.950	1.909	30.356	11.335	.000
	Lower-bound	57.950	1.000	57.950	11.335	.002
Error(TYPE)	Sphericity Assumed	199.383	78	2.556		
	Greenhouse-Geisser	199.383	71.144	2.803		
	Huynh-Feldt	199.383	74.451	2.678		
	Lower-bound	199.383	39.000	5.112		

T-Test for target 3

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	T3FF - T334	-.5750	2.5433	.4021	-1.3884	.2384	-1.430	39	.161
Pair 2	T3FF - T3GEN	1.1000	1.9289	.3050	.4831	1.7169	3.607	39	.001
Pair 3	T334 - T3GEN	1.6750	2.2689	.3588	.9494	2.4006	4.669	39	.000

Anova for target 4

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	101.267	2	50.633	19.708	.000
	Greenhouse-Geisser	101.267	1.917	52.812	19.708	.000
	Huynh-Feldt	101.267	2.000	50.633	19.708	.000
	Lower-bound	101.267	1.000	101.267	19.708	.000
Error(TYPE)	Sphericity Assumed	200.400	78	2.569		
	Greenhouse-Geisser	200.400	74.782	2.680		
	Huynh-Feldt	200.400	78.000	2.569		
	Lower-bound	200.400	39.000	5.138		

T-Tests for target 4

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	T4FF - T434	.5000	2.2646	.3581	-.2242	1.2242	1.396	39	.170
Pair 2	T4FF - T4GEN	2.1500	2.0544	.3248	1.4930	2.8070	6.619	39	.000
Pair 3	T434 - T4GEN	1.6500	2.4631	.3894	.8623	2.4377	4.237	39	.000

T-Test on order of construction

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	FIRST	4.0656	16	1.5329	.3832
	SECOND	3.7406	16	1.5604	.3901

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	FIRST - SECOND	.3250	1.5402	.3851	-.4957	1.1457	.844	15	.412

Experiment 1: Identification

Friedman Test

Ranks

	Mean Rank
FF	2.09
TQ	2.05
GEN	1.86

Test Statistics[a]

N	40
Chi-Square	2.952
df	2
Asymp. Sig.	.229

a. Friedman Test

Experiment 2: Stage 1
Cochran Test

Frequencies

	Value	
	0	1
full face best composite	23	9
full face intermediate composite	26	6
best full face and threequarter composites together	20	12
intermediate full face and three quarter composites together	15	17

Test Statistics

N	32
Cochran's Q	8.426[a]
df	3
Asymp. Sig.	.038

a. 1 is treated as a success.

McNemarTests

Test Statistics[b]

	full face best composite & full face intermediate composite	full face intermediate composite & best full face and threequarter composites together	best full face and threequarter composites together & intermediate full face and three quarter composites together	full face intermediate composite & intermediate full face and three quarter composites together	full face best composite & best full face and threequarter composites together
N	32	32	32	32	32
Exact Sig. (2-tailed)	.375[a]	.180[a]	.359[a]	.019[a]	.629[a]

a. Binomial distribution used.

b. McNemar Test

Was the 2-view effect due to the fact that we just added in a better image?

Paired Samples Test

	Paired Differences					t	df	ig. (2-tailed
				95% Confidence Interval of the Difference				
	Mean	td. Deviatio	Std. Erro Mean	Lower	Upper			
Pair ' FULLFACE - THR	.3687	1.9306	.6826	-1.2452	1.9827	.540	7	.606

Experiment 2: Array task

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
FULLFACE	48	.5208	.5049	.00	1.00
THREQUAR	48	.7292	.5739	.00	2.00
GENERATE	48	.4167	.4982	.00	1.00

Friedman test

Test Statistics[a]

N	48
Chi-Square	7.078
df	2
Asymp. Sig.	.029

a. Friedman Test

Wilcoxon Signed Ranks Tests

Test Statistics[c]

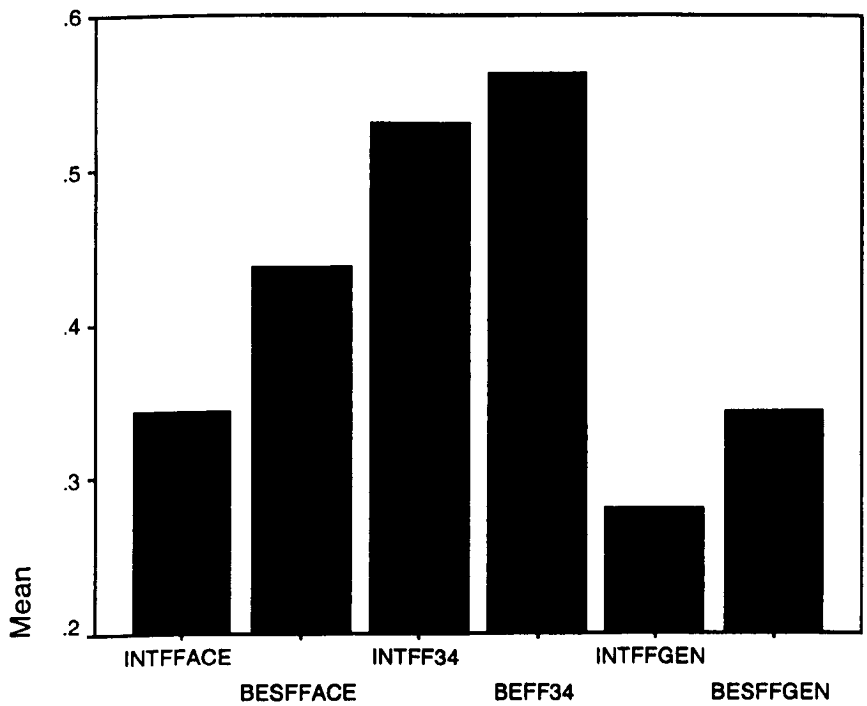
	THREQUAR - FULLFACE	GENERATE - FULLFACE	GENERATE - THREQUAR
Z	-1.826[a]	-.898[b]	-2.694[b]
Asymp. Sig. (2-tailed)	.068	.369	.007

a. Based on negative ranks.

b. Based on positive ranks.

c. Wilcoxon Signed Ranks Test

Graph showing performance by type of composite and quality



Wilcoxon Signed Ranks Tests

Test Statistics[c]

	INTFF34 - INTFFACE	BEFF34 - BESFFACE	INTFFGEN - INTFFACE	BESFFGEN - BESFFACE	INTFFGEN - INTFF34	BESFFGEN - BEFF34
Z	-1.500[a]	-1.000[a]	-.471[b]	-.655[b]	-1.886[b]	-1.698[b]
Asymp. Sig. (2-tailed)	.134	.317	.637	.513	.059	.090

a. Based on negative ranks.

b. Based on positive ranks.

c. Wilcoxon Signed Ranks Test

Experiment 3: Encoding Specificity Likeness Ratings

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
FFFF	22	1.88	6.06	3.7074	1.1922
FF34	22	1.38	6.00	3.5909	1.2236
FFALL	22	3.25	6.75	4.6023	1.0269
THFF	22	1.25	5.75	3.4886	1.1883
TH34	22	1.50	6.25	3.8807	1.1986
THALL	22	1.95	6.00	3.7564	1.0857
Valid N (listwise)	22				

Repeated measures anova- 2 (type- ff,3/4) by 3 (view -video)

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	6.100	1	6.100	5.356	.031
	Greenhouse-Geisser	6.100	1.000	6.100	5.356	.031
	Huynh-Feldt	6.100	1.000	6.100	5.356	.031
	Lower-bound	6.100	1.000	6.100	5.356	.031
Error(TYPE)	Sphericity Assumed	23.913	21	1.139		
	Greenhouse-Geisser	23.913	21.000	1.139		
	Huynh-Feldt	23.913	21.000	1.139		
	Lower-bound	23.913	21.000	1.139		
VIEW	Sphericity Assumed	40.790	2	20.395	19.242	.000
	Greenhouse-Geisser	40.790	1.901	21.462	19.242	.000
	Huynh-Feldt	40.790	2.000	20.395	19.242	.000
	Lower-bound	40.790	1.000	40.790	19.242	.000
Error(VIEW)	Sphericity Assumed	44.517	42	1.060		
	Greenhouse-Geisser	44.517	39.912	1.115		
	Huynh-Feldt	44.517	42.000	1.060		
	Lower-bound	44.517	21.000	2.120		
TYPE * VIEW	Sphericity Assumed	22.252	2	11.126	16.187	.000
	Greenhouse-Geisser	22.252	1.952	11.400	16.187	.000
	Huynh-Feldt	22.252	2.000	11.126	16.187	.000
	Lower-bound	22.252	1.000	22.252	16.187	.001
Error(TYPE*VIEW)	Sphericity Assumed	28.868	42	.687		
	Greenhouse-Geisser	28.868	40.988	.704		
	Huynh-Feldt	28.868	42.000	.687		
	Lower-bound	28.868	21.000	1.375		

View and full-face composites

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
VIEW	Sphericity Assume	53.893	2	26.946	32.208	.000
	Greenhouse-Geiss	53.893	1.668	32.315	32.208	.000
	Huynh-Feldt	53.893	1.794	30.033	32.208	.000
	Lower-bound	53.893	1.000	53.893	32.208	.000
Error(VIEW)	Sphericity Assume	35.139	42	.837		
	Greenhouse-Geiss	35.139	35.022	1.003		
	Huynh-Feldt	35.139	37.683	.932		
	Lower-bound	35.139	21.000	1.673		

View and three-quarter-view composites

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
VIEW	Sphericity Assumed	9.150	2	4.575	5.024	.011
	Greenhouse-Geisser	9.150	1.731	5.286	5.024	.015
	Huynh-Feldt	9.150	1.872	4.887	5.024	.013
	Lower-bound	9.150	1.000	9.150	5.024	.036
Error(VIEW)	Sphericity Assumed	38.246	42	.911		
	Greenhouse-Geisser	38.246	36.348	1.052		
	Huynh-Feldt	38.246	39.320	.973		
	Lower-bound	38.246	21.000	1.821		

T-Tests on view and type of composite

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	TH34 - THALL	.1243	.8061	.1719	-.2331	.4817	.723	21	.478
Pair 2	THFF - THALL	-.2678	.4957	.1057	-.4875	-4.80E-02	-2.534	21	.019
Pair 3	THFF - TH34	-.3920	.7737	.1650	-.7351	-4.90E-02	-2.377	21	.027
Pair 4	FF34 - FFALL	-1.0114	.6835	.1457	-1.3144	-.7083	-6.941	21	.000
Pair 5	FFFF - FFALL	-.8949	.7416	.1581	-1.2237	-.5661	-5.660	21	.000
Pair 6	FFFF - FF34	.1165	.4877	.1040	-9.97E-02	.3327	1.120	21	.275
Pair 7	FFFF - THFF	.2188	.7412	.1580	-.1099	.5474	1.384	21	.181
Pair 8	FF34 - TH34	-.2898	.5167	.1102	-.5188	-6.07E-02	-2.631	21	.016
Pair 9	FFALL - THALL	.8459	.5946	.1268	.5822	1.1095	6.672	21	.000

Experiment 3: Encoding Specificity Array Task

Friedman Test on collapsed data

Ranks

	Mean Rank
FULLFACE	1.84
THRQUART	1.97
BOTH	2.19

Test Statistics[a]

N	64
Chi-Square	6.049
df	2
Asymp. Sig.	.049

a. Friedman Test

Wilcoxon Signed Ranks Test

Test Statistics[b]

	BOTH - FULLFACE	BOTH - THRQUART	THRQUART - FULLFACE
Z	-2.354[a]	-1.733[a]	-.661[a]
Asymp. Sig. (2-tailed)	.019	.083	.508

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

Cochran Test on type and view

Test Statistics

N	64
Cochran's Q	22.688[a]
df	8
Asymp. Sig.	.004

a. 0 is treated as a success.

Cochran Test on full-face composites and view

Frequencies

	Value	
	0	1
FFFF	55	9
FF34	54	10
FFALL	41	23

Test Statistics

N	64
Cochran's Q	11.806[a]
df	2
Asymp. Sig.	.003

a. 0 is treated as a success.

McNemar Tests for full-face composites and view

Test Statistics[b]

	FF34 & FFALL	FFFF & FFALL	FFFF & FF34
N	64	64	64
Exact Sig. (2-tailed)	.011[a]	.004[a]	1.000[a]

a. Binomial distribution used.

b. McNemar Test

Cochran Test on 3/4 view comps and view

Test Statistics

N	64
Cochran's Q	3.161[a]
df	2
Asymp. Sig.	.206

a. 0 is treated as a success.

McNemar tests for 3/4 view composites and view

Test Statistics[c]

	TH34 & THALL	THFF & THALL	THFF & TH34
N	64	64	64
Chi-Square[a]	.000		
Asymp. Sig.	1.000		
Exact Sig. (2-tailed)		.210[b]	.065[b]

a. Continuity Corrected

b. Binomial distribution used.

c. McNemar Test

Pairwise comparisons

Test Statistics[d]

	'34 - FFI	ALL - FF	ALL - FF	I34 - THI	ALL - TH	ALL - TH	THALL FFALL	I34 - FF	IFF - FF
Z	-.243[a]	-2.985[a]	-2.711[a]	-2.111[a]	-1.460[a]	.000[b]	-.898[c]	-1.886[a]	-.500[a]
Asymp. Sig.	.808	.003	.007	.035	.144	1.000	.369	.059	.617

a. Based on negative ranks.

b. The sum of negative ranks equals the sum of positive ranks.

c. Based on positive ranks.

d. Wilcoxon Signed Ranks Test

Experiment 4: Likeness Ratings

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TARGET	Sphericity Assumed	153.992	3	51.331	36.785	.000
	Greenhouse-Geisser	153.992	2.776	55.465	36.785	.000
	Huynh-Feldt	153.992	3.000	51.331	36.785	.000
	Lower-bound	153.992	1.000	153.992	36.785	.000
Error(TARGET)	Sphericity Assumed	75.438	57	1.323		
	Greenhouse-Geisser	75.438	52.751	1.430		
	Huynh-Feldt	75.438	57.000	1.323		
	Lower-bound	75.438	19.000	3.970		
TYPE	Sphericity Assumed	52.613	1	52.613	38.061	.000
	Greenhouse-Geisser	52.613	1.000	52.613	38.061	.000
	Huynh-Feldt	52.613	1.000	52.613	38.061	.000
	Lower-bound	52.613	1.000	52.613	38.061	.000
Error(TYPE)	Sphericity Assumed	24.489	19	1.289		
	Greenhouse-Geisser	24.489	19.000	1.289		
	Huynh-Feldt	24.489	19.000	1.289		
	Lower-bound	24.489	19.000	1.289		
TARGET * TYPE	Sphericity Assumed	20.523	3	6.841	7.224	.001
	Greenhouse-Geisser	20.523	2.647	7.755	7.224	.002
	Huynh-Feldt	20.523	3.000	6.841	7.224	.001
	Lower-bound	20.523	1.000	20.523	7.224	.021
Error(TARGET*TYPE	Sphericity Assumed	61.657	57	1.082		
	Greenhouse-Geisser	61.657	50.285	1.226		
	Huynh-Feldt	61.657	57.000	1.082		
	Lower-bound	61.657	19.000	3.245		

Anova for target 1

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	187.633	2	93.817	55.386	.000
	Greenhouse-Geisser	187.633	1.954	96.048	55.386	.000
	Huynh-Feldt	187.633	2.000	93.817	55.386	.000
	Lower-bound	187.633	1.000	187.633	55.386	.000
Error(TYPE)	Sphericity Assumed	64.367	38	1.694		
	Greenhouse-Geisser	64.367	37.117	1.734		
	Huynh-Feldt	64.367	38.000	1.694		
	Lower-bound	64.367	19.000	3.388		

T-test

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1 BES - WOR	4.15	1.93	.43	3.25	5.05	9.631	19	.000
Pair 2 BES - MOR	1.00	1.89	.42	.11	1.89	2.364	19	.029
Pair 3 WOR - MOR	-3.15	1.69	.38	-3.94	-2.36	-8.314	19	.000

Anova for target 2

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	14.033	2	7.017	3.479	.041
	Greenhouse-Geisser	14.033	1.976	7.103	3.479	.042
	Huynh-Feldt	14.033	2.000	7.017	3.479	.041
	Lower-bound	14.033	1.000	14.033	3.479	.078
Error(TYPE)	Sphericity Assumed	76.633	38	2.017		
	Greenhouse-Geisser	76.633	37.538	2.041		
	Huynh-Feldt	76.633	38.000	2.017		
	Lower-bound	76.633	19.000	4.033		

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1 BES - WOR	1.00	1.97	.44	.64E-02	1.92	2.266	19	.035
Pair 2 BES - MOR	.00E-02	2.11	.47	-1.04	.94	-.106	19	.917
Pair 3 WOR - MOR	-1.05	1.93	.43	-1.95	-.15	-2.430	19	.025

Anova for target 3

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	159.100	2	79.550	34.003	.000
	Greenhouse-Geisser	159.100	1.368	116.327	34.003	.000
	Huynh-Feldt	159.100	1.438	110.645	34.003	.000
	Lower-bound	159.100	1.000	159.100	34.003	.000
Error(TYPE)	Sphericity Assumed	88.900	38	2.339		
	Greenhouse-Geisser	88.900	25.986	3.421		
	Huynh-Feldt	88.900	27.321	3.254		
	Lower-bound	88.900	19.000	4.679		

T-Test

Paired Samples Test

	Paired Differences					t	df	ig. (2-tailed
				95% Confidence Interval of the Difference				
	Mean	td. Deviatio	Std. Erro Mean	Lower	Upper			
Pair 1 BES - WOR	3.35	2.41	.54	2.22	4.48	6.211	19	.000
Pair 2 BES - MOR	-.20	1.24	.28	-.78	.38	-.721	19	.479
Pair 3 WOR - MOR	-3.55	2.58	.58	-4.76	-2.34	-6.142	19	.000

Anova for target 4

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	243.900	2	121.950	95.026	.000
	Greenhouse-Geisser	243.900	1.573	155.051	95.026	.000
	Huynh-Feldt	243.900	1.691	144.276	95.026	.000
	Lower-bound	243.900	1.000	243.900	95.026	.000
Error(TYPE)	Sphericity Assumed	48.767	38	1.283		
	Greenhouse-Geisser	48.767	29.888	1.632		
	Huynh-Feldt	48.767	32.120	1.518		
	Lower-bound	48.767	19.000	2.567		

T-Test

Paired Samples Test

	Paired Differences					t	df	ig. (2-tailed
	Mean	td. Deviatio	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1 BES - WOR	4.20	1.61	.36	3.45	4.95	11.672	19	.000
Pair 2 BES -MOR	-.15	1.18	.26	-.70	.40	-.567	19	.577
Pair 3 WOR - MOR	-4.35	1.93	.43	-5.25	-3.45	-10.096	19	.000

Experiment 4: Identification

Cochran test

Frequencies

	Value	
	0	1
BEST	25	7
WORST	27	5
MORPH	23	9
FOUR	21	11

Test Statistics

N	32
Cochran's Q	3.750[a]
df	3
Asymp. Sig.	.290

a. 0 is treated as a success.

Experiment 4: 6AFC
Cochran Test on overall correct responses

Test Statistics

N	64
Cochran's Q	4.914[a]
df	3
Asymp. Sig.	.178

a. 0 is treated as a success.

McNemar Tests

Test Statistics[b]

	MORPH & BEST	MORPH & WORST	MORPH & FOUR
N	64	64	64
Chi-Square[a]	1.633	1.441	3.704
Asymp. Sig.	.201	.230	.054

a. Continuity Corrected

b. McNemar Test

Cochran Test on overall incorrect responses

Test Statistics

N	64
Cochran's Q	5.122[a]
df	3
Asymp. Sig.	.163

a. 1 is treated as a success.

Experiment 5: Likeness Ratings

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
COMP	24	1.63	6.13	3.5938	1.1739
MINI	24	1.56	6.31	3.7917	1.1341
MORPH	24	2.00	7.50	5.5104	1.4437
Valid N (listwise)	24				

Repeated Measures Anova

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TARGET	Sphericity Assumed	83.421	3	27.807	22.099	.000
	Greenhouse-Geisser	83.421	2.537	32.882	22.099	.000
	Huynh-Feldt	83.421	2.878	28.989	22.099	.000
	Lower-bound	83.421	1.000	83.421	22.099	.000
Error(TARGET)	Sphericity Assumed	86.821	69	1.258		
	Greenhouse-Geisser	86.821	58.350	1.488		
	Huynh-Feldt	86.821	66.187	1.312		
	Lower-bound	86.821	23.000	3.775		
TYPE	Sphericity Assumed	229.038	2	114.519	56.596	.000
	Greenhouse-Geisser	229.038	1.191	192.315	56.596	.000
	Huynh-Feldt	229.038	1.219	187.907	56.596	.000
	Lower-bound	229.038	1.000	229.038	56.596	.000
Error(TYPE)	Sphericity Assumed	93.078	46	2.023		
	Greenhouse-Geisser	93.078	27.392	3.398		
	Huynh-Feldt	93.078	28.034	3.320		
	Lower-bound	93.078	23.000	4.047		
TARGET * TYPE	Sphericity Assumed	17.550	6	2.925	2.863	.012
	Greenhouse-Geisser	17.550	3.088	5.684	2.863	.041
	Huynh-Feldt	17.550	3.621	4.846	2.863	.033
	Lower-bound	17.550	1.000	17.550	2.863	.104
Error(TARGET*TYPE	Sphericity Assumed	141.006	138	1.022		
	Greenhouse-Geisser	141.006	71.020	1.985		
	Huynh-Feldt	141.006	83.289	1.693		
	Lower-bound	141.006	23.000	6.131		

Repeated Measures Anova on target 1

Within-Subjects Factors

Measure: MEASURE_1

TYPE	Dependent Variable
1	T1COMP
2	T1MINI
3	T14MORPH

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	105.231	2	52.615	47.634	.000
	Greenhouse-Geisser	105.231	1.532	68.668	47.634	.000
	Huynh-Feldt	105.231	1.620	64.955	47.634	.000
	Lower-bound	105.231	1.000	105.231	47.634	.000
Error(TYPE)	Sphericity Assumed	50.811	46	1.105		
	Greenhouse-Geisser	50.811	35.246	1.442		
	Huynh-Feldt	50.811	37.262	1.364		
	Lower-bound	50.811	23.000	2.209		

T-Tests on target 1

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	T1COMP	3.8750	24	1.1563	.2360
	T1MINI	4.6979	24	1.3752	.2807
Pair 2	T1MINI	4.6979	24	1.3752	.2807
	T14MORPH	6.7500	24	1.5948	.3255
Pair 3	T1COMP	3.8750	24	1.1563	.2360
	T14MORPH	6.7500	24	1.5948	.3255

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1 T1COMP - T1MINI	-.8229	1.0121	.2066	-1.2503	-.3955	-3.983	23	.001
Pair 2 T1MINI - T14MORF	-2.0521	1.5965	.3259	-2.7262	-1.3779	-6.297	23	.000
Pair 3 T1COMP - T14MOI	-2.8750	1.7477	.3567	-3.6130	-2.1370	-8.059	23	.000

Anova on target 2

Within-Subjects Factors

Measure: MEASURE_1

TYPE	Dependent Variable
1	T2COMP
2	T2MINI
3	T24MORPH

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	50.106	2	25.053	14.276	.000
	Greenhouse-Geisser	50.106	1.264	39.652	14.276	.000
	Huynh-Feldt	50.106	1.303	38.448	14.276	.000
	Lower-bound	50.106	1.000	50.106	14.276	.001
Error(TYPE)	Sphericity Assumed	80.727	46	1.755		
	Greenhouse-Geisser	80.727	29.064	2.778		
	Huynh-Feldt	80.727	29.974	2.693		
	Lower-bound	80.727	23.000	3.510		

T-Test on target 2

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	T2COMP	3.1354	24	1.4577	.2976
	T2MINI	2.9271	24	1.5331	.3129
Pair 2	T2COMP	3.1354	24	1.4577	.2976
	T24MORPH	4.7917	24	2.0212	.4126
Pair 3	T2MINI	2.9271	24	1.5331	.3129
	T24MORPH	4.7917	24	2.0212	.4126

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1 T2COMP - T2MINI	.2083	.9659	.1972	-.1995	.6162	1.057	23	.302
Pair 2 T2COMP - T24MO	-1.6563	2.3310	.4758	-2.6406	-.6719	-3.481	23	.002
Pair 3 T2MINI - T24MOR	-1.8646	2.0403	.4165	-2.7261	-1.0030	-4.477	23	.000

Anova on target 3

Within-Subjects Factors

Measure: MEASURE_1

TYPE	Dependent Variable
1	T3COMP
2	T3MINI
3	T34MORPH

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	47.978	2	23.989	18.590	.000
	Greenhouse-Geisser	47.978	1.429	33.574	18.590	.000
	Huynh-Feldt	47.978	1.497	32.044	18.590	.000
	Lower-bound	47.978	1.000	47.978	18.590	.000
Error(TYPE)	Sphericity Assumed	59.360	46	1.290		
	Greenhouse-Geisser	59.360	32.868	1.806		
	Huynh-Feldt	59.360	34.437	1.724		
	Lower-bound	59.360	23.000	2.581		

T-Test on target 3

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	T3MINI	3.4896	24	1.2302	.2511
	T34MORPH	5.2500	24	2.0270	.4138
Pair 2	T3COMP	3.5486	24	1.4214	.2901
	T34MORPH	5.2500	24	2.0270	.4138
Pair 3	T3COMP	3.5486	24	1.4214	.2901
	T3MINI	3.4896	24	1.2302	.2511

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	T3MINI - T34MORP	-1.7604	1.7654	.3604	-2.5059	-1.0149	-4.885	23	.000
Pair 2	T3COMP - T34MOR	-1.7014	1.9098	.3898	-2.5078	-.8950	-4.364	23	.000
Pair 3	T3COMP - T3MINI	.903E-02	.9892	.2019	-.3587	.4767	.292	23	.773

Anova on target 4

Within-Subjects Factors

Measure: MEASURE_1

TYPE	Dependent Variable
1	T4COMP
2	T4MINI
3	T44MORPH

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	43.273	2	21.636	23.046	.000
	Greenhouse-Geisser	43.273	1.290	33.542	23.046	.000
	Huynh-Feldt	43.273	1.334	32.437	23.046	.000
	Lower-bound	43.273	1.000	43.273	23.046	.000
Error(TYPE)	Sphericity Assumed	43.186	46	.939		
	Greenhouse-Geisser	43.186	29.672	1.455		
	Huynh-Feldt	43.186	30.683	1.407		
	Lower-bound	43.186	23.000	1.878		

T-Test on target 4

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	T4MINI	4.0521	24	1.3000	.2654
	T44MORPH	5.2500	24	1.9167	.3913
Pair 2	T4COMP	3.3750	24	1.3290	.2713
	T44MORPH	5.2500	24	1.9167	.3913
Pair 3	T4COMP	3.3750	24	1.3290	.2713
	T4MINI	4.0521	24	1.3000	.2654

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	T4MINI - T44MORP	-1.1979	1.4218	.2902	-1.7983	-.5975	-4.127	23	.000
Pair 2	T4COMP - T44MO	-1.8750	1.7336	.3539	-2.6070	-1.1430	-5.299	23	.000
Pair 3	T4COMP - T4MINI	-.6771	.7784	.1589	-1.0058	-.3484	-4.261	23	.000

Experiment 5: 6AFC

Cochran Test

Frequencies

	Value	
	0	1
comp 8 subs per book	18	14
bestmini 8 subs per book	18	14
worst mini 8 subs per book	21	11
4 morph 8 subs per book	17	15

Test Statistics

N	32
Cochran's Q	1.227[a]
df	3
Asymp. Sig.	.746

a. 0 is treated as a success.

By target analyses
Cochran Test

Frequencies

	Value	
	0	1
T1COMP	2	6
T1BMINI	4	4
T1WMINI	5	3
T14MORPH	4	4
T2COMP	3	5
T2BMINI	4	4
T2WMINI	5	3
T24MORPH	4	4
T3COMP	7	1
T3BMINI	6	2
T3WMINI	6	2
T34MORPH	7	1
T4COMP	6	2
T4BMINI	4	4
T4WMINI	5	3
T44MORPH	2	6

Test Statistics

N	8
Cochran's Q	19.861[a]
df	15
Asymp. Sig.	.177

a. 1 is treated as a success.

For target 1
Cochran Test

Frequencies

	Value	
	0	1
T1COMP	2	6
T1BMINI	4	4
T1WMINI	5	3
T14MORPH	4	4

Test Statistics

N	8
Cochran's Q	3.000[a]
df	3
Asymp. Sig.	.392

a. 1 is treated as a success.

For target 2
Cochran Test

Frequencies

	Value	
	0	1
T2COMP	3	5
T2BMINI	4	4
T2WMINI	5	3
T24MORPH	4	4

Test Statistics

N	8
Cochran's Q	1.200[a]
df	3
Asymp. Sig.	.753

a. 0 is treated as a success.

For target 3
Cochran Test

Frequencies

	Value	
	0	1
T3COMP	7	1
T3BMINI	6	2
T3WMINI	6	2
T34MORPH	7	1

Test Statistics

N	8
Cochran's Q	.750[a]
df	3
Asymp. Sig.	.861

a. 0 is treated as a success.

For target 4
Cochran Test

Frequencies

	Value	
	0	1
T4COMP	6	2
T4BMINI	4	4
T4WMINI	5	3
T44MORPH	2	6

Test Statistics

N	8
Cochran's Q	5.000[a]
df	3
Asymp. Sig.	.172

a. 0 is treated as a success.

Experiment 6: Identification

Cochran Test

Frequencies

	Value	
	0	1
BEST	15	17
WORST	20	12
ALL4	12	20
FFACE	20	12

Test Statistics

N	32
Cochran's Q	5.243[a]
df	3
Asymp. Sig.	.155

a. 0 is treated as a success.

False positive ID data
Cochran Test

Frequencies

	Value	
	0	1
SINGLE	24	8
BEST4	24	8
WORST4	23	9
ALL4	27	5

Test Statistics

N	32
Cochran's Q	1.543[a]
df	3
Asymp. Sig.	.672

a. 0 is treated as a success.

By target and type
Cochran Test

Frequencies

	Value	
	0	1
T1SINGLE	8	0
T1BEST4	0	8
T1WORST4	3	5
T1ALL4	1	7
T2SINGLE	4	4
T2BEST4	6	2
T2WORST4	8	0
T2ALL4	3	5
T3SINGLE	8	0
T3BEST4	8	0
T3WORST4	8	0
T3ALL4	8	0
T4SINGLE	0	8
T4BEST4	1	7
T4WORST4	1	7
T4ALL4	0	8

Test Statistics

N	8
Cochran's Q	84.860[a]
df	15
Asymp. Sig.	.000

a. 0 is treated as a success.

For target 1

Cochran Test

Frequencies

	Value	
	0	1
T1SINGLE	8	0
T1BEST4	0	8
T1WORST4	3	5
T1ALL4	1	7

Test Statistics

N	8
Cochran's Q	16.286[a]
df	3
Asymp. Sig.	.001

a. 0 is treated as a success.

McNemar Test

Test Statistics[b]

	T1WORST4 & T1ALL4	T1BEST4 & T1ALL4	T1SINGLE & T1ALL4	T1SINGLE & T1BEST4	T1BEST4 & T1WORST4	T1SINGLE & T1WORST4
N	8	8	8	8	8	8
Exact Sig. (2-tailed)	.625[a]	1.000[a]	.016[a]	.008[a]	.250[a]	.063[a]

a. Binomial distribution used.

b. McNemar Test

For target 2
Cochran Test

Test Statistics

N	8
Cochran's Q	7.080[a]
df	3
Asymp. Sig.	.069

a. 1 is treated as a success.

McNemar Test

Test Statistics[b]

	T2WORST4 & T2ALL4	T2BEST4 & T2ALL4	T2SINGLE & T2ALL4	T2SINGLE & T2BEST4	T2SINGLE & T2WORST4	T2BEST4 & T2WORST4
N	8	8	8	8	8	8
Exact Sig. (2-tailed)	.063[a]	.375[a]	1.000[a]	.688[a]	.125[a]	.500[a]

a. Binomial distribution used.

b. McNemar Test

For target 4
Cochran Test

Frequencies

	Value	
	0	1
T4SINGLE	0	8
T4BEST4	1	7
T4WORST4	1	7
T4ALL4	0	8

Test Statistics

N	8
Cochran's Q	2.000[a]
df	3
Asymp. Sig.	.572

a. 1 is treated as a success.

No of faces used to make an ID (strategy)
Best set of 4- Cochran Test

Frequencies

	Value	
	0	1
BEST1F	13	16
BEST2F	27	2
BEST3F	28	1
BESTALL4	28	1

Test Statistics

N	29
Cochran's Q	32.400[a]
df	3
Asymp. Sig.	.000

a. 1 is treated as a success.

McNemar Test

Test Statistics[b]

	BEST1F & BEST2F	BEST1F & BEST3F	BEST1F & BESTALL4	BEST3F & BESTALL4	BEST2F & BEST3F	BEST2F & BESTALL4
N	29	29	29	29	29	29
Exact Sig. (2-tailed)	.001[a]	.000[a]	.000[a]	1.000[a]	1.000[a]	1.000[a]

a. Binomial distribution used.

b. McNemar Test

For worst sets of 4- Cochran Test

Frequencies

	Value	
	0	1
WORST1F	19	10
WORST2F	24	5
WORST3F	29	0
WORSTAL4	24	5

Test Statistics

N	29
Cochran's Q	10.000[a]
df	3
Asymp. Sig.	.019

a. 0 is treated as a success.

McNemar Test

Test Statistics[b]

	WORST1F & WORST2F	WORST1F & WORST3F	WORST1F & WORSTAL4	WORST2F & WORST3F	WORST2F & WORSTAL4	WORST3F & WORSTAL4
N	29	29	29	29	29	29
Exact Sig. (2-tailed)	.302[a]	.002[a]	.302[a]	.063[a]	1.000[a]	.063[a]

a. Binomial distribution used.

b. McNemar Test

For the sets of 4 from 4 different witnesses
Cochran Test

Frequencies

	Value	
	0	1
ALL41F	20	9
ALL42F	23	6
ALL43F	28	1
ALL4ALL4	23	6

Test Statistics

N	29
Cochran's Q	6.000[a]
df	3
Asymp. Sig.	.112

a. 0 is treated as a success.

Strategy and accuracy

Friedman Test

Ranks

	Mean Rank
ONEFACE	3.14
TWOFACES	2.38
THREEFAC	1.96
ALL4	2.52

Test Statistics[a]

N	28
Chi-Square	18.113
df	3
Asymp. Sig.	.000

a. Friedman Test

Wilcoxon Signed Ranks Test

Test Statistics[c]

	ALL4 - THREEFAC	ALL4 - TWOFACES	ALL4 - ONEFACE	THREEFAC - TWOFACES	THREEFAC - ONEFACE	TWOFACES - ONEFACE
Z	-2.496[a]	-.440[a]	-1.939[b]	-2.111[b]	-3.586[b]	-2.297[b]
Asymp. Sig. (2-tailed)	.013	.660	.050	.035	.000	.022

a. Based on negative ranks.

b. Based on positive ranks.

c. Wilcoxon Signed Ranks Test

Overall differences in identification accuracy for each set of 4 depending on strategy used

Cochran Test

Frequencies

	Value	
	0	1
looked at 1 face for best set of 4	17	11
2 faces for best set of 4	27	1
3 faces for best set of 4	27	1
all 4 faces for best set of 4	27	1
1 face for worst set of 4	24	4
2 faces for worst set of 4	25	3
3 faces for worst set of 4	28	0
all 4 faces for worst set of 4	24	4
1 face for set of 4 from different witnesses	20	8
2 faces for set of 4 from different	23	5
3 faces for set of 4 from different	27	1
4 faces for set of 4 from different	22	6

Test Statistics

N	28
Cochran's Q	35.466[a]
df	11
Asymp. Sig.	.000

a. 0 is treated as a success.

Best set of 4- Cochran Test

Frequencies

	Value	
	0	1
looked at 1 face for best set of 4	17	11
2 faces for best set of 4	27	1
3 faces for best set of 4	27	1
all 4 faces for best set of 4	27	1

Test Statistics

N	28
Cochran's Q	21.429[a]
df	3
Asymp. Sig.	.000

a. 0 is treated as a success.

McNemar Test

Test Statistics[b]

	looked at 1 face for best set of 4 & 2 faces for best set of 4	looked at 1 face for best set of 4 & 3 faces for best set of 4	looked at 1 face for best set of 4 & all 4 faces for best set of 4	2 faces for best set of 4 & 3 faces for best set of 4	2 faces for best set of 4 & all 4 faces for best set of 4	3 faces for best set of 4 & all 4 faces for best set of 4
N	28	28	28	28	28	28
Exact Sig. (2-tailed)	.006[a]	.006[a]	.006[a]	1.000[a]	1.000[a]	1.000[a]

a. Binomial distribution used.

b. McNemar Test

For worst sets of 4
Cochran Test

Frequencies

	Value	
	0	1
1 face for worst set of 4	24	4
2 faces for worst set of 4	25	3
3 faces for worst set of 4	28	0
all 4 faces for worst set of 4	24	4

Test Statistics

N	28
Cochran's Q	3.909[a]
df	3
Asymp. Sig.	.271

a. 0 is treated as a success.

McNemar Test

Test Statistics[b]

	1 face for worst set of 4 & 2 faces for worst set of 4	1 face for worst set of 4 & 3 faces for worst set of 4	1 face for worst set of 4 & all 4 faces for worst set of 4	2 faces for worst set of 4 & 3 faces for worst set of 4	2 faces for worst set of 4 & all 4 faces for worst set of 4	3 faces for worst set of 4 & all 4 faces for worst set of 4
N	28	28	28	28	28	28
Exact Sig. (2-tailed)	1.000[a]	.125[a]	1.000[a]	.250[a]	1.000[a]	.125[a]

a. Binomial distribution used.

b. McNemar Test

For sets of 4 from 4 different witnesses
Cochran Test

Frequencies

	Value	
	0	1
1 face for set of 4 from different witnesses	20	8
2 faces for set of 4 from different	23	5
3 faces for set of 4 from different	27	1
4 faces for set of 4 from different	22	6

Test Statistics

N	28
Cochran's Q	5.200[a]
df	3
Asymp. Sig.	.158

a. 0 is treated as a success.

McNemar tests

Test Statistics[b]

	3 faces for set of 4 from different & 4 faces for set of 4 from different	2 faces for set of 4 from different & 4 faces for set of 4 from different	1 face for set of 4 from different witnesses & 4 faces for set of 4 from different	1 face for set of 4 from different witnesses & 2 faces for set of 4 from different	2 faces for set of 4 from different & 3 faces for set of 4 from different	1 face for set of 4 from different witnesses & 3 faces for set of 4 from different
N	28	28	28	28	28	28
Exact Sig. (2-tailed)	.125[a]	1.000[a]	.791[a]	.581[a]	.219[a]	.039[a]

a. Binomial distribution used.

b. McNemar Test

Experiment 7: PROfit Identification

Friedman Test

Ranks

	Mean Rank
SINGLE	2.25
SKETCH	2.30
EVOFIT	2.69
ALL4	2.76

Test Statistics[a]

N	40
Chi-Square	7.686
df	3
Asymp. Sig.	.053

a. Friedman Test

Wilcoxon Signed Ranks Test-pairwise comparisons

Test Statistics[c]

	ALL4 - EVOFIT	ALL4 - SKETCH	ALL4 - SINGLE	SKETCH - SINGLE	EVOFIT - SINGLE	EVOFIT - SKETCH
Z	-.882[a]	-2.153[a]	-2.782[a]	.000[b]	-1.705[a]	-1.661[a]
Asymp. Sig. (2-tailed)	.378	.031	.005	1.000	.088	.097

a. Based on negative ranks.

b. The sum of negative ranks equals the sum of positive ranks.

c. Wilcoxon Signed Ranks Test

PROfit false positives
Friedman Test

Ranks

	Mean Rank
PROFIT	2.54
PROSKETC	2.36
PROEVO	2.44
ALL4	2.66

Test Statistics[a]

N	40
Chi-Square	2.599
df	3
Asymp. Sig.	.458

a. Friedman Test

E-Fit - Friedman Test

Ranks

	Mean Rank
EFIT	2.26
ESKETCH	2.30
EEVOFIT	2.66
EALL4	2.78

Test Statistics[a]

N	40
Chi-Square	7.971
df	3
Asymp. Sig.	.047

a. Friedman Test

Wilcoxon Signed Ranks Test

Test Statistics[b]

	EALL4 - ESKETCH	EALL4 - EFIT	EEVOFIT - ESKETCH	EEVOFIT - EFIT	EALL4 - EEVOFIT
Z	-2.459[a]	-2.327[a]	-1.978[a]	-1.806[a]	-.783[a]
Asymp. Sig. (2-tailed)	.014	.020	.048	.071	.434

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

E-Fit false positives
Friedman Test

Ranks

	Mean Rank
EFIT	2.69
EFITSKET	2.49
EFITEVO	2.33
ALL4	2.50

Test Statistics[a]

N	40
Chi-Square	5.105
df	3
Asymp. Sig.	.164

a. Friedman Test

Experiment 8: Likeness ratings

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
PROFIT	40	.00	2.00	.3500	.5335
PROSKETC	40	.00	2.00	.2750	.5541
PROEVO	40	.00	2.00	.3250	.5723
ALL4	40	.00	2.00	.4250	.5943
Valid N (listwise)	40				

Repeated Measures Anova

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TARGET	Sphericity Assumed	280.015	9	31.113	10.196	.000
	Greenhouse-Geisser	280.015	5.738	48.797	10.196	.000
	Huynh-Feldt	280.015	8.503	32.930	10.196	.000
	Lower-bound	280.015	1.000	280.015	10.196	.005
Error(TARGET)	Sphericity Assumed	521.818	171	3.052		
	Greenhouse-Geisser	521.818	109.029	4.786		
	Huynh-Feldt	521.818	161.563	3.230		
	Lower-bound	521.818	19.000	27.464		
TYPE	Sphericity Assumed	37.103	2	18.552	9.805	.000
	Greenhouse-Geisser	37.103	1.673	22.182	9.805	.001
	Huynh-Feldt	37.103	1.815	20.440	9.805	.001
	Lower-bound	37.103	1.000	37.103	9.805	.005
Error(TYPE)	Sphericity Assumed	71.897	38	1.892		
	Greenhouse-Geisser	71.897	31.781	2.262		
	Huynh-Feldt	71.897	34.490	2.085		
	Lower-bound	71.897	19.000	3.784		
TARGET * TYPE	Sphericity Assumed	185.930	18	10.329	5.930	.000
	Greenhouse-Geisser	185.930	8.127	22.877	5.930	.000
	Huynh-Feldt	185.930	14.766	12.592	5.930	.000
	Lower-bound	185.930	1.000	185.930	5.930	.025
Error(TARGET*TYPE	Sphericity Assumed	595.737	342	1.742		
	Greenhouse-Geisser	595.737	154.420	3.858		
	Huynh-Feldt	595.737	280.558	2.123		
	Lower-bound	595.737	19.000	31.355		

T-Test between types of composite

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	EVOFIT	3.7050	20	1.2373	.2767
	MORPH	3.7100	20	1.3222	.2957
Pair 2	PROFIT	3.1800	20	1.0405	.2327
	MORPH	3.7100	20	1.3222	.2957
Pair 3	PROFIT	3.1800	20	1.0405	.2327
	EVOFIT	3.7050	20	1.2373	.2767

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1 EVOFIT - MOR	5.00E-03	.4817	.1077	-.2305	.2205	-.046	19	.963
Pair 2 PROFIT - MOR	-.5300	.6174	.1381	-.8189	-.2411	-3.839	19	.001
Pair 3 PROFIT - EVO	-.5250	.7225	.1616	-.8631	-.1869	-3.250	19	.004

Anova on Target 1

Within-Subjects Factors

Measure: MEASURE_1

TYPE	Dependent Variable
1	T1PRO2
2	T1EVO2
3	T1MOR2

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	16.033	2	8.017	4.864	.013
	Greenhouse-Geisser	16.033	1.915	8.371	4.864	.014
	Huynh-Feldt	16.033	2.000	8.017	4.864	.013
	Lower-bound	16.033	1.000	16.033	4.864	.040
Error(TYPE)	Sphericity Assumed	62.633	38	1.648		
	Greenhouse-Geisser	62.633	36.390	1.721		
	Huynh-Feldt	62.633	38.000	1.648		
	Lower-bound	62.633	19.000	3.296		

T-Test for target 1

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	T1PRO2	4.0500	20	1.8771	.4197
	T1EVO2	2.8000	20	1.7045	.3811
Pair 2	T1PRO2	4.0500	20	1.8771	.4197
	T1MOR2	3.6000	20	1.9841	.4437
Pair 3	T1EVO2	2.8000	20	1.7045	.3811
	T1MOR2	3.6000	20	1.9841	.4437

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1 T1PRO2 - T1EVO	1.2500	1.9702	.4405	.3279	2.1721	2.837	19	.011
Pair 2 T1PRO2 - T1MOR	.4500	1.6376	.3662	-.3164	1.2164	1.229	19	.234
Pair 3 T1EVO2 - T1MOR	-.8000	1.8238	.4078	-1.6536	5.357E-02	-1.962	19	.065

Anova for target 2

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	10.533	2	5.267	3.790	.032
	Greenhouse-Geisser	10.533	1.954	5.391	3.790	.033
	Huynh-Feldt	10.533	2.000	5.267	3.790	.032
	Lower-bound	10.533	1.000	10.533	3.790	.066
Error(TYPE)	Sphericity Assumed	52.800	38	1.389		
	Greenhouse-Geisser	52.800	37.122	1.422		
	Huynh-Feldt	52.800	38.000	1.389		
	Lower-bound	52.800	19.000	2.779		

T-Test for target 2

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1 T2PRO1 - T2EV	.7000	1.5594	.3487	2.98E-02	1.4298	2.008	19	.059
Pair 2 T2PRO1 - T2MO	-.3000	1.7800	.3980	-1.1331	.5331	-.754	19	.460
Pair 3 T2EVO1 - T2MO	-1.0000	1.6543	.3699	-1.7743	-.2257	-2.703	19	.014

Anova for target 3

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	94.033	2	47.017	36.737	.000
	Greenhouse-Geisser	94.033	1.864	50.436	36.737	.000
	Huynh-Feldt	94.033	2.000	47.017	36.737	.000
	Lower-bound	94.033	1.000	94.033	36.737	.000
Error(TYPE)	Sphericity Assumed	48.633	38	1.280		
	Greenhouse-Geisser	48.633	35.424	1.373		
	Huynh-Feldt	48.633	38.000	1.280		
	Lower-bound	48.633	19.000	2.560		

T-Test for target 3

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1 T3PRO1 - T3EV	-2.9500	1.6694	.3733	-3.7313	-2.1687	-7.903	19	.000
Pair 2 T3PRO1 - T3MC	-.7500	1.3717	.3067	-1.3920	-.1080	-2.445	19	.024
Pair 3 T3EVO1 - T3MC	2.2000	1.7351	.3880	1.3880	3.0120	5.670	19	.000

Anova for target 4

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	.933	2	.467	.639	.533
	Greenhouse-Geisser	.933	1.420	.657	.639	.484
	Huynh-Feldt	.933	1.501	.622	.639	.492
	Lower-bound	.933	1.000	.933	.639	.434
Error(TYPE)	Sphericity Assumed	27.733	38	.730		
	Greenhouse-Geisser	27.733	26.972	1.028		
	Huynh-Feldt	27.733	28.523	.972		
	Lower-bound	27.733	19.000	1.460		

Anova for target 5

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	10.033	2	5.017	3.044	.059
	Greenhouse-Geisser	10.033	1.478	6.789	3.044	.077
	Huynh-Feldt	10.033	1.573	6.379	3.044	.074
	Lower-bound	10.033	1.000	10.033	3.044	.097
Error(TYPE)	Sphericity Assumed	62.633	38	1.648		
	Greenhouse-Geisser	62.633	28.081	2.230		
	Huynh-Feldt	62.633	29.883	2.096		
	Lower-bound	62.633	19.000	3.296		

Anova for target 6

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	20.033	2	10.017	3.833	.030
	Greenhouse-Geisser	20.033	1.933	10.363	3.833	.032
	Huynh-Feldt	20.033	2.000	10.017	3.833	.030
	Lower-bound	20.033	1.000	20.033	3.833	.065
Error(TYPE)	Sphericity Assumed	99.300	38	2.613		
	Greenhouse-Geisser	99.300	36.728	2.704		
	Huynh-Feldt	99.300	38.000	2.613		
	Lower-bound	99.300	19.000	5.226		

T-Test for target 6

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1 T6PRO2 - T6EV	-1.2500	2.4895	.5567	-2.4151	3.49E-02	-2.246	19	.037
Pair 2 T6PRO2 - T6MC	-1.2000	2.1667	.4845	-2.2141	-.1859	-2.477	19	.023
Pair 3 T6EVO2 - T6MC	000E-02	2.1879	.4892	-.9740	1.0740	.102	19	.920

Anova for target 7

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	11.200	2	5.600	3.980	.027
	Greenhouse-Geisser	11.200	1.938	5.778	3.980	.028
	Huynh-Feldt	11.200	2.000	5.600	3.980	.027
	Lower-bound	11.200	1.000	11.200	3.980	.061
Error(TYPE)	Sphericity Assumed	53.467	38	1.407		
	Greenhouse-Geisser	53.467	36.828	1.452		
	Huynh-Feldt	53.467	38.000	1.407		
	Lower-bound	53.467	19.000	2.814		

T-Test for target 7

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1 T7PRO2 - T7EV	-1.0000	1.5218	.3403	-1.7122	-.2878	-2.939	19	.008
Pair 2 T7PRO2 - T7MC	-.2000	1.7351	.3880	-1.0120	.6120	-.515	19	.612
Pair 3 T7EVO2 - T7MC	.8000	1.7652	.3947	2.61E-02	1.6261	2.027	19	.057

Anova for target 8

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	29.100	2	14.550	10.195	.000
	Greenhouse-Geisser	29.100	1.930	15.077	10.195	.000
	Huynh-Feldt	29.100	2.000	14.550	10.195	.000
	Lower-bound	29.100	1.000	29.100	10.195	.005
Error(TYPE)	Sphericity Assumed	54.233	38	1.427		
	Greenhouse-Geisser	54.233	36.673	1.479		
	Huynh-Feldt	54.233	38.000	1.427		
	Lower-bound	54.233	19.000	2.854		

T-Test for target 8

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	T8PRO2 - T8RV	-1.6500	1.8432	.4122	-2.5126	-.7874	-4.003	19	.001
Pair 2	T8PRO2 - T8MC	-1.2000	1.6092	.3598	-1.9531	-.4469	-3.335	19	.003
Pair 3	T8RVO2 - T8MC	.4500	1.6051	.3589	-.3012	1.2012	1.254	19	.225

Anova for target 9

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	7.500	2	3.750	1.282	.289
	Greenhouse-Geisser	7.500	1.691	4.436	1.282	.287
	Huynh-Feldt	7.500	1.838	4.081	1.282	.288
	Lower-bound	7.500	1.000	7.500	1.282	.272
Error(TYPE)	Sphericity Assumed	111.167	38	2.925		
	Greenhouse-Geisser	111.167	32.122	3.461		
	Huynh-Feldt	111.167	34.920	3.183		
	Lower-bound	111.167	19.000	5.851		

Anova for target 10

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	23.633	2	11.817	4.725	.015
	Greenhouse-Geisser	23.633	1.625	14.542	4.725	.022
	Huynh-Feldt	23.633	1.756	13.462	4.725	.019
	Lower-bound	23.633	1.000	23.633	4.725	.043
Error(TYPE)	Sphericity Assumed	95.033	38	2.501		
	Greenhouse-Geisser	95.033	30.878	3.078		
	Huynh-Feldt	95.033	33.356	2.849		
	Lower-bound	95.033	19.000	5.002		

T-Test for target 10

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1 T10PRO1 - T10EV	-1.4000	2.6636	.5956	-2.6466	-.1534	-2.351	19	.030
Pair 2 T10PRO1 - T10MC	-1.2500	1.7130	.3830	-2.0517	-.4483	-3.263	19	.004
Pair 3 T10EVO1 - T10MC	.1500	2.2308	.4988	-.8940	1.1940	.301	19	.767

Experiment 8: Array task

Friedman Test

Ranks

	Mean Rank
PROFIT	1.96
EVOFIT	2.01
MORPH	2.03

Test Statistics[a]

N	40
Chi-Square	.146
df	2
Asymp. Sig.	.930

a. Friedman Test

By target
Cochran Test

Test Statistics

N	8
Cochran's Q	39.371[a]
df	29
Asymp. Sig.	.095

a. 0 is treated as a success.

Experiment 9: Amount of information recalled

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TARGET	Sphericity Assumed	2.148	3	.716	.247	.861
	Greenhouse-Geisser	2.148	1.641	1.309	.247	.750
	Huynh-Feldt	2.148	3.000	.716	.247	.861
	Lower-bound	2.148	1.000	2.148	.247	.653
Error(TARGET)	Sphericity Assumed	26.070	9	2.897		
	Greenhouse-Geisser	26.070	4.922	5.296		
	Huynh-Feldt	26.070	9.000	2.897		
	Lower-bound	26.070	3.000	8.690		
FEATURE	Sphericity Assumed	207.367	7	29.624	12.325	.000
	Greenhouse-Geisser	207.367	2.303	90.036	12.325	.005
	Huynh-Feldt	207.367	7.000	29.624	12.325	.000
	Lower-bound	207.367	1.000	207.367	12.325	.039
Error(FEATURE)	Sphericity Assumed	50.477	21	2.404		
	Greenhouse-Geisser	50.477	6.909	7.305		
	Huynh-Feldt	50.477	21.000	2.404		
	Lower-bound	50.477	3.000	16.826		
TARGET * FEATURE	Sphericity Assumed	41.789	21	1.990	1.093	.379
	Greenhouse-Geisser	41.789	2.249	18.584	1.093	.396
	Huynh-Feldt	41.789	9.309	4.489	1.093	.400
	Lower-bound	41.789	1.000	41.789	1.093	.373
Error(TARGET*FEATURE	Sphericity Assumed	114.742	63	1.821		
	Greenhouse-Geisser	114.742	6.746	17.009		
	Huynh-Feldt	114.742	27.926	4.109		
	Lower-bound	114.742	3.000	38.247		

Anova for target 1

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
FEATURE	Sphericity Assumed	53.719	7	7.674	5.673	.001
	Greenhouse-Geisser	53.719	2.023	26.557	5.673	.041
	Huynh-Feldt	53.719	6.233	8.618	5.673	.002
	Lower-bound	53.719	1.000	53.719	5.673	.097
Error(FEATURE)	Sphericity Assumed	28.406	21	1.353		
	Greenhouse-Geisser	28.406	6.068	4.681		
	Huynh-Feldt	28.406	18.699	1.519		
	Lower-bound	28.406	3.000	9.469		

T-Test for target 1

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	T1SHAPE - T1HAI	1.2500	1.7078	.8539	-1.4675	3.9675	1.464	3	.239
Pair 2	T1SHAPE - T1BR(	2.7500	2.0616	1.0308	-.5304	6.0304	2.668	3	.076
Pair 3	T1SHAPE - T1EYE	2.5000	2.6458	1.3229	-1.7100	6.7100	1.890	3	.155
Pair 4	T1SHAPE - T1NOS	2.5000	3.1091	1.5546	-2.4473	7.4473	1.608	3	.206
Pair 5	T1SHAPE - T1MO	3.0000	2.4495	1.2247	-.8977	6.8977	2.449	3	.092
Pair 6	T1SHAPE - T1EAF	4.0000	1.4142	.7071	1.7497	6.2503	5.657	3	.011
Pair 7	T1HAIR - T1BROV	1.5000	.5774	.2887	.5813	2.4187	5.196	3	.014
Pair 8	T1HAIR - T1EYES	1.2500	.9574	.4787	-.2735	2.7735	2.611	3	.080
Pair 9	T1HAIR - T1NOSE	1.2500	1.7078	.8539	-1.4675	3.9675	1.464	3	.239
Pair 10	T1HAIR - T1MOUT	1.7500	.9574	.4787	.2265	3.2735	3.656	3	.035
Pair 11	T1HAIR - T1EARS	2.7500	.5000	.2500	1.9544	3.5456	11.000	3	.002

Anova for target 2

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
FEATURE	Sphericity Assumed	29.500	7	4.214	3.505	.012
	Greenhouse-Geisser	29.500	2.272	12.982	3.505	.087
	Huynh-Feldt	29.500	7.000	4.214	3.505	.012
	Lower-bound	29.500	1.000	29.500	3.505	.158
Error(FEATURE)	Sphericity Assumed	25.250	21	1.202		
	Greenhouse-Geisser	25.250	6.817	3.704		
	Huynh-Feldt	25.250	21.000	1.202		
	Lower-bound	25.250	3.000	8.417		

T-Test for target 2

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	T2SHAPE - T2HAIR	.2500	2.0616	1.0308	-3.0304	3.5304	.243	3	.824
Pair 2	T2SHAPE - T2BROW	1.2500	1.5000	.7500	-1.1368	3.6368	1.667	3	.194
Pair 3	T2SHAPE - T2EYES	1.2500	1.5000	.7500	-1.1368	3.6368	1.667	3	.194
Pair 4	T2SHAPE - T2NOSE	1.0000	2.1602	1.0801	-2.4374	4.4374	.926	3	.423
Pair 5	T2SHAPE - T2MOUTI	1.7500	2.2174	1.1087	-1.7783	5.2783	1.578	3	.213
Pair 6	T2SHAPE - T2EARS	3.0000	1.4142	.7071	.7497	5.2503	4.243	3	.024
Pair 7	T2HAIR - T2BROWS	1.0000	2.1602	1.0801	-2.4374	4.4374	.926	3	.423
Pair 8	T2HAIR - T2EYES	1.0000	2.1602	1.0801	-2.4374	4.4374	.926	3	.423
Pair 9	T2HAIR - T2NOSE	.7500	1.2583	.6292	-1.2522	2.7522	1.192	3	.319
Pair 10	T2HAIR - T2MOUTH	1.5000	2.3805	1.1902	-2.2879	5.2879	1.260	3	.297
Pair 11	T2HAIR - T2EARS	2.7500	1.8930	.9465	-.2621	5.7621	2.905	3	.062

Anova for target 3

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
FEATURE	Sphericity Assumed	98.219	7	14.031	4.230	.005
	Greenhouse-Geisser	98.219	1.780	55.180	4.230	.081
	Huynh-Feldt	98.219	4.197	23.404	4.230	.021
	Lower-bound	98.219	1.000	98.219	4.230	.132
Error(FEATURE)	Sphericity Assumed	69.656	21	3.317		
	Greenhouse-Geisser	69.656	5.340	13.044		
	Huynh-Feldt	69.656	12.590	5.533		
	Lower-bound	69.656	3.000	23.219		

T-Test for target 3

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	T3SHAPE - T3HAIF	3.2500	3.9476	1.9738	-3.0315	9.5315	1.647	3	.198
Pair 2	T3SHAPE - T3BRO	4.0000	2.4495	1.2247	.1023	7.8977	3.266	3	.047
Pair 3	T3SHAPE - T3EYE	4.2500	4.3493	2.1747	-2.6708	11.1708	1.954	3	.146
Pair 4	T3HAIR - T3NOSE	1.0000	2.1602	1.0801	-2.4374	4.4374	.926	3	.423
Pair 5	T3SHAPE - T3NOS	4.2500	3.9476	1.9738	-2.0315	10.5315	2.153	3	.120
Pair 6	T3SHAPE - T3MOU	4.2500	1.5000	.7500	1.8632	6.6368	5.667	3	.011
Pair 7	T3SHAPE - T3EAR	6.5000	3.6968	1.8484	.6175	12.3825	3.517	3	.039
Pair 8	T3HAIR - T3BROW	.7500	2.5000	1.2500	-3.2281	4.7281	.600	3	.591
Pair 9	T3HAIR - T3EYES	1.0000	1.1547	.5774	-.8374	2.8374	1.732	3	.182
Pair 10	T3HAIR - T3MOUTI	1.0000	2.4495	1.2247	-2.8977	4.8977	.816	3	.474
Pair 11	T3HAIR - T3EARS	3.2500	.5000	.2500	2.4544	4.0456	13.000	3	.001

Anova for target 4

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
FEATURE	Sphericity Assumed	67.719	7	9.674	4.848	.002
	Greenhouse-Geisser	67.719	1.514	44.741	4.848	.079
	Huynh-Feldt	67.719	2.728	24.828	4.848	.034
	Lower-bound	67.719	1.000	67.719	4.848	.115
Error(FEATURE)	Sphericity Assumed	41.906	21	1.996		
	Greenhouse-Geisser	41.906	4.541	9.229		
	Huynh-Feldt	41.906	8.183	5.121		
	Lower-bound	41.906	3.000	13.969		

T-Test for target 4

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	T4SHAPE - T4HAI	.0000	2.1602	1.0801	-3.4374	3.4374	.000	3	1.000
Pair 2	T4SHAPE - T4BR(	3.0000	3.7417	1.8708	-2.9538	8.9538	1.604	3	207
Pair 3	T4SHAPE - T4EYI	.7500	2.2174	1.1087	-2.7783	4.2783	.676	3	.547
Pair 4	T4SHAPE - T4NO	3.5000	2.8868	1.4434	-1.0935	8.0935	2.425	3	.094
Pair 5	T4SHAPE - T4MO	1.7500	1.2583	.6292	-.2522	3.7522	2.782	3	.069
Pair 6	T4SHAPE - T4EAI	3.0000	2.1602	1.0801	-.4374	6.4374	2.777	3	.069
Pair 7	T4HAIR - T4BROV	3.0000	2.0000	1.0000	-.1824	6.1824	3.000	3	.058
Pair 8	T4HAIR - T4EYES	.7500	1.5000	.7500	-1.6368	3.1368	1.000	3	.391
Pair 9	T4HAIR - T4NOSE	3.5000	1.0000	.5000	1.9088	5.0912	7.000	3	.006
Pair 10	T4HAIR - T4MOU	1.7500	1.5000	.7500	-.6368	4.1368	2.333	3	.102
Pair 11	T4HAIR - T4EARS	3.0000	1.4142	.7071	.7497	5.2503	4.243	3	.024

T-Test for target 4 - (eyes and mouth)

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	T4SHAPE - T4MOUTH	1.7500	1.2583	.6292	-.2522	3.7522	2.782	3	069
Pair 2	T4HAIR - T4MOUTH	1.7500	1.5000	.7500	-.6368	4.1368	2.333	3	102
Pair 3	T1NOSE - T1MOUTH	.5000	1.2910	.6455	-1.5543	2.5543	.775	3	495
Pair 4	T1BROWS - T1EYES	-.2500	.9574	.4787	-1.7735	1.2735	-.522	3	.638
Pair 5	T1EYES - T1NOSE	.0000	1.1547	.5774	-1.8374	1.8374	.000	3	1.000
Pair 6	T1EYES - T1MOUTH	.5000	.5774	.2887	-.4187	1.4187	1.732	3	182
Pair 7	T1EYES - T1EARS	1.5000	1.2910	.6455	-.5543	3.5543	2.324	3	.103
Pair 8	T1HAIR - T1EYES	1.2500	.9574	.4787	-.2735	2.7735	2.611	3	080
Pair 9	T1SHAPE - T1EYES	2.5000	2.6458	1.3229	-1.7100	6.7100	1.890	3	155
Pair 10	T1BROWS - T1MOUTH	.2500	1.2583	.6292	-1.7522	2.2522	.397	3	718
Pair 11	T1MOUTH - T1EARS	1.0000	1.1547	.5774	-.8374	2.8374	1.732	3	182

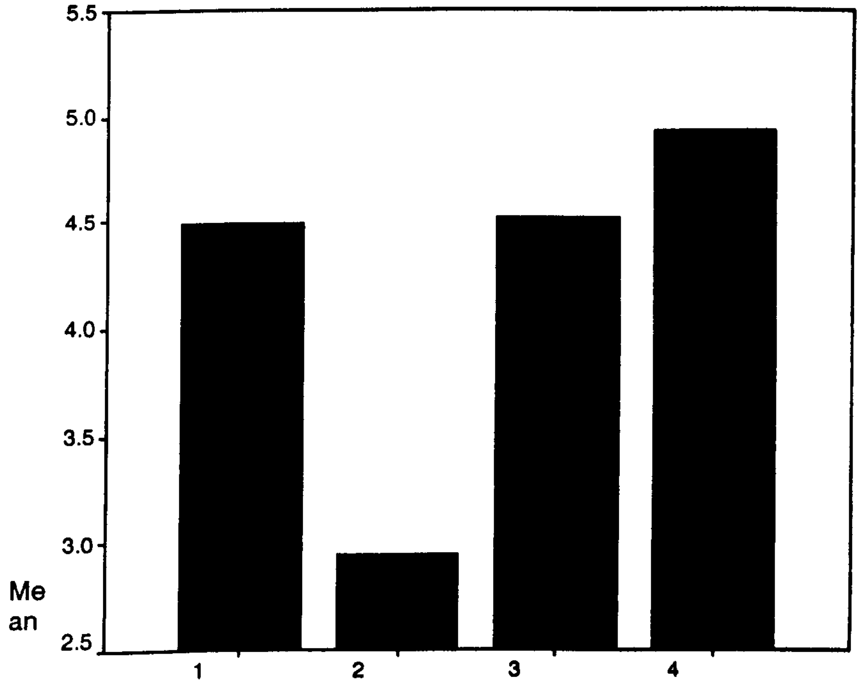
Likeness ratings

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TARGET	Sphericity Assumed	45.977	3	15.326	24.011	.000
	Greenhouse-Geisser	45.977	2.235	20.568	24.011	.000
	Huynh-Feldt	45.977	2.547	18.048	24.011	.000
	Lower-bound	45.977	1.000	45.977	24.011	.000
Error(TARGET)	Sphericity Assumed	36.382	57	.638		
	Greenhouse-Geisser	36.382	42.472	.857		
	Huynh-Feldt	36.382	48.402	.752		
	Lower-bound	36.382	19.000	1.915		

Graph



T-Tests on likeness ratings

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	TARGET1 - TARG	1.5500	.8095	.1810	1.1712	1.9288	8.563	19	.000
Pair 2	TARGET1 - TARG	2.50E-02	1.2352	.2762	-.6031	.5531	-.091	19	.929
Pair 3	TARGET1 - TARG	-.4375	.8424	.1884	-.8318	4.32E-02	-2.322	19	.031
Pair 4	TARGET2 - TARG	-1.5750	1.4489	.3240	-2.2531	-.8969	-4.861	19	.000
Pair 5	TARGET2 - TARG	-1.9875	1.0339	.2312	-2.4714	-1.5036	-8.597	19	.000
Pair 6	TARGET3 - TARG	-.4125	1.2651	.2829	-1.0046	.1796	-1.458	19	.161

Correlations

Correlations

		negative descriptors	positive descriptors	total recalled	mean ratings
negative descriptors	Pearson Correlation	1.000	.123	.586*	.234
	Sig. (2-tailed)	.	.649	.017	.384
	N	16	16	16	16
positive descriptors	Pearson Correlation	.123	1.000	.876**	.202
	Sig. (2-tailed)	.649	.	.000	.454
	N	16	16	16	16
total recalled	Pearson Correlation	.586*	.876**	1.000	.278
	Sig. (2-tailed)	.017	.000	.	.297
	N	16	16	16	16
mean ratings	Pearson Correlation	.234	.202	.278	1.000
	Sig. (2-tailed)	.384	.454	.297	.
	N	16	16	16	16

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

By target

Correlations

		1RECAL	T1RATE	2RECAL	T2RATE	3RECAL	T3RATE	4RECAL	T4RATE
T1RECA	Pearson Corr	1.000	-.149	1.000**	-.149	-.908	-.871	-.283	-.608
	Sig. (2-tailed)	.	.851	.000	.851	.092	.129	.717	.392
	N	4	4	4	4	4	4	4	4
T1RATE	Pearson Corr	-.149	1.000	-.149	1.000**	.196	.433	.904	-.138
	Sig. (2-tailed)	.851	.	.851	.000	.804	.567	.096	.862
	N	4	4	4	4	4	4	4	4
T2RECA	Pearson Corr	1.000**	-.149	1.000	-.149	-.908	-.871	-.283	-.608
	Sig. (2-tailed)	.000	.851	.	.851	.092	.129	.717	.392
	N	4	4	4	4	4	4	4	4
T2RATE	Pearson Corr	-.149	1.000**	-.149	1.000	.196	.433	.904	-.138
	Sig. (2-tailed)	.851	.000	.851	.	.804	.567	.096	.862
	N	4	4	4	4	4	4	4	4
T3RECA	Pearson Corr	-.908	.196	-.908	.196	1.000	.653	.477	.852
	Sig. (2-tailed)	.092	.804	.092	.804	.	.347	.523	.148
	N	4	4	4	4	4	4	4	4
T3RATE	Pearson Corr	-.871	.433	-.871	.433	.653	1.000	.361	.169
	Sig. (2-tailed)	.129	.567	.129	.567	.347	.	.639	.831
	N	4	4	4	4	4	4	4	4
T4RECA	Pearson Corr	-.283	.904	-.283	.904	.477	.361	1.000	.274
	Sig. (2-tailed)	.717	.096	.717	.096	.523	.639	.	.726
	N	4	4	4	4	4	4	4	4
T4RATE	Pearson Corr	-.608	-.138	-.608	-.138	.852	.169	.274	1.000
	Sig. (2-tailed)	.392	.862	.392	.862	.148	.831	.726	.
	N	4	4	4	4	4	4	4	4

**Correlation is significant at the 0.01 level (2-tailed).

Experiment 10: Amount of description

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
amount of info recalled	16	22.00	42.00	33.2500	6.1482
% correct matches	16	25.00	100.00	62.7500	22.3622
mean ratings	16	2.30	6.15	4.4812	1.0855
Valid N (listwise)	16				

Repeated Measures Anova

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig
TARGET	Sphericity Assumed	10.188	3	3.396	.553	.659
	Greenhouse-Geisser	10.188	1.707	5.969	.553	.580
	Huynh-Feldt	10.188	3.000	3.396	.553	.659
	Lower-bound	10.188	1.000	10.188	.553	.511
Error(TARGET)	Sphericity Assumed	55.250	9	6.139		
	Greenhouse-Geisser	55.250	5.121	10.790		
	Huynh-Feldt	55.250	9.000	6.139		
	Lower-bound	55.250	3.000	18.417		
FEATURE	Sphericity Assumed	407.500	7	58.214	21.661	.000
	Greenhouse-Geisser	407.500	1.288	316.374	21.661	.009
	Huynh-Feldt	407.500	1.841	221.319	21.661	.003
	Lower-bound	407.500	1.000	407.500	21.661	.019
Error(FEATURE)	Sphericity Assumed	56.438	21	2.688		
	Greenhouse-Geisser	56.438	3.864	14.606		
	Huynh-Feldt	56.438	5.524	10.217		
	Lower-bound	56.438	3.000	18.813		
TARGET * FEATURE	Sphericity Assumed	47.312	21	2.253	1.086	.386
	Greenhouse-Geisser	47.312	2.617	18.080	1.086	.401
	Huynh-Feldt	47.312	21.000	2.253	1.086	.386
	Lower-bound	47.312	1.000	47.312	1.086	.374
Error(TARGET*FEATURE	Sphericity Assumed	130.750	63	2.075		
	Greenhouse-Geisser	130.750	7.851	16.655		
	Huynh-Feldt	130.750	63.000	2.075		
	Lower-bound	130.750	3.000	43.583		

For target 1

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
FEATURE	Sphericity Assumed	92.969	7	13.281	7.933	.000
	Greenhouse-Geisser	92.969	1.969	47.216	7.933	.021
	Huynh-Feldt	92.969	5.699	16.312	7.933	.000
	Lower-bound	92.969	1.000	92.969	7.933	.067
Error(FEATURE)	Sphericity Assumed	35.156	21	1.674		
	Greenhouse-Geisser	35.156	5.907	5.952		
	Huynh-Feldt	35.156	17.098	2.056		
	Lower-bound	35.156	3.000	11.719		

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	T1SHAPE - T1HAI	1.7500	2.0616	1.0308	-1.5304	5.0304	1.698	3	.188
Pair 2	T1SHAPE - T1BR	3.2500	2.2174	1.1087	-.2783	6.7783	2.931	3	.061
Pair 3	T1SHAPE - T1EY	3.0000	2.5820	1.2910	-1.1085	7.1085	2.324	3	.103
Pair 4	T1SHAPE - T1NO	2.7500	2.6300	1.3150	-1.4348	6.9348	2.091	3	.128
Pair 5	T1SHAPE - T1MO	3.5000	1.2910	.6455	1.4457	5.5543	5.422	3	.012
Pair 6	T1SHAPE - T1EA	5.5000	1.2910	.6455	3.4457	7.5543	8.521	3	.003
Pair 7	recall by target - T1SHAPE	-5.5000	2.3805	1.1902	-9.2879	-1.7121	-4.621	3	.019
Pair 8	T1HAIR - T1BROW	1.5000	.5774	.2887	.5813	2.4187	5.196	3	.014
Pair 9	T1HAIR - T1EYES	1.2500	2.2174	1.1087	-2.2783	4.7783	1.127	3	.342
Pair 10	T1HAIR - T1NOSE	1.0000	1.6330	.8165	-1.5985	3.5985	1.225	3	.308
Pair 11	T1HAIR - T1MOUT	1.7500	.9574	.4787	.2265	3.2735	3.656	3	.035
Pair 12	T1HAIR - T1EARS	3.7500	.9574	.4787	2.2265	5.2735	7.833	3	.004

For Target 2

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
FEATURE	Sphericity Assumed	102.500	7	14.643	7.834	.000
	Greenhouse-Geisser	102.500	1.847	55.510	7.834	.025
	Huynh-Feldt	102.500	4.669	21.952	7.834	.001
	Lower-bound	102.500	1.000	102.500	7.834	.068
Error(FEATURE)	Sphericity Assumed	39.250	21	1.869		
	Greenhouse-Geisser	39.250	5.540	7.085		
	Huynh-Feldt	39.250	14.008	2.802		
	Lower-bound	39.250	3.000	13.083		

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	T2GEN - T2SHAF	-4.0000	1.8257	.9129	-6.9052	-1.0948	-4.382	3	.022
Pair 2	T2SHAPE - T2HA	-.7500	.9574	.4787	-2.2735	.7735	-1.567	3	.215
Pair 3	T2SHAPE - T2BF	.5000	1.0000	.5000	-1.0912	2.0912	1.000	3	.391
Pair 4	T2SHAPE - T2EY	1.0000	1.6330	.8165	-1.5985	3.5985	1.225	3	.308
Pair 5	T2SHAPE - T2NC	3.0000	.8165	.4082	1.7008	4.2992	7.348	3	.005
Pair 6	T2SHAPE - T2MC	1.7500	2.6300	1.3150	-2.4348	5.9348	1.331	3	.275
Pair 7	T2SHAPE - T2EA	4.5000	1.2910	.6455	2.4457	6.5543	6.971	3	.006
Pair 8	T2GEN - T2HAIR	-4.7500	.9574	.4787	-6.2735	-3.2265	-9.922	3	.002
Pair 9	T2HAIR - T2BRO	1.2500	1.5000	.7500	-1.1368	3.6368	1.667	3	.194
Pair 10	T2HAIR - T2EYES	1.7500	1.5000	.7500	-.6368	4.1368	2.333	3	.102
Pair 11	T2HAIR - T2NOS	3.7500	.9574	.4787	2.2265	5.2735	7.833	3	.004
Pair 12	T2HAIR - T2MOU	2.5000	3.1091	1.5546	-2.4473	7.4473	1.608	3	.206
Pair 13	T2HAIR - T2EARS	5.2500	.5000	.2500	4.4544	6.0456	21.000	3	.000

For target 3

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
FEATURE	Sphericity Assumed	86.375	7	12.339	4.516	.003
	Greenhouse-Geisser	86.375	2.015	42.871	4.516	.063
	Huynh-Feldt	86.375	6.150	14.045	4.516	.005
	Lower-bound	86.375	1.000	86.375	4.516	.124
Error(FEATURE)	Sphericity Assumed	57.375	21	2.732		
	Greenhouse-Geisser	57.375	6.044	9.492		
	Huynh-Feldt	57.375	18.449	3.110		
	Lower-bound	57.375	3.000	19.125		

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	T3GEN - T3SHAPE	-3.7500	3.2016	1.6008	-8.8444	1.3444	-2.343	3	.101
Pair 2	T3SHAPE - T3HAIR	.0000	1.4142	.7071	-2.2503	2.2503	.000	3	1.000
Pair 3	T3SHAPE - T3BRO	.7500	.9574	.4787	-.7735	2.2735	1.567	3	.215
Pair 4	T3SHAPE - T3EYE	.0000	1.4142	.7071	-2.2503	2.2503	.000	3	1.000
Pair 5	T3SHAPE - T3NOS	1.5000	.5774	.2887	.5813	2.4187	5.196	3	.014
Pair 6	T3SHAPE - T3MOU	1.0000	2.7080	1.3540	-3.3091	5.3091	.739	3	.514
Pair 7	T3SHAPE - T3EAR	4.5000	.5774	.2887	3.5813	5.4187	15.588	3	.001
Pair 8	T3GEN - T3HAIR	-3.7500	1.8930	.9465	-6.7621	-.7379	-3.962	3	.029
Pair 9	T3HAIR - T3BROW	.7500	2.3629	1.1815	-3.0099	4.5099	.635	3	.571
Pair 10	T3HAIR - T3EYES	.0000	2.1602	1.0801	-3.4374	3.4374	.000	3	1.000
Pair 11	T3HAIR - T3NOSE	1.5000	1.0000	.5000	9.12E-02	3.0912	3.000	3	.058
Pair 12	T3HAIR - T3MOUTH	1.0000	2.4495	1.2247	-2.8977	4.8977	.816	3	.474
Pair 13	T3HAIR - T3EARS	4.5000	1.2910	.6455	2.4457	6.5543	6.971	3	.006

For target 4

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
FEATURE	Sphericity Assumed	172.969	7	24.710	9.365	.000
	Greenhouse-Geisser	172.969	2.378	72.746	9.365	.009
	Huynh-Feldt	172.969	7.000	24.710	9.365	.000
	Lower-bound	172.969	1.000	172.969	9.365	.055
Error(FEATURE)	Sphericity Assumed	55.406	21	2.638		
	Greenhouse-Geisser	55.406	7.133	7.767		
	Huynh-Feldt	55.406	21.000	2.638		
	Lower-bound	55.406	3.000	18.469		

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	T4GEN - T4SHAPE	-6.7500	2.2174	1.1087	-10.2783	-3.2217	-6.088	3	.009
Pair 2	T4SHAPE - T4HAIR	2.5000	1.0000	.5000	.9088	4.0912	5.000	3	.015
Pair 3	T4SHAPE - T4BRO	2.7500	2.6300	1.3150	-1.4348	6.9348	2.091	3	.128
Pair 4	T4SHAPE - T4EYE	3.2500	1.7078	.8539	.5325	5.9675	3.806	3	.032
Pair 5	T4SHAPE - T4NOS	5.0000	2.1602	1.0801	1.5626	8.4374	4.629	3	.019
Pair 6	T4SHAPE - T4MOU	5.5000	1.2910	.6455	3.4457	7.5543	8.521	3	.003
Pair 7	T4SHAPE - T4EAR	7.5000	1.7321	.8660	4.7439	10.2561	8.660	3	.003
Pair 8	T4GEN - T4HAIR	-4.2500	1.5000	.7500	-6.6368	-1.8632	-5.667	3	.011
Pair 9	T4HAIR - T4BROW	.2500	3.0957	1.5478	-4.6759	5.1759	.162	3	.882
Pair 10	T4HAIR - T4EYES	.7500	1.8930	.9465	-2.2621	3.7621	.792	3	.486
Pair 11	T4HAIR - T4NOSE	2.5000	1.2910	.6455	.4457	4.5543	3.873	3	.030
Pair 12	T4HAIR - T4MOUT	3.0000	1.8257	.9129	.484E-02	5.9052	3.286	3	.046
Pair 13	T4HAIR - T4EARS	5.0000	1.4142	.7071	2.7497	7.2503	7.071	3	.006

Accuracy of description

Friedman test

Ranks

	Mean Rank
TARGET1	2.00
TARGET2	3.38
TARGET3	1.88
TARGET4	2.75

Test Statistics[a]

N	8
Chi-Square	7.944
df	3
Asymp. Sig.	.047

a. Friedman Test

Wilcoxon signed ranks test

Test Statistics[c]

	TARGET2 - TARGET1	TARGET3 - TARGET1	TARGET4 - TARGET1	TARGET3 - TARGET2	TARGET4 - TARGET2	TARGET4 - TARGET3
Z	-1.930[a]	-.991[b]	-1.406[a]	-2.232[b]	-1.342[b]	-1.802[a]
Asymp. Sig. (2-tailed)	.054	.322	.160	.026	.180	.072

a. Based on negative ranks.

b. Based on positive ranks.

c. Wilcoxon Signed Ranks Test

Composite likeness ratings

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
mean ratings	16	2.30	6.15	4.4812	1.0855
Valid N (listwise)	16				

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
T1COMP	20	3.00	8.50	5.0375	1.4079
T2COMP	20	2.00	7.00	4.5000	1.4600
T3COMP	20	1.00	8.00	3.6125	1.7631
T4COMP	20	1.75	7.50	4.7750	1.3473
Valid N (listwise)	20				

Anova

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TARGET	Sphericity Assumed	23.016	3	7.672	15.983	.000
	Greenhouse-Geisser	23.016	2.259	10.190	15.983	.000
	Huynh-Feldt	23.016	2.579	8.925	15.983	.000
	Lower-bound	23.016	1.000	23.016	15.983	.001
Error(TARGET)	Sphericity Assumed	27.359	57	.480		
	Greenhouse-Geisser	27.359	42.914	.638		
	Huynh-Feldt	27.359	48.998	.558		
	Lower-bound	27.359	19.000	1.440		

T-Tests

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	T1COMP	5.0375	20	1.4079	.3148
	T2COMP	4.5000	20	1.4600	.3265
Pair 2	T1COMP	5.0375	20	1.4079	.3148
	T3COMP	3.6125	20	1.7631	.3942
Pair 3	T1COMP	5.0375	20	1.4079	.3148
	T4COMP	4.7750	20	1.3473	.3013
Pair 4	T2COMP	4.5000	20	1.4600	.3265
	T3COMP	3.6125	20	1.7631	.3942
Pair 5	T2COMP	4.5000	20	1.4600	.3265
	T4COMP	4.7750	20	1.3473	.3013
Pair 6	T3COMP	3.6125	20	1.7631	.3942
	T4COMP	4.7750	20	1.3473	.3013

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1 T1COMP - T2CO	.5375	1.1188	.2502	387E-02	1.0611	2.148	19	.045
Pair 2 T1COMP - T3CO	1.4250	.8777	.1963	1.0142	1.8358	7.261	19	.000
Pair 3 T1COMP - T4CO	.2625	.7715	.1725	9.86E-02	.6236	1.522	19	.145
Pair 4 T2COMP - T3CO	.8875	1.2044	.2693	.3238	1.4512	3.296	19	.004
Pair 5 T2COMP - T4CO	-.2750	.7691	.1720	-.6349	493E-02	-1.599	19	.126
Pair 6 T3COMP - T4CO	-1.1625	1.0490	.2346	-1.6535	-.6715	-4.956	19	.000

Correlations

Correlations

		amount of info recalled	% correct matches	mean ratings
amount of info recalled	Pearson Correlation	1.000	-.349	-.123
	Sig. (2-tailed)	.	.186	.651
	N	16	16	16
% correct matches	Pearson Correlation	-.349	1.000	.314
	Sig. (2-tailed)	.186	.	.237
	N	16	16	16
mean ratings	Pearson Correlation	-.123	.314	1.000
	Sig. (2-tailed)	.651	.237	.
	N	16	16	16

Correlations

		RECAL	ACCU	RATIN	RECAL	ACCU	RATIN	RECAL	ACCU	RATIN	RECAL	ACCU	RATIN
T1REC	Pearson C	1.000	-.686	.176	-.083	-.330	.962*	-.599	-.462	.527	.944	-.528	-.455
	Sig. (2-taile	.	.314	.824	.917	.670	.038	.401	.538	.473	.056	.472	.545
	N	4	4	4	4	4	4	4	4	4	4	4	4
T1ACC	Pearson C	-.686	1.000	.363	.487	.913	-.694	.983*	-.192	-.559	-.844	.198	-.279
	Sig. (2-taile	.314	.	.637	.513	.087	.306	.017	.808	.441	.156	.802	.721
	N	4	4	4	4	4	4	4	4	4	4	4	4
T1RAT	Pearson C	.176	.363	1.000	-.206	.569	.333	.529	-.942	-.672	.124	-.838	-.884
	Sig. (2-taile	.824	.637	.	.794	.431	.667	.471	.058	.328	.876	.162	.116
	N	4	4	4	4	4	4	4	4	4	4	4	4
T2REC	Pearson C	-.083	.487	-.206	1.000	.585	-.322	.390	.062	.443	-.389	.569	-.151
	Sig. (2-taile	.917	.513	.794	.	.415	.678	.610	.938	.557	.611	.431	.849
	N	4	4	4	4	4	4	4	4	4	4	4	4
T2ACC	Pearson C	-.330	.913	.569	.585	1.000	-.361	.939	-.509	-.430	-.566	-.039	-.618
	Sig. (2-taile	.670	.087	.431	.415	.	.639	.061	.491	.570	.434	.961	.382
	N	4	4	4	4	4	4	4	4	4	4	4	4
T2RAT	Pearson C	.962*	-.694	.333	-.322	-.361	1.000	-.571	-.558	.304	.972*	-.717	-.498
	Sig. (2-taile	.038	.306	.667	.678	.639	.	.429	.442	.696	.028	.283	.502
	N	4	4	4	4	4	4	4	4	4	4	4	4
T3REC	Pearson C	-.599	.983*	.529	.390	.939	-.571	1.000	-.357	-.653	-.748	.013	-.422
	Sig. (2-taile	.401	.017	.471	.610	.061	.429	.	.643	.347	.252	.987	.578
	N	4	4	4	4	4	4	4	4	4	4	4	4
T3ACC	Pearson C	-.462	-.192	-.942	.062	-.509	-.558	-.357	1.000	.385	-.350	.847	.977*
	Sig. (2-taile	.538	.808	.058	.938	.491	.442	.643	.	.615	.650	.153	.023
	N	4	4	4	4	4	4	4	4	4	4	4	4
T3RAT	Pearson C	.527	-.559	-.672	.443	-.430	.304	-.653	.385	1.000	.419	.442	.272
	Sig. (2-taile	.473	.441	.328	.557	.570	.696	.347	.615	.	.581	.558	.728
	N	4	4	4	4	4	4	4	4	4	4	4	4
T4REC	Pearson C	.944	-.844	.124	-.389	-.566	.972*	-.748	-.350	.419	1.000	-.592	-.279
	Sig. (2-taile	.056	.156	.876	.611	.434	.028	.252	.650	.581	.	.408	.721
	N	4	4	4	4	4	4	4	4	4	4	4	4
T4ACC	Pearson C	-.528	.198	-.838	.569	-.039	-.717	.013	.847	.442	-.592	1.000	.720
	Sig. (2-taile	.472	.802	.162	.431	.961	.283	.987	.153	.558	.408	.	.280
	N	4	4	4	4	4	4	4	4	4	4	4	4
T4RAT	Pearson C	-.455	-.279	-.884	-.151	-.618	-.498	-.422	.977*	.272	-.279	.720	1.000
	Sig. (2-taile	.545	.721	.116	.849	.382	.502	.578	.023	.728	.721	.280	.
	N	4	4	4	4	4	4	4	4	4	4	4	4

*Correlation is significant at the 0.05 level (2-tailed).

Experiment 11: Combined presentations

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
COMP	128	.00	1.00	.3906	.4898
DESCR	128	.00	1.00	.4922	.5019
CDS	128	.00	1.00	.5391	.5004
CDT	128	.00	1.00	.5469	.4998
Valid N (listwise)	128				

Cochran Test

Frequencies

	Value	
	0	1
COMP	78	50
DESCR	65	63
CDS	59	69
CDT	58	70

Test Statistics

N	128
Cochran's Q	7.938[a]
df	3
Asymp. Sig.	.047

a. 1 is treated as a success.

McNemar Tests

Test Statistics[b]

	CDS & CDT	DESCR & CDT	COMP & CDT	COMP & DESCR	DESCR & CDS	COMP & CDS
N	128	128	128	128	128	128
Chi-Square[a]	.000	.610	5.014	2.286	.379	5.143
Asymp. Sig.	1.000	.435	.025	.131	.538	.023

a. Continuity Corrected

b. McNemar Test

By target
Cochran Test

Test Statistics

N	32
Cochran's Q	58.636[a]
df	15
Asymp. Sig.	.000

a. 1 is treated as a success.

For target 1- Cochran Test

Test Statistics

N	32
Cochran's Q	7.165[a]
df	3
Asymp. Sig.	.067

a. 1 is treated as a success.

Test Statistics[b]

	T1C & T1D	T1C & T1CDS	T1C & T1DCT	T1D & T1CDS	T1D & T1DCT	T1CDS & T1DCT
N	32	32	32	32	32	32
Exact Sig. (2-tailed)	.424[a]	.057[a]	.022[a]	.454[a]	.332[a]	1.000[a]

a. Binomial distribution used.

b. McNemar Test

For target 2
Cochran Test

Test Statistics

N	32
Cochran's Q	5.872[a]
df	3
Asymp. Sig.	.118

a. 1 is treated as a success.

McNemar Test

Test Statistics[b]

	T2C & T2D	T2C & T2CDS	T2C & T2CDT	T2D & T2CDS	T2D & T2CDT	T2CDS & T2CDT
N	32	32	32	32	32	32
Exact Sig. (2-tailed)	.454[a]	.022[a]	.092[a]	.383[a]	.549[a]	.824[a]

a. Binomial distribution used.

b. McNemar Test

For target 3
Cochran Test

Test Statistics

N	32
Cochran's Q	7.737[a]
df	3
Asymp. Sig.	.052

a. 0 is treated as a success.

McNemar Test

Test Statistics[b]

	T3C & T3D	T3C & T3CDS	T3C & T3CDT	T3CDS & T3CDT	T3D & T3CDT	T3D & T3CDS
N	32	32	32	32	32	32
Exact Sig. (2-tailed)	.022[a]	.180[a]	.021[a]	.607[a]	1.000[a]	.481[a]

a. Binomial distribution used.

b. McNemar Test

For target 4
Cochran Test

Test Statistics

N	32
Cochran's Q	1.277[a]
df	3
Asymp. Sig.	.735

a. 0 is treated as a success.

Experiment 12

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
composite	40	.00	2.00	.7250	.8469
description	40	.00	2.00	1.0500	.7143
witness composite and description combined	40	.00	2.00	1.2000	.7232
perfect description	40	.00	2.00	1.1500	.8022
combined perfect description and witness composite	40	.00	2.00	1.1750	.5943
Valid N (listwise)	40				

Friedman test

Ranks

	Mean Rank
composite	2.46
description	2.99
witness composite and description combined	3.21
perfect description	3.14
combined perfect description and witness composite	3.20

Test Statistics[a]

N	40
Chi-Square	8.387
df	4
Asymp. Sig.	.078

a. Friedman Test

Wilcoxon tests

Test Statistics

	escription composite	witness composite and descriptior ombined composite	perfect escription composite	combined perfect descriptior nd witness omposite composite	witness composite and descriptior ombined lescriptior	perfect escription lescriptior	combined perfect lescriptior nd witness omposite lescriptior	perfect escription witness composite and lescriptior combined	combined perfect lescriptior nd witness omposite witness composite and lescriptior combined	combined perfect lescriptior nd witness omposite perfect lescriptior
Z	-1.669[a]	-2.526[a]	-2.646[a]	-2.921[a]	-.868[a]	-.632[a]	-.939[a]	-.334[b]	-.114[b]	-.156[a]
Asymp. Sig. (2	.095	.012	.008	.003	.385	.528	.348	.738	.909	.876

a. Based on negative ranks.

b. Based on positive ranks.

c. Wilcoxon Signed Ranks Test

For target 1
Cochran Test

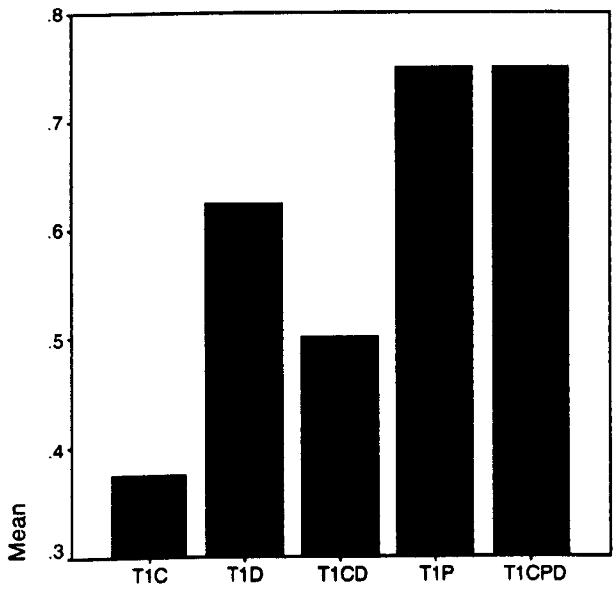
Frequencies

	Value	
	0	1
T1C	5	3
T1D	3	5
T1CD	4	4
T1P	2	6
T1CPD	2	6

Test Statistics

N	8
Cochran's Q	4.533[a]
df	4
Asymp. Sig.	.339

a. 1 is treated as a success.



For target 2

Frequencies

	Value	
	0	1
T2C	3	5
T2D	4	4
T2CD	4	4
T2P	4	4
T2CPD	2	6

Test Statistics

N	8
Cochran's Q	1.455[a]
df	4
Asymp. Sig.	.835

a. 0 is treated as a success.

For target 3
Cochran Test

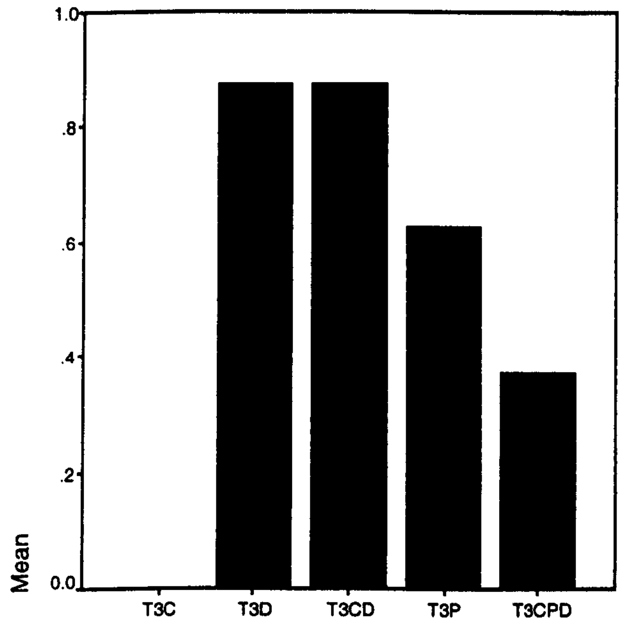
Frequencies

	Value	
	0	1
T3C	8	0
T3D	1	7
T3CD	1	7
T3P	3	5
T3CPD	5	3

Test Statistics

N	8
Cochran's Q	16.762[a]
df	4
Asymp. Sig.	.002

a. 0 is treated as a success.



McNemar Test

Test Statistics[b]

	3C & T3I	3C & T3C	3C & T3	3C & T3CP	3D & T3C	3D & T3	3D & T3CP	3CD & T3	T3CD & T3CPD	3P & T3CP
N	8	8	8	8	8	8	8	8	8	8
Exact Sig. (2-	.016[a]	.016[a]	.063[a]	.250[a]	1.000[a]	.625[a]	.125[a]	.500[a]	.125[a]	.625[a]

a. Binomial distribution used.

b. McNemar Test

For target 4

Frequencies

	Value	
	0	1
T4C	4	4
T4D	5	3
T4CD	6	2
T4P	1	7
T4CPD	2	6

Test Statistics

N	8
Cochran's Q	8.190[a]
df	4
Asymp. Sig.	.085

a. 1 is treated as a success.

Test Statistics[b]

	T4D & T4CD	T4C & T4CD	T4P & T4CPD	T4CD & T4P
N	8	8	8	8
Exact Sig. (2-tailed)	1.000[a]	.500[a]	1.000[a]	.063[a]

a. Binomial distribution used.

b. McNemar Test

Target 5
Cochran Test

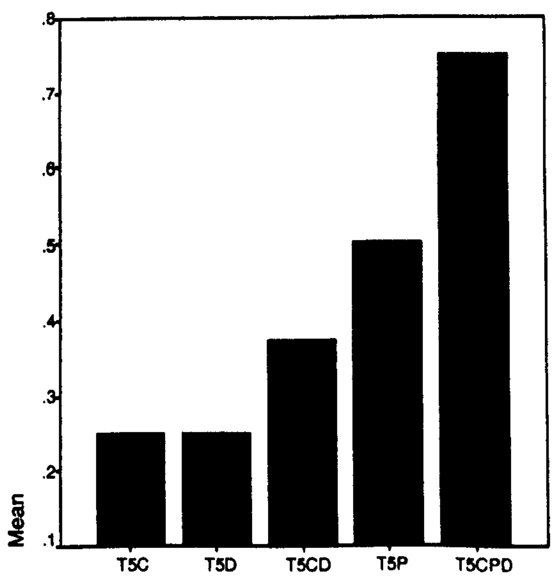
Frequencies

	Value	
	0	1
T5C	6	2
T5D	6	2
T5CD	5	3
T5P	4	4
T5CPD	2	6

Test Statistics

N	8
Cochran's Q	4.870[a]
df	4
Asymp. Sig.	.301

a. 1 is treated as a success.



For target 6
Cochran Test

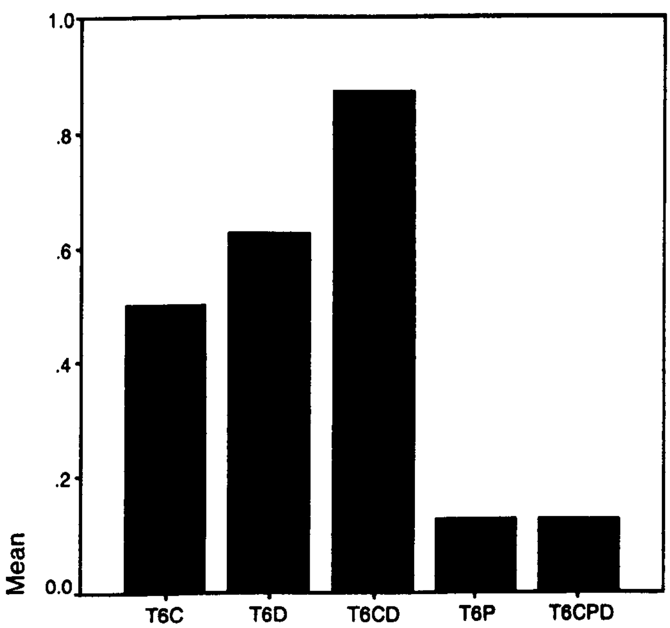
Frequencies

	Value	
	0	1
T6C	4	4
T6D	3	5
T6CD	1	7
T6P	7	1
T6CPD	7	1

Test Statistics

N	8
Cochran's Q	11.826[a]
df	4
Asymp. Sig.	.019

a. 0 is treated as a success.



McNemar Test

Test Statistics[b]

	6C & T6	C & T6C	6C & T6	C & T6CP	D & T6C	6D & T6	D & T6CP	CD & T6	T6CD & T6CPD	P & T6CP
N	8	8	8	8	8	8	8	8	8	8
Exact Sig. (	1.000[a]	.250[a]	.250[a]	.375[a]	.625[a]	.219[a]	.125[a]	.031[a]	.070[a]	1.000[a]

a Binomial distribution used.

b McNemar Test

For target 7
Cochran Test

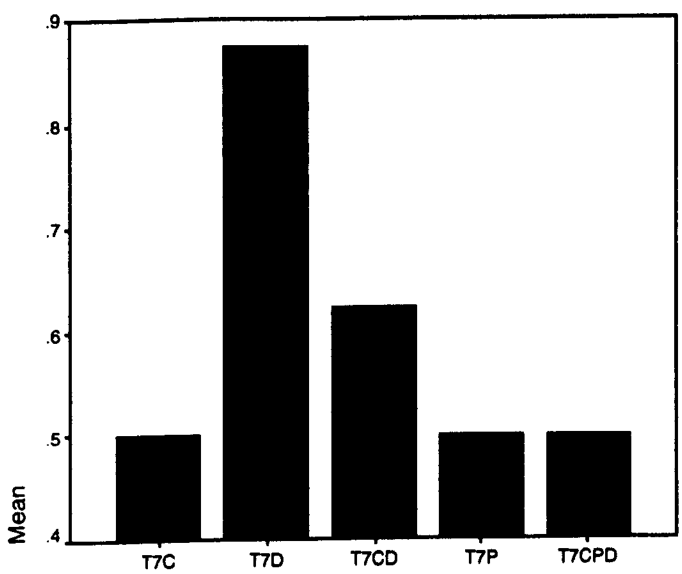
Frequencies

	Value	
	0	1
T7C	4	4
T7D	1	7
T7CD	3	5
T7P	4	4
T7CPD	4	4

Test Statistics

N	8
Cochran's Q	3.400[a]
df	4
Asymp. Sig.	.493

a. 1 is treated as a success.



For target 8
Cochran Test

Frequencies

	Value	
	0	1
T8C	5	3
T8D	8	0
T8CD	5	3
T8P	2	6
T8CPD	3	5

Test Statistics

N	8
Cochran's Q	10.095[a]
df	4
Asymp. Sig.	.039

a. 0 is treated as a success.

McNemar Test

Test Statistics

	8C & T8	8C & T8C	8C & T8	C & T8CP	D & T8C	8D & T8	D & T8CP	CD & T8	T8CD & T8CPD	P & T8CP
N	8	8	8	8	8	8	8	8	8	8
Exact Sig. (2	.250[a]	1.000[a]	.375[a]	.625[a]	.250[a]	.031[a]	.063[a]	.375[a]	.688[a]	1.000[a]

a. Binomial distribution used.

b. McNemar Test

For target 9
Cochran Test

Frequencies

	Value	
	0	1
T9C	5	3
T9D	1	7
T9CD	0	8
T9P	0	8
T9CPD	1	7

Test Statistics

N	8
Cochran's Q	14.333[a]
df	4
Asymp. Sig.	.006

a. 0 is treated as a success.

Test Statistics

	9C & T9	C & T9C	9C & T9	C & T9CP	D & T9C	9D & T9	D & T9CP	T9CD & T9CPD	P & T9CP
N	8	8	8	8	8	8	8	8	8
Exact Sig. (	.125[a]	.063[a]	.063[a]	.125[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]

a. Binomial distribution used.

b. McNemar Test

McNemar Test

For target 10
Cochran Test

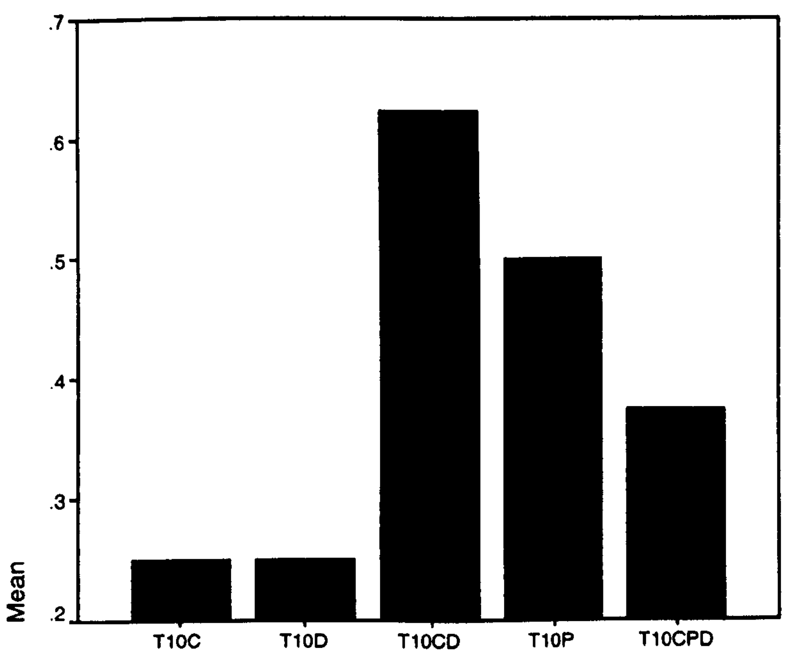
Frequencies

	Value	
	0	1
T10C	6	2
T10D	6	2
T10CD	3	5
T10P	4	4
T10CPD	5	3

Test Statistics

N	8
Cochran's Q	3.091[a]
df	4
Asymp. Sig.	.543

a. 1 is treated as a success.



Amount of information

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
amount of info recalled	10	35.00	68.00	48.5000	10.3414
Valid N (listwise)	10				

Standard Description Sheet

Participant No: Date: Start time: Gender: Target No:

Description of Facial Features

SHAPE

HAIR

BROWS

EYES

NOSE

MOUTH

EARS

End time: