Running head: FACE COMPOSITE PRODUCTION

# 'It's the Thought that Counts': Face Composite Production Can Hamper Recognition Performance

Kate-Victoria Kempen
Department of Psychology
University of Cape Town

Supervisor: Colin Tredoux

Word Count:

Abstract: [276]

Main Body: [9 906]

Abstract

Eyewitness testimony in forensic settings is viewed as a valid technique, yet 77% of exonerated prisoners in the U.S.A. were misidentified due to mistaken eyewitness testimonies over the last 25 years. These eyewitnesses usually identified the suspects after constructing a composite likeness of the suspect face. This study investigated whether composite production had a detrimental effect on later recognition accuracy. Experiment 1 randomly assigned 124 participants into a control group, a composite construction group, and a yoked-composite exposure group. Participants were exposed to a target and were required to select the same face from a lineup 2 days later. Results showed a significant difference between accuracy for the controls (65.12% accuracy) and the constructs (44.19% accuracy), which supports the finding by Wells, Charman, & Olson (2005). The viewers obtained the lowest score (36.84% accuracy). However, when biased lineups were corrected for, the difference between controls and constructors was no longer significant. The view group remained significantly different when compared to the controls. In Experiment 2, the composites created were blended with the target face to form a graded lineup of similar faces. Both groups performed just above chance levels, and there was no significant difference between groups. There was a significant difference between White and Other participants who obtained the correct answers, with the latter group selecting the target more often. The White faces were more likely to choose a blend of the faces than the Other participants. People encode other-race faces featurally, and since composite construction is a featural process, the Other participants obtained higher accuracies. Results suggest that it is mere exposure to a composite, and not construction itself, that contaminates the memory trace.

*Keywords*: composite production; face construction; face recognition; identification; simultaneous lineups; photographic lineup; eyewitness testimony.

## Weak Evidence for Eyewitness Testimony Accuracy

When an offence is committed, suspects are usually apprehended through the discovery of tangible evidence. Often, where there is no visible evidence to tie a suspect to a crime, law enforcement relies on the testimony of an eyewitness, who identifies the suspect from a lineup. On the surface, the notion of an eyewitness seems reliable. Tangible evidence, such as strands of hair, can be planted at the scene of a crime. Thus, having an eyewitness to the crime would truly tie the suspect to the incident.

However, eyewitness' perceptions can be distorted when witnessing an event. At the time of this writing, 235 prisoners, including 17 inmates on death-row, have been exonerated through DNA evidence in the last two decades in the United States of America. It has been found that 77% of these exonerations were due to eyewitness misidentification testimony. The average time spent imprisoned was 12 years (Innocence Project, n.d.). These misidentifications result in two major problems in that not only does the guilty perpetrator remain free, but an innocent individual is convicted (Wells, Memon, & Penrod, 2006). It is due to reasons such as these that research regarding facial recognition should be drastically increased, so that unequivocal results may begin to form. This could ultimately affect the way law enforcement evaluate eyewitness testimony.

## Composite Production Systems

Many of the exonerated inmates mentioned above had been apprehended and misidentified after an eyewitness had constructed a composite from memory. A composite is a pictorial likeness of a face, in which features (such as hair, eyes, nose and mouth) are selected to match those of the suspect.

Police traditionally used sketch artists who were trained in drawing specific and detailed suspect faces. In more contemporary settings, a variety of composite systems are used, such as the Identi-Kit (Laughery & Fowler, 1980), Photofit (Christie & Ellis, 1981), Mac-A-Mug (Wogalter, Laughery, & Thompson, 1986), and FACES (Wells, Charman, & Olson, 2005). The latter two have reported greater composite likeness, due to the amount of detail and the number of facial features available for selection within each system (Wells et al.).

Viewing or constructing a composite after the crime or incident may interfere with the original memory trace. Thus, research in forensic settings should commit itself to analysing the processes that occur during a construction interference phase. Information resulting from these studies can have a considerable impact on the composite-implementation techniques used by law enforcement systems. This can result in a more effective way to gain knowledge from eyewitnesses without confounding later recognition.

## Consequences of Composite Production on Recognition

Most face composite research takes the form of three phases: (a) encoding, (b) interference, and (c) recognition. It is during the interference phase in which the construction of composites appears to impact upon later recognition. The literature surrounding the detrimental consequences of face construction on future recognition will be discussed below, with regards to theories and claims that possibly underpin these notions.

### *Detrimental Consequences*

One study found hampered identification accuracies when the participants worked with sketch artists (Hall, 1976). Sketch artists are able to construct and alter an unlimited number of features. Thus, this decrement in recognition ability cannot be attributed to the limited number of features within composite systems.

The Identi-Kit operates by superimposing features on transparent overlays to form a face. One study found increased witness conservativeness when participants used the Identi-Kit to construct composites. Thus, participants were less likely to select *any* photograph from a lineup (Yu & Geiselman, 1993). Participants who constructed a composite were more uncertain about their memory of the original target. It appeared that they would rather reject the lineup completely than select a foil. This study also introduced a group who viewed the target and were required to provide a written description of the face. Research has found verbalisation of an image to interfere with the original memory trace (Loftus & Loftus, 1980; Schooler & Engstler-Schooler, 1990). The detrimental effects in recognition when participants worked with sketch artists (Hall, 1976) could be attributed to the need for excessive verbalisation of features. This ‘verbal overshadowing effect’ is found to impair processing of a face by inducing participants to focus on specific features (Fallshore & Schooler, 1995). A meta-analysis further showed verbalisation of a suspect’s features largely overshadows the original image in later identification (Meissner & Brigham, 2001b).

A study using the Mac-A-Mug program to construct composites showed that as detail increases in the systems, so recognition accuracy decreases (Wogalter et al., 1986, Experiment 2). Mac-A-Mug does not require the assistance of an intervening technician, which therefore eliminates any verbal description. However, results still indicated a negative effect in identification after constructing a composite.

Another study showed that participants merely exposed to filler-mugshots after they had viewed the target, obtained decreased recognition accuracies (Brown, Deffenbacher, &

Sturgill, 1977). The researchers asked participants to view a lineup of strangers who posed as criminals. Three days later, the same participants were shown a set of 15 mugshots. One week later, the participants were required to select the suspect from a lineup. Participants in the mugshot-viewing group misidentified the target more often than the controls who did not view the mugshots.

Another study also demonstrated a poor performance in recognition ability after participants were required to construct a composite (Comish, 1987). Subsequent to construction, the participants viewed a lineup which contained foils that resembled the errors made during construction. Comish found that participants who constructed a composite but were not exposed to the yoked-foil lineup were more likely to reject the lineup. This result is consistent with the Yu and Geiselman (1993) finding that construction increases witness conservativeness.

Although a meta-analysis (Meissner & Brigham, 2001b) suggests that construction has a positive effect on later recognition performance, this research predates a more recent study conducted by Wells et al. (2005). Wells et al. used the FACES 3.0 composite system, and found dramatic decreases in recognition accuracy after participants constructed a composite (10% accuracy). This poor accuracy was compared to the control group (84% accuracy). Their study utilised three groups: a control, a composite production, and a yoked-composite group. This implementation of a yoked-composite group aimed to similarly depict the Brown et al. (1977) condition where participants were exposed to mugshots. Wells et al. aimed to eliminate the possibility that it was merely viewing a composite that resulted in decreased recognition. The authors would then be able to attribute any error to the construction itself. Wells et al. also postulated that composite-exposure was externally provided information, as opposed to a self-generated, internalised image. Thus, these yoked-composite exposure participants would be able to separate this composite memory from their own original visual memory.

Unusually, Wells et al. (2005) used a 180- second exposure time when they presented the target face. This is drastic compared to the usual time of 16 seconds that is typically implemented (Maskow, Schmidt, Tredoux, & Nunez, 2007). This is also unusual as prolonged exposure to a target has been found to strengthen recognition performance (Shapiro & Penrod, 1986). During these 180 seconds, participants were required to rate the target on 10 traits: traits: attractive, intelligent, warm, aggressive, kind, happy, foolish, humourous, studious, and likeable. The traits were rated on a scale that ranged from

0 (not at all) to 10 (very). A previous study found that trait-encoding resulted in better identification accuracy (Wells & Hryciw, 1984).

Wells et al. (2005) had participants write a verbal description of the face after viewing had taken place. Verbal description of a face relies on accessing features and the verbal memory trace. Due to the verbal overshadowing effect, this has been found to impair identification. It seems unusual that Wells et al. would make use of a trait-encoding procedure to enhance recognition, followed by a feature-based description session to impair recognition.

Construction participants were instructed on how to use FACES to construct a target face. Yoked-composite exposure participants were shown a composite produced by the composite builders. The yoked-composite, composite-construction, and control groups were dismissed and told to return for a follow-up session in two days' time.

The participants returned to the laboratory and attempted to identify the target from a simultaneous lineup, consisting of five fillers and the target. The participants were given the instruction that the suspect may or may not be present. If the participant made no identification, they were asked to choose a suspect if they were forced to choose.

The Wells et al. (2005) results indicated that construction resulted in less accurate identification of the target (10%), compared to the other two groups. This is less than chance accuracy (16.67%).

Wells et al. (2005) concluded that constructing a composite was detrimental to the original memory trace. Also, 58% of participants in the composite construction group rejected the lineup, making no identification. These participants became more conservative in their selection and decided to make no choice than possibly choose incorrectly. When forced to choose, the same group only yielded 30% accuracy, almost double than if the participants had chosen at random. If conservativeness is taken away, construction has a small facilitating effect in that accuracy is higher than taking a blind guess.

The yoked-composite condition had a higher accuracy of 44% and a 50% rate of lineup rejection. These results indicated that composite exposure (for both construction group and yoked-composite group) was increased witness conservativeness. Unlike the composite construction group, the yoked-composite group obtained 82% accuracy when forced to choose. The yoked-viewers were more likely than the constructors to identify the target when conservativeness was removed. Both groups were similar in their likeliness to reject the lineup (58% and 50%, respectively). However, when making a forced choice, the yoked-composite group obtained 2.5 times higher accuracy rates than the construction group

(82% and 30%, respectively). This could be attributed to composite construction, as opposed to composite exposure. Composite construction resulted in increased internalisation of the face. Participants who viewed the composite perceived the face holistically, as opposed to the feature-by-feature construction.

Two papers have attempted to replicate the Wells et al. (2005) results (Dumbell, 2008; Maskow et al., 2007). Maskow et al. conducted their Experiment 1 closely to the Wells et al. study, yet obtained dramatic ceiling effects in which almost all the participants selected the correct target. In Experiment 2, in an attempt to moderate these ceiling effects, the exposure time was decreased to 16 seconds, yet ceiling effects were still obtained. Maskow et al. employed a Guided Memory Interview in one of the conditions prior to recognition. They hypothesised that if negative results comparable to the Wells et al. study were obtained, the Guided Memory Interview would eliminate these effects. The inclusion of this condition in the experiment required that each participant be tested alone in the laboratory. This differed from the Wells et al. study where participants were tested together. Experimenter effects may have given each subject a feeling of importance, which resulted in increased performance.

Dumbell (2008) decreased the exposure time to two seconds, yet found no significant difference in accuracies between conditions. Only 76 participants were obtained, which did not offer much statistical power. Wells et al. (2005) tested over 300 participants and possibly had a more homogenous lineup than did Maskow et al. (2007) and Dumbell. This could have made the Wells et al. lineup more difficult, which would have resulted in decreased accuracies.

## Accounting for Composite Interference

### *Quality of Composites*

It has been found that these composites tend to poorly resemble the target face (Ellis, Davies, & Shepherd, 1978). Comish (1987) postulated that viewing a composite competes with the original visual memory. The new image formed could either merge or blend to form a combination of both faces, or the composite face could override the original image completely (Wells et al., 2005).

A study found that participants who constructed a composite from memory performed more poorly than groups who constructed a composite with the target face present (Ellis, Shepherd, & Davies, 1975).

Another study found that witnesses who were shown composites created using Mac-A-Mug, were unable to match them to photographs (Kovera, Penrod, Pappas, & Thill, 1997). The same study also found that composites created from more contemporary systems tend to poorly reflect the target faces they are intended to mirror, regardless of the participants' familiarity with the target. This eliminated the possibility that poor construction results from the 'newness' of the target face not being properly encoded. These inadequate composite-reflections increase the number of misidentifications of innocent individuals.

*Feature-based encoding*

The problem may lie in the feature-integrated construction of composites. We perceive faces in a configural, holistic way. Inferior recognition could result from having to break down a holistic face into independent features, which is not a natural, automatic activity. Thus, when working manually on an Identi-Kit, for example, one does not have to work as hard to recreate features. This results in less internalisation of the image. This applies to the hundreds of features on Mac-A-Mug and FACES, which could account for greater disruption of recognition.

Research that implemented trait-encoding as opposed to feature-processing, found more accurate recognition performance (Wells & Hryciw, 1984). It appeared that trait-encoding facilitated holistic processing of faces. This study, which utilised the Identi-Kit, showed superior recognition for feature-encoding when feature-based recognition was implemented. This suggested that if a target was encoded featurally, then accuracy would be higher if the participant used featural systems for recognition. Wogalter et al. (1986, p. 12) state:

> Apparently procedures which result in more accurate representations, sketch artists and Mac-A-Mug, lead to less accurate subsequent recognition. Other procedures which require relatively little detail, Identi-Kit and FIS, may result in lower quality images while improving later recognition. Law enforcement agencies should be sensitive to this trade-off and use procedures appropriate to the circumstances of the case.

## Rationale for Research

One should expect this forensic procedure of believable eyewitnesses to be valid. However, research and the number of exonerations suggests otherwise. Central to this may be the construction of composites which research has shown to have both a facilitating and a detrimental effect on recognition performance.

Future studies should therefore attempt to replicate the dramatic results obtained by Wells et al. (2005) which have enjoyed much media coverage, particularly due to the significant, influential findings that other studies have been unable to obtain (Munger, 2006; Roth, 2007) and the implications bearing on these results. If law systems are engaging in detrimental eyewitness techniques, this needs to be rectified before more innocent people are incarcerated. It is essential that these findings be replicated. If similar results are obtained, then this may be an important finding with regards to eyewitness application. The police would need to re-examine methods used for obtaining suspect descriptions without adversely impacting upon later recognition. If, however, the Wells et al. results are not replicated, a non-significant result is also useful in strengthening findings around composite production. It could be postulated that composite-building is useful in identifying target suspects and should continue to be used. Thus, with the ambivalent literature surrounding this topic, and with wrongful incarcerations, research is drastically needed in this field, and replication is therefore indispensable.

## Specific Aims and Hypotheses

The research surrounding composite production is equivocal. A meta-analysis showed only a slight facilitating effect (Meissner & Brigham, 2001b). However, this result has recently been overshadowed by the dramatic results obtained by Wells et al. (2005). It seems unusual that only one study has managed to achieve such indisputable findings.

This research aims to re-investigate this experiment by replication. Two previous attempts were unsuccessful (Dumbell, 2008; Maskow et al., 2007), and possible reasons for this have been discussed. This research will therefore recreate the conditions outlined in Wells et al. (2005), however, FACES 4.0 will be used, as it is a more detailed model of its predecessor and will allow the participants to construct a composite with detailed features.

This study hypothesises that composite production will have a negative effect on identification accuracy (Experiment 1). It will also investigate the forced-choice decisions of participants once lineup-rejection responses have been removed. The construction groups should continue to perform worse in accuracy. Bias will be calculated and examined in order to ensure target and lineup fairness.

Experiment 2 will investigate why the constructors perform worse in recognition. This will be examined through selection of lineup members that are a percentage 'blend' of the target and composite. The construction groups should select a face that is a blend of the original target and the interfering composite.

Wells et al. (2005) used a predominantly White sample, which does not reflect the demographics at the University of Cape Town (UCT). Recognition for same-race suspects is more accurately remembered than for other-race targets (Meissner & Brigham, 2001a). Wells et al. did not account for that in his experiment, and possible effects of race will be investigated.

Lastly, quality of composites will be rated in a post-hoc study. This will investigate whether composites resemble the target from which they are created. It is hypothesised that better-quality composites will do worse in accuracy than the poorer composites, as people are likely to confuse two similar-looking faces, as found in Wogalter (1986).

## Experiment 1

### *Method*

*Design.* A randomised, 2 x 3 design included the independent variable 'Condition' (control, yoked-composite view, construct), with 'Correct' (correct, incorrect) as the dependent variable. The participants were randomised according to race (White, Other), group (control group, yoked-composite group, and construction group), target photograph (1 to 6), and target position (position 2 or 5). Participants were placed into these groups by random assignment, with at least one participant in each cell, using the randomisation function on MS Excel 2003.

Five computers were used in each session, and testing took place at UCT's Psychology Department, in a quiet, comfortable computer laboratory.

*Participants.* This study obtained 124 undergraduate students of 18 years or older from UCT. They signed up through the Student Research Participation Program as part of a course requirement. The sign-up sheet had the title *"Fun with Faces"* as a misleading title. This ensured that participants were not primed to remember faces. It was clearly stated that participants would need to return two days later for the follow-up session. They were not harmed or distressed, and were debriefed at the end of their second session (see Appendix K for ethical considerations).

### *Materials*

*Lineup construction and photographs.* A simultaneous lineup consisting of six colour photographs of White, female faces was constructed on computer with the target face present. There was one target face and five filler photographs in one lineup. Another lineup formed the Target Absent group and contained six filler photographs.

A further five lineups (for each Target Present and Target Absent) were constructed in this way. Six targets were used to counter any effects that may result from uniqueness in one target face (Wells et al., 2005). Six targets and 30 fillers were used in total for the Target Present groups. However, some of the fillers appropriately fit more than one lineup and were used multiple times. As only one lineup was shown to each participant, this did not confound the results. The Target Absent group contained 36 filler photographs.

The target faces in the encoding phase were standardised to 17.61 cm x in height and 15.14 cm in width. They were adjusted for brightness and colour differences. The lineup photographs for recognition were standardised to 8.59 cm in height and 6.87 cm in width. Each photograph was “embedded in a different, solid, colour background in order to distract participants from differences in colour, brightness and saturation across the images” (Maskow et al., 2007, p. 14) and then edited to grayscale.

The target and fillers did not have any distinguishing features which would have made them easily identifiable. The target face appeared in frontal position in the encoding phase. The lineup faces appeared in three-quarter pose during the recognition phase. Research has shown that target photographs presented at encoding phase that are identical at recognition phase results in ceiling effects (Wells et al., 2005).

The target face appeared in one of two different placement positions amongst the 12 different lineups. This ensured that the particular placement of a photograph did not inadvertently affect participant response. A target appeared either in position 2 or 5, and participants were randomly assigned to these conditions.

A selection of 371 photographs of White, female students from several university campuses was used. These individuals had given their permission for these photographs to be used for research purposes. They were from a cohort which the present undergraduates did not recognise. The filler photographs were similar to the target in age, race, sex, and appearance. The appropriateness of filler photographs was judged by a pre-study in which 10 independent observers wrote a description of each of the targets. If a feature appeared more than five times in all the descriptions, it was used as a predictor to select filler faces.

Three hundred and sixty pre-experiment lineups were administered around UCT. This mock witness technique was used to assess the fairness of the lineup. This standard practice involved individuals selecting the target face from the lineup based only on the description (Tredoux, 1999; Wells et al., 2005; Maskow et al., 2007; Oswald & Coleman, 2007). If the lineup was fair, and the fillers adequately resembled the target face, the individuals should not have chosen the target above chance levels (16.67%). Faces that were chosen most frequently in their specific lineups were removed and used as misleading suspects in the Target Absent lineups.

*Presentation materials.* All instructions and photographs appeared on a MS PowerPoint presentation. There was sufficient time between slides for the participant to read the instructions. The target photograph was shown for five seconds, and a prior instruction warned the participant to pay attention to the upcoming face.

*Answering materials*. Participants recorded their demographic information on printed sheets. Answer sheets had the letters of the photograph positions (a – f), which the participant circled if the target was present. If the target was not present, the participant circled an option “N”. A further instruction asked the participant to choose a face if they were forced to choose. The Target Present lineup was always the first lineup.

Participants who constructed a face indicated the percentage they thought their composite resembled the target face, from 0% (not at all) to 100% (extremely). They also rated their confidence regarding their accuracy response from 0% (not confident) to 100% (extremely confident).

*Composite construction software.* Participants in the construction condition were instructed in constructing a face using FACES 4.0: The Ultimate Composite Picture (Cote, 2005). FACES 4.0 contains more detailed features than the other composite systems previously discussed. This detailed composite software includes “440 hairstyles, 221 head shapes, 831 sets of eyebrows, 934 pairs of eyes, 1154 noses, 915 lips, 927 jaws, 855 beards, 122 chins, 63 hats, 161 glasses, 106 moles, 3 scars, 6 piercings, 8 earrings and 9 tattoos” (Maskow et al., 2007, p. 14). Users begin by selecting a face shape and then select various facial features in sequence. Features can be easily selected from an interface on the right by clicking on the feature to be added. The size, position, and colour can be adjusted according to the participants’ discretion.

## Procedure

### *Pilot session*

A pilot study was conducted to determine appropriate target-exposure time, to prevent ceiling effects. Ten participants were obtained and were tested as a control condition. Each participant was exposed to one of the randomly assigned target photographs, for five seconds. After participants had viewed the face, completed a rating task, and played a distracter game, they were required to select the target out of a lineup. Results indicated that 60% of the pilot participants chose correctly. The exposure time was left at five seconds. If the accuracy was similar amongst controls in the real experiment, then construction could either facilitate or contaminate accuracy, without resulting in ceiling effects.

### *First session*

*Encoding phase, all groups.* Participants were seated in the computer laboratory according to the previously allocated randomisation of group condition. Each workstation contained a computer, printed instruction and answer sheet, and consent form. Participants were instructed to fill in their demographic information, read the confidentiality request (see Appendix A), and sign the consent form, giving their permission to take part in the study.

A PowerPoint presentation was started, which forewarned the participants that they would be shown a face. The participants were exposed to the target for five seconds, which was pre-timed on the slideshow. After they had viewed the target, all participants were instructed to complete the same trait-encoding task (Appendix B) used by Wells et al. (2005) to ensure deliberate encoding. All participants were then instructed to play a distracter game, obtained from the supervisor, for 10 minutes. Participants in the control group were then thanked, dismissed from the study, and reminded of their session in two days' time.

*Composite production phase, construction group only.* The construction group was instructed on the use of FACES 4.0 (Cote, 2005) prior to their PowerPoint presentation, and was given 20 minutes to create a practice face. They then viewed the slideshow, completed the trait-encoding task, and played the distracter game, and constructed the target face. No participant took longer than 20 minutes. They then answered two questions asking about the participant's perceived accuracy of the construction (see Appendix C). Once they had completed their task, the picture was saved, the participants thanked, dismissed from the study, and reminded of their session in two days' time.

*Viewing composite phase, yoked-composite group only.* At the end of their encoding phase, the yoked-composite group was shown a composite created on FACES. They were told that another group had attempted to reconstruct the target face. Once they had viewed the face, the participants were thanked, dismissed from the study, and reminded of their session in two days' time.

*Second session*

*Recognition phase, all groups*. The participants returned two days later and were seated at the same computers with their same answer sheet. Instructions on the slideshow indicated they would be required to select the target face from a lineup of six faces (see Appendix D for one example). A following instruction emphasised that the target face may or may not be present. The simultaneous lineup was then displayed with all six faces in three-quarter view. No time restrictions were given during identification, and the participants indicated their choice on the sheet provided (see Appendix E).

Once all the participants had answered, they were then debriefed and told the purpose of the study.

## Results

Participants' responses were recorded as follows:

- *Correct.* Participants accurately selected the target face in the Target Present condition.
- *Incorrect.* Participants incorrectly selected a filler photograph in the Target Present condition.
- *Incorrect rejection.* Participants in the Target Present condition incorrectly rejected the lineup.
- *Correct rejection.* Participants in the Target Absent condition correctly rejected the lineup.

Data analysis was conducted using the software program STATISTICA 8 (StatSoft, Inc., 2008). Although multiple comparisons testing increases Type I error, the alpha level was left at $p < 0.05$, as exact $p$-values between proportions were computed.

### *Descriptive Statistics*

Table 1 depicts participants' correct responses for the lineups.

Table 1

*Correct Choice of Target*

| Condition | Correct Choice | | Total |
|---|---|---|---|
| | White | Other-Race | |
| | White | Other-Race | |
| Control | 46.51% (20) | 53.49% (23) | 65.12% (28) |
| View | 39.47% (15) | 60.53% (23) | 36.84% (19) |
| Construct | 44.19% (19) | 55.81% (24) | 44.19% (61) |
| Total | 55.56% (30) | 44.29% (31) | |

Note. Reported results in percentages (*n)*

Initial inspection of results indicated the controls performed as the pilot session predicted, obtaining an accuracy of 65.12%. The construction condition performed worse than the controls, with an accuracy of 44.19%. Interestingly, the view group had the most incorrect decisions. The differences between frequencies in correct and incorrect decisions are displayed graphically in *Figure 1*.

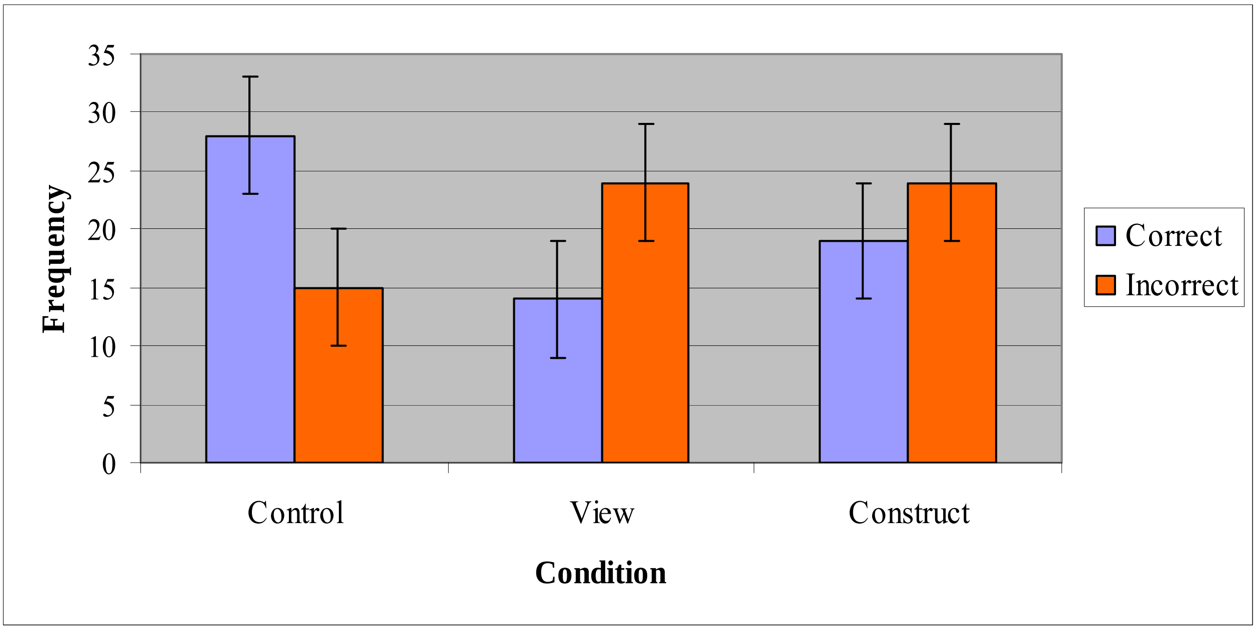



*Figure 1.* Comparison of correct and incorrect responses in Target Present lineups.

*Contingency Testing*

An examination of the percentages supported the main hypothesis that composite production harms face recognition. A contingency test on unforced accuracy (3 x 2) was conducted and supported this observation, showing a significant difference in the results obtained across the groups, $\chi^2$ (2, $N$ = 124) = 7.11, $p$ = .029. Error bars indicated significance between: the control and construct conditions; and the control and view conditions, but no significance between the view and construct conditions. Cramer's V = 0.24 indicated a moderate relationship as a measure of effect size. There was no significant difference between the accuracies between Race, $p$ = 0.107.

The $p$-values were calculated using difference tests between proportions. The results are displayed in Table 2.

Table 2

*Significance amongst Conditions (N = 124)*

| | Control | View | Construct |
|---|---|---|---|
| Control | | 0.015* | 0.026* |
| View | | | 0.251 |
| Construct | | | |

*Significant at $p$ < 0.050

*Forced Decisions*

Of the 124 participants, 63 got the answer incorrect. This result was made up of 12 filler selections, and 51 "N" selections. It could be possible that the composite made the participants more likely to reject the lineup. To eliminate this conservativeness, the participants were asked to indicate a choice if they were forced to choose. A breakdown of the 51 forced-choice "N" selections is displayed in Table 3.

Table 3

*Forced-Choice Responses for "N" Selectors Only (N = 51)*

| | Forced-Choice | |
|---|---|---|
| | Correct | Incorrect |
| Control | 40% (4) | 60% (6) |
| View | 45% (9) | 55% (11) |
| Construct | 61.90% (13) | 38.10% (8) |
| Total | 100% (51) | |

Note. Reported results in percentages (*n*)

Participants in the composite construction group had facilitated recognition responses (61.90%) when compared to the controls (40%). The performance seen in Table 1 and *Figure 1* could be due to increased conservativeness amongst the construction group. Although these figures did not reach a level of significance, $p = 0.126$, if we take chance level to be 16.67% (1 target out of 6 possible selections), the forced-choice construction group chose approximately 3.7 times greater than chance. The forced-choice responses with the absence of "N" are displayed in Table 4.

Table 4

*Forced-Choice Responses for All Participants (N = 124)*

| | Forced-Choice | | |
|---|---|---|---|
| | Correct | Incorrect | *N* |
| Control | 74.42% (32) | 25.58% (11) | 43 |
| View | 60.53%% (23) | 39.47% (15) | 38 |
| Construct | 74.42% (32) | 25.58% (11) | 43 |
| Total | 87 | 37 | 124 |

Note. Reported results in percentages (*n*)

When the forced-choice responses are analysed, the construct group performed the same as the controls (74.42% for both). The view condition performed slightly worse (60.53%). When conservativeness was removed, construction appeared to have no effect on facilitating or contaminating recognition when compared to controls, $p = 0.500$.

*Target Absent Lineups*

Target Absent lineups were shown to the participant after the Target Present selection. Each Target Present lineup had a corresponding Target Absent lineup with similar-looking fillers. The correct answer was “N”. All three conditions obtained near ceiling effects, as seen in Table 5. Composite construction did not hamper recognition performance compared to controls, $p = 0.199$, and obtained 90.70% accuracy.

Table 5

*Target Absent Reponses (N = 124)*

| Condition | Target Absent | | *N* |
|---|---|---|---|
| | Correct ("N") | Incorrect | |
| Control | 95.35% (41) | 4.65% (2) | 43 |
| View | 86.84% (33) | 13.16% (5) | 38 |
| Construct | 90.70% (39) | 9.30% (4) | 43 |
| Total | 91.13% (113) | 8.87% (11) | 124 |

Note. Reported results in percentages (*n)*

*Lineup Bias*

A problem with the Dumbell (2008) study was that it only contained one lineup and did not account for lineup bias due to unseen peculiarities, such as a specific feature on a target face. For this research, 12 lineups were used (six Target Present lineups plus six Target Absent lineups). In order to calculate lineup bias and assess the fairness of a lineup, 205 mock witnesses selected the target based on a description (see Appendix F for an example). However, as 124 of these participants had participated in Experiment 1, each participant was given only 10 lineups. They were not shown the two lineups they would have seen during their previous identification stage.

The frequencies for each target and filler photograph were totaled and the effective size “*E*” for each lineup was calculated. This value indicated the fully appropriate number of members of the lineup. Lineup bias was also calculated from the frequency of correct responses over the total lineup size. Bias and effective size are displayed in Table 6.

Table 6

*Proportions of Bias and Effectiveness in Each Lineup*

| Lineup | *N* | Target | Bias | E | Confidence Intervals |
|---|---|---|---|---|---|
| 1 | 184 | Present | 0.00 | 2.84 | 2.54-3.23 |
| 2 | 184 | Absent | 0.09 | 3.04 | 2.65-3.56 |
| 3 | 184 | Present | 0.62 | 2.30 | 1.99-2.74 |
| 4 | 184 | Absent | 0.01 | 1.25 | 1.14-1.39 |
| 5 | 185 | Present | 0.14 | 1.46 | 1.31-1.65 |
| 6 | 184 | Absent | 0.15 | 4.22 | 3.85-4.65 |
| 7 | 182 | Present | 0.40 | 2.35 | 2.15-2.58 |
| 8 | 183 | Absent | 0.10 | 2.48 | 2.12-2.98 |
| 9 | 182 | Present | 0.05 | 3.22 | 2.76-3.87 |
| 10 | 182 | Absent | 0.19 | 2.83 | 2.54-3.18 |
| 11 | 185 | Present | 0.05 | 2.02 | 1.73-2.42 |
| 12 | 185 | Absent | 0.18 | 3.90 | 3.52-4.38 |

Lineup 6, for example, contains just over four appropriate members of the lineup. The target was selected at approximately chance. If the expected proportion for choosing a subject by chance is 0.167, then lineups which have high bias > 0.30 should be removed. This would suggest the target is standing out and may be selected above chance due to some idiosyncratic property. Lineups 3 and 7 had high bias, so the responses for corresponding Target Absent lineups 2 and 4 were deleted and the analyses were re-run without this biasing data.

*Re-analysis of Unbiased Data*

A 3 x 2 contingency test was re-run on the edited data for the Target Present question, and the percentages and frequencies are displayed in Table 7 in Appendix G. The results indicated $\chi^2$ (2, $N = 85$) = 4.75, $p = .093$, Cramer's V = 0.24. The overall test was no longer significant. Differences tests were conducted between the proportions. Results are displayed in Table 8.

Table 8

*Significance amongst Conditions (N = 85)*

| | Control | View | Construct |
|---|---|---|---|
| Control | | 0.015* | 0.156 |
| View | | | 0.218 |
| Construct | | | |

*Significant at $p < 0.050$

The control condition decreased in accuracy from 65.12% to 61.29%, and the construction group increased in accuracy from 44.19% to 48.28%. However, there was no significant difference in recognition accuracy when constructing a composite, when compared to controls, $p = 0.156$. The view condition remained unchanged in significance.

*Race Effects*

Dumbell (2008) suggested that not obtaining the Wells et al. (2005) results was due to 'race effect' which suggests that same-race participants perform better at recognition with same-race targets. Although Wells did not include demographics of his sample, "Other" participants make up 9.6% of the total population at the University of Iowa (http://www.uiowa.edu/admissions/undergrad/diversity/students.htm). The "Other" participants constituted 54.12% of this experiment. When the unbiased, edited data was re-analysed with White participants (closer to the Wells et al. procedure), the results indicated no significant difference amongst any of the conditions (see Table 9).

Table 9

*Statistics for White Participants in Unbiased, Edited Data*

| | Correct % (*n*) | *p* | | |
|---|---|---|---|---|
| | | Control | View | Construct |
| Control | 75% (12) | | 0.096 | 0.056 |
| View | 50% (5) | | | 0.427 |
| Construct | 46.15% (6) | | | |

*Significant at $p < 0.050$

*Quality of Composites*

To investigate the similarity of the composites to the original target, 33 White females signed up for a face-sorting task. They were required to sort the 43 constructed composites into piles which contained the target lineups. Each composite had to be matched to the target face from which it was constructed. The proportion of correct sorting for each composite can be seen in Table 10. The target was present in the six lineups. *Still do analysis for poor/moderate run analysis*

Table 10

*Quality of Composites*

| Composite Quality | Range | *N* |
|---|---|---|
| Poor | 0.03 - 0.17 | 35 |
| Moderate | 0.18 - 0.32 | 8 |
| *M* (*SD)* | 0.199 (0.11) | |

Most of the participants (81.40%) sorted the faces at chance levels or below. The participants were only able to match the composites to the lineups slightly above chance, $M = 0.199$. These composites constructed do not appear to resemble their corresponding targets. A 2 x 2 contingency test was run on the unbiased accuracies between the groups that constructed 'poor' and 'moderate' composites. There was no significant difference in accuracies between the groups, $\chi^2$ (1, $N$ = 29) = 1.01, $p$ = .316, $\varphi$ = 0.186. The proportions for each response can be seen in Table 11.

Table 11

*Percentage Selections between Composite Construction Groups*

| Quality | Correct | Incorrect | Total |
|---|---|---|---|
| Poor | 36.36% (4) | 63.64% (7) | 100% (11) |
| Moderate | 55.56% (10) | 44.44% (8) | 100% (18) |
| Total | 14 | 15 | 29 |

Note. Reported results in percentages *(n)*

The 'moderate' composite quality constructors were able to construct a face that was able to be matched to the target 55.56% of the time. This is over three times greater than chance levels. However, the 'poor' composite constructors had their composites matched just over two times chance levels (36.36%). Although there was no significant difference in the accuracies between groups, the 'moderate' constructors obtained slightly higher recognition accuracies. The 'moderate' constructors still had low matching rates, with correct matches made slightly more than chance. This is further evidence that composites to not resemble the faces they are intended to mirror.

## Discussion

The initial results of this experiment suggest that construction does have a contaminating effect on recognition accuracy. However, although the construction condition only obtained 44.19% accuracy, it still offered 2.65 times greater chance of selecting the target accurately, as opposed to random guessing.

Interestingly, when the "N" option was investigated alone, the constructors obtained greater accuracy (61.90%) when compared to the controls. Although the comparison was not significant, it appeared that construction increased conservativeness which resulted in the participant rejecting the lineup completely. When the responses of all the participants were re-analysed under forced-choice conditions, the control and composite conditions performed the same.

Mock witness techniques were used to determine the bias of the lineups, and two lineups were excluded, and the data re-analysed. The exclusion of two targets in the data resulted in a non-significant difference between groups, suggesting that composite construction did not hamper recognition accuracy. Although a rough pre-study mock witness technique ($N$ = 360) was implemented to find appropriate foils, a more thorough analysis should have been conducted on that data to dissolve any problematic, biasing lineups.

Effective size of each lineup was calculated and the average *E* = 2.66, which suggested that each lineup contained less than three suitable members. For future experiments, the effective size should be worked out, and those unsuitable lineups adjusted.

An interesting theme throughout the results was the poor results obtained by the view condition. If, as Wells et al. (2005) had proposed, constructing a face with different features is an unnatural act that should interfere with our ingrained holistic processing, then the construction group should have performed worse than the view group. However, the view group in this experiment achieved the lowest accuracies throughout all comparisons. This could have been due to the lack of personal investment in the composite face. The construction group had more time to invest in the face when they pieced features together. The view group merely obtained an impression of the constructor's impression of the target face and may have had "no particular reason to believe that it was a serious attempt" (Wells et al., 2005, p. 151).

Thus, perhaps it is not the construction of the composite itself that has a contaminating effect, as Wells et al. (2005) proposed. If we take into account the generally low accuracies of the view and construct groups, there appears to be a hampering effect with exposure to a composite. The post-hoc composite-quality experiment showed that the composites barely resemble the target, with a match only being made at chance levels. However, since the 'moderate' composite constructors obtained slightly higher accuracies than the 'poor' constructors, if the sample size was bigger, there may be a greater difference between the proportions of these two composite groups. This could be evidence that composite quality is directly proportional to recognition accuracy, and not inversely proportional, as previously hypothesised.

## Experiment 2

### *Method*

*Design.* A randomised, 2 x 2 design included the independent variable 'Condition' (morph-view, morph-construct), with 'Correct' (correct, incorrect) as the dependent variable. The participants were randomised according to the same variables in Experiment 1. They were randomly assigned into these groups, with at least one participant in each cell. A control group was not implemented as the controls in Experiment 1 differed significantly from both contaminated groups. This experiment aimed to investigate if constructors incorrectly select graded-blends of target faces.

*Participants.* This study obtained 81 undergraduate students of 18 years or older from UCT. The same sign-up sheet, requirements, and venue were used as for Experiment 1. The participants were not harmed or distressed, and were debriefed at the end of their second session (see Appendix K for ethical considerations).

*Materials*

*Lineup construction and photographs.* Six composite targets were created from the six, White, female targets used in Experiment 1 (see Appendix H for an example). They closely resembled the six targets and therefore ensured the same variety and difference amongst targets that Experiment 1 displayed. The lineup was sequential, customised to each participant, and consisted of six black-and-white morphed composites of White, female faces constructed using FACES.

After the participant had constructed the composite target, and been dismissed, an individual sequential lineup was created by the researcher. This was done by using the features that the participant had selected in FACES, and adding them to the original target in a stepwise progression. Thus, the first face in the six-picture lineup would be the original target, and the last picture would be the created FACES construction. The second face would be the original target, with the participant's selected eyebrows, facial lines, and face shape. The third face would then be the second face with participant's selected nose and lips added; fourth face contained added hairstyle; and the fifth face had the composite's eyes (see Appendix I for an example of one lineup). Thus, the original target was gradually morphed by incorporating facial features from the participant's constructed composite.

During the encoding phase, the target faces for encoding and identification were standardised to 17.61 cm x in height and 15.14 cm in width, and were adjusted for brightness and colour differences. The lineup photographs were shown in sequence, i.e., one after the other and shown in frontal view, as that was the only position that FACES offers.

The target face appeared in two different placement positions throughout the six different lineups. This ensured that the particular placement of a photograph did not inadvertently affect participant response. A target appeared either in position 2 or 5, and was randomised according to blocking.

*Presentation materials.* All instructions and photographs appeared on a MS PowerPoint presentation. There was sufficient time between slides for the participant to read the instructions. The target photograph was shown for five seconds, and a prior instruction warned the participant to pay attention to the upcoming face.

*Answering materials*. Participants used the same answering materials as Experiment 1. However, as the lineups were manually constructed by feature-incorporation, these groups did not have a Target Absent lineup. An "N" option was recorded as incorrect, as the target was always present. An instruction still asked the participant to select a face if they were forced to choose.

*Composite construction software.* Participants in the construction condition were instructed in constructing a face using FACES 4.0: The Ultimate Composite Picture (Cote, 2005).

Procedure

*First session*

*Encoding phase, all groups.* Participants were seated in the computer laboratory according to the previously allocated randomisation of group condition. They followed the same procedure as the view and construct conditions in Experiment 1, but differed on target shown.

*Morph-composite production group only.* The construction group was instructed on the use of FACES 4.0 (Cote, 2005) prior to their PowerPoint presentation, and was given 20 minutes to create a practice face, in order to explore the features. They then viewed the slideshow, completed the trait-encoding task, played the distracter game, and constructed the target face. No participant took longer than 20 minutes. They then answered two questions asking about the participant's perceived accuracy of the construction (see Appendix C). Once they had completed their task, the picture was saved, the participants thanked, dismissed from the study, and reminded of their session in two days' time.

*Morph-view group only.* At the end of their encoding phase, the morph-view group was shown a composite created on FACES. They were told that another group had attempted to reconstruct the target face. Once they had viewed the face, the participants were thanked, dismissed from the study, and reminded of their session in two days' time.

*Second session*

*Recognition phase, both groups*. The participants returned two days later and were seated at the same computers with their same answer sheet. Instructions on the slideshow indicated to the participants that they would have to choose the target face seen two days earlier, from a lineup of six faces. The morph-view group was shown the lineup that corresponded to the yoked-target face that they viewed two days earlier. The sequential lineup was then displayed with all 6 faces in frontal view, shown one after the other. The participants were given as much time as needed to identify the target and to go through the sequence of faces and indicate their choice on the sheet provided (see Appendix E). A following instruction emphasised that the target face may or *may not* be present.

Once the participants had answered, they were debriefed, told the purpose of the study and were asked if they had any questions pertaining to the experiment.

Results

Participants' responses were recorded in the same way as Experiment 1 (correct and incorrect coding). Data analysis was conducted using the software program STATISTICA 8 (StatSoft, Inc., 2008). Although multiple comparisons testing increases Type I error, the alpha level was left at $p < 0.05$, as exact $p$-values between proportions were computed

*Descriptive Statistics*

Table 12 depicts the frequencies of participants' answers for the unforced lineup.

Table 12

*Unforced-choice Frequencies for All Participants*

| | Unforced Choice | | | |
|---|---|---|---|---|
| | Correct | Incorrect | | |
| Condition | | Filler | "N" (Not Present) | Total |
| Morph-View | 28.95% (11) | 36.84% (14) | 34.21% (13) | 100% (38) |
| Morph-Construct | 23.26% (10) | 39.53% (17) | 37.21% (16) | 100% (43) |
| Total | 25.93% (21) | 38.27% (31) | 35.80% (29) | 100% (81) |

Initial inspection of the results indicates both conditions did poorly, obtaining approximately 1.5 times greater than chance guessing. The constructors performed worse (23.26%) than the viewers. The view group had the most correct decisions, which contradicts the results obtained in Experiment 1. The differences between correct and incorrect decisions are displayed graphically in *Figure 2*.

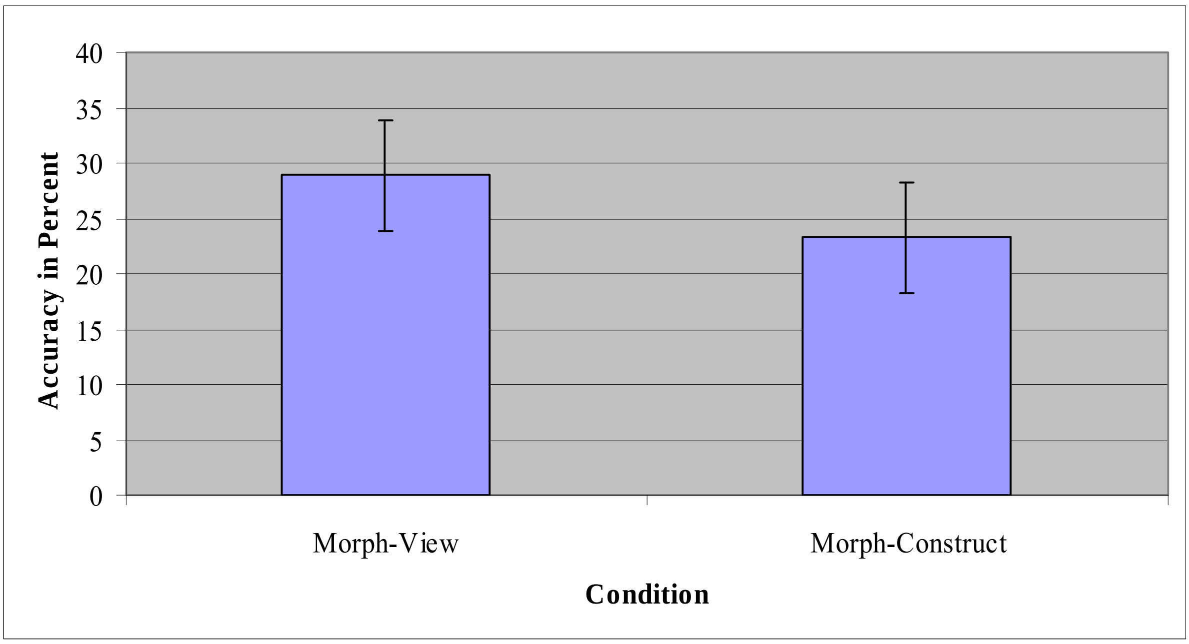



*Figure 2.* Percentage accuracy (target chosen) between conditions.

*Contingency Testing*

A contingency test of the effect of unforced accuracy (2 x 2) was conducted and supported this observation, showing no significant difference, $\chi^2$ (1, $N$ = 81) = 0.34, $p$ = .560, $\varphi$ = -0.065. This indicates no significant difference in accuracy between the two conditions.

*Forced Decision*

Forced decision responses were investigated by removing the "N" option, and analysing the second-choice answer. Depending on the participant's face selection, they were given a score of difference from the original target, ranging from 0% (original target) to 100% (composite), to investigate where participants were incorrectly choosing. The frequencies of forced selections (excluding correct responses) are seen in *Figure 3*.

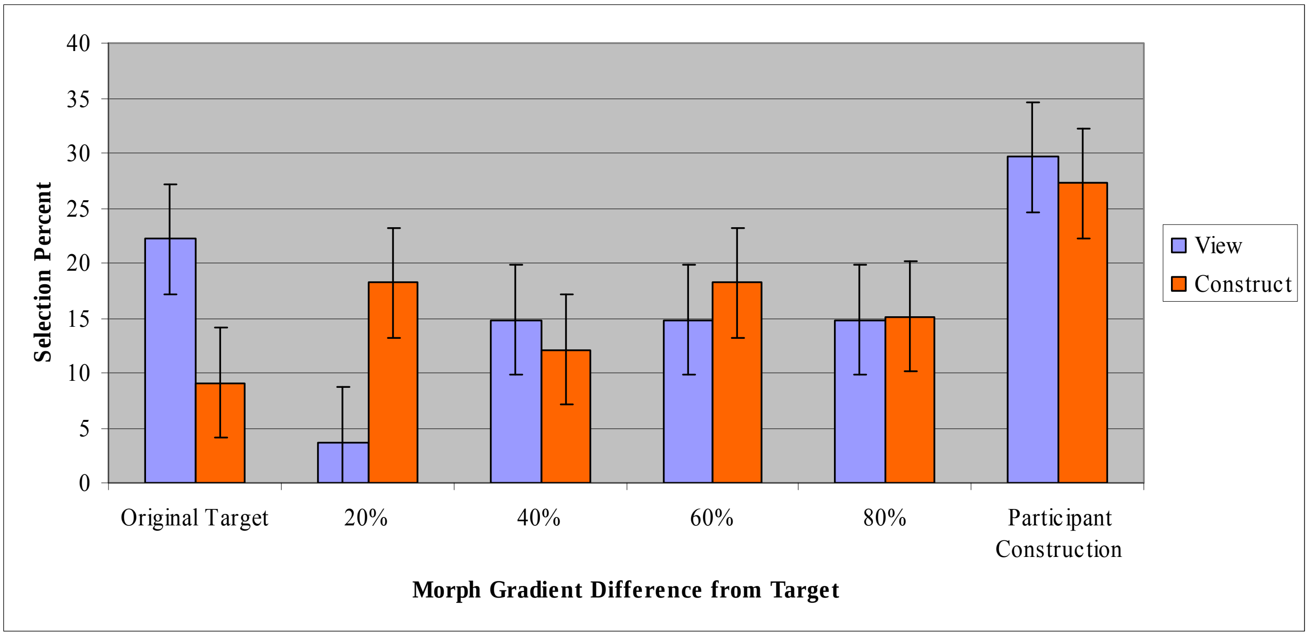



*Figure 3.* Comparison of morphed composite selection between conditions.

*Figure 4* depicts the selection responses of participants who got the incorrect answer and those who got the correct answer, following "N". The view group obtained the highest accuracy in selection the original target (22.22%). However, they also obtained the highest incorrect responses and chose the constructed face (29.63%). The morph-construction group selected the fewest original targets (9.09%) less than chance, and was more likely to select the composite instead of the target, $p < 0.001$. Viewers were just as likely to select the target as selecting the composite, $p = 0.276$.

The view group had the lowest selections at 20%, but the selections increase until the composite. The equal frequency between 40% and 80% could suggest that the morph-view group have contaminated memory for the blend of features, and are unable to distinguish between the stages.

Excluding target and composite, participants selected the morph at 60% difference more frequently, which contained roughly half of the original target features, and half of the composite features.

A one-way ANOVA was run on the morph gradient-position between both conditions. The morph-view group were slightly more likely to choose a face closer to the original target ($M$ = 2.03, $SD$ = 0.33), however there was no significant difference between the morphed stages of faces chosen, $F$ (1, 79) = 0.26, $p$ = 0.612 (see *Figure 4* in Appendix J).

*Race Effects*

There was a significant difference between the accuracies of races, $p$ = 0.025. Although there was no significance between conditions, there was significance difference in accuracy between the races within each condition (see Table 13).

Table 13

*Breakdown of Correct Unforced Choices across Race*

| Condition | Unforced Choice | | |
|---|---|---|---|
| | Other | White | Total |
| Morph View | 81.82% | 18.18% | 100% |
| *n* | 9 | 2 | 11 |
| *p* | 0.002* | | |
| Morph Construct | 70.00% | 30.00% | 100% |
| *n* | 7 | 3 | 10 |
| *p* | 0.037* | | |

*Significant at $p < 0.050$

The Other group performed significantly better in the morph-view condition than the White group, $p$ = 0.002. The Other group also performed significantly better in the morph-construct condition than the White group, $p$ = 0.037. This result suggests that other-race participants are better at identifying faces in a lineup after being exposed to a constructed composite. *Figure 5* displays the percentages of morph selections between White and Other participants.

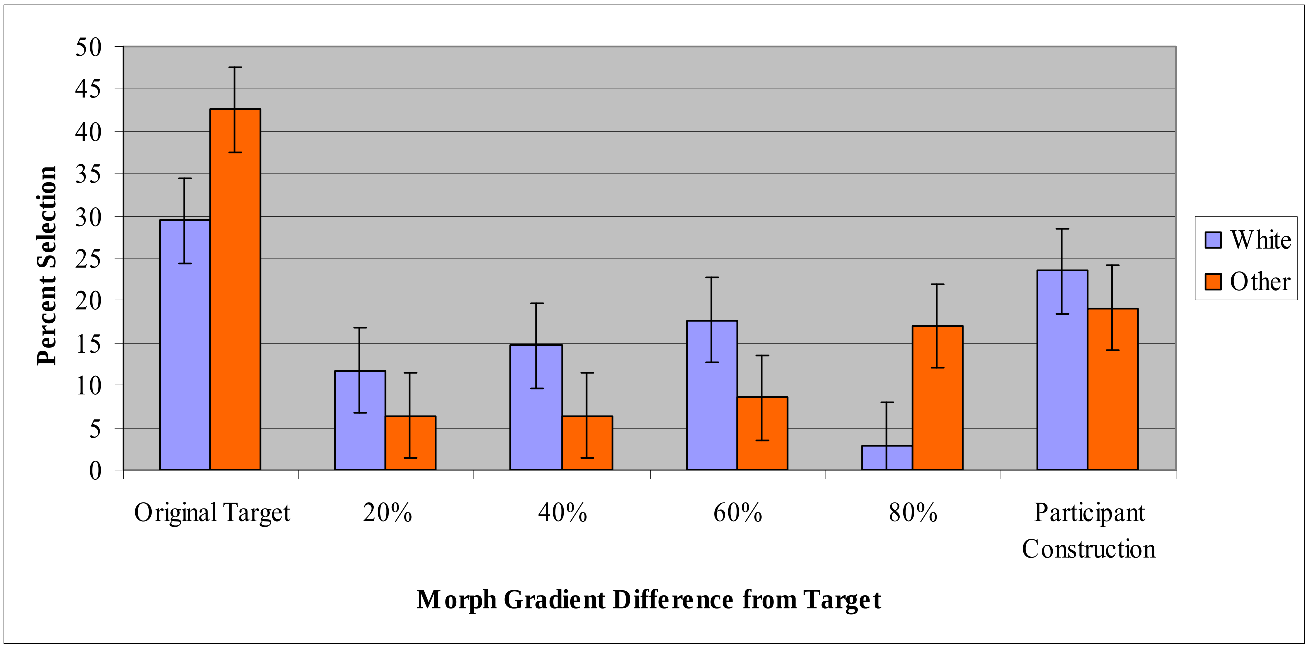



*Figure 5.* Comparison of composite morph selections between White and Other participants.

The Other group selected the target significantly better than the White group, $p = 0.041$. They more frequently selected the target than the incorrect composite, $p < 0.001$. They selected the blends (20% to 60%) at relatively low frequencies. The White group selected the target only slightly more than the composite, although this result was not significant, $p = 0.198$.

## Discussion

This experiment investigated the selection of participants' responses to determine why composite construction causes interference. Wells et al. (2005) proposed that either the original memory and/or the composite remained, or competed, in memory, or that a new memory blend was formed.

There was no significant difference in accuracy between the morph-view and the morph-construct group. Both groups only selected the composite at less than two times chance. This is only slightly better than guessing. Thus, this is consistent with the finding in Experiment 1 which suggested that it was not the action of the construction that contaminated memory, but rather the composite itself.

If we analyse which "N" or a filler face was selected, both conditions tended to select the composite at higher frequencies. The construction group obtained the least accuracy, thus suggesting that construction does contaminate memory to some extent. Both groups were likely to select a blended morphed face at chance.

Members of other-race tend to be better at processing features on a face, whilst same-race members perceive a same-race face holistically (Meissner, 2001). The Other group selected the composite at approximately the same frequency as the White group. The White condition were more likely to select a blend of the two faces, as they still retain a memory trace for both faces, whereas the Other group appear to retain the original memory. Seventy percent of the participants who got the answer correct were Other participants. Thus, the significant difference in accuracy between races is possibly due to the way each encodes, stores, and retrieves the face.

## General Discussion

This research set out to determine whether constructing a composite had a detrimental effect on recognition, and attempted to replicate the Wells et al. (2005) results. The results obtained in Experiment 1 appeared to support the findings. The construction group performed significantly worse than the controls, which suggested that contamination was taking place. The view group performed worse than the construction group, in opposition with Wells et al. This suggested that exposure to the composite may contaminate memory, and not the construction itself. Wells et al. hypothesised that it was the construction that contaminated memory. These results appear to fit the results obtained by Comish (1987), who also speculated that composite exposure hampered recognition performance.

Two lineups (3 and 7) were found to be biased. These lineups were removed and the contingency tests were re-run on the data. The unbiased analysis found no significant difference between the control and construct groups, which is in accordance with other research conducted (Dumbell, 2008; Maskow et al., 2007). Although the difference between the results was not significant, they were still low at 48.28%. Cramer’s V as a measure of effective size remained at 0.24 which suggested that some effect was still taking place. However, this is approximately three times greater than chance and does not come close to the low (10%) result Wells et al. (2005) obtained.

When the forced-choice “N” responses were analysed, 61.90% of the constructors got the answer correct. This supports the Yu and Geiselman (1993) study which suggested that composite production increases eyewitness conservativeness and makes participants more likely to reject the lineup entirely. This could suggest that the construction results in participants becoming more uncertain in their selection. However, as eyewitnesses in forensic settings are allowed to reject a lineup if they are feeling uncertain, the fact that they may know who the target is, but that it is masked by conservativeness, should alert law enforcement to change their practices. Wells et al. (2005) did not analyse their forced-choice responses, even though 58% of the participants in the construction group rejected the lineup. Merely viewing the composite seemed to hamper recognition, which supports the finding by Brown, Deffenbacher, and Sturgill (1977), that composite-viewing is enough to interfere with the memory trace.

Wells et al. (2005) also did not account for race within their study. Even when the biased lineups and other-race participants were removed from this study for an analysis, closely resembling the Wells et al. study, there was no significant difference between any of the conditions, although there the control and construction groups were nearing significance, $p < 0.056$.

A post-hoc study investigated the quality of the composites. The mean quality was found to be 0.19. This indicated that blind-raters were only able to match the composite to the target just above chance. The composites thus did not resemble the targets. Previous research suggested that more accurate representations lead to less accurate subsequent recognition (Wogalter, 1986).

The effective size was calculated to determine the number of sufficient members of the lineup. The mean score was 2.66 which was low, and possibly compromised the validity of the lineup. This indicated that the lineups only contained just over two possible appropriate selections. Thus, the participant was really only facing a two-person lineup. Due to Wells et al.'s (2005) low accuracy, this could possibly be due to the homogeneity of his lineups. This research tried to counter a biased study by utilising several lineups. This could prevent any one target having a noticeable property and causing the suspect to stand out, which would dramatically affect the results. However, participants did not obtain ceiling affects, which would have happened if the lineups were biased with inappropriate foils. The controls performed as expected, and the construction group obtained only two times greater than chance. The lineups could have been biased against the target, in which the fillers were too similar.

This problem may have been encountered in Experiment 2, which could account for the poor results obtained by both conditions. Brigham, Meissner and Wasserman (1999) suggest that too much similarity creates negative bias, resulting in the appearance of "clones". Furthermore, the fillers all look like one another, but the target and the composite look nothing alike, with no similar features. This could possibly be why participants mostly chose at either ends of the gradient-scale, regardless of the mixed order of the sequential lineup. Ebbesen and Flowe (2002) suggest that viewing a sequential lineup results in participants responding more conservatively. This therefore affects the bias of the lineup, not the discriminability. However, the differences between simultaneous and sequential lineups cannot be investigated further within this study as each experimented utilised different types of pictures.

There was no significant difference between the morph-view and morph-construct groups, although both groups performed poorly. It is possible that Experiment 2 consisted of extremely similar faces, making the experiment difficult. This cannot be compared to a control group, as this experiment did not make use of one, due to the individualised lineups that the constructors and yoked-viewers received. One would not be able to ensure the controls all obtaining the same lineup, and would have to have a yoked-control condition, which would be yoked to a yoked-view condition. Furthermore, implementation of a control group would have required too many participants which may have been unattainable. A future study would definitely address this issue and make use of a control group. This could investigate if the lineups were poor due to similar lineup members, or whether exposure to the composite, as opposed to the construction, hampers recognition. The results obtained by the view group could be compared to the control group. Further research would also investigate whether there is a relationship between quality of composite and the percentage blend chosen in the morphed-gradient lineups.

Interestingly, this study showed other-race participants were more likely than same-race participants to obtain the correct answer. This could be attributed to how each race perceives their own, and other-race's, faces. It was suggested that own-race processing involves holistic processing, whereas other-race processing involves featural processing. Thus, since composite systems make use of featural procedures, this explains why the other-race participants performed better. It is not clear, however, if this may possibly be due to other-race participants simply constructing poorer composites, as the morphed faces were not rated as the composites in Experiment 1 were.

## Conclusion

Some limitations pertaining to this research have been discussed, such as homogeneity of lineups, which appear to be no match for Wells et al.'s (2005) lineups which achieved such low accuracies. This study was able to obtain a significant difference result between control, view, and construct conditions which the Maskow et al. (2007) and Dumbell (2008) studies were unable to achieve. Wells et al. did not correct for bias, which could have resulted in the high control group accuracy. Although this result changed when biased lineups were corrected for, the view group remained significantly different when compared to the controls.

It appears as though constructions systems are not wholly useful when it comes to recognition of faces. Firstly, the composites constructed do not appear to resemble their targets. Secondly, composite systems increase eyewitness conservativeness, making them more conservative about selecting a face, even when they do know the correct answer. Thirdly, they appear counterintuitive to how people normally encode and process faces, holistically and configurally.

Although some of the results between groups were not significant, the groups still performed poorly. This could be attributable to the composite systems, which require featural-encoding which is a difficult, unnatural task. Having to reconstruct a face featurally does not allow for internalisation of the image, as people do not have to work as hard to recreate the face.

Does face composite production hamper recognition? If we take the initial analysis, then yes. If the unbiased data is analysed, then it does not contaminate memory sufficiently differently from the control group. However, as the composites constructed are of bad quality, the participants who viewed them performed the worst. We can therefore state that composites, more than their production, appear to hamper eyewitness recognition and testimony.

References

Brigham, J., Meissner, C., & Wasserman, A. (1999). Applied issues in the construction and expert testament of photo lineups. *Applied Cognitive Psychology, 13,* 73-92.

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

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

Comish, S. (1987). Recognition of facial stimuli following an intervening task involving the Identi-Kit. *Journal of Applied Psychology, 72*, 488-491.

Dumbell, E. (2008). *Face composite production effects on witness memory: Case not closed.* Unpublished Honours' thesis, University of Cape Town.

Ebbesen, E., & Flowe, H. (2002). Simultaneous v. sequential lineups: What do we really know? Unpublished manuscript, University of California, San Diego.

Ellis, H., Davies, G., & Shepherd, J. (1978). A critical examination of the Photo-fit system for recalling faces. *Ergonomics, 21*, 297–301.

Ellis, H., Shepherd, J., & Davies, G. (1975). An investigation of the use of the Photo-fit technique for recalling faces. *British Journal of Psychology, 66,* 29-37.

Fallshore, M., & Schooler, J. (1995). Verbal vulnerability of perceptual expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1608–1623.

Hall, D. (1976). Obtaining eyewitness identifications in criminal investigations: Applications of social and experimental psychology (Doctoral dissertation, Ohio State University, 1976). *Dissertation Abstracts International, 37*, 2569. Abstract retrieved 3 March, 2009, from FirstSearch/Dissertation Abstracts International database.

Innocence Project. (n.d.). *Facts on Post-Conviction DNA Exonerations*. Retrieved April 4, 2009, from http://www.innocenceproject.org/Content/351.php

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

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

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

Maskow, N., Schmidt, H., Tredoux, C., & Nunez, D. (2007). *Face composite production does not affect eyewitness identification accuracy.* Unpublished manuscript, University of Cape Town.

Meissner, C. (2001). The influence of perceptual learning in memory for faces: A process-tracing approach to investigating the own-race bias. Unpublished manuscript. Florida State University, Department of Psychology, Florida.

Meissner, C., & Brigham, J. (2001a). Thirty years of investigating the own-race bias in memory for faces – A meta-analytic review. *Psychology, Public Policy, and Law, 7,* 3-35.

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

Munger, D. (2006). When crime-fighting tools go bad: Problems with the face-composite system. *Cognitive Daily,* September 27, 2006. Retrieved on 11 May 2009 from http://scienceblogs.com/cognitivedaily/2006/09/when_crimefighting_tools_go_ba.php

Roth, M. (2007). Why police composites don't always hit mark. *Pittsburgh Post-Gazette*, March 25, 2007. Retrieved on 11 May 2009 from http://www.post-gazette.com/pg/07084/772371-84.stm

Oswald, K., & Coleman, M. (2007). Memory demands on facial composites identification. *Applied Cognitive Psychology, 21,* 345-360.

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

Shapiro, P., & Penrod, S. (1986). A meta-analysis of facial identification studies. *Psychological Bulletin, 100*, 139-156.

Tredoux, C. (1999). Statistical considerations when determining measures of lineup size and lineup bias. *Applied Cognitive Psychology, 13,* 9-26.

University of Iowa. (n.d.). *Student Demographics*. Retrieved October 23, 2009, from http://www.uiowa.edu/admissions/undergrad/diversity/students.htm

Wells, G., Charman, S., & Olson, E. (2005). Building face composites can harm lineup identification performance. *Journal of Experimental Psychology: Applied, 11,* 147-156.

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

Wells, G., Memon, A., & Penrod, S. (2006). Eyewitness evidence – Improving its probative value. *Psychological Science in the Public Interest, 7*, 45-75.

Wogalter, M., Laughery, K., & Thompson, B. (1986). *Composite construction on subsequent recognition performance.* Unpublished manuscript. University of Richmond.

Yu, C., & Geiselman, R. (1993). Effects of constructing Identi-Kit composites on photospread identification performance. *Criminal Justice and Behavior, 20*, 280-292.

# Appendix A

## Participant Consent Form

Thank you for participating in this study! Some demographic information is required which can be filled in the spaces below. Please remember that all information will be kept strictly confidential and only results will be reported in the research project. Your student number should be given to the researcher in order to obtain your SRPP credits, but this information will not be reported in the research.

**Student No.:**

**Age:**

**Sex (please tick):** MALE | FEMALE

**Race (please tick):** WHITE | BLACK | COLOURED | INDIAN | OTHER

For your time and participation, you will receive **90 minutes** (3 units) towards your SRPP requirement.

Please note that at any time during study if you feel uncomfortable or experience any distress, you are free to leave.

You should not experience any mental, physical or emotional distress, but if you do, please notify the researcher who will be more than happy to relieve any uncomfortable feelings.

**THE STUDY:** This study is concerned with the morphing of faces and how people perceive these faces. Your participation is instrumental in explaining how humans perceive and recognise other human faces. Please note that more information will be given to you at the end of the study and you will be debriefed and will be told the expected results and hypotheses.

If you would like more information regarding the study, you can talk to the researcher at any time, on 072 394 8193.

**PLEASE REMEMBER YOUR SECOND SESSION IN TWO DAYS' TIME!**

By signing this form, I hereby give consent to (1) participate in this study and (2) for the responses and results in the study to be used. I acknowledge that I have read through the description above and filled in the required information. I am aware that any personal information will not be distributed.

**SIGNATURE:**___________________________ **DATE:**_______________

Appendix B

Trait-Encoding Sheet Including Distracter Game Score

**ANSWER SHEET ONE – Session One**

**1.** Please rate the face you have just seen: (0 – not at all; 10 – extremely)

**a)** 0 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 - 10

**b)** 0 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 - 10

**c)** 0 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10

**d)** 0 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10

**e)** 0 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10

**f)** 0 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10

**g)** 0 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10

**h)** 0 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10

**i)** 0 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10

**j)** 0 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10

**2.** Please write your hockey score (overall total games won and lost) in the boxes below:

**WON:** ☐ **LOST:** ☐

# Appendix C

## Composite Construction Perceived Accuracy Questions

**3.** To what extent, in your opinion, does the face you created accurately resemble and depict the face you saw earlier?
(0 – not at all; 10 – extremely) *(please circle)*

**0 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10**

**4.** How confident are you with the answer you gave above?
(0 – not at all; 10 – extremely) *(please circle)*

**0 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10**

**PLEASE MAKE SURE YOUR FACE IS SAVED TO THE DESKTOP WITH YOUR STUDENT NUMBER AS THE FILENAME!**

Appendix D

Example of Identification Lineup during Encoding Phase (*Figure 1*) and Idenification Phase (*Figure 2)* (Target 5 in Position 2)

*Figure 1*. Target face during encoding phase (in colour).


*Figure 2*. Lineup during identification phase.

Appendix E

Second Session Answer Sheet

**ANSWER SHEET TWO – Session Two**

**1.** (a) (b) (c) (d) (e) (f) (N)

**2.** (a) (b) (c) (d) (e) (f)

**3.** (a) (b) (c) (d) (e) (f) (N)

**4.** (a) (b) (c) (d) (e) (f)

**Face Rating Task:**

**5.** (a) (b) (c) (d) (e) (f)

**6.** (a) (b) (c) (d) (e) (f)

**7.** (a) (b) (c) (d) (e) (f)

**8.** (a) (b) (c) (d) (e) (f)

**9.** (a) (b) (c) (d) (e) (f)

**10.** (a) (b) (c) (d) (e) (f)

**11.** (a) (b) (c) (d) (e) (f)

**12.** (a) (b) (c) (d) (e) (f)

**13.** (a) (b) (c) (d) (e) (f)

**14.** (a) (b) (c) (d) (e) (f)

**15.** (a) (b) (c) (d) (e) (f)

**16.** (a) (b) (c) (d) (e) (f)

# Appendix F

## An Example of a Mock Witness Lineup Description Task: Lineup 3 (Target Present)

- You will now be required to view several line-ups and choose which person you think best fits the written description above.
- Please circle your selection (a) – (f) on Question 5 to Question 16.
- Press <spacebar> after each line-up to view the next set of photographs.

**<u>Question 7:</u>** White, blonde, female, loose hair, arched eyebrows.

## Appendix G

Frequencies and Percentages for Unbiased and Edited Data

Table 7

*Percentages of Responses for Edited Data (N = 85)*

| | Target Present | |
|---|---|---|
| | Correct | Incorrect |
| Control | 61.29% (19) | 38.71 (12) |
| View | 32% (8) | 68% (17) |
| Construct | 48.28% (14) | 51.72% (15) |
| Total | 100% (85) | |

Note. Reported results in percentages (*n)*

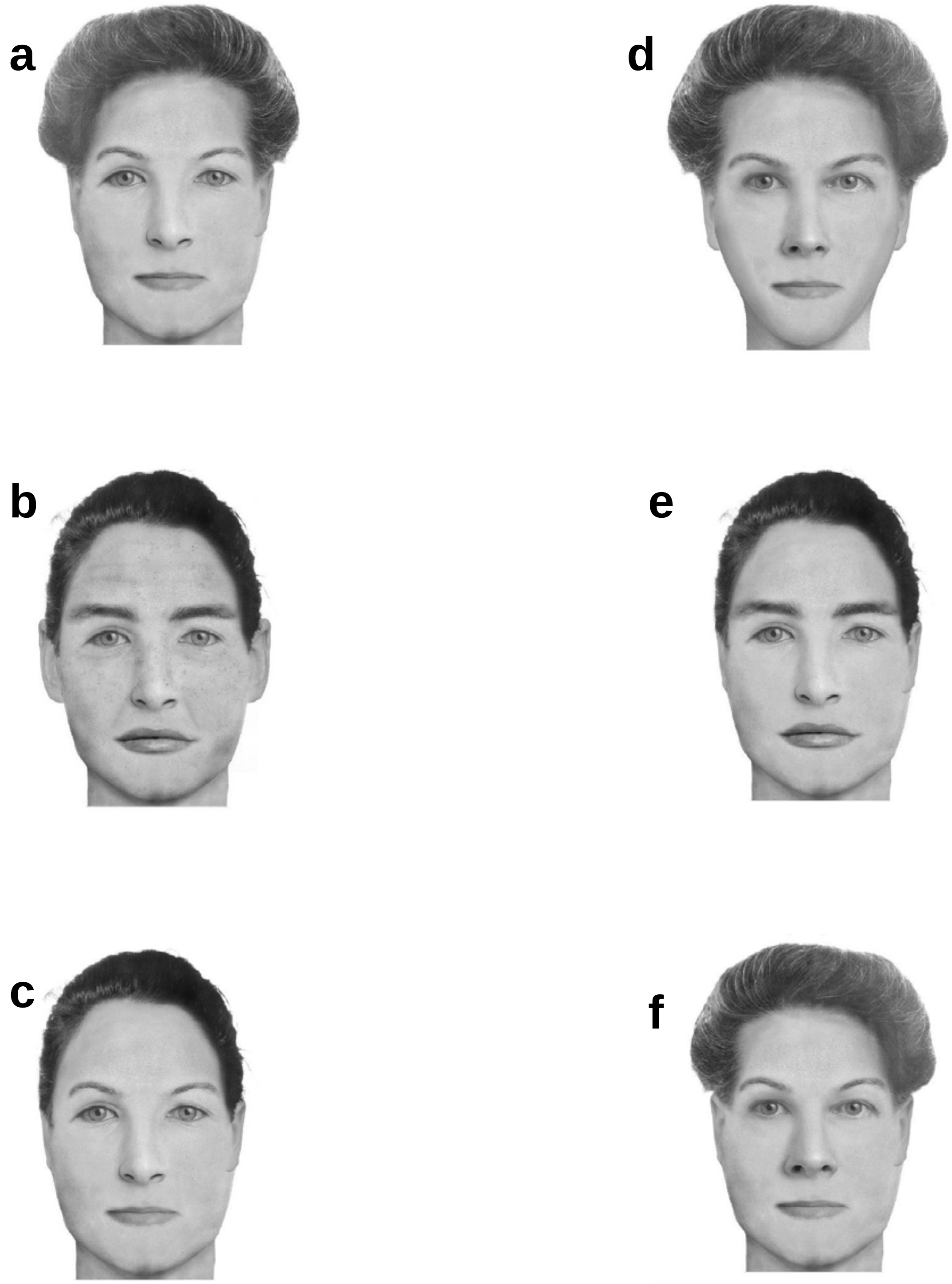
Appendix H

Example of Composite Created in Experiment 2 from Target Face in Experiment 1

# Appendix I

## Researcher-Created Customised Lineup for a Participant

"b" is the correct original target selection, and the participant created face "d". A gradient of incorporation of facial features from the original target results in the participant's constructed face.



Appendix J

Graph Depicting the Means Between the Morph-View and Morph-Construct Groups

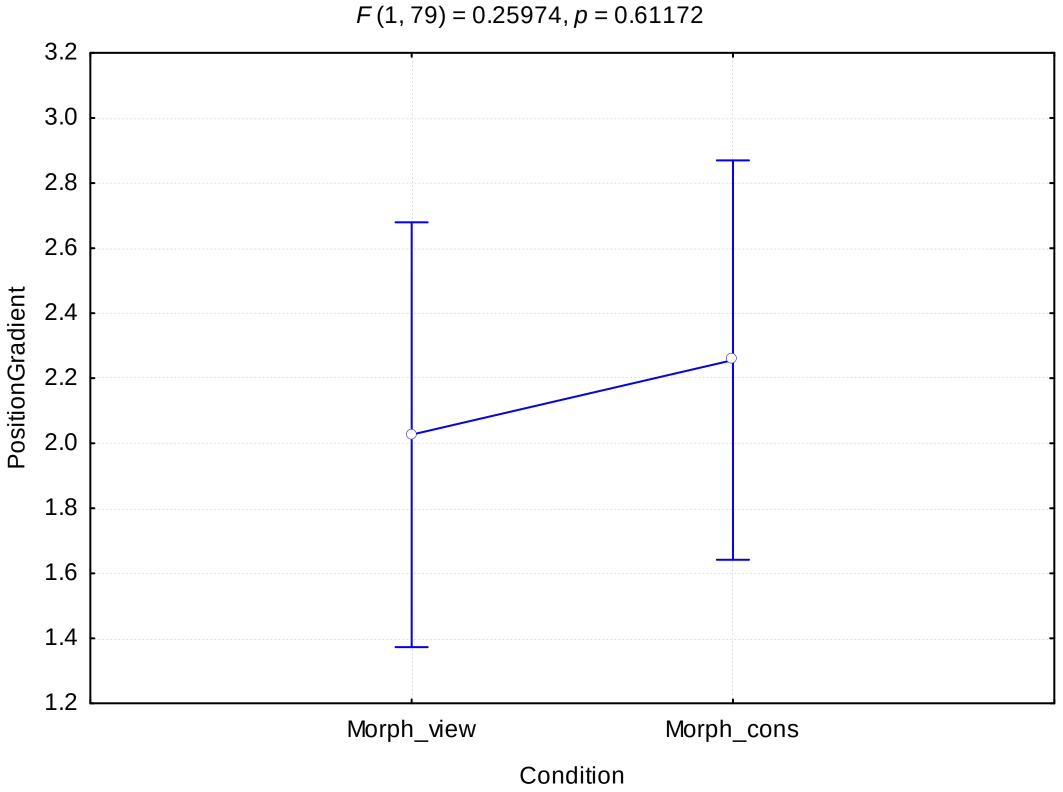



*Figure 4.* Non-significant difference between the morph conditions during identification.

## Appendix K

### Ethical Considerations

This study followed the ethical guidelines as set out by the UCT Codes for Research. Ethical approval was granted by the Research Ethics Committee of the UCT Department of Psychology. Participation was voluntary, and students chose to sign up. Only participants of 18 years or older were allowed to participate in the study as 18 years is the legal age of majority.

The results obtained did not make use of the participants' names, and all the answer sheets remained anonymous, aside from their demographic information. These participants gave their permission to participate in the study by signing a consent form which allowed the use of their responses for this study only. The participants were informed that they were free to leave the study if they wished, and that attendance, although appreciated, was not compulsory.

The participants were debriefed once the experiment was completed and told the purpose of the study. There were no foreseeable emotional or physical risks associated with participating in this study. However, some of the participants may have felt 'duped' by the study as it was not wholly related to the title to which they signed up. This was carefully explained to them and the notion of expectation effects if the study were titled *"Encoding and Recognising Faces"* was emphasised.

**PLAGIARISM DECLARATION**

1. I know that plagiarism is wrong. Plagiarism is using another’s work and to pretend that it is ones own.

2. I have used the American Psychological Association (APA) as the convention for citation and referencing. Each significant contribution to, and quotation in, this essay/report/project/…from the work, or works of other people has been attributed and
has cited and referenced.

3. This essay/report/project… is my own work.

4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

5. I acknowledge that copying someone else's assignment or essay, or part of it, is wrong, and declare that this is my own work

SIGNATURE: __________________________

DATE: _________________