

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Running head: FEATURAL COMPOSITES AND LINEUPS

“What Big Teeth You Have” – Red Riding Hood and the Face Recognition Failure:
The Effects of Isolated Featural and Configural Composite Construction on Recognition
Accuracy

Kate-Victoria Kempen

KMPKAT002

A minor dissertation submitted in partial fulfillment of the requirements for the award of the
degree of Master of Arts in Psychological Research

Faculty of Humanities

University of Cape Town

2012

COMPULSORY DECLARATION

This work has not been previously submitted in whole, or in part, for the award of any degree. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work, or works, of other people has been attributed, and has been cited and referenced.

Signature: _____

Date: _____

Acknowledgements

I would like to extend my gratitude to the following people who helped and supported me throughout my work:

To my supervisor, Professor Colin Tredoux, for the patience, invaluable assistance, wisdom, and repartee that he has given me throughout these years. The investment he puts in all his students makes me strive to be a better researcher, for myself, for him, and for our team.

To my friends, who supported me, put up with my endless talking about my research, kept me sane, and cheered me on.

To my mother, who supported me, brought me endless cups of tea, offered words of wisdom, and brought peace into my life. Thank you for always being my number-one fan.

University of Cape Town

Table of Contents

ACKNOWLEDGEMENTS	2
TABLE OF CONTENTS	3
LIST OF FIGURES	7
LIST OF TABLES	8
ABSTRACT	9
CHAPTER ONE: INTRODUCTION	10
Cognitive Processes in Encoding, Construction, and Recognition	13
Different Streams of Processing	13
Top-down processing	13
Bottom-up processing	14
Evidence from the Visual System	15
Processing Faces Featurally and Configurally	16
The Quality of Composites	20
Composite Production Systems	21
Sketch artists	21
Manual systems	22
Computer software systems	24
Eigenface algorithms	25
The Effect of Composite Production	29
No significant effect	30
Facilitating effect	31
Detrimental effect	32
Theories Accounting for Composite Interference	37
Recoding interference	37

FEATURAL COMPOSITES AND LINEUPS	4
Transfer-inappropriate processing shift	39
Misinformation effect	41
Criterion shift	42
Summary	45
Encoding, Interference, and Recognition Failure	47
Rationale for Research	48
Specific Aims and Hypotheses	51
CHAPTER TWO: EXPERIMENT 1	53
Method	53
Participants	53
Materials	54
Target stimuli	54
Presentation and answering materials	54
Lineup construction, whole-face lineups	54
Lineup construction, feature lineups	56
Procedure	56
Encoding phase	56
Recognition phase	57
Results	59
Contingency Testing	59
Odds Ratios	60
Reaction Times of Overall Responses	61
Reaction Times of Correct and Incorrect Responses	63
Discussion	65

FEATURAL COMPOSITES AND LINEUPS	5
CHAPTER THREE: EXPERIMENT 2A	69
Method	69
Participants	69
Materials	70
Target stimuli	70
Lineup construction	71
Presentation and answering materials	73
Composite construction software	73
Procedure	74
Encoding phase	74
Composite production phase	74
Recognition phase	75
Results	79
Descriptive Statistics	79
Contingency Testing	81
Hits and Misses Across Targets	83
Forced Decision	83
Target Absent Lineups	85
Lineup Bias	86
Reasons for Lineup Rejection	88
Meta-Cognitive Questions	90
Reaction Time	92
Discussion	94
CHAPTER FOUR: EXPERIMENT 2B	102
Method	102

FEATURAL COMPOSITES AND LINEUPS	6
Participants	102
Materials	102
Target stimuli and composites	102
Presentation and answering materials	103
Slideshow and task construction	103
Procedure	103
Results	105
Composite Selections	105
Reaction Times	106
Discussion	108
CHAPTER FIVE: GENERAL DISCUSSION	110
Limitations and Future Research	110
Discussion	113
Conclusion	119
REFERENCES	124
APPENDIX A	134
APPENDIX B	135
APPENDIX C	141
APPENDIX D	142
APPENDIX E	144
APPENDIX F	145
APPENDIX G	146

List of Figures

Figure 1.1. Various composites of the Station Strangler	12
Figure 1.2. A comparison of E-FIT and EvoFIT composites	26
Figure 1.3. Comparison of all composite system methods	28
Figure 1.4. Meta-analysis effect size plot of composite construction studies	36
Figure 1.5. Signal detection theory model	44
Figure 2.1. Whole face lineup with standardised gridlines	56
Figure 2.2. Percentage accuracy across whole face and featural lineups	60
Figure 2.3. Mean reaction times for whole face and featural lineups	63
Figure 3.1. Targets used at encoding	71
Figure 3.2. Experimental procedure for Experiment 2A	78
Figure 3.3. Percentage accuracy across conditions	81
Figure 3.4. Percentage of unforced and forced choice decisions across groups	85
Figure 3.5. Percentage accuracy for biased and unbiased lineups	88
Figure 3.6. Chi-square graph of reaction time for which accuracy is highest	93
Figure 4.1. Example of composite matching task	104
Figure 4.2. Percentage of composite selections	106
Figure 4.3. Mean reaction time across composite type	107

List of Tables

Table 2.1. Percentage correct and incorrect selections in lineup task	59
Table 2.2. Odds ratios for correct and incorrect selections between lineups	61
Table 2.3. Reaction times for all lineups	62
Table 2.4. Reaction times for correct and incorrect responses across lineup type	63
Table 3.1. Modal descriptions for all targets	72
Table 3.2. Percentage correct and incorrect selections for unforced decisions	80
Table 3.3. Percentage correct and incorrect choices across groups	83
Table 3.4. Correct and incorrect selections for forced decisions – target present	84
Table 3.5. Correct and incorrect selections for unforced decisions – target absent	86
Table 3.6. Proportions of lineup bias and effective size	87
Table 3.7. Proportions of selection reasons across conditions	89
Table 3.8. Mean responses for meta-cognitive questions	90
Table 3.9. Descriptive statistics for confidence values across groups	91
Table 3.10. Mean reaction times for correct and incorrect responses across groups	92

Abstract

When a crime is committed, law enforcement typically relies on the testimony of an eyewitness. However, eyewitness testimony is often susceptible to contamination. Eyewitnesses are usually required to construct a composite of the perpetrator. Research has suggested that eyewitnesses who construct composites are more likely to misidentify the target in a later recognition task (Comish, 1987; Wells, Charman, & Olson, 2005; Yu & Geiselman, 1993). This hampering effect may occur because composite programs utilise a featural selection strategy, which is in opposition to configural and holistic processing that is used when faces are encoded. This series of experiments aimed to investigate the featural aspect of composite construction. Two types of lineups were used in Experiment 1: a whole-face lineup, and a set of six feature lineups. Participants in each group encoded a target face and were then required to either select the whole target from the lineup, or identify the correct features. The whole-face lineup group performed significantly better than the featural lineup group. Participants in the featural lineups could only select the correct features at chance levels. This suggests that participants are unable to identify correct features, thus, participants may not be able to select appropriate features within the composite systems. Experiments 2A and 2B attempted to separate the featural process of construction. Participants selected the composite features one-at-a-time, and never assembled the features into a whole-face composite. A second group (a featural-assembly group) selected the individual features and assembled them once all the features were chosen. This was compared to a control group, and a standard, whole-face construction group. Results indicated that the isolated-featural group obtained significantly fewer hits than the control group. However, the isolated-featural group rejected the lineup, and did not incorrectly select fillers. When forced-choice results were analysed, all participants performed at the same level as the control group. This offers evidence towards a criterion shift, which is discussed.

CHAPTER ONE:

Introduction

When a crime is committed, law enforcement agencies typically rely on an eyewitness's construction of a composite. This composite, or facial replica, is then distributed to the public to aid in recognition and apprehension of the culprit. However, research has shown that composites can have a detrimental effect on memory and recognition (Kempen, 2009; Wells, Charman, & Olson, 2005; Wogalter, Laughery, & Thompson, 1986, Experiment 2). At the time of this writing, 235 prisoners, including 17 inmates on death-row, have been exonerated through DNA evidence in the United States of America. 77% of these exonerations were due to eyewitness misidentifications (Innocence Project, n.d., <http://www.innocenceproject.org/Content/351.php>).

It is not known how many of these eyewitnesses were required to construct a composite. However, typically when law enforcement only has the testimony of an eyewitness, composites are usually constructed to aid in the apprehension of a suspect. This may be detrimental for two reasons. Firstly, the composite may be contaminating memory for the original target. Secondly, the composite may resemble an innocent suspect who is then arrested. This happened to Kenneth Wyniemko (Innocence Project, n.d., http://www.innocenceproject.org/Content/Kenneth_Wyniemko.php). Wyniemko was charged with the rape of a woman and sentenced to prison for 40-60 years. The victim in question was required to construct a composite of the perpetrator. She later admitted that she did not see the perpetrator very well and that the composite was only 60% accurate. Wyniemko was at the police station because of an unrelated misdemeanor when a policeman noticed he resembled the composite. He was arrested and placed in a lineup where the victim identified him as the rapist. He spent over eight years in prison until he was exonerated due to DNA evidence.

Another case of mistaken identification is that of Willie Williams (Innocence Project, n.d., http://www.innocenceproject.org/Content/Willie_Williams.php). Williams was charged with the rape and abduction of two women. The first woman constructed a composite after being raped. A few days later, another woman was raped by the same perpetrator. A policeman showed the second victim the composite, and she agreed that it was her attacker. A few weeks later, Williams was spotted near the area of the attacks, and a policeman thought he resembled the constructed composite. He was arrested and both victims positively identified him from a lineup. He spent nearly 22 years in prison before the real perpetrator was convicted and sentenced.

One notable case in South African history is that of the “Station Strangler” who was wanted for the murder of 22 young boys. The only evidence law enforcement had were the composites constructed by eyewitnesses who had spotted the Strangler with various children. Norman Afzal Simons was a primary school teacher who had admitted himself to a psychiatric clinic. A nurse at the clinic thought that Simons resembled one of the composites and telephoned the police. As can be seen in Figure 1.1 the composites that were constructed to resemble the perpetrator are all of varying quality. Simons was later arrested and charged with murder. He was convicted by the testimony of one eyewitness who hesitantly picked him out of a lineup. She had only seen Simons from a distance and for a brief instant of time. He has currently served 17 years of his life imprisonment. Could this also be a case of mistaken eyewitness testimony?



Figure 1.1. Various composites constructed by eyewitnesses to resemble the Station Strangler and distributed through various forms of media. Norman Afzal Simons (top left) was later apprehended and charged for murder.

Wells (1978) makes the distinction between estimator and system variables which may affect memory for the crime. Estimator variables are variables which cannot be changed or influenced by law enforcement once the crime has been committed. Such intrinsic estimator variables include ethnicity (Meissner & Brigham, 2001b) and age (Bartlett & Memon, 2007) of the perpetrator and victim, stress (Deffenbacher, 1983), presence of a weapon (Stebly, 1992), as well as direct involvement in the crime (as opposed to being a bystander) (Stanny & Johnson, 2000). Extrinsic variables include length of exposure to the target or perpetrator (Shepherd, Gibling, & Ellis, 1991), distance between the perpetrator and victim (Loftus & Harley, 2005), and length of delay between encoding and recognition (Shapiro & Penrod, 1986). System variables, however, can be altered by law enforcement

following the event. Such variables may include how the suspect is presented to the witness, for example, by means of a simultaneous lineup, a sequential lineup, a mug book, or a show-up. Another factor which may influence an eyewitness's memory is the use of composites. Composites are facial likenesses or attempted reconstructions of a face. Eyewitnesses are typically required to construct a composite of the suspect when tangible evidence is not available. The eyewitness will work together with the police until a satisfactory pictorial likeness of the suspect is constructed. This composite may be distributed amongst the public to aid in suspect recognition. Once a suspect is apprehended, the same eyewitness may be required to select the guilty perpetrator from a lineup.

Does composite construction cause eyewitnesses to select an innocent suspect? As seen in the case of Willie Williams, exposure to a constructed composite (constructed by the first victim) may have influenced the second victim's selection from the lineup. The practice of composite construction may contaminate memory for the original event and may influence the way in which faces are perceived and recognised.

Cognitive Processes in Encoding, Construction, and Recognition

An important question in the eyewitness literature is whether faces are perceived featurally, or holistically (configurally), and the impact this has on subsequent tasks. Featural processing in this paper will refer to the processing of individual, separate parts, or structures of a face or object. Configural and holistic processing will be used interchangeably to refer to the processing of spaces and relationships of features and the overall holistic impression that it gives to a face or object.

Different Streams of Processing

Top-down processing. Memory can be highly susceptible to contamination. Schacter (1996) argues that memories are not veridical representations of past events, but are records of our experience of them. He suggests that memories are encoded in neural networks that are

shaped by previous experiences in the world. In an experiment by Bartlett (1932, as cited in Groome, 2006), participants were required to read a short Native American folk tale about ghosts and then write down as much as they could remember. Bartlett found that participants systematically changed the story to be more rational and fitting with their previous experiences. He concluded that participants altered and rationalised the story to fit with their expectations of how things operate in the world. Our pre-existing knowledge therefore influences how new memories are stored. Configural, perceptual experiences of depth, size, and speed (for example) are derived from cues in our environment. We can deduce the depth of a swimming pool when we see a person standing in it. We are able to do this due to our knowledge about average height. We frequently rely on our memories of how things operate to “fill in the blanks”. We make use of these constructivist strategies to assemble our perception, as sensory information can be limited.

For example, the Müller-Lyer illusion is a drawing of two parallel lines with arrows on the ends that either converge to, or diverge from, the lines. The illusion is that the line with the diverging arrows appears longer, when both lines are, in fact, the same length. Gregory (1966) suggests that we get tricked by this illusion as each line resembles the corner of a wall; the diverging lines come towards us, appearing closer and therefore bigger, while the converging lines make the wall look further away. Knowledge of walls, corners, and viewpoints of structures influences our perception of line length. If we were using a feature-extraction method, we would not be influenced by these top-down processes. Top-down, holistic processing may therefore be a result of the previous knowledge and experience we have with an object or event and not only the direct experience of the object.

Bottom-up processing. An opposing theory to the constructivist approach is the Gibsonian view (Gibson, 1950) which aims to find *why*, as opposed to *how*, perception works. Gibson suggests that perception must allow us to interact within our world, and, as our

surroundings are rich with sensory information, a constructivist, top-down processing approach is not needed. He further explains that we do not require stored knowledge that explains what objects *are* because objects “afford their use” (Groome, 2006). Thus, the use of an object brings about its purpose, as opposed to top-down processing which suggests that knowing *what* an object is leads to its use.

Evidence from the Visual System

Evidence for both constructivist and Gibsonian views can be found when examining the visual system. Shapley (1995) found that there appears to be two distinct streams running from the visual cortex in the occipital lobe. These are: the ventral stream, which runs to the inferotemporal cortex, and the dorsal stream, which leads to the parietal cortex. These two separate streams appear to be highly specialised. The ventral stream appears to be associated with knowledge-based representations, object-centered identification, recognition, and fine detail (Baizer, Ungerleider, & Desimone, 1991; Goodale & Milner, 1992). In contrast, the dorsal stream is concerned with visually-guided behaviour for motion-processing from a viewer-centered approach (Milner & Goodale, 1995). Evidence suggests that the ventral stream is concerned with recognising objects, while the dorsal stream is more suited to detecting change in the environment (Beck, Muggleton, Walsh, & Lavie, 2005; Zeki, 2003). These distinctions have been confirmed through neurological studies which show that cells in the inferotemporal cortex are more selective for individual identity (Kanwisher, McDermott, & Chun, 1997). Conversely, cells in the intraparietal sulcus are more specialised towards spatial perception and spatial memory (Haxby, Hoffman, & Gobbini, 2000). Thus, when exposed to a previously-seen target, the dorsal stream will initially detect if there is a change, and the ventral stream will work to discover what has changed.

We therefore have two distinct visual processing streams, one for bottom-up, featural, stimulus-driven processes, and one for top-down, holistic, knowledge-based representations.

Our visual system for perception therefore works in concert when encoding an object or an event (Haxby et al., 2000). This, therefore, affects how we perceive and recognise faces.

Processing Faces Featurally and Configurally

In order to investigate how faces are recognised, it is necessary to understand how they are processed. Research has suggested that we process faces using featural and configural processing modes. Featural processing refers to the processing of the isolated, individual *parts* or features of a face, such as eyes, nose, mouth, and eyebrows. Contrastingly, configural processing takes into account the spatial relationships between the features. Thus, featural processing is, by implication, a component of configural processing. In an experiment by Tanaka and Farah (1993), participants were required to select a target's feature, either in isolation, or embedded in a face. Participants were worse at accurately selecting the feature when it was presented in isolation (62%) than when it was shown within a face-space (73%). Other participants were shown another face with scrambled features. Participants were better at accurately selecting the feature in isolation (71%) than when it was tested in the whole, scrambled face (64%). This is presumably because the scrambling of features disrupts the relationship between the features, and thus, the configural processing of the face. However, the features used in this study were from composite images from the Mac-a-Mug program. This program uses greyscale images of line-drawn features. These features are not wholly featural, as the eyes (for example) included the eyebrows. Thus, the stimuli used in the experiment retained the configural information between the features.

Another study, which confirms the holistic processing of faces, used chimeric stimuli which consisted of the top and bottom halves of photographs of familiar (famous) faces and unfamiliar faces (Young, Hellawell, & Hay, 1987). When the faces were aligned (and made to form a whole face), participants took longer to recognise each half. The authors suggest that this was because the two halves created the illusion of a new face which induced a

holistic and configural processing mode of the entire face. Participants were quicker to respond when the halves were misaligned (placed horizontally off-center to each other). This effect disappeared when the faces were inverted (turned upside down). That is, participants took the same amount of time to recognise halves of the face when the faces were upside down. The distinction in this experiment is not wholly configural versus featural; the halves that Young et al. (1987) used were not entirely featural, as the configurations in the upper and lower halves of the faces were preserved. Instead, the experiment showed that when faces are aligned to their natural position, we automatically try to process the whole face as a new gestalt. Research suggests that the encoding of certain parts of faces is influenced by the relations and context of other components of the face (Searcy & Bartlett, 1996; Tanaka & Farah, 1993; Young et al., 1987). This may be a reason as to why recognition is impaired when faces are inverted.

The work by Tanaka and Farah (1993) forms one side of the extreme, in which features do not sum to create the whole. The authors suggest that faces are processed holistically as templates. On the other side of the extreme, another study has found that participants are able to select individual features above chance levels (Macho & Leder, 1998). Another study which corroborates this viewpoint suggests features are responsible for 91% of the variance in recognition of upright faces (Rakover & Teucher, 1997). A more moderate view finds that both featural and holistic processing is necessary for face recognition (Bartlett & Searcy, 1993; Cabeza & Kato, 2000; Sergent, 1984; Tanaka & Sengco, 1997).

Neurophysiology and neuropsychology have found that a group of neurons located in the superior temporal sulcus of the monkey responds discriminately towards faces and face parts. Monkeys were fitted with a recording chamber which was implanted in the temporal sulcus. Responses were measured in reaction to the presentation of various stimuli.

When a feature of a face is removed, the responses are not reduced; however, responses are greatly lessened when the features of a face are scrambled (Desimone, Albright, Gross, & Bruce, 1984). This fits with evidence from the visual system (discussed earlier) which suggests that the ventral stream runs from the visual cortex in the occipital lobe to the inferotemporal cortex. The ventral stream is concerned with recognition and object-centered identification. Brain imaging studies have confirmed that people have a fusiform face area (FFA), more specifically, the fusiform gyrus, located on the ventral side of the temporal lobe. This area is specialised for face recognition (Kanwisher et al., 1997). Research has suggested that people are more expert at recognising faces than other types of stimuli (Tanaka & Farah, 1993; Yin, 1969). The case of patient CK, who had visual agnosia following brain damage, indicated that faces and objects are double dissociated (Moscovitch, Winocur, & Behrmann, 1997). CK, who had been a collector of toy soldiers, was unable to correctly distinguish between the soldiers. However, recognition for human faces was preserved. Another patient, with prosopagnosia, had impairment in face recognition yet was able to accurately distinguish between novel stimuli (Duchaine, Dingle, Butterworth, & Nakayama, 2004).

We are better at recognising faces than recognising other objects (Hayward, Rhodes, & Schwaninger, 2008). However, the *own-race advantage* suggests that people recognise faces of their *own* race better than faces of other races (see Meissner & Brigham, 2001b, for a review). Studies utilising whole and inverted faces lead to the conclusion that own-race faces are processed configurally, as recognition for own-race faces is impaired when faces are inverted (Valentine & Bruce, 1986). A study by Tanaka and Farah (2003) found that White participants demonstrated a greater configural advantage for own-race faces than for Asian faces. Interestingly, the Asian participants had an equal advantage for faces of both races. Asian participants were more likely to have lived in White-dominated countries than vice-versa, and would be more familiar with White people.

One cross-race study utilised the chimeric face method employed by Young et al. (1987) (Michel, Caldara, & Rossion, 2006). Participants were either Belgian or Asian, and this corresponded with the race of the composite halves. Participants were required to determine whether the top half of the composite was from their own race, or from the other race. Participants took longer to respond for faces of their own race when the halves were aligned, than when the halves were misaligned.

A recent study, in which processing orientation was either to featural or configural aspects of a face, found that participants who engaged in configural processing showed a greater own-race recognition advantage (Greenberg & MacGregor-Hannah, 2010). Interestingly, participants who engaged in featural processing performed equally well on recognition for both races. This suggests that featural processing does not rely on stored knowledge or representations, as recognition was not biased towards a particular race. Sporer (2001) proposes a model which suggests that we fare better at recognising faces of our own race (in-group) because when we see features that we recognise, the face is automatically processed holistically. This results in better face recognition. However, when a feature is different to that of our in-group, we adopt a different processing mode to encode that out-group. This ventral stream processing therefore allows the individual to detect featural changes, devoid of familiarity and semantic biasing. Contrastingly, global and holistic aspects of the face are embedded in configural processing. As previously discussed, the latter relies on stored knowledge and familiarity, both of which are by-products of an experienced event.

Tulving's *encoding specificity principle* suggests that we recognise stimuli better if our mode of processing at recognition is the same as it was at encoding (Tulving, 1983; Tulving & Thomson, 1973). Thus, if faces are encoded holistically, retrieval is better if a holistic mode of processing is also induced at the retrieval stage. Likewise, at an interference stage (composite construction, for example), the mode of processing should match that at

encoding and recognition, in order to facilitate retrieval cues. Featural and configural processing are of fundamental importance in understanding how faces and composites are processed, and the effect they have on subsequent tasks. Thus, it is necessary to understand the properties and processes underlying composites and the methods of their construction.

The Quality of Composites

Composites, although widely used, frequently do not resemble the original targets they are intended to mirror (Davies & Valentine, 2007; Frowd et al., 2005). Composite quality tasks typically involve participants constructing a composite of a target face, either from memory, or with the target face in view. These composites are then compared to the original target face by matching or naming tasks. One study found that participants in a naming task were only able to correctly name 2.8% of a collection of 50 famous faces that had been constructed by other participants (Frowd et al., 2005). Another study found that participants were unable to name composites that were constructed to resemble people they knew personally (Kovera, Penrod, Pappas, & Thill, 1997). This suggests that even if the constructor is familiar with the target, composite production still produces a poor likeness of the intended face. However, studies have shown that composites constructed with the target in view are more accurately matched than those constructed from memory (Davies & Valentine, 2007; Ellis, Davies, & Shepherd, 1978).

The consequences of poor quality composites (as suggested by Maskow, Schmidt, Tredoux, & Nunez, 2007) are twofold: Firstly, poor quality composites may lead to the mistaken apprehension of an innocent individual (a false positive), as the composite may coincidentally resemble them. Secondly, misinformation in a poor quality composite may interfere and contaminate the eyewitness's memory for the original target. When an incorrect memory of the target face is recalled during construction, development and learning take place during this retrieval process. The constructor may consolidate these contaminated face-

traces with the original memory, resulting in incorrect selections during a subsequent recognition task. Thus, if composites *are* contaminating memory traces for remembered faces, the problem may lie in the systems used to construct these lookalikes.

Composite Production Systems

Composite construction techniques have developed considerably over the last few decades. At present, there are four generations of composite systems (Davies & Valentine, 2007). These range from sketch artists' impressions to synthesised, computerised composites, which will be discussed below.

The debate still remains regarding the effectiveness of composites, with a recent study suggesting they may have contaminating and detrimental effects on eyewitness memory (Wells et al., 2005). Composite construction involves the assembly of individual facial features to form a whole face. Various systems utilising different cognitive processes have been used in composite production. Their development and implementation has influenced eyewitness memory and recognition accuracy.

Sketch artists. Police traditionally used sketch artists' drawings of a perpetrator. These drawings were based on a description given by an eyewitness. Sketch artists were able to draw an unlimited number and variety of features, and were able to make very fine alterations to composite sketches. The sketch artist would work with the witness and draw the outline of the face, add features, alter details, and even change the expression of the face until the witness was satisfied that it represented the most accurate impression of the suspect. These techniques, useful though they were, proved limiting in that they required trained artists who were skilled in accurate facial drawings. Artists' composites could take between one and three hours to construct (Taylor, 2001). Literature evaluating the accuracy of sketch artists' composites and their impact on later recognition is sparse. This is primarily due to the nature in which sketch artists work, using a variety of different methods of construction,

making it difficult to standardise any composites for use in a laboratory study (Davies & Valentine, 2007).

Manual systems. Due to the extensive time required, and the limited availability of sketch artists, law enforcement developed other systems which could construct faces from already-made drawings of facial features. The first of these systems was the Identi-kit, which consisted of line-drawings of features that were printed on acetate. The witness would work with the trained police operator and describe the previously seen face. Features that best matched the witness's description would be selected in a piecemeal fashion. All the selected features were then superimposed over one another to form a composite available for distribution amongst the public.

The lack of realism within Identi-kit's line-drawings led to the development of a new system, Photofit, which used photographs of features. In this way, operators and witnesses were able to construct a face based on a wide variety of available lifelike features. However, a study showed that verbal descriptions of a target resulted in more accurate identifications compared to composites constructed using Photofit (Christie & Ellis, 1981).

Although the manual systems had dealt with the time constraints that hampered sketch artists, these systems also had their limitations. Whilst the Identi-kit and Photofit contained hundreds of features, it was still impossible to have every feature available. They both relied on the assumption that people naturally parse features in a piecemeal fashion to form a face. Research has shown that we encode and perceive faces holistically and configurally (Bruce, 1988; Tanaka & Farah, 1993; Valentine & Bruce, 1986). Once a holistic impression of a face is made, it becomes very difficult, and unnatural, to try and individuate the separate features.

Furthermore, the presence of a helping operator may influence the eyewitness's later recognition accuracy for the suspect. Research has suggested that having to verbalise an image (as the artist drawings and mechanical systems required) can have a detrimental effect

on one's memory for the target (Loftus & Loftus, 1980; Schooler & Engstler-Schooler, 1990). Artists and manual system operators are required to call for exhaustive descriptions from the participant. This *verbal overshadowing effect* (VOE) is found to impair processing of a full-face by focusing purely on specific features (Fallshore & Schooler, 1995).

Verbalisation requires the participant to access features of the memory trace, which is contrary to how it was encoded (holistically). Indeed, a meta-analysis showed that verbalisation of features impairs the original memory trace in later recognition (Meissner & Brigham, 2001a). However, research by Meissner, Sporer, and Schooler (2007) suggests that face descriptions may draw on distinctly different processes than those used in face recognition. The authors suggest that face description benefits from focusing on individual features, whereas face recognition is facilitated by global processing of a face. They further note that nonverbal operations such as face recognition are predominantly right hemisphere functions, as opposed to verbal left hemisphere operations. Thus, the verbalisation of a face should not hamper recognition; it may instead be the distinct modes of processing engaged during encoding and a description task. This will be discussed later with regards to facial processing. Indeed, several attempts to replicate the VOE have been unsuccessful (see Meissner and Brigham, 2001a, for a review). A recent article by Schooler (2011) reported that when tests are repeated, the statistically significant effects appear to diminish. He proposed that this *decline effect* may be due to regression to the mean. Schooler suggests that "if early results are most likely to be reported when errors combine to magnify the apparent effect, then published studies will show systematic bias towards initially exaggerated findings, which are subsequently statistically self-corrected" (2011, p. 437). He further suggests that differences in methodologies, researcher enthusiasm, experimenter bias, and the difficulties in publishing failed replications all contribute to the decline effect.

A study by Laughery and Fowler (1980) demonstrated that sketch artists provided better composites than the manual systems, although both performed poorly. The sketch artist group may have performed better due to an unlimited amount of features. However, both groups were hampered by the presence of a third-party operator.

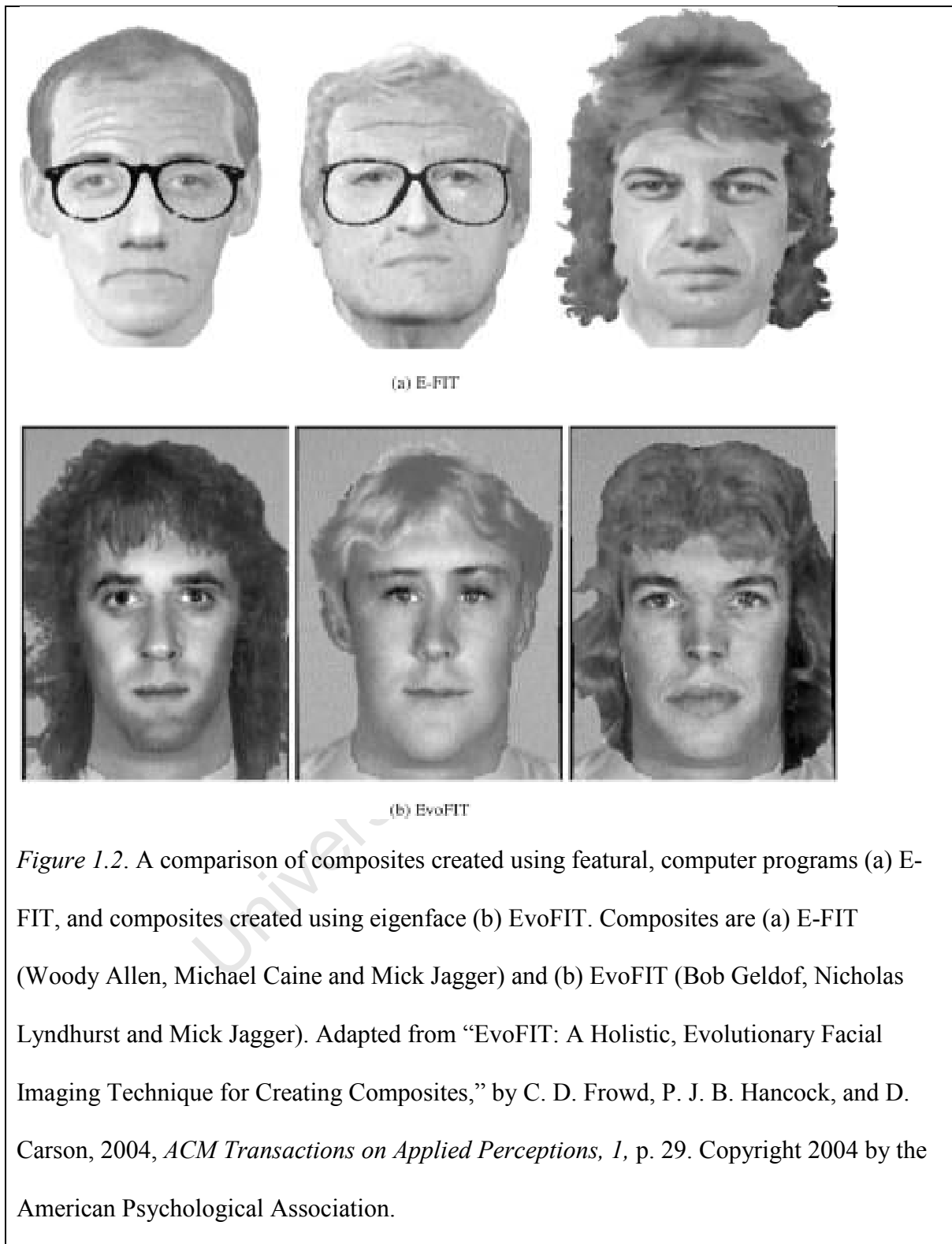
Computer software systems. Thus, the need for a participant-operable system arose, and this was met by the development of computerised composite software. These programs, namely Mac-a-Mug Pro (Shaharazam, 1986), E-FIT (Aspley Limited, 1993), and FACES (Cote, 2005), allowed participants to independently construct a composite using a computer. These programs consisted of hundreds of features which the participant could go through until they found the most appropriate features. The benefit of these software systems was that it allowed the user to click on the resembling feature and it would automatically embed itself into the whole face, appearing on one side of the screen. In this way, features formed a smooth, merged, whole face, unlike the blocked-fitting features of the manual systems. Furthermore, the software allowed the user to configurally manipulate the features within the program. The size, colour, and distance between features, could all be adjusted. This was not previously achievable with the manual acetate prints.

However, a study using Mac-a-Mug (which contains more detailed features than the manual systems) showed that participants who constructed composites using computerised software were unable to match them to photographs (Kovera et al., 1997). The same study also found that composite builders were unable to construct composites of faces that were well-known to them. This suggests that an increase in available features and alterations does not necessarily correspond with increased composite quality. Another study, which compared the quality of composites constructed using either Photofit or E-FIT, found no significant difference between the two systems (Davies, van der Willik, & Morrison, 2000). Furthermore, the naming rate for both systems was low (17%). The naming rate for

composites created using E-FIT increased to 49% when the composites were constructed with the target face in view. No difference was found when the Photofit group constructed the composite with the target in view. The lack of significant findings between mechanical and computerised systems has been found in other studies (Green & Geiselman, 1989; Kovera et al., 1997). Computerised systems only have the ability to construct more accurate composites when the target face is in view. Both groups perform poorly when target faces are matched to composites created from memory.

Although more contemporary, computerised, systems have increased flexibility and number of features, this is only beneficial when the target face is present and can be referenced. Thus, is there something inherent to the computerised systems that render it no more accurate than the manual ones? A possible answer could be that the Mac-a-Mug and E-FIT software utilise a featural-based selection and configuration strategy that the manual systems operate on. Computerised software was developed to eliminate the need for an operator, and thus, verbalisation. However, when a face is remembered, features are recalled which are then verbalised in one's own head. The act of visualising blue eyes leads to telling oneself (although silently) that the eyes were blue. Thinking occurs in language, and it is thus impossible to ever eliminate any form of slight verbalisation. Thus, it may be possible that trying to individuate the separate features in one's mind could lead to a visual overshadowing effect. Access to the entire holistic face, together with its configurations, is blocked because of focusing specifically on piecemeal parts of a face.

Eigenface algorithms. The latest development in the composite production software is the move away from featural selection programs towards a more holistic construction of faces. These eigenface programs, such as EvoFIT (Hancock, 2000) and ID (Tredoux, Rosenthal, Nunez, & da Costa, 1999), operate by blending or evolving a selection of whole, artificial faces into a synthesised, holistic version of the target.

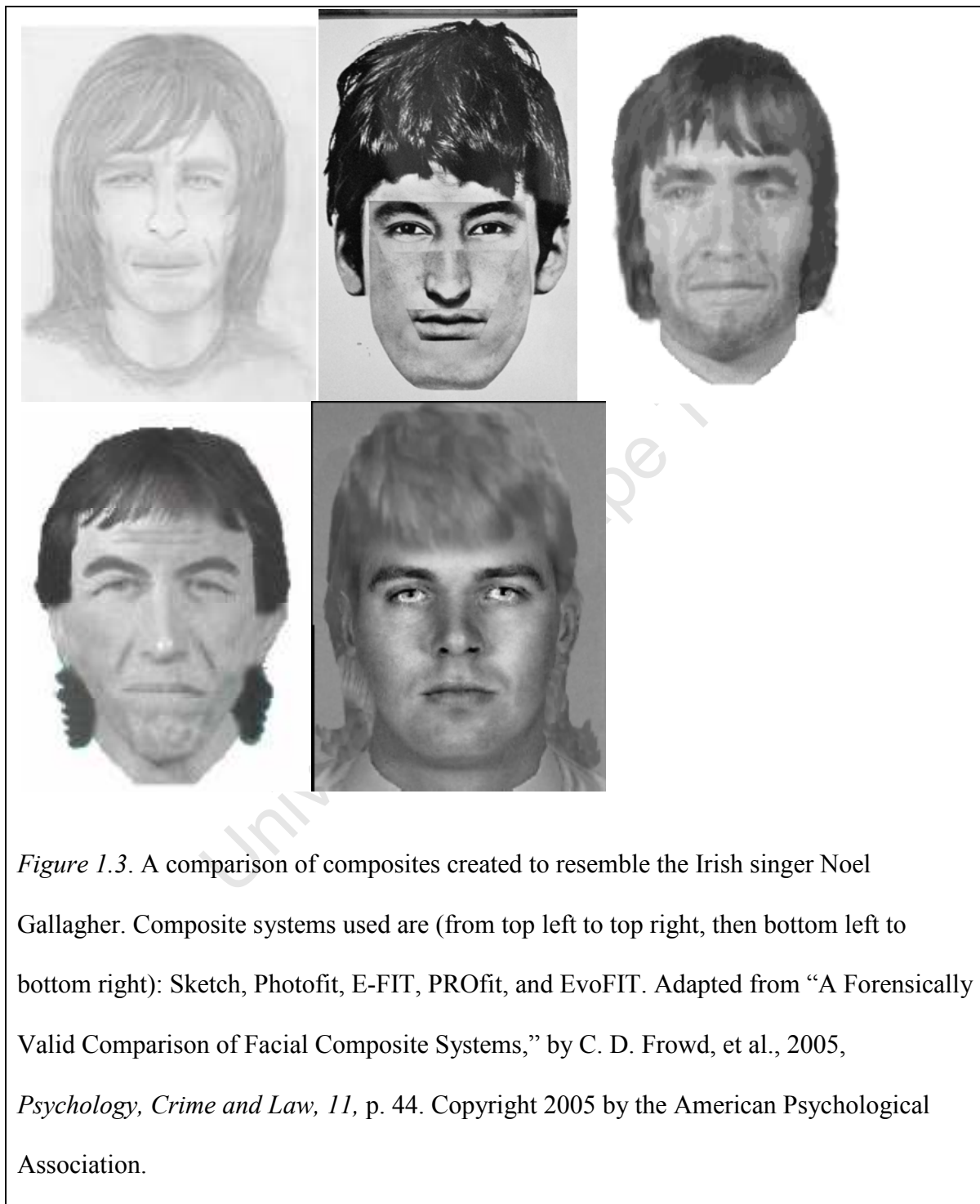


These programs use a method called Principal Components Analysis (PCA), similar to factor analysis. Hundreds of faces in a population are landmarked within the system, such

that all the features of the face lie in the same position on each face. Faces are represented by a system of coordinates based on average eigenfaces. The differences between faces are modeled and used to create reference faces. These synthetic faces are created by the coding of variance across all the faces, as opposed to the individual features, which gives them their holistic nature. Participants are exposed to a random, varying assortment of faces. The initial generated faces differ in varying degrees of resemblance. The faces that look most similar to the target are selected. The program then “breeds” them together, forming a holistic face which begins to resemble the perpetrator. These faces are evolved together to create a blend of the selected faces. These steps are repeated until the differences between the generated faces become slight and the composites begin to look more and more alike. A final composite face is then selected.

A study that examined ID constructions found that participants were able to produce recognisable composites if the target face was in view, or if the target was highly familiar and well memorised (Tredoux et al., 1999). The authors also found that unfamiliar faces were only produced at chance levels. Another study found increased recognition for FACES composites over ID composites when constructed with a target in view (Tredoux, Nunez, Oxtoby, & Prag, 2006). However, ID outperformed FACES when constructions were created from memory. Although ID seemed to fare better than FACES for constructions from memory, the overall quality of the composites were “rated below the midpoint of the matching scale” (Tredoux et al., 2006, p. 8). Holland, Otzen, and Sporer (1994) showed that participants who produced poor quality composites were less likely to identify the target (as cited in Maskow et al., 2007). Poor quality composites may therefore contaminate the original memory trace during reconstruction.

A study by Frowd, Hancock, and Carson (2004) found that naming rates for composites constructed using EvoFIT (an eigenface program) were lower than composites constructed using E-FIT (a featural, computer software system).



*Figure 1.3. A comparison of composites created to resemble the Irish singer Noel Gallagher. Composite systems used are (from top left to top right, then bottom left to bottom right): Sketch, Photofit, E-FIT, PROfit, and EvoFIT. Adapted from “A Forensically Valid Comparison of Facial Composite Systems,” by C. D. Frowd, et al., 2005, *Psychology, Crime and Law*, 11, p. 44. Copyright 2005 by the American Psychological Association.*

A later study by Frowd et al. (2005) examined all types of composite construction. The study compared composites created by sketch artists, Photofit (mechanical system), E-FIT (computer system), ProFIT (computer system), and EvoFit (holistic, eigenface system). The results indicated that E-FIT composites were named more often than the other systems. However, ProFIT composites were named at higher rates when the target was more distinctive. Interestingly, Photofit and EvoFIT composites were named least often. There therefore appeared to be no significant difference between the older, featural, mechanical systems, and the newer, holistic, eigenface computer systems.

Although this later generation of composite systems was developed to better simulate the way in which we naturally process faces (holistically), these programs still innately utilise some form of featural construction. It may be possible that during these “holistic” composite constructions, “breeder” faces are selected based on the featural overlap they share with the memory for the target. For example, generated faces may be chosen because they have similar eyes to the perpetrator, whilst another eigenface is selected because of a similar hairstyle. Although these holistic programs may closer approximate our natural processing of faces, it still appears as though constructing a composite from memory is too difficult a task. However, as previously discussed, composite construction techniques are not wholly featural or configural, they are a combination of both. This report will now examine the sparse literature surrounding the effects of composite production on recognition.

The Effect of Composite Production

Much of the research conducted on the effects of composite production takes the form of three phases: encoding, interference, and recognition. Participants are typically exposed to a target (or number of targets) for a variable amount of time. Participants in the experimental condition are then required to construct a composite using a particular construction method. This acts as the interference phase. After a variable time delay, participants are then required

to select the target in a recognition test. The sparse literature surrounding the effects of composite production on later recognition is equivocal. There are three classes of studies which have investigated this forensic technique. The first class of studies finds no significant effect of composite construction on later recognition; the second group finds a facilitating or positive effect on composite production, and the third collection of literature finds a detrimental effect on later memory.

No significant effect. A study conducted by Davies et al. (1978, Experiment 2) found no significant effect on recognition accuracy between participants in a control condition and participants who were required to construct a composite using Photofit. Photofit consists of printed photographs divided into five face parts: forehead and hair, eyebrows and eyes, nose, mouth, and chin and cheeks. These photographed parts are then assembled with the help of an operator and mounted on a board. Davies et al. found that participants were able to accurately select the target after a delay of up to three weeks. However, participants in the study were all exposed to the target at the same time, and were then scheduled one-at-a-time to construct their composite. Composites constructed with Photofit need to be assembled with the aid of an operator. As such, some participants constructed their composite immediately, whilst other participants were scheduled to construct their composites up to 48 hours later. However, Davies et al. thought that this was not of significant importance. This seems to be an experimental flaw. The composite construction condition should have been standardised so that all construction participants could build their composite after the same delay. Furthermore, participants in this experiment were exposed to the target for 15 seconds. They were also warned that they would have to remember the face and answer questions about it. All participants (in both conditions) were able to accurately select the target after two trials of photograph presentation in the recognition task. This may suggest that ceiling effects were obtained due to the lengthy exposure time. Participants were also primed to encode and

remember the target. This could possibly account for the lack of significant difference between the two groups. However, the authors did find that participants, who had initially failed to select the target after the first trials, had lower quality composites. A second study conducted by Holland, Otzen and Sporer (1994) (as cited in Maskow, et al., 2007) also found no significant effect when participants constructed a composite. However, they found that participants who constructed good quality composites were more likely to accurately select the target than participants who constructed low quality composites.

Facilitating effect. A second group of studies found a facilitating effect on a later recognition task. An experiment by Mauldin and Laughery (1981) found that participants who constructed a composite using Identi-kit performed significantly better than a control group and verbalisation group. In the recognition task, participants were shown 130 slides and had to make a yes-no decision on each photograph. The target was always present in position 125. The control group were required to complete Rotter's I-E scale questionnaire which measures an internal-external locus of control. The verbalisation group were required to complete a facial feature description questionnaire. Participants in the control condition achieved 60% accuracy, compared to composite construction participants who obtained 90% accuracy. The authors state that these results are similar to other results seen in face recognition literature. However, they did not have their composites rated. It is unclear if the composites were of such poor quality that they provided no interference, and thus could not have masked the memory for the original face. Also, Mauldin and Laughery used 128 participants in their study, and each participant was shown a different target. This attempted to eliminate any spurious effects that may have resulted from unforeseen idiosyncrasies present in any one face. Since all participants were given the same sequential lineup, and 128 targets were used, it is doubtful that the lineup members would have looked similar to each other. Another study by Wogalter et al. (1986, Experiment 1) which used the FIS (Field

Identification System) found similar results. The FIS is similar to the Identi-kit in that both techniques consist of line drawings of features assembled into a book. However, FIS, unlike the Identi-kit, may be operated without the assistance of an operator. Wogalter et al. attempted to minimise participants' verbalisation in this experiment. Participants in the construction group selected the target more often, compared to the control group. The authors found a correlation between composite quality and recognition performance. They suggested that the FIS, like the Identi-kit, makes use of less detail than Photofit (for example), which found no significant effect between groups. They propose that composite systems with less detail available, i.e., line-drawing systems that are less detailed than photograph systems, have less of an intervening effect. A meta-analysis found that composite production did have a facilitating effect on later recognition (Meissner & Brigham, 2001a). This meta-analysis indicated that composite construction (compared to a control group) made participants 1.56 times more likely to accurately select the target. However, this meta-analysis predates other literature which has found composite construction to have detrimental consequences on memory, which will be discussed below.

Detrimental effect. Another study by Wogalter et al. (1986, Experiment 2), which used Mac-a-Mug, found composite production to have a hampering effect on memory in a subsequent recognition task. Mac-a-Mug is a computer-based composite system which contains a wide range of photographs. It eliminates the need for an assisting operator and also minimises the amount of verbalisation required during construction. The authors found that participants in the composite construction condition had a lower proportion of correct hits than participants in the control condition. This further evidenced Wogalter et al.'s claim that composite systems with increased detail result in participants performing poorer than a control group. Indeed, one study found that participants, who worked with sketch artists to construct a composite, performed worse in the recognition task, even when the target was not

present in the lineup (Hall, 1976, as cited in Yu & Geiselman, 1993). Sketch artists, unlike other composite systems, have the ability to draw an unlimited number of features and are able to edit composites in a way that other composite programs are unable to do. Indeed, Wogalter et al., (1986, p. 11) suggest that “an a priori analysis of the different techniques indicates that in order of increasing detail requirements the techniques would be FIS, Identi-kit, Photo-fit, Mac-a-Mug, and sketch artists.” Another study exposed participants to a composite-face (Comish, 1987). Once participants had constructed their composite, their lineups were individually built in such a way that all the fillers looked like the target, but incorporated incorrect features from the participants’ composites. Participants who had constructed a composite were more likely to incorrectly select a foil containing some of their own errors. Participants were also more likely to incorrectly reject the lineup and say that the target was not present. This was an indication that misinformation present within the composite may be “recoded” as being accurate, and remembered and recalled during recognition. Comish (1987) found that participants who constructed a face with Identi-kit were less accurate at recognising the target in the lineup, when compared with a control group (22% and 44%, respectively). Another study found participants that constructed a composite using Identi-kit were less likely to select any face from the lineup (Yu & Geiselman, 1993). They were less certain about their decision and rejected the lineup rather than incorrectly select a foil. This suggests that composite construction may increase selection conservativeness. This finding is corroborated in another study which found that constructors performed worse than the control group; when these incorrect answers were analysed, 51 out of 63 “incorrect” answers were mistaken lineup rejections (Kempen, 2009).

One study, which has since received wide media coverage, found large detrimental effects of composite construction on later recognition (Wells et al., 2005). In this experiment, the authors exposed participants to a target face for 180 seconds. This is a surprisingly long

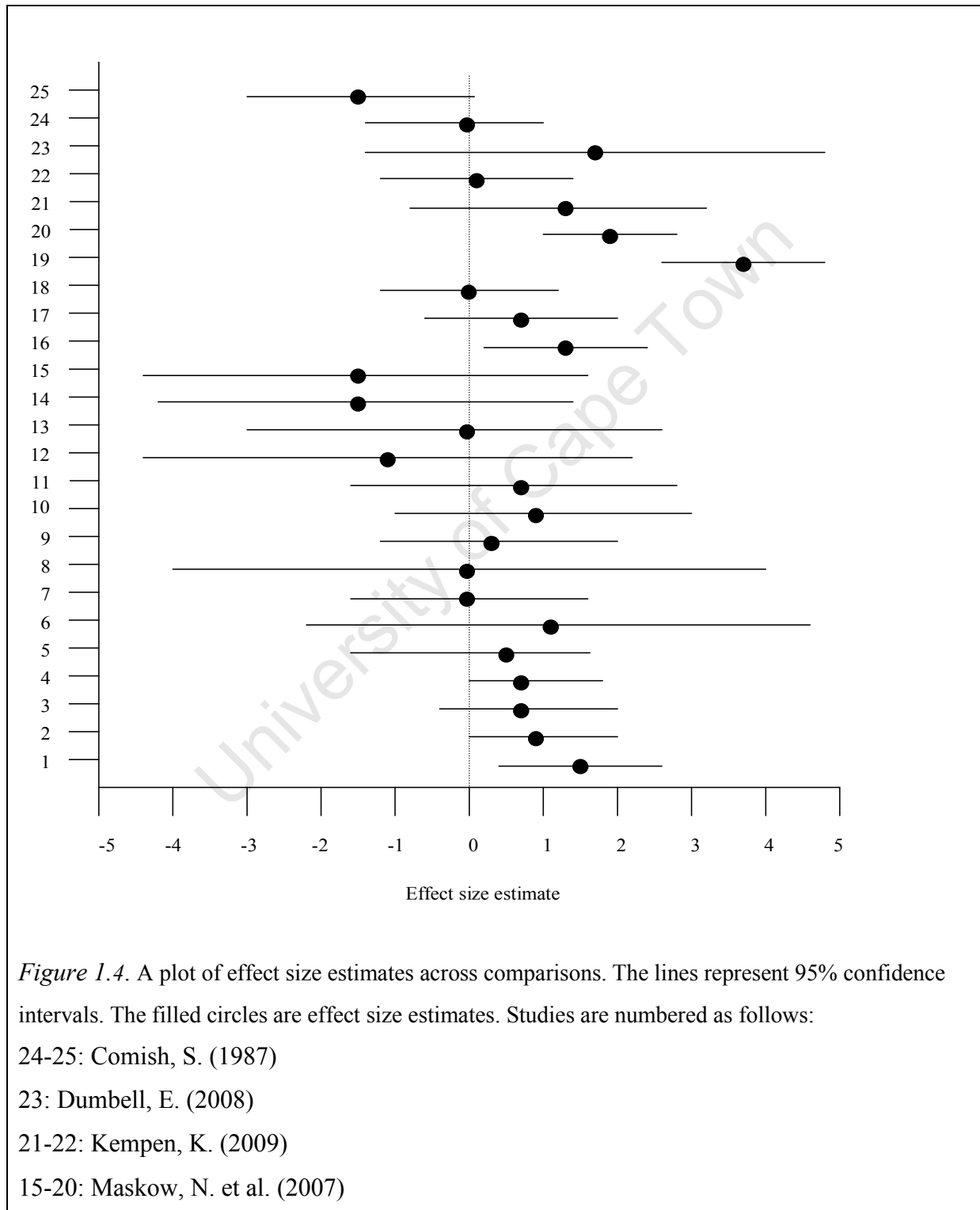
exposure time compared to the 16 second encoding time that is typically implemented (Maskow et al., 2007). Wells et al. (2005) used three groups in their study: a control group, a composite production group, and a yoked composite-exposure group. After the target was encoded, the yoked composite-exposure group was shown a composite constructed by another participant. The control group did not engage in any intervening activity or filler task. Wells et al. used FACES 3.0, a computerised composite software system, which can be operated without the assistance of a trained operator. It contains hundreds of features which can be resized and configured around the face. Participants in the Wells et al. composite production group were required to construct a composite. All participants returned two days later to select the target from a six-person lineup. Results showed that participants in the control group achieved 84% accuracy in comparison to the construction group which obtained 10% recognition accuracy. The yoked composite-exposure group attained a recognition accuracy of 44%. Participants in composite construction and yoked composite-exposure group were also more likely to reject the lineup completely (58% and 50%, respectively). However, when these two groups were forced to make a selection, the hit rate of the target under forced identification was 30% and 82%, respectively. This unambiguous finding was not in keeping with the equivocal literature that had preceded it. The Wells et al. results suggested that composite exposure and construction can have negative carryover effects on a later recognition task. The composites in this study were matched to their targets at levels just above chance rates.

In light of this unequivocal finding which no other study had achieved, the Wells et al. (2005) paper was well published and reported in popular media (as cited in Maskow et al., 2007). If these results are an accurate reflection of the effect of composite production, then this would have serious implications for the forensic techniques used in law enforcement. It was therefore necessary that this study be replicated to corroborate these findings. A study by

Maskow et al. (2007) attempted to replicate the results. They also implemented an exposure time of 180 seconds, but achieved ceiling effects for all the groups. In a second experiment, Maskow et al. reduced the exposure time to 16 seconds. The construction group obtained 73.3% accuracy. Even with a reduced exposure time, this accuracy still exceeded that of the Wells et al. composite construction accuracy (10%). Maskow et al. concluded that such a lengthy exposure time resulted in an encoding that was robust against composite interference. A second, attempted replication (Dumbell, 2008) decreased the exposure time to two seconds, yet was still unable to achieve significant results. However, a third replication was attempted by Kempen (2009) and managed to obtain a significant difference between the control group and the construction group (65.12% and 44.19%, respectively). Interestingly, the yoked composite-exposure condition performed worse than the construction group, achieving only 36.84% accuracy. This study followed the Wells et al. procedure but decreased the exposure time to five seconds. However, of the 63 incorrect selections, only 12 of the decisions were false alarms (filler selections). The other 51 selections were incorrect lineup rejections. Participants were more conservative and less confident about their selection and decided to reject the lineup rather than make an incorrect selection. When the forced-choice responses were analysed, along with the initial selections, all the significant differences disappeared. The control and construction groups performed in exactly the same way (both achieved 74.42% accuracy) and the yoked composite-exposure group increased in recognition response to 60.53%. The finding of conservative lineup decisions supports the results discussed earlier by Comish (1987) and Yu and Geiselman (1993).

Thus, the literature surrounding composite production on lineup recognition remains largely equivocal. All the studies, to date, have found ambiguous results. One study (Wells et al., 2005) found large effects, yet three attempted replications (Dumbell, 2008; Kempen, 2009; Maskow et al., 2007), although unpublished, have yet to achieve similar findings. It is

therefore necessary that for more research to be done on the effects of composite construction. If Wells et al.'s composite construction hit rate of 10% is accurate, there are many innocent individuals at risk for mistaken identification and wrongful incarceration.



- 11-14: Mauldin, M. A. & Laughery, K. R. (1981)
8-10: Topp, L. (2006)
6-7: Wells, G. L., Charman, S. D., & Olson, E. A. (2005)
3-5: Wogalter, M., Laughery, K., & Thomson, B. (1987, Experiment 2)
2: Yu, C. J., & Geiselman, R. E. (1993)
1: Wogalter, M., Laughery, K., & Thomson, B. (1987, Experiment 1)

However, a recent, unpublished meta-analysis on the effect of composite production found that participants who constructed composites had fewer target hits than the control groups (Tredoux, et al., 2010). The meta-analysis reported an odds-ratio of 1.91, indicating that the control groups were 1.91 times more likely to select the target than the composite construction groups. The average weighted effect size between control and construction groups was 0.65, 95% CI [0.38, 0.92]. The meta-analysis found that the type of composite system used, delay after encoding, and length of target exposure, all moderated the effect. For the purposes of this review, the effect sizes and confidence intervals for all the studies that compared control groups and composite construction groups were calculated. The results were plotted in a graph in Figure 1.4 above. The effect sizes occurring to the right of the graph indicate a significant effect where composite construction had a hampering effect on recognition accuracy. Values occurring on the left hand side depict the studies where composite construction had a facilitating effect. As can be seen, composite construction appears to have a hampering effect when all studies are taken into account.

Theories Accounting for Composite Interference

In order to fully investigate the hampering effect of composite production on later recognition, it is necessary to examine possible theories that may account for this decrement in accuracy and how they relate to face recognition.

Recoding interference. The verbal overshadowing effect (VOE), a theory first devised by Schooler and Engstler-Schooler (1990), suggests that verbal description of a face may hamper later recognition of that face. The authors conducted a study in which

participants were either placed in a control condition, or a target-description condition. After encoding a 30-second video of a simulated crime, participants in the description condition were required to give a verbal description of the perpetrator. Schooler and Engstler-Schooler found that participants who had constructed a description of the perpetrator were significantly more likely to perform worse in a recognition task than participants in a control group (39% vs. 64%, respectively). This result has since been replicated by other studies (Dodson, Johnson, & Schooler, 1997; Finger & Pezdek, 1999). A meta-analysis found a significant negative effect for the VOE on later recognition (Meissner & Brigham, 2001a). Macrae and Lewis (2002) suggest that this effect occurs because it induces a featural mode of processing which is not conducive to the holistic processing we use to encode faces.

Finger and Pezdek (1999) found that the VOE is only apparent if tested immediately following verbalisation. The authors instructed participants to remember every detail of the face and also to provide descriptive detail about all the features. They found that participants experienced a greater VOE than participants who were instructed to engage in free recall. The study also found that a “release” of the VOE occurred after a delay. Furthermore, they also found that when participants exhibited self-generated misinformation about the face, the VOE prevailed even after the delay. This suggests that when participants generate elaborate descriptions containing misinformation, participants may internalise the description as being veridical and perhaps recode the misinformation as belonging to the original source. Indeed, Ackil and Zaragoza (1998) suggest that self-generated misinformation may result in a memory that was too similar to the original event and render the participant unable to distinguish between the original event and the misinformation.

Meissner, Brigham, and Kelley (2001) found that participants who generated more elaborate descriptions showed increased recognition impairment than those who gave free recall descriptions. This may suggest that forcing participants to verbalise more descriptors

increases the chance for confabulation and recoding of that misinformation. The authors also found that when fewer incorrect features were generated, participants were more likely to have fewer misidentifications than participants who generated more incorrect features. This finding supports the similar results obtained by Schooler and Engstler-Schooler (1990).

It is possible, therefore, that requiring someone to make a composite is much like forcing a participant to generate an elaborate description. This, in turn, would support the earlier discussion that composite systems with increased detail may result in increased misidentifications. For example, in highly-detailed composite systems, there are numerous individual features which can be selected. It is unlikely that participants will be able to recall every isolated feature. Therefore, being presented with an “ear” or “chin” category which they do not remember may result in confabulation of that feature. This may account for the earlier facilitating studies which used the Identi-kit and FIS. These systems consist of a booklet of line-drawings divided into four face sections, with fewer face parts to recall and account for. Thus, it may not be the actual act of verbalisation that accounts for interference in recognition, but rather the recoding of elaborated misinformation. As discussed earlier, Meissner et al. (2007) suggest that face description and face recognition may be distinctly different processes; thus, describing a face may not hamper face recognition. However, Meissner et al. discussed verbalisation with regards to face descriptions, and not composites. It is possible that a visual overshadowing effect (Kempen, this publication) occurs when features are recognised and embedded in a composite.

However, the recoding interference theory does not explain why the VOE occurs when participants are required to verbalise a different face (Dodson et al., 1997), but does not occur when participants are required to generate a description of a face from another ethnicity (Fallshore & Schooler, 1995). This may occur because faces of an out-group ethnicity are distinctly different from one’s own in-group faces. As discussed earlier, out-group faces are

processed featurally. Thus, describing a face of another ethnicity may not elicit a contaminating VOE, whereas describing an in-group face, which relies on configural and holistic processes, may be susceptible to the VOE. However, as mentioned earlier, Schooler (2011) suggests that the decline effect may be attributed to unpublished studies resulting from a failure to replicate the VOE.

Transfer-inappropriate processing shift. The transfer-inappropriate processing shift (TIPS) was first described by Schooler (2002). It is a re-theorising of the transfer inappropriate retrieval (TIR) first theorised by Diamond and Carey (1986), which in turn stems from transfer-appropriate processing (TAP) (Fisher & Craik, 1977). These theories build upon the work by Tulving (1972) (as cited in Groome, 2006). This earlier work, which they termed the encoding specificity principle (ESP), suggested that recognition at the retrieval stage is facilitated if it possesses some of the original information that was present when it was encoded. It suggested that the amount of feature overlap present at encoding and retrieval will facilitate the retrieval of an item from memory. Indeed, deeper connections and traces between a memory at encoding and retrieval can be formed if semantic elaboration occurs (Craik & Tulving, 1975). This allows us to process the item and all its cues on a deeper, semantic level, as opposed to a structural, featural level. A study on face recognition found that participants were able to identify the target more frequently if the target was initially rated on semantic characteristics of the face, as opposed to superficial features of the face (Winograd, 1976). Although semantic and structural processing of items may occur in parallel, semantic processing results in a deeper, longer-lasting memory trace (Craik & Tulving, 1975). Support for the ESP and TAP are shown in a study in which percentage recognition was higher for semantic orienting tasks which had semantic cues available at retrieval (Fisher & Craik, 1997).

One study oriented participants to a mode of featural processing after encoding a target face (Macrae & Lewis, 2002). The authors induced automatic and controlled processing in participants by exposing them to a task using Navon letters. Since encoding a face is an automatic task, participants who engaged in controlled processing of the Navon letters performed worse in a recognition task. This is because the modes of processing immediately following encoding were not the same. Thus, the argument may follow for composite production: that engaging in a featural mode for construction is incongruent with the holistic and configural mode of encoding and retrieval. However, a meta-analysis by Meissner and Brigham (2001a) found that the VOE only occurred with a delay of up to 30 minutes. Thus, it is not likely that a featural mode of processing is induced for the entire length of construction, along with the delay up until retrieval.

Misinformation effect. Post-event information can have an influence on eyewitness testimony. This is typically the reason why, following a crime, law enforcement will ask all the witnesses not to share stories until they have individually told their account of the event to the police. Misinformation can retroactively influence one's perception of an event. A study by Loftus and Palmer (1974) showed participants a film of two cars colliding. The questions that followed either asked the participants, in different groups, to estimate the speed at which the cars were travelling when the two cars either, hit into each other, or *smashed* into each other. The insertion of the word *smashed* caused participants to estimate the cars' speed much higher than the other group who were questioned about the cars *hitting* each other. This overestimation continued even until one week later, when participants confabulated they had even seen smashed glass in the video. Thus, as testimony progresses with time, it becomes less reliable and more susceptible to contamination.

Studies have shown that merely viewing a composite may contaminate memory (Brown, Deffenbacher, & Sturgill, 1977; Comish, 1987; Kempen, 2009; Wells et al., 2005).

In these experiments, participants exposed to a misleading composite performed significantly worse than a control group in the recognition task. Another study demonstrated that even reading a false description of the original target face can negatively influence memory for faces (Loftus & Greene, 1980). In the subsequent recognition session, participants were more likely to select a foil (or incorrect lineup member) which matched the misleading description. Hence, poor quality composites may contaminate our original memory for the target. Our memory for faces is therefore highly susceptible to interference and contamination. This was discussed in the study by Comish (1987) in which lineup fillers resembled errors from the participants' composites. Participants who had constructed a composite were more likely to select a foil containing one of these errors. Thus, if participants are forced to engage in descriptive recall when constructing a composite, features may be misremembered and inserted into the composite-face. This whole-face may then be recoded and misremembered as a veridical representation of the original target.

Criterion shift. One of the theories accounting for composite interference is that of a criterion decision shift, which suggests an increased reluctance to choose anyone from the lineup. Participants are more likely to incorrectly reject the lineup rather than mistakenly select a foil. Thus, when results are analysed, it would appear as though the construction condition has a decrease in hits. The *type* of error made is therefore very important. This was noted in a study in which participants who constructed a composite made 50% misses (incorrect rejections), compared to the control condition which only attained 33% misses (Yu & Geiselman, 1993). Interestingly, however, the construction condition only obtained 10% false alarms (foil selections), compared to the no-treatment control group which selected a filler 30% of the time. This suggests that the construction condition incorrectly selected fillers at rates which were less than those of the control condition. This effect was also evidenced in a study by Kempen (2009) in which the majority (51 out of 63) of incorrect answers were

lineup rejections. When these lineup rejections were analysed, there appeared to be no difference between the control group and the composite construction condition. Indeed, in the study by Wells et al. (2005), 58% of the composite construction condition made no selection, compared to 10% of the control condition. Participants thus appear to become more conservative in their decision criterion; they begin to doubt their perceived confidence against the difficulty of the task (Clare & Lewandowsky, 2004). Participants are more likely to doubt their perceived ability to recall the face and this increases their reluctance to choose a member from the lineup. Thus, if participants perceive the composite they created to be of poor quality and a low resemblance match to the target face, they may doubt their ability to accurately remember the target at all. However, when participants are forced to make a selection, they are able to select the target (Kempen, 2009; Wells et al., 2005). Another study also found that participants who engaged in verbalisation of a target were more reluctant to select a lineup member (Clare & Lewandowsky, 2004, Experiment 1). The same study divided the verbalisation into separate groups: holistic and featural verbalisation. They found that participants who engaged in holistic verbalisation of a target (e.g., rating characteristics and traits) were more likely, than participants who engaged in featural verbalisation (describing facial features), to incorrectly reject the lineup (36.12% vs. 19.2%, respectively). Furthermore, the featural verbalisation group were more likely, than the holistic condition, to accurately select the target (69.2% vs. 57.4%, respectively). However, both verbalisation groups were more likely to correctly reject the lineup in a target-absent condition than the control group (52% for both verbalisation groups, vs. 22.7% for the control group). The authors noted that the holistic verbalisation condition experienced the largest criterion shift. Thus, the more conservative criterion shift worked in favour of the verbalisation conditions with regards to the target-absent lineups. It would appear as though they were correctly rejecting the lineup; however, they were increasing their response threshold for making any

decision. Indeed, in a second experiment by Clare and Lewandowsky (2004) in which the authors removed the “may or may not be present” instruction, there was no difference between the control and verbalisation conditions on proportion of hits, and all groups obtained near ceiling effects (control = 86%, holistic verbalisation = 81%, and featural verbalisation = 84.4%). This suggests that when participants are forced to choose a lineup member, they respond with similar rates to a non-verbalisation condition.

Criterion shift forms a part of signal detection theory (SDT). SDT determines the ability of an individual to distinguish a stimulus/signal from a background of distracting noise or random activity. It explains how individuals manoeuvre and reason their way through decisions of uncertainty. There are four typical stimulus responses: a hit, a miss, a false alarm, and a correct rejection. Each individual has an internal response which determines where their threshold for making a decision will be. The threshold is affected by many psychological factors, such as expectation and experience. The threshold is the deciding point above which an individual will make a decision. When an individual’s internal response is greater than the threshold, the individual will respond “yes”. If the internal response is lower than the criterion threshold, a “no” decision will be made. We can apply this theory to lineup selection, see Figure 1.5 below.

		Suspect	
		Present	Absent
Response	Yes	HIT	FALSE ALARM
	No	MISS	CORRECT REJECTION

Figure 1.5 Signal detection theory diagram, with reference to lineup selection.

We can assume that the police have apprehended a suspect, but we are unsure if the correct perpetrator is present in the lineup. If we select the correct perpetrator, a hit will be made. If the perpetrator is not present in the lineup, and the police have apprehended an innocent suspect, a false alarm will be made. If the correct perpetrator is present in the lineup

and no one is selected, a miss will be made. If the correct perpetrator is not present in the lineup and no decision is made, that decision will be a correct rejection. SDT tries to explain how people attempt to maximise the chance of getting either a hit (identifying the suspect when he is guilty) or a correct rejection (correctly saying the suspect is not there when the lineup contains only innocent members), and decrease the chance of getting a false alarm (identifying a innocent individual) or a miss (rejecting the lineup when the suspect is guilty). If a criterion threshold is set very high, witnesses will only respond yes to very few stimuli. Although this will decrease the chances of obtaining a false alarm, it will also decrease the chance of sending a guilty criminal to prison. If the criterion is set very low, and the witness responds yes to more stimuli, it will increase the chances of apprehending the suspect, but will also increase the chances of sending an innocent individual to jail. This is the decision that has to be made. Is it better to set guilty criminals free at the expense of apprehending an innocent individual? Or, can that sacrifice be made if it means that there are fewer criminals on the streets? Each individual determines their own criterion and their threshold at which they will respond yes or no.

It may be possible that the VOE is explained by a criterion shift, in such a way that witnesses set their criterion threshold very high and are more conservative about their selection; they would thus rather respond no to all stimuli. This will decrease the chance of obtaining a false alarm, but will also decrease the chance of getting hits and increase the chance of a miss (and will look like a decrement in recognition accuracy). This would be evident in target absent lineups. If participants or witnesses set their threshold very low, and respond yes to all stimuli, they will have an increase of hits in target present lineups. However, the correct response for a target absent lineup is to reject the lineup, and if participants respond yes, they will incur the error of obtaining an increase in false alarms. Thus, the criterion shift theory suggests that giving a verbal description of a target (or

constructing a composite) will cause participants to set their threshold higher and will be displayed in the results as a decrease in hits.

Summary. The above section examined four theoretical arguments that may account for decrements in memory. Firstly, the VOE suggests that participants who are required to construct a composite perform poorer than their controls because engaging in verbalisation of features is not a natural task. Focusing on individual features overshadows or masks the image of the original whole face. However, research has shown that although a slight VOE may occur, it may also release after 30 minutes, and thus would not account for studies that utilise a longer delay (Meissner & Brigham, 2001a). Other research surrounding recoding interference suggests that participants who engage in more elaborate descriptions of a target typically perform worse than participants who are instructed to give a free recall of the target (Meissner et al., 2001). Thus, the nature of composite construction may be to elicit and force participants to engage in an elaborate recall of features. This would explain why simpler composite systems with less detail do not find a negative carryover effect from composite construction. The TIPS suggests that interference occurs because the modes of processing at encoding and retrieval are not appropriate. Thus, if participants engage in a featural interference task when the encoding and retrieval mode is holistic or configural, they will perform worse than when the processing mode is consistent. This was evidenced by Macrae and Lewis (2002). The third theory is that of the misinformation effect. Much research has suggested that when participants are exposed to descriptions or lineup fillers containing errors, participants will be more likely to respond in favour of the misinformation. Furthermore, as previously discussed, composites tend to be of poor quality. Therefore, if a poor quality composite is constructed, then recoded holistically, this will be recalled during the retrieval phase. The fourth theoretical account discussed criterion shift, particularly with regards to confidence. Participants who are required to verbalise a description, or construct a

composite, are more likely to incorrectly reject the lineup. However, these responses are typically coded as incorrect responses. Research has indicated that when the forced-choice responses are analysed, participants in the verbalisation conditions fare no worse than their control counterparts. Interestingly, Clare and Lewandowsky (2004) found that participants who engaged in holistic verbalisation were most likely to reject the lineup; they became more conservative and increased their criterion threshold. If participants perceive their composites to be of poor quality, they will also misjudge their ability to accurately select the target, and will instead reject the lineup.

Encoding, Interference, and Recognition Failure

The question arises, if all composite systems generate similar results, does the problem of poor identification lie within the construction itself? Davies et al. (2000) suggest that mechanical and computerised systems are unable to generate an accurate composite from memory. They further conclude that more recent, computerised systems only fare better than mechanical techniques when the composite is constructed with the aid of a target photograph that can be referenced. Contemporary systems have the potential to create better quality facial likenesses. Participants are only able to construct a good quality composite with the target in view.

Previous research suggests that featural construction impairs later memory. However, I propose an inverse speculation: Featural construction may increase recognition accuracy. A recent study which compared PRO-fit (a computerised program) and EvoFIT found the latter produced more recognisable composites (4.2% and 24.5%, respectively) (Frowd et al., 2010). The authors had two distinct groups within the EvoFIT constructions: a blur group, in which the external features (hair, face shape, ears, and neck) were blurred; and a non-blur group, in which participants constructed standard composites. Blurring of the external features allowed the participants to focus specifically on the internal (eyes, eyebrows, nose, and mouth)

features of the face. Interestingly, the study found increased naming accuracy for composites that had blurred external features. This may suggest that participants are able to construct accurate composites when focusing on features of the face. If accurate composites increase recognition (Holland et al., 1994, as cited in Maskow et al., 2007), featural construction, which increases composite accuracy, may lead to superior recognition. Thus, it may be the featural references from a target that increase composite quality. Davies et al. (2000) state that “participants fail to produce convincing likenesses from memory, but with the face present, they can analyze it in terms of features and reproduce it accurately” (p. 123). The study by Frowd et al. (2010) confirms composite constructions that focus on internal facial features, and “blur” the external, configural parts of the face, produce better quality composites. Therefore, composite quality and recognition seem to fail at the construction phase where participants are required to retrieve a face from memory.

While encoding and identification phases are distinct stages on their own, the construction phase acts as a post-encoding and a pre-recognition stage. Tulving (1972) (as cited in Groome, 2006) suggested that the ease with which we could retrieve a memory trace is dependent on the amount of feature overlap present at encoding and retrieval. Retrieving a memory of the target during construction is a process of encoding *and* recognition.

Participants re-encode the composite they have created. As composites created from memory are of poor quality, later recognition could fail due to limited feature overlap. Construction also serves as an identification stage as participants are required to recall and recognise target features. Tulving’s ESP states that recognition is superior to recall because the amount of feature overlap in recognition tests is greater than that in recall tests. Recall tests require participants to spontaneously generate the features from memory. Features are likely to be misremembered when they are recalled during construction. This results in poor quality

composites, which are re-encoded. This may further result in poor recognition accuracy as participants remember a more-recently encoded lookalike.

The problem with poor identifications and constructions, therefore, may stem from how we encode faces and the cognitive processes we undergo during the construction phase. Do people construct faces *only* featural or *only* holistically? Or do these processes work in concert during encoding, reconstruction and recognition?

Rationale for Research

Are composite systems really only featural (such as Identikit and FACES) or only holistic (such as ID)? My suggestion is that they are not just one or the other. They are a combination of two strategies that work together in unison. “Featural” techniques, namely mechanical and computerised systems, require parts to be selected in a piecemeal fashion. However, judgments are made regarding the placement, location, and size of features, making it a configural process. Likewise, in eigenface programs, participants select the faces that best resemble the target. This technique may utilise some aspects of featural construction. In EvoFIT, for example, the random generation of faces is initially based on choosing a set of external features, then a set of internal features. The program generates a selection of faces; participants then select the best overall likenesses, which are then bred together. However, I suggest that judging a synthetic face for its likeness requires some sort of feature evaluation. For example, one face may bear a resemblance to the target because he has bushy eyebrows. Another may be similar in appearance because of his full lips. These “parent” faces are evolved to create combinations that begin to resemble the target. This requires feature differentiation and selection.

The evidence discussed earlier indicates that:

1. Composites tend to be of poor quality;

2. Faces containing misinformation (e.g., poor quality composites) can contaminate memory (Holland et al., 1994, as cited in Maskow et al., 2007);
3. Thus, more accurate composites may lead to increased recognition;
4. It was previously suggested that poor quality composites may be due to systems engaging in unnatural featural parsing;
5. However, “featural” composite systems tend to utilise some aspects of configural and holistic construction, by ordering and placing features;
6. Likewise, “holistic” systems, as suggested above, may engage in featural differentiation and selection;
7. Composites constructed from memory are of worse quality than those constructed with the target face in view;
8. It is suggested that in-view constructions have reference points that can be referred to during construction;
9. Blurring external features and focusing on internal features (Frowd et al., 2010) leads to increased composite recognition;
10. I propose that it may be that featural referencing produces better quality composites;
11. If featural referencing produces better quality composites, it may increase recognition accuracy;
12. Holistic verbalisation increases lineup rejection (Clare and Lewandowsky, 2004);
13. Thus, I suggest that configural and holistic processes could be resulting in poor quality composites and decreased recognition.

What impact does holistic construction have on later recognition? Although it may be difficult to conceptualise such a system, it is possible to conduct a more refined featural construction. This study proposes that featural construction can be achieved by eliminating

feature placement in relation to other parts of a face. Participants would construct a composite by selecting individual features without seeing them in the context of a whole face. Each feature would be saved before moving on to the next feature subset. Examining the differences in recognition accuracy between standard composite constructors and featural constructors may demonstrate the effect holistic construction has on later memory.

While other studies have documented the effects of featural- and configural-orientation on later recognition, no studies have examined the effects of an intervening construction task (which may act as contaminating post-event information), divided into separate processing channels, on later recognition. If isolated featural construction induces superior recognition to featural-holistic construction, this may have implications for composite construction techniques in law enforcement agencies. The cognitive processes in composite construction need to be understood. If, as suggested, focusing on inner features increases recognition, featural construction may increase recognition accuracy. Creating a whole face in context exposes you to the entire misremembered composite. This incorrect face may be recalled during the recognition task. It may be that the configural and holistic aspects of a face are contaminating our memory.

Specific Aims and Hypotheses

As discussed above, composites are of poor quality, and constructing them may involve featural and configural processes. However, if focusing on internal features increases composite recognition, the contaminating culprit may be the holistic and configural nature of construction. Featural construction is a distinctly different process to the holistic encoding of a face. For this reason, it is hypothesised that featural construction (of a composite) may allow for superior recognition, as configuring parts into a whole face may disrupt our memory; holding two separate faces in memory may result in remembering some parts of the one, and certain aspects of the other. By constructing a face featurally, it may be difficult to

parse the individual features together in one's head; thus, no whole face will be interfering with the original face trace. As no previous studies have investigated this particular method of featural construction, it is hypothesised that participants who construct these featural composites will show greater recognition accuracy in the whole face lineups. When these results are obtained, they can be tentatively subtracted from the standard composite constructors' responses. This may indicate the extent of isolated configural contamination.

In the recognition stage, participants will be shown either a whole-face lineup, or separate lineups of individual features. It is hypothesised that participants who construct a featural composite will also perform better on the featural lineup. This may suggest how faces are encoded and processed; increased accuracy on the isolated feature lineup could suggest that faces are encoded featurally. Accurate responses and the reaction times will indicate more saliently-encoded features of the face. Results of a post hoc study may indicate that featural composites, as opposed to standard composites, bear greater resemblance to the original target.

CHAPTER TWO:

EXPERIMENT 1

Method

If feature selection is an inherent component of composite systems, and people process faces in a configural and holistic way, how accurate are people at correctly identifying isolated features? According to the encoding specificity principle, items will be better recalled if the mode of processing at recognition matches that at encoding. However, if participants engage in a featural construction task, they may perform better at a featural recognition task, compared with participants who construct a composite in a more configural way. Previous research has suggested that people are unable to select features in isolation (Searcy & Barlett, 1996; Tanaka & Farah, 1993). If composite production *does* result in decreased hits, then the control group should accurately select the target at higher rates. It was necessary to see if participants in a control group were able to accurately select the target features, compared to a standard whole-face lineup. There was a suspicion that the featural lineups may be too difficult for participants, so two control groups were used (one for whole-face lineups, and one for featural lineups). If participants (who did not engage in any composite production) were able to select the target features at adequate levels, those lineups would be used for the composite construction experiment. This experiment aimed to compare recognition accuracy between whole-face lineups and isolated-featural lineups. This experiment acted as a pilot study for the featural lineups. If a control condition was able to accurately select isolated features from a lineup, it would then be used in Experiment 2A.

Participants

A total of 36 participants was used in this study. Participants were undergraduate Psychology students between the ages of 18 and 22 years ($M = 19.97$, $SD = 1.03$). Students participated in exchange for course credit. Ten (27.78%) students were male and 26

(72.22%) students were female. As the stimuli used were of White, female faces, the race distribution of participants was also recorded. A total of 23 (63.89%) participants identified themselves as “White”, and 13 (36.11%) participants identified themselves as “Black” or “Coloured”. Testing took place in a computer laboratory in the Psychology Department at the University of Cape Town.

Materials

Target stimuli. Two target images were used in this experiment. The images were of White females. The targets were both greyscale images in the frontal positions and were standardised to 7.62 cm in width and 11.36 cm in height, with a resolution of 450 x 671 pixels (8-bit). As the lineups also contained fillers (and targets) in frontal position, the targets at encoding were flipped horizontally (along a vertical axis) to create some difference between images at encoding and recognition.

Presentation and answering materials. The encoding and recognition task were both presented using E-Prime 2.0. This enabled participants’ reaction times, as well as their responses, to be recorded. Participants were required to indicate their answers using the number keypad. The entire task was computer-based. This attempted to minimise experimenter effects and to standardise the experimental procedure. The slideshows were presented on a computer screen with a resolution of 1024 x 768 pixels.

Lineup construction, whole-face lineups. Lineups were constructed following the guidelines set out by Malpass, Tredoux, and McQuiston-Surrett (2007). After a five-second exposure time, 10 independent observers wrote down a verbal description of all the targets from memory. Once all the descriptions were collected, a modal description was created using the descriptors that appeared most often. If more than five participants described the same feature, it was used in the final, modal, target description. A contact sheet was created which consisted of thumbnails of hundreds of White, female, colour photographs. The

thumbnails were presented in rows and columns of four so that there were 16 thumbnails per page. The thumbnails were all standardised to 263 pixels in width, and 394 pixels in height. The modal description and contact sheet was then distributed to 10 different participants who were required to select five fillers that best matched the description. The fillers that were chosen most often were used in the final lineups. This was repeated for the second target.

The target and filler photographs were all edited in such a way that a black shirt covered all of their clothes. This was done in order to eliminate the possibility that participants may recognise the target due to the clothes that the target was wearing. The target and filler faces did not have any distinguishable and identifiable marks or features, such as tattoos, piercings, or scars. Accurate selections should be based on person identification, as opposed to attire identification. A study by Lindsay, Wallbridge, and Drennan (1987) found that participants were more likely to select a suspect if the lineup was biased to contain the same attire at encoding. Distinctive blemishes and jewellery were all edited and removed from the photographs. The backgrounds were all edited to have a grey, shadowed, stippled wall background. All of the images were in colour. Each lineup image was standardised to 6.49 cm in width and 9.74 cm in height, with a resolution of 189 x 283 pixels (8-bit). The images were all standardised to have the same inter-ocular distance (38 pixels between the pupils of each lineup member), and were all standardised to be displayed on the same horizontal plane (see Figure 2.1).

Position effects were balanced by using two positions. The target was placed in either position 3 or position 5. The target was always present in the lineup (see Appendix A for whole face lineup). Target absent lineups were not used in this experiment.¹ The featural lineup condition had to view six lineups. Adding more lineups (in the form of target absent

¹ Depending on the results of this experiment, it would be useful to investigate the results of a target absent lineup in future research.

lineups) might have made them more confused and susceptible to guessing. This will be addressed in more detail in the Discussion section.



Figure 2.1. Whole face lineup, with gridlines (not shown in experiment), inter-ocular eye distance standardised. This was done for the other targets, and for the featural lineups.

Lineup construction, feature lineups. The features of the targets and fillers were used from the whole-face condition. The individual features from each face were cut out, making a lineup for each feature. Thus, six lineups were created from six features: eyes, nose, mouth, hair, eyebrows, and chin/jaw. The target feature was either in position 3 or 5 and this was counterbalanced across participants (see Appendix B for all feature lineups).

Procedure

Encoding phase, both groups. Participants were randomly assigned into conditions. They were seated in the computer laboratory and were unable to view each other's

computers. Each workstation contained a computer and a consent form. Participants were required to complete the consent form pertaining to their demographic information, confidentiality request, and permission to take part in the study (see Appendix C).

The E-Prime slideshow was opened. A slide alerted the participant that they were going to be exposed to a face. After a five second exposure, participants were required to rate the previously seen face on various traits: warm, attractive, aggression, distinctive, intelligent, kindness, foolish, humourous, studious, and likeable. This was similar to the trait-rating task that Wells et al. (2005) used to facilitate holistic, deeper encoding. Participants were required to indicate their answer on the number keypad, from 0 (not at all) to 9 (extremely). After the target had been rated, participants were required to play a distracter game for ten minutes. The distracter task, called "QWOP" is a keyboard-based game in which participants have to manoeuvre an athlete's legs using the letters "Q", "W", "O", and "P". This game can be freely downloaded (www.foddy.net/Athletics.html). The distracter task was used, firstly to induce a delay after encoding, and also to prevent deliberate and mindful rehearsal of the target.

Recognition phase, whole-face group only. The recognition task was administered 10 minutes later. An E-Prime slideshow was opened. Instructions notified the participant that their task was to select the target they had seen in the earlier slideshow, from a lineup of six faces. A following instruction warned that the target face may or may not be present. A simultaneous lineup was then displayed with all six faces in frontal view. Participants were given as much time as they needed to make their selection. Once they had decided, participants had to indicate their selection on the number keypad from "1-6", which corresponded to numbers above members of the lineup. If they thought the target was not present, participants had to select "0".

Featural group only. Participants in the featural group were required to select the target feature from a lineup of filler features. An instruction notified the participant that they would be required to view six different feature lineups, and to select the correct target feature. The target feature was always present. The order in which the feature lineups were shown was randomised. After completing their recognition task, all participants were debriefed and told the purpose of the study.

Results

This experiment investigated whether participants were able to correctly select individual features from featural lineups, compared to a whole-face standard lineup. Participants in the featural lineup group had to select the correct target feature from a series of six lineups, corresponding to six features, i.e., eyebrows, eyes, hair, jaw, mouth, and nose. Conversely, participants in the whole-face lineup condition had to select the whole target face from the lineup. Participants were significantly more likely to correctly select the target in a whole-face lineup than from a featural lineup. This is discussed below.

Contingency Testing

The frequencies of correct responses (hits) and incorrect responses (misses) are shown in Table 2.1.

Table 2.1

Percentage Correct and Incorrect Choices in Lineup Task

Lineup Type	Response		Total
	Correct	Incorrect	
Eyebrows	18.75% (3)	81.25% (13)	100% (16)
Eyes	18.75% (3)	81.25% (13)	100% (16)
Hair	12.50% (2)	87.50% (14)	100% (16)
Jaw	18.75% (3)	81.25% (13)	100% (16)
Mouth	31.25% (5)	68.75% (11)	100% (16)
Nose	18.75% (3)	81.25% (13)	100% (16)
Whole Face	81.25% (13)	18.75% (3)	100% (16)

Note. Numbers in parentheses are frequency counts.

Each participant in the featural lineup group had to make selections from six featural lineups. All the separate responses per feature are shown in Table 2.1. As can be seen, the featural group performed poorly, and obtained floor effects. Contrastingly, the whole-face group achieved near ceiling effects (81.25% accuracy). Interestingly, the hair lineups obtained the lowest correct responses, 12.50%. The mouth lineups achieved the highest

correct responses, 31.25%. This is interesting as it has been suggested that hair is one of the most salient and recognisable features of a face, followed closely by forehead/hair, eyes, mouth, chin, and nose (Davies, Ellis, and Shepherd, 1977). Hair and eyes obtained the lowest and second lowest responses, respectively. There was no significant difference between the selections across the featural lineups, $\chi^2 (h, N = 96) = 1.90, p < 0.86, \text{Cramer's } V = 0.14$.

Figure 2.2 displays the averaged featural correct responses and the whole-face selections.

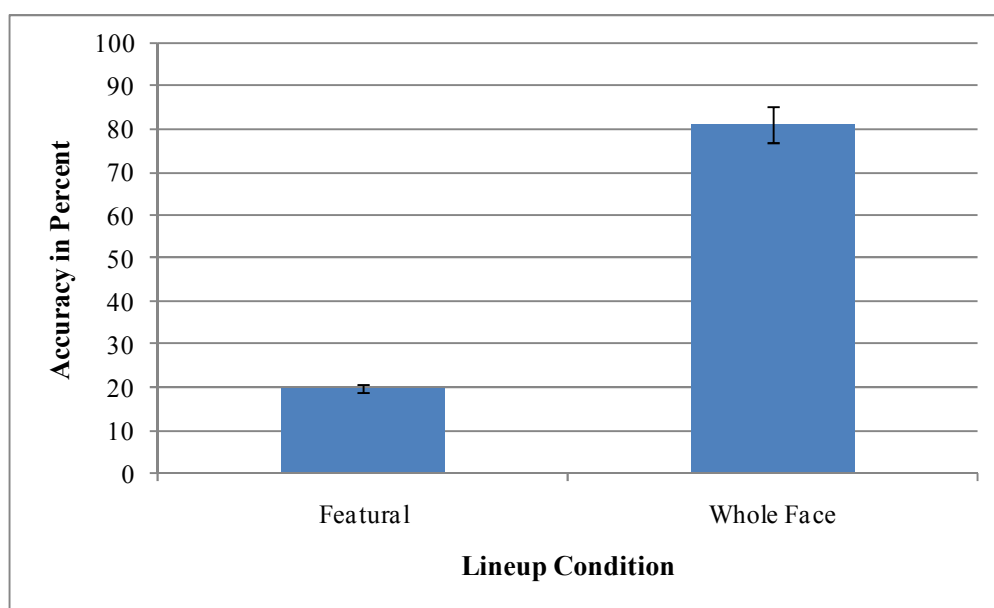


Figure 2.2. Percentage accuracy choices, per group, in the lineup task. I-bars are 95% confidence intervals.

The average correct response for the featural groups was 19.79%, which differed significantly from the whole-face group, $\chi^2 (1, N = 112) = 25.38, p < 0.01, \text{Cramer's } V = 0.48$. An odds-ratio conducted between these two groups suggested that the whole-face lineup group were 17.56 times more likely to correctly select the target from the lineup than the featural group were at selecting a correct feature from any one lineup.

Odds Ratios

None of the featural lineups showed any superiority effect over each other. Participants appeared equally likely to perform poorly in all of the featural lineups. The odds-ratios between the feature lineups and the whole-face lineup are displayed in Table 2.2 below.

Table 2.2

Odds Ratios for Accurate and Inaccurate Selections Between Lineups

	Eyebrows	Eyes	Hair	Jaw	Mouth	Nose	Whole Face
Eyebrows	-						
Eyes	1.00	-					
Hair	1.62	1.62	-				
Jaw	1.00	1.00	0.62	-			
Mouth	0.51	0.51	0.31	0.51	-		
Nose	1.00	1.00	0.62	1.00	1.97	-	
Whole Face	0.05	0.05	0.03	0.05	0.10	0.05	-

The vertical column should be read first, in Table 2.2. For example, the odds-ratio between eyebrows and eyes is 1. This suggests that these two featural lineups have an equal chance of achieving the same correct responses. Conversely, the eyebrows and whole-face groups have a ratio of 0.05. This suggests the eyebrow group are 0.05 times more likely than the whole-face group to correctly select the target from the lineup. Although the contingency test done earlier between the feature groups indicated no significant difference, the odds-ratio table suggests that participants in the mouth group were almost three times more likely to select the correct answer than participants in the hair group.

Reaction Times

The reaction times for all selection responses were recorded and averaged across participants. The times (in seconds) are displayed in Table 2.3 below.

Table 2.3

Mean Reaction Times and Standard Deviations (in Seconds) for All Lineup Responses

Lineup Type	Reaction Time
Eyebrows	11.16 (5.99)
Eyes	14.49 (8.52)
Hair	10.36 (6.68)
Jaw	10.72 (5.39)
Mouth	10.72 (10.03)
Nose	12.00 (13.21)
Whole Face	9.99 (5.01)

Note. Numbers in parentheses are standard deviations.

The whole-face group were the quickest to respond in the lineup task, making their selections in an average of 9.99 seconds. A t-test conducted between the average featural lineup reaction times ($M = 11.58$, $SD = 8.61$) and the whole-face lineup group ($M = 9.99$, $SD = 5.01$) also yielded no significant differences in reaction times, $t(110) = 0.83$, $p = 0.41$, Cohen's $d = 0.23$. Although the whole-face group performed better in the lineup task than the featural lineup groups, they were not quicker to make their responses. This result is displayed graphically in the box-and-whisker plot in Figure 2.3 below.

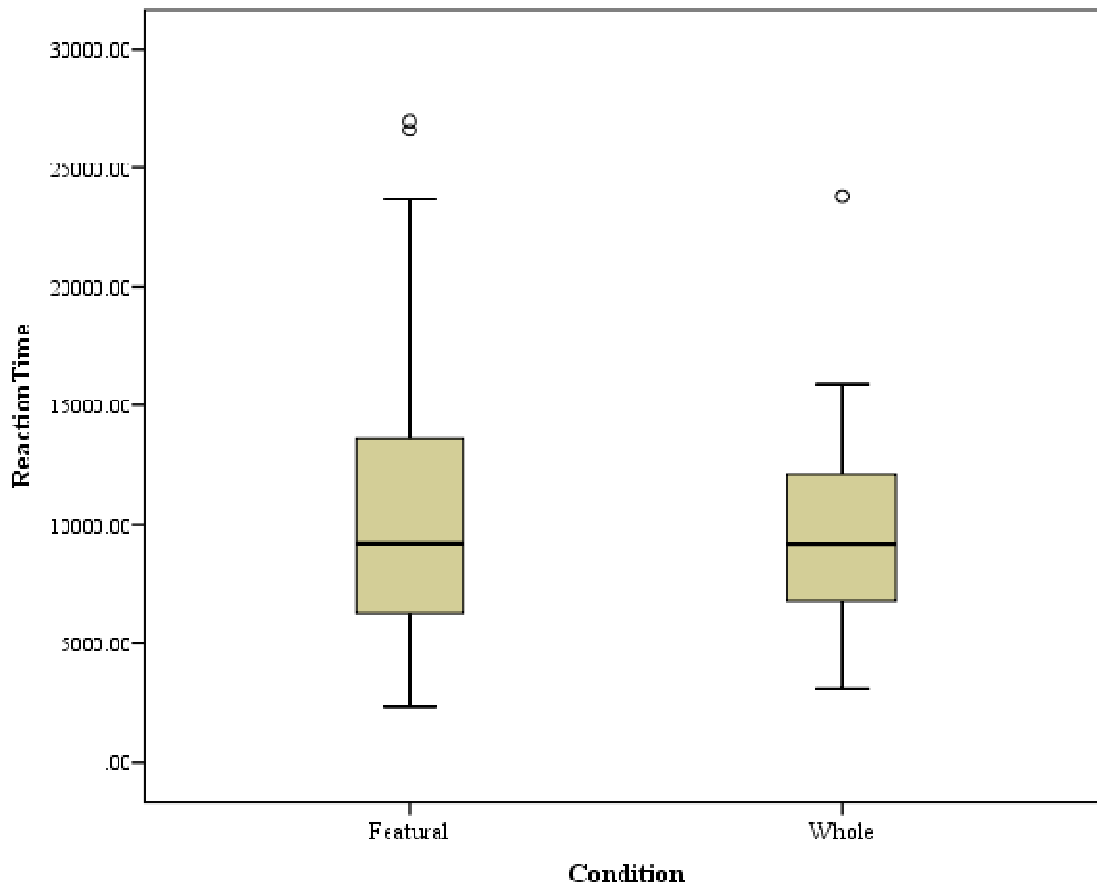


Figure 2.3. Mean reaction times per group, in the lineup task. O-circles are outliers.

Reaction Times of Correct and Incorrect Responses

The reaction times for correct and incorrect selection responses were recorded and averaged across participants. The average times and standard deviations (in seconds) are displayed in Table 2.4 below.

Table 2.4

Mean Reaction Times and Standard Deviations (in Seconds) for Correct and Incorrect Responses Across Lineup Type

Lineup Type	Response	
	Correct	Incorrect
Whole-Face	8.54 (3.20)	16.29 (7.33)
Featural	14.30 (12.43)	11.22 (7.35)

Note. Numbers in parentheses are standard deviations.

The total reaction times were divided into correct and incorrect selections. As can be seen in Table 2.4, the whole-face group ($M = 8.54$, $SD = 3.20$) responded almost twice as quickly for correct responses than for incorrect responses ($M = 16.29$, $SD = 7.33$). An independent t-test found a significant difference between the means, $t(14) = 2.98$, $p = 0.01$, Cohen's $d = -1.37$. There was no significant difference between the means of correct featural responses ($M = 14.30$, $SD = 12.43$) and incorrect featural responses ($M = 11.22$, $SD = 7.35$), $t(94) = -1.41$, $p = 0.16$, Cohen's $d = -0.30$. Participants in the whole-face group responded more quickly when making accurate selections, than when making inaccurate selections. Conversely, the featural group responded in the same amount of time for correct and incorrect selections.

Discussion

This is the first experiment to date that has used isolated featural lineups. The study by Tanaka and Farah (1993) had participants select features out of only two possible options. The features were line drawings and only the eyes, nose, and mouth were used. The eyes also included the eyebrows, which contained configural and holistic information. For example, the shape of eyebrows is able to give expression to a face. This experiment aimed to investigate whether participants are able to remember and select the correct features of a face. Construction of a composite is very difficult; if participants are unable to select features from a lineup (a six-option, multiple-choice task), then they may not be able to find an accurate feature out of hundreds within the composite systems.

The initial results of this experiment indicate that participants are unable to correctly select isolated features from a lineup. There was a significant difference between the average correct feature selections (19.79%) and the whole-face selections (81.25%). The odds-ratio indicated that participants are 17.56 times more likely to accurately select a whole face than a feature, 95% CI [4.54, 67.88]. If chance is 16.67% (100% divided by six lineup options), the features were chosen barely above rates of random guessing.

The features were all selected at equal rates. This result goes against other literature which found hair and eyes to be the most salient and recognised features (Comish, 1987; Yarmey, 2004). Indeed, Yarmey (2004) suggested that hair and eyes were the most important cues in face recognition, while Comish (1987) suggested that eyebrows and lips were of less importance. However, in this experiment, participants were only able to select these features at chance levels (a floor effect). No feature superiority effect was seen. It is possible that participants encoded the target face holistically and then had to rate the face on a variety of traits and characteristics. These results therefore may offer some evidence towards the TIPS and ESP theories. Since the face was encoded semantically, there are not many featural

overlaps and cues available at the featural, or structural, retrieval stage. Winograd (1976) found that faces that were rated on semantic characteristics were identified more often than faces that were encoded on structural characteristics. All participants in this experiment had to rate the target on a variety of traits. This deep semantic encoding may have resulted in poor recognition results on the featural lineups.

Evidence for criterion shift could not be detected as participants were not given a “not present” option, or a target-absent lineup. It was hypothesised that this task would be very difficult for participants in the featural condition; thus, by giving them a “not present” option, it would have masked all the results as being incorrect. Since participants were given six lineups (for six features), if all lineups had a corresponding target-absent equivalent and “not present” option, participants may have started utilising a cognitive strategy in which they always reject the first lineup because they know they will get a second chance to make a selection in the “forced-to-choose” lineup. As it turned out, participants were choosing at chance rates from the lineup. Inserting another target-absent lineup would have resulted in the task being even more difficult, with participants now having to select one pair of eyes from 12 pairs.

There was no significant difference between the overall selection reaction times between participants. Although the whole-face group performed better than the featural group, they were both making selections at equal rates. There are two possible reasons for this. It is possible that participants in the whole-face group automatically knew who the target was and responded quickly. Likewise, in the featural group, because the task was difficult, participants resorted to random selection. However, the average overall reaction times for the whole-face and featural groups are 9.99 and 11.58 seconds, respectively.

A t-test conducted between the two reactions times found an effect size of 0.23. Dunning and Perretta (2002) suggest that a time boundary of 10 to 12 seconds is able to best

distinguish between accurate and inaccurate lineup responses. They suggest that this time limit delineates participants' automatic responses. Participants selecting at greater reaction times tend to exhibit more controlled and effortful decision-making. Both the whole-face and featural group fell within this accuracy bracket. Sporer (1993) suggests that accurate responses are made more quickly than incorrect responses. There was no significant difference in accuracy response time between the whole-face and featural groups.

Interestingly, when the correct and incorrect responses were analysed across the reaction times, the whole-face group responded twice as quickly when making correct selections, than when making incorrect selections. This fits with the research by Sporer. However, participants in the featural group did not differ significantly in response time between accurate and inaccurate selections. Both responses were selected at equal times. The featural groups were only able to accurately select the features at an average rate of 19.79%. This is almost equal to chance levels of selection (16.67%). Thus, the lack of significant difference between reaction times across correct and incorrect responses may suggest that participants were guessing.

This experiment did not give participants the forced-to-choose instruction when they rejected the lineup. It was not hypothesised that a control group should experience a criterion shift similar to the way in which construction and verbalisation groups experience it.

Furthermore, the featural lineup condition was shown six lineups. The whole face lineup group was shown only one lineup. Having a forced-to-choose instruction would work for a whole face lineup as participants are forced to make one extra decision. However, with the featural lineup group, if that option was present after every lineup, it might result in some different cognitive strategy. Participants may know that they will get a second chance to choose after every featural lineup. For this reason, a target absent lineup was not utilised.

After seeing multiple versions of a "mouth" lineup, participants may realise that they would

get multiple chances with the “eyes” lineup. Although this might have been a limitation of this experiment, it was a trade-off that had to be made.

In summary, participants are not able to correctly detect individual features in a lineup. Composite systems assume that participants are able to remember features, but this experiment indicates that they are not able to. This suggests that participants might not be able to select features out of a composite system. Previous research suggests that isolated features are able to be selected above chance rates (Macho & Leder, 1998; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). However, the result in this experiment indicated otherwise. Previous studies by Tanaka and Farah (1993), and Tanaka and Sengco (1997), only used a two-choice recognition task. Participants in the isolated featural condition achieved 62% and 65% respectively. The featural lineups in this publication may have been more difficult because they used a six-option lineup. The chance of accurately selecting the target by guessing is 16.67%. Furthermore, six different features were used in this experiment, whereas Tanaka and Farah, and Tanaka and Sengco, only used three features (eyes, nose, and mouth). Adding more lineups also increases the chances of obtaining an incorrect answer. Tanaka and Farah found a feature superiority effect in which eyes were recognised most often (80%), compared to the nose (62%) and mouth (63%) "lineups".² This experiment showed no feature superiority; all features were selected at equal rates. It may be that the lineups used in this experiment were too difficult, evidenced by the chance rate of selection, and the similar reaction times between correct and incorrect featural selections.

² Lineups is in inverted commas because Tanaka and Farah (1993) and Tanaka and Sengco (1997) only offered participants a two-choice option.

CHAPTER THREE:

EXPERIMENT 2A

Method

This experiment aimed to investigate the specific processes involved in composite construction, namely the featural and configural components inherent in the production. Is it the act of selecting individual features that causes a decrease in recognition? Or is recognition hampered by the configural, holistic placement of features and re-encoding of a new face? Experiment 1 indicated that participants are unable to accurately select isolated features from a six-option lineup. As the control condition was not able to accurately select the isolated features from the lineup, only the whole face lineups were used in this experiment. It was hypothesised that standard, configural composite construction should impair performance on the recognition task. Faces are encoded using configural, holistic processes; thus, a configural intervening task (such as constructing and re-encoding a poor quality composite) may contaminate memory due to congruent processing modes. Likewise, a composite constructed using a featural method may be less likely to contaminate memory. The whole-face control group in Experiment 1 achieved 81.25% accuracy. These whole-face lineups (plus a third lineup) were used in this experiment. This gave the construction groups in this experiment enough of a range to perform at poorer recognition rates.

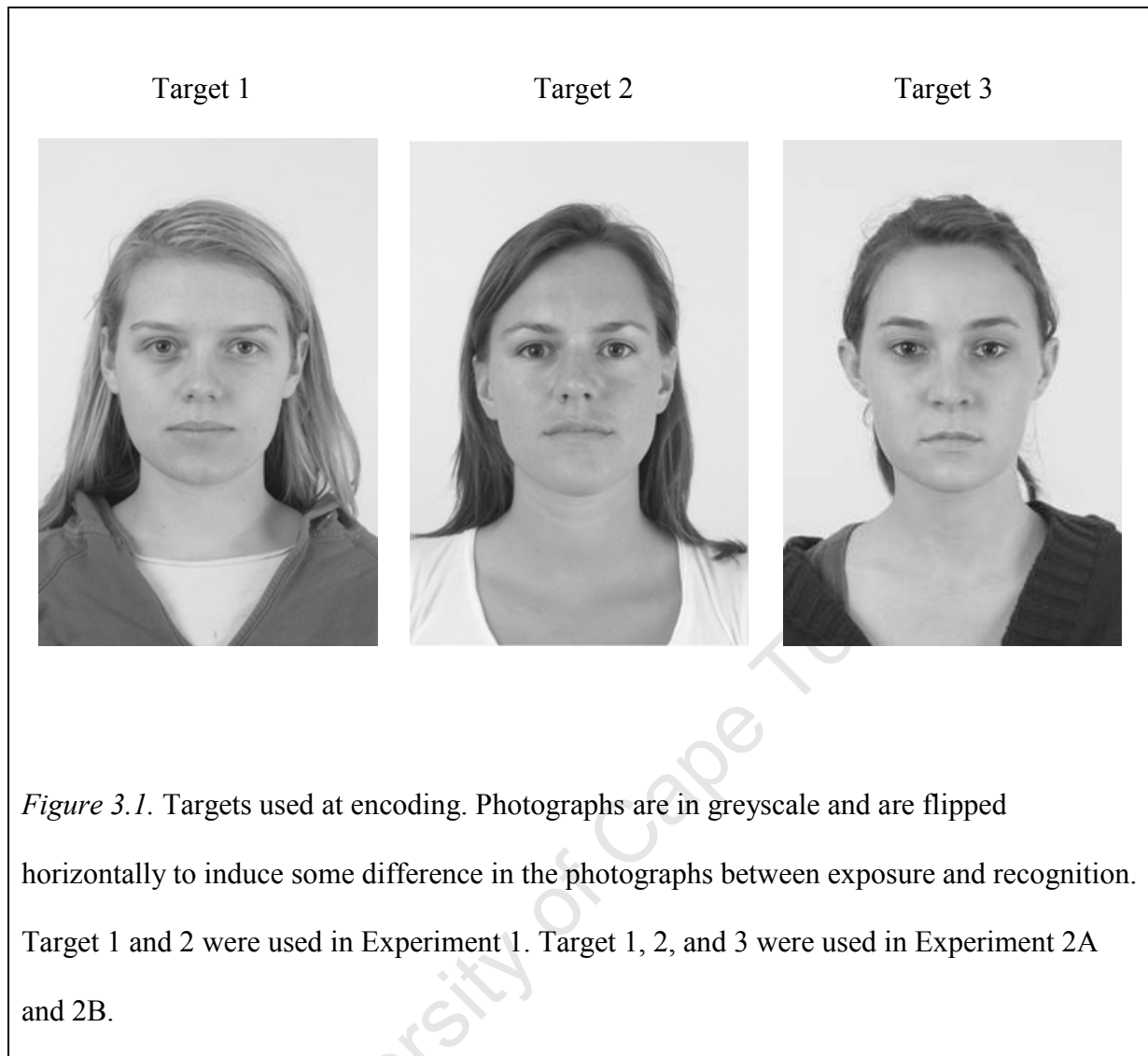
Participants

A total of 144 participants (114 females and 29 males) was used in this study. Participants were undergraduate Psychology students between the ages of 18 and 32 years ($M = 20.38$, $SD = 1.89$). Students participated in exchange for course credit. As the stimuli used were of White, female faces, only White participants were used in this experiment. Testing took place in a computer laboratory in the Psychology Department at the University of Cape Town.

Materials

Target stimuli. Three target faces were used in this experiment (see Figure 3.1). Two of the targets used were from Experiment 1. A third target face was also used. Three targets were used as the use of multiple targets prevents face recognition on the basis of an overlooked distinctive property (Wells et al., 2005). There might be some idiosyncratic or unusual property about the target that may not have been noticed. If only one target was used, that chosen target may have a unique or distinctive property about it. The use of three targets could balance out any difference effects between the targets themselves. The targets were all greyscale images in the frontal positions and were standardised to 7.62 cm in width and 11.36 cm in height, with a resolution of 450 x 671 pixels (8-bit). As the lineups also contained fillers (and targets) in frontal position, the targets at encoding were flipped horizontally (along a vertical axis) to create some difference between images at encoding and recognition.³

³ It is possible that flipping a photograph may remove some cues that are present in the inherent symmetry of the face. This decision was made, however, to avoid photograph recognition if the target was in the exact same pose as at exposure.



Lineup construction. Lineups were constructed following the guidelines set out by Malpass et al. (2007). Fillers for the lineup were selected in the same way as in Experiment 1. After a five second exposure time, 10 independent observers wrote down a verbal description of all the targets from memory. Thereafter, a modal description was created by using the descriptors most frequently reported. If more than five participants described the same feature, it was used in the final, modal, target description. A contact sheet was created which consisted of thumbnails of hundreds of White, female photographs. The thumbnails were presented in rows and columns of four so that there were 16 thumbnails per page. The thumbnails were all standardised to 263 pixels in width, and 394 pixels in height. The modal

description and contact sheet was then distributed to 10 different participants who were required to select 11 fillers that best matched the description (see Table 3.1 below for the modal descriptions).

Table 3.1

Modal descriptions for each target

Target	Modal Description
1	Female, blonde hair, blue eyes, rounded jaw, fair skin
2	Female, brown hair, brown eyes, rounded jaw, round nose
3	Female, light brown hair tied back, brown eyes, thinnish/small lips, fair skin, oval-shaped face

Note. The modal descriptions for target 1 and target 2 were used in Experiment 1.

Eleven fillers were selected because five fillers were necessary for the target present lineup, and six fillers were required for the target absent lineup. The fillers that were chosen most often were used in the final lineups. The filler that was chosen at the highest frequency was used as a “suspect” in the target absent lineups. This procedure was repeated for the other two targets.

The target and filler photographs were all edited in such a way that a black shirt covered all of their clothes. This was done in order to eliminate the possibility that participants may recognise the target due to the clothes the target was wearing. The target and filler faces did not have any distinguishable and identifiable marks or features, such as tattoos, piercings, or scars. Distinctive blemishes and jewellery were all edited and removed from the photographs. The backgrounds were all edited to have a grey, shadowed, stippled wall background. All of the images were in colour. Each lineup image was standardised to 6.49 cm in width and 9.74 cm in height, with a resolution of 189 x 283 pixels (8-bit).

Position effects were balanced by using two positions. The target was placed in either position 3 or position 5 within the target present and target absent lineups. The target position was counterbalanced across conditions. This was done to ensure that photograph placement did not inadvertently affect participant response. Participants saw both target present and

target absent lineups, but the order in which they saw them (e.g., either target present first, followed by target absent, or vice-versa) was also counterbalanced across conditions.

Presentation and answering materials. The encoding and recognition task were both presented using E-Prime 2.0. This enabled participants' reaction times, as well as their responses, to be recorded. Participants were required to indicate their answers using the number keypad from "1-6". If they thought the target was not there during the recognition task, they had to indicate "0" on the number keypad. A second instruction forced them to make a selection. As this experiment aimed to investigate the theories underpinning diminished recognition, questions were asked regarding why participants rejected the lineup. Finding reasoning behind lineup rejection might shed some light on the criterion shift theory. Thus, another slide asked participants for a reason for their selection. If participants had selected one of the lineup members, they pressed "N" on the keyboard. If they rejected the lineup, they had to answer "A", "B", "C", or "D". Participants had to indicate the letter on the keyboard corresponding to their reason for their lineup rejection decision:

- (N) - I did select a person from the lineup
- (A) - The original person was definitely not there
- (B) - I cannot remember what the original person looked like
- (C) - I could not decide between multiple members of the lineup
- (D) - I had a slight idea of who the person was, but I was not very confident in my decision

If participants selected "A", "B", "C", or "D", the slideshow automatically took them to a forced-choice lineup where they were required to make a selection.

The slideshows were presented on a computer screen with a resolution of 1024 x 768 pixels.

Composite construction software. Participants in the composite production conditions were instructed in constructing a face using FACES 4.0: The Ultimate Composite

Picture (Cote, 2005). FACES 4.0 contains detailed features and allows for adjustments of size, position, and colour. Composites are created by clicking any one of the facial features on the right of the interface.

Procedure

Encoding phase, all groups. Participants were seated in the computer laboratory and were unable to view each other's computers. Each workstation contained a computer and a consent form (see Appendix C). Participants were required to complete the consent form pertaining to their demographic information, confidentiality request, and permission to take part in the study.

A graphical representation of the procedure for the encoding and recognition sessions across groups is displayed at the end of this procedure.

The E-Prime slideshow was opened. A slide warned the participant that they were going to be exposed to a face. After a five second exposure, participants were required to rate the previously seen face on various traits: warm, attractive, aggression, distinctive, intelligent, kindness, foolish, humourous, studious, and likeable. Participants were required to indicate their answer on the number keypad, from 0 (not at all) to 9 (extremely). After the target had been rated, participants were required to play a distracter game for 10 minutes. This was the same distracter task used in Experiment 1. The control group was then thanked, dismissed from the study, and reminded of their second session in two days' time.

Composite production phase, whole-face construction group only. A breakdown of the different methods of construction is displayed in Appendix D. After target exposure and rating, the whole-face composite construction group was instructed in the use of FACES 4.0. A slideshow instructed them how to construct a face in the "standard" way – by assembling all of the features together in the whole face. They were given as much time as

was needed to construct a composite that best resembled the target. Once they had finished and saved their composite, participants were thanked and dismissed from the study.

Composite production phase, isolated-featural construction group only. The isolated-featural group was instructed on how to construct a composite using a piecemeal, featural method. A slideshow instructed them to select one feature, save the feature as “*participant number + feature*”, and then close the window. They were required to do this with all the other features they thought were present in the face. Thus, the isolated-featural group never saw all the features assembled together within the composite face-space. After they had selected a particular feature, they were required to indicate the number of the feature selected on a sheet provided (see Appendix E). The number appears near the top-left of the feature (see Appendix F for an example of the number reference).

Composite production phase, featural-assembly construction group only. The featural-assembly group followed the same procedure as the isolated-featural construction group. However, when participants had selected each separate feature and indicated each feature on the sheet provided, they were then required to assemble only *those* features back into a whole face. Thus, they had to reference each feature number and select only the features they had previously selected. Participants had to configure and spatially arrange all the features into a whole face. This was necessary as composite quality was to be tested in a later post hoc experiment. It was done so that the different types of construction methods (whole-face versus featural) could be compared against each other.

All participants were then thanked, dismissed from the study, and reminded of their second session in two days' time. The first testing session took approximately one hour to complete.

Second session, recognition phase, all groups. All participants returned two days later for the recognition task. An E-Prime slideshow was opened. Instructions notified the

participant that their task was to select the target they had seen in the previous session, from a lineup of six faces. A further instruction stressed that the target face may or may not be present. A simultaneous lineup was then displayed with all six faces in frontal view. Participants were given as much time as they needed to make their selection. Participants had to indicate their selection on the number keypad from “1-6”, which corresponded to numbers above members of the lineup. If they thought the target was not present, they had to select “0”. Participants then had to indicate a reason for their decision. If participants did not make a selection in the lineup (and thought the target was not present), the lineup was displayed again and participants were forced to make a selection. This procedure was repeated for the target absent lineups. The order of target present and target absent lineups was counterbalanced.

Meta-cognitive questions were administered to all groups at the end of each lineup task, regarding participants' selection confidence, the difficulty of the task, and the similarity of the lineup members. For example, it would be unclear if participants are accurately selecting the target because they remember what the target looked like or if the lineup task was too easy and there was no similarity between the lineup members. Sporer (1993) suggests that meta-cognitive questions help to examine aspects of participants' decision-making processes. A meta-analysis found that the confidence-accuracy correlation was higher for participants who made an accurate selection than participants who made an inaccurate selection (Sporer, Penrod, Read, & Cutler, 1995). Sporer further suggests that confidence is a good predictor of accuracy amongst choosers (participants who make a selection). Furthermore, it is useful to study confidence levels if investigating a criterion-shift due to composite construction. If composite production results in a decrease in recognition, is this due to a more conservative and uncertain selection? Participants answered the meta-cognitive questions using the number keypad on the keyboard, on a scale from 0 (not at all

confident/difficult/similar) to 9 (extremely confident/difficult/similar). When the lineup task was completed, participants were debriefed and told the purpose of the study.

The procedure of Experiment 2A is displayed on the next page.

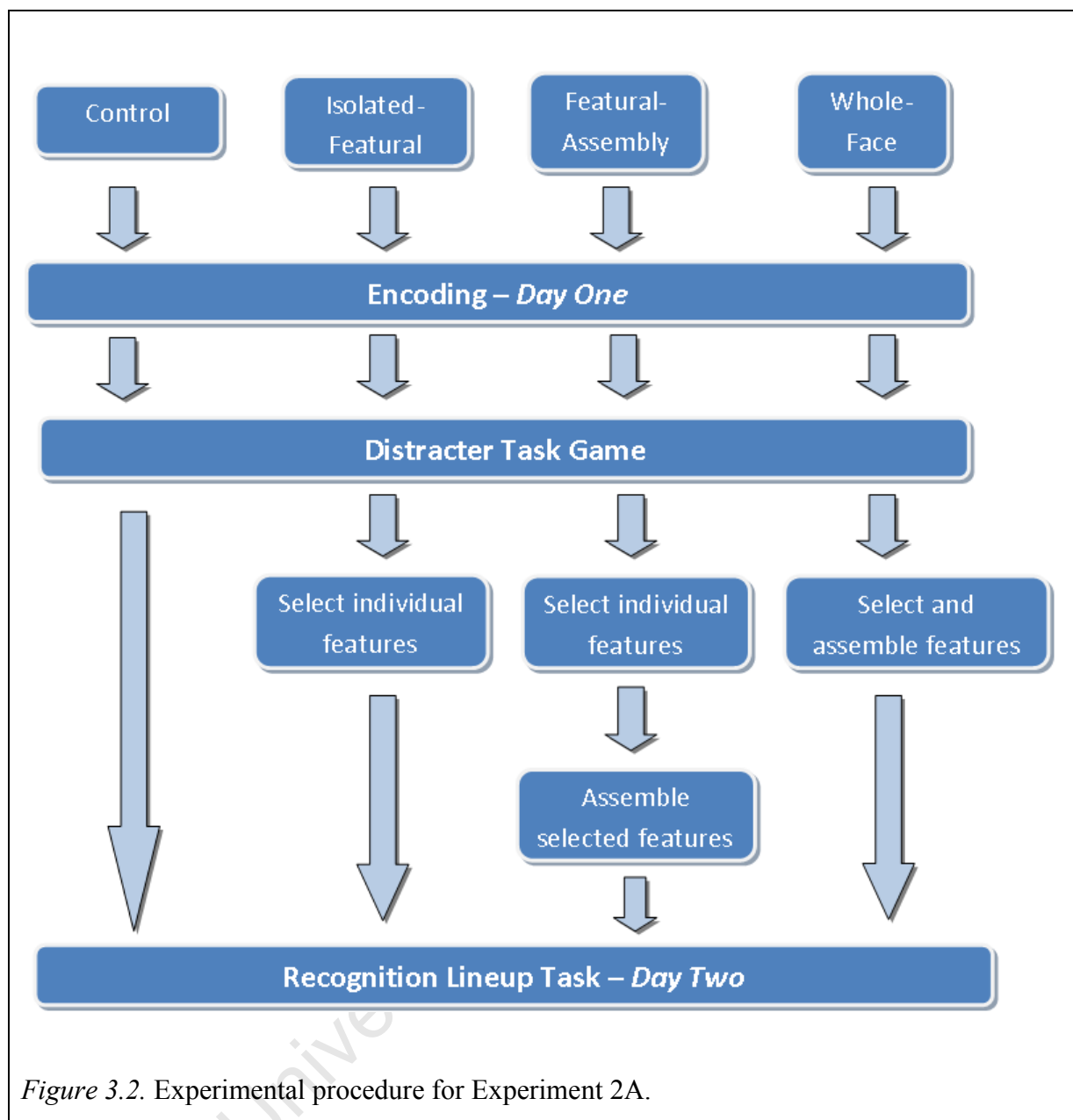


Figure 3.2. Experimental procedure for Experiment 2A.

Results

This experiment was designed to investigate whether participants in the isolated-featural construction group performed differently to a standard method of construction. It also aimed to investigate whether composite construction does, in fact, have a negative effect on recognition performance. For the initial analyses, participants' responses were coded as correct if the participant correctly selected the target. Participants' responses were coded as incorrect if the participant selected a filler, or rejected the lineup (in target present lineups). The initial analyses looked at "unforced" selections in target present lineups, that is, participants were not forced to make a selection from the lineup. The descriptive results of the groups were tested using contingency tests. An alpha level of 0.05 was used for all analyses. The differences between forced and unforced decisions were looked at, as well as differences between target present and target absent lineups. As the results below indicate, participants who constructed isolated-featural composites performed significantly worse than the control group. However, there were no significant differences between the different types of composite construction. Interestingly, when the forced-choice selections were analysed, there were no significant differences between any of the conditions. This is further discussed below.

Descriptive Statistics

Table 3.2 displays the frequencies of correct and incorrect selections in the target present lineups. The incorrect selections are further separated between participants who selected a filler, and participants who made no selection.

Table 3.2

Percentage Correct and Incorrect Choices in Lineup Task: Unforced Decisions

Condition	Unforced Choice						Total
	Correct		Incorrect				
	%	95% CI	Filler		"N" (Not Present)		
			%	95% CI	%	95% CI	
Control	83.33% (30)	[0.68, 0.93]	2.78% (1)	[<0.01, 0.15]	12.89% (5)	[0.06, 0.29]	100% (36)
Isolated-Featural	52.78% (19)	[0.37, 0.68]	0.00% (0)	[<0.01, 0.08]	47.22% (17)	[0.32, 0.63]	100% (36)
Featural-Assembly	65.71% (23)	[0.49, 0.79]	0.00% (0)	[<0.01, 0.09]	34.29% (12)	[0.21, 0.51]	100% (35)
Whole-Face	69.44% (25)	[0.53, 0.82]	5.56% (2)	[<0.01, 0.20]	25.00% (9)	[0.14, 0.41]	100% (36)
Total	67.83% (97)	[0.60, 0.75]	2.10% (3)	[<0.01, 0.06]	30.07% (43)	[0.23, 0.38]	100% (143)

Note. Numbers in parentheses are frequency counts. CI = confidence interval.

Upon initial inspection, the control group performed better than the isolated-featural, featural-assembly, and whole-face groups (83.33%, compared to 52.78%, 65.71%, and 69.44%, respectively). The isolated-featural group performed the worst, whilst the featural-assembly and whole-face group (which both assembled and configured the composite into a whole face) performed almost the same way in accuracy. Interestingly, the isolated-featural and featural-assembly groups, although performing the worst in accuracy (52.78% and 65.71%, respectively), did not incorrectly select any fillers. However, both had greater frequencies of lineup rejections. The correct selections are displayed graphically with error bars in Figure 3.3.

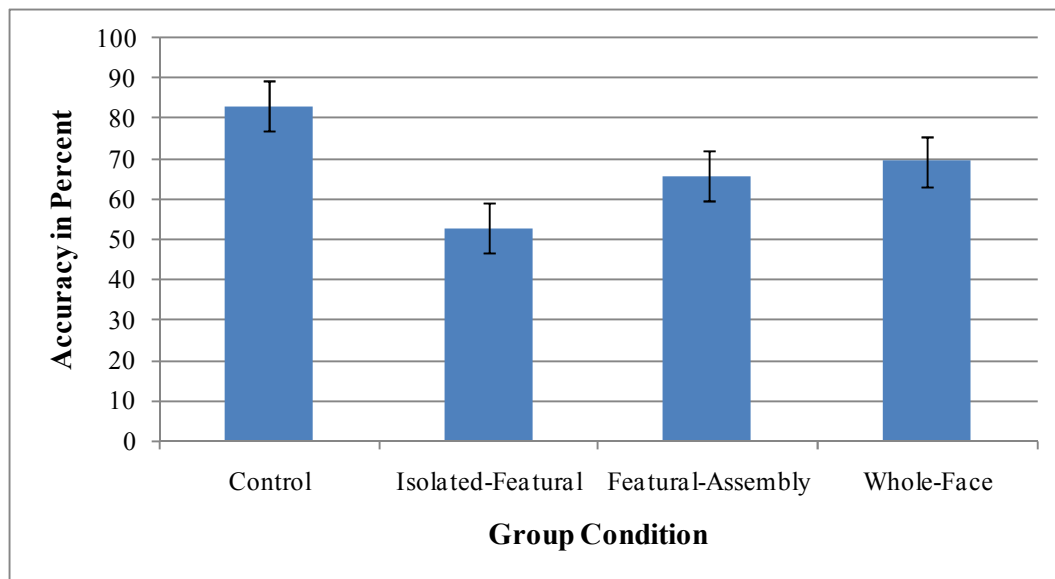


Figure 3.3. Percentage accuracy choices, per group, in the lineup task. I-bars are 95% confidence intervals.

Contingency Testing

A contingency test found an overall significant difference between correct responses across the conditions, $\chi^2(3, N = 143) = 7.82, p = 0.05$. Cramer's $V = 0.23$ indicated a moderate relationship as a measure of effect size. Pairwise comparisons showed a significant difference between the control and isolated-featural groups, $\chi^2(1, N = 72) = 7.73, p = 0.005$, Cramer's $V = 0.33$. Thus, the isolated-featural group performed significantly worse on the lineup task than the control group. However, there was no significant difference between the control and featural-assembly groups, $\chi^2(1, N = 71) = 2.91, p = 0.09$, Cramer's $V = 0.20$, and between the control and whole-face groups, $\chi^2(1, N = 72) = 1.93, p = 0.17$, Cramer's $V = 0.16$. There were also no significant differences between the isolated-featural and featural-assembly groups, $\chi^2(1, N = 71) = 1.23, p = 0.27$, Cramer's $V = 0.13$, between the isolated-featural and whole-face groups, $\chi^2(1, N = 72) = 2.10, p = 0.14$, Cramer's $V = 0.17$, or between the featural-assembly and whole-face groups, $\chi^2(1, N = 71) = 0.11, p = 0.74$, Cramer's $V = 0.04$.

However, the problem with multiple comparisons is that it increases the chance of making a Type 1 error and falsely rejecting the null hypothesis. To correct for this, the Bonferroni correction is used to reduce the error rate. The alpha (0.05) significance level is divided by the number of comparisons (six). The corrected significance level is $p = 0.008$. The only comparison that remains significantly different is between the control and the isolated-featural group. However, the Bonferroni correction is a conservative test, and also increases the chance of making a Type 2 error and incorrectly accepting the null hypothesis when significant differences are present.

There was an overall significant difference for lineup rejections, $\chi^2(3, N = 143) = 10.26, p = 0.02$, Cramer's $V = 0.27$. Pairwise comparisons showed a significant difference between the control and isolated-featural condition, $\chi^2(1, N = 72) = 9.43, p = 0.002$, Cramer's $V = 0.36$, between the control and featural-assembly groups, $\chi^2(1, N = 71) = 4.05, p = 0.04$, Cramer's $V = 0.24$, and between isolated-featural and the whole-face groups, $\chi^2(1, N = 72) = 3.85, p = 0.05$, Cramer's $V = 0.23$. However, there was no significant difference in number of lineup rejections between the control and whole-face groups, $\chi^2(1, N = 72) = 1.42, p = 0.23$, Cramer's $V = 0.14$, and between the featural-assembly and whole-face groups, $\chi^2(1, N = 71) = 0.74, p = 0.39$, Cramer's $V = 0.10$. Thus, the isolated-featural and featural-assembly conditions were incorrectly rejecting the lineup more frequently than the control group. The participants were not incorrectly selecting a filler, but were rather refusing to make any selection at all. If we use the Bonferroni method to correct for Type 1 error, there are no significant differences between the lineup rejections across groups. Thus, the groups all rejected the lineups at equal rates. However, as mentioned earlier, the Bonferroni correction is a conservative method and risks accepting the null hypothesis when differences are present. It should therefore be analysed with caution.

Hits and Misses Across Targets

Table 3.3

Percentage Correct and Incorrect Choices Per Condition Across Targets

	Target 1		Target 2		Target 3	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Control	17.02% (8)	8.51% (4)	25.00% (12)	0.00% (0)	20.83% (10)	4.17% (2)
Isolated-Featural	17.02% (8)	8.51% (4)	10.42% (5)	14.58% (7)	12.50% (6)	12.50% (6)
Featural-Assembly	17.02% (8)	6.38% (3)	14.58% (7)	10.42% (5)	16.67% (8)	8.33% (4)
Whole-Face	19.15% (9)	6.38% (3)	16.67% (8)	8.33% (4)	16.67% (8)	8.33% (4)
Sub-Total	70.21% (33)	29.79% (14)	66.67% (32)	33.33% (16)	66.67% (32)	33.33% (16)
Total	100.00% (47)		100% (48)		100% (48)	

Note. Numbers in parentheses are frequency counts.

As can be seen in Table 3.3, there were no significant differences between correct and incorrect responses across targets, $\chi^2 (2, N = 143) = 0.18, p = 0.91$, Cramer's $V = 0.04$. This suggests that accuracy responses across the targets were equally distributed and not biased. Thus, it was not easier to respond to one target over the others.

Forced Decisions

Initial analyses indicated that 46 participants made an incorrect selection. However, 43 of the 46 incorrect responses were "not present" selections. Participants incorrectly rejected the lineup. Thus, this suggests that participants were more uncertain of the correct target, and rejected the lineup. The forced-choice responses were analysed and the results re-run. A contingency test found no overall significance between the groups, $\chi^2 (3, N = 143) = 0.77, p = 0.86$, Cramer's $V = 0.07$. Table 3.4 shows the percentage of correct and incorrect choices for forced decisions.

Table 3.4

Percentage Correct and Incorrect Choices in Target Present Lineup Task: Forced

Decisions

Condition	Forced Choice		Total
	Correct	Filler	
Control	94.44% (32)	2.78% (1)	100% (36)
Isolated-Featural	86.11% (31)	0.00% (0)	100% (36)
Featural-Assembly	91.43% (32)	0.00% (0)	100% (35)
Whole-Face	91.67% (33)	5.56% (2)	100% (36)
Total	89.51% (128)	10.49% (15)	100% (143)

Note. Numbers in parentheses are frequency counts.

As can be seen, in the table above, most of the participants obtained the correct answer. The isolated-featural group increased in accuracy from 52.78% to 86.11%. There was no longer a significant difference between the control group and the isolated-featural group, $\chi^2(1, N = 72) = 0.13, p = 0.72$, Cramer's $V = 0.04$. Thus, participants were not choosing incorrectly, as previously thought, but instead were being more conservative in their selection. When forced to make a decision, participants did know the correct answer. Figure 3.4 displays the unforced and forced accuracies across conditions.

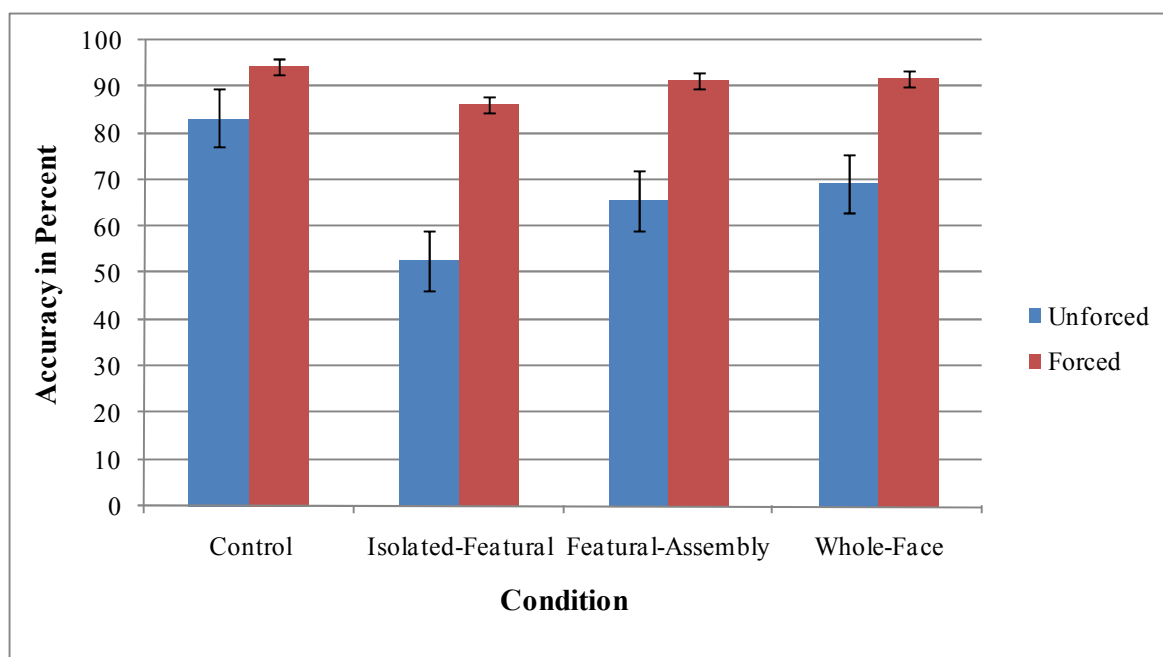


Figure 3.4. Percentage accuracy choices for unforced and forced decisions, per group, in the lineup task. I-bars are 95% confidence intervals.

The isolated-featural group had a significant increase from 52.78% accuracy to 86.11% accuracy, $\chi^2(1, N = 72) = 9.43, p = 0.002$, Cramer's $V = 0.36$. The featural-assembly group also experienced a significant increase in number of correct identifications from 65.71% to 91.43% accuracy, $\chi^2(1, N = 70) = 6.87, p = 0.009$, Cramer's $V = 0.31$. The whole-face construction group also achieved a significant increase in hits, from 67.83% to 91.67% accuracy, $\chi^2(1, N = 72) = 5.68, p = 0.02$, Cramer's $V = 0.28$. Although the control group increased slightly, they showed no significant increase in accuracy, from 83.33% to 94.44% accuracy, $\chi^2(1, N = 72) = 0.47, p = 0.50$, Cramer's $V = 0.80$. Thus, apart from the control group, all the other conditions showed a significant increase in correct identifications when they were forced to make a selection.

Target Absent Lineups

The target absent lineups were shown to all participants, either before, or after, the target present lineups. This was counterbalanced across conditions. The correct response was "not present". The results are displayed in Table 3.5.

Table 3.5

Percentage Correct and Incorrect Choices in Target Absent Lineup Task: Unforced

Decisions

Condition	Unforced Choice		Total
	Correct	Filler	
Control	88.89% (32)	11.11% (4)	100% (36)
Isolated-Featural	88.89% (32)	11.11% (4)	100% (36)
Featural-Assembly	100.00% (35)	0.00% (0)	100% (35)
Whole-Face	97.22% (35)	2.78% (1)	100% (36)
Total	93.71% (134)	6.29% (9)	100% (143)

Note. Numbers in parentheses are frequency counts.

All of the groups obtained near ceiling effects, with 93.71% of participants correctly rejecting the lineup. The control and isolated-featural groups performed the worst (both obtaining 88.89%), although there were no significant differences between the groups, $\chi^2(3, N = 143) = 5.94, p = 0.12$, Cramer's $V = 0.20$.

Lineup Bias

One of the caveats of the Wells et al. (2005) study is that it never calculated the amount of bias within the lineups. In a post hoc study, 80 mockwitnesses had to select the target from the lineup, based only on a description (see Appendix G for an example of the mockwitness task). If the lineup is fair, and all the fillers adequately resemble the target, no lineup member should be chosen at a rate greater than chance (0.167). The effective size of a lineup " E " is an estimate of the number of appropriate foils in the lineup, see Table 3.6.

Table 3.6

Proportions of Bias and Effective Size in Each Lineup

Lineup	N	Target	Bias	E	95% CI of E
1	80	Present	0.08	2.90	[2.37, 3.73]
1	80	Absent	0.21	4.23	[3.62, 5.07]
2	80	Present	0.03	5.04	[4.57, 5.62]
2	80	Absent	0.53	2.74	[2.25, 3.51]
3	80	Present	0.36	3.59	[3.01, 4.46]
3	80	Absent	0.44	3.12	[2.70, 3.60]

Thus, lineup 2, target present, has just over five possible lineup filler selections. The bias of the lineup indicates the proportion of target selections, based on the description. If chance is 0.167, then any lineup that has a bias greater than 0.167, should be removed. Lineup 2, target absent, has a bias of 0.53 (almost three times greater than chance), however, since the target is not present in the lineup, there is no need to remove this from the analyses. Lineup 3, target present, has a bias two times greater than chance, and the target absent lineup is almost three times greater than chance levels. Thus, both of these lineups were removed and the analyses re-run without the biasing data. A contingency test conducted between all the groups, without Target 3 (target present and target absent), found no significant difference between the groups, $\chi^2(3, N = 95) = 4.90, p = 0.18$, Cramer's $V = 0.23$. The overall test was no longer significant. However, pairwise comparisons showed a significant difference between the control and isolated-featural groups, $\chi^2(1, N = 48) = 4.75, p = 0.03$, Cramer's $V = 0.31$. The isolated-featural group continued to perform significantly worse on the lineup task than the control group, even with the bias removed. There was no significant difference between the control and featural-assembly groups, $\chi^2(1, N = 47) = 2.03, p = 0.15$, Cramer's $V = 0.21$, and between the control and whole-face groups, $\chi^2(1, N = 48) = 1.06, p = 0.30$, Cramer's $V = 0.15$. Thus, the biased lineup had a minimal influence on the rates of accuracy selection. Figure 3.5 below displays a bar-graph of the original "biased" lineup accuracies compared to the 'unbiased' selection accuracies.

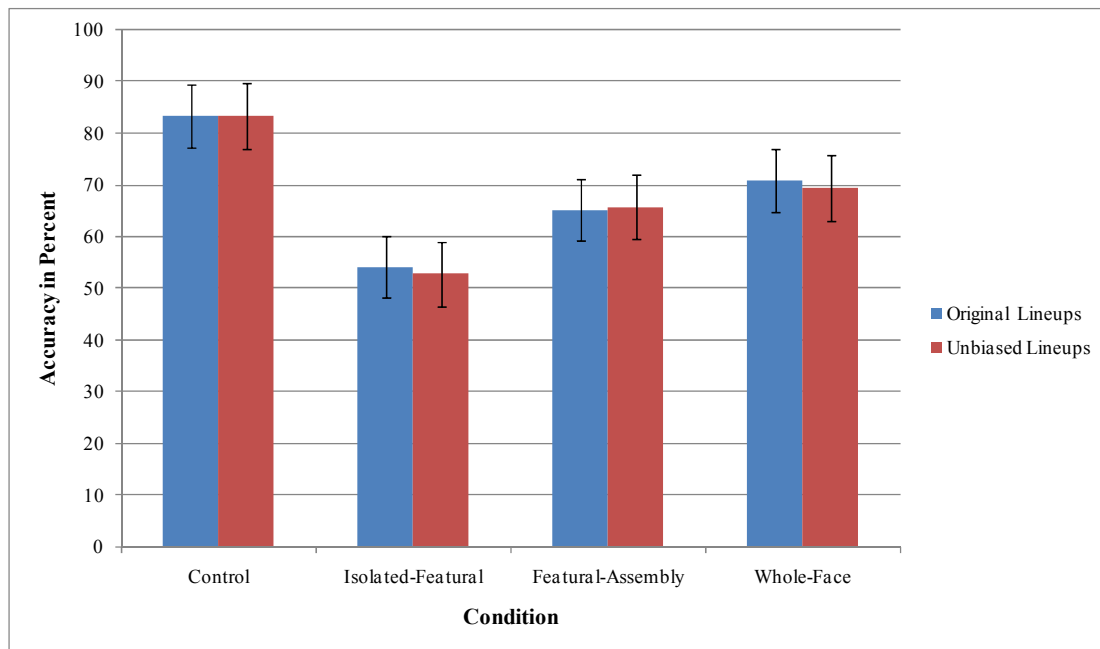


Figure 3.5. Percentage accuracy choices for biased and unbiased lineups, per group, in the lineup task. I-bars are 95% confidence intervals.

Reasons for Lineup Rejection

After participants had either selected a lineup member or rejected the lineup, they were required to indicate a reason for their decision. Participants were required to respond with one of the following responses:

- (N) - I did select a person from the lineup
- (A) - The original person was definitely not there
- (B) - I cannot remember what the original person looked like
- (C) - I could not decide between multiple members of the lineup
- (D) - I had a slight idea of who the person was, but I was not very confident in my decision

Table 3.7 indicates the proportion of participants' responses across conditions in target present lineups.

Table 3.7

Proportions of Selection Reasons Across Conditions in Target Present Lineups

	Control	Isolated-Featural	Featural-Assembly	Whole-Face
N (Did select)	83.33% (30)	52.78% (19)	57.14% (20)	77.78% (28)
A (Not present)	8.33% (3)	33.33% (12)	17.14% (6)	11.11% (4)
B (Could not remember)	0.00% (0)	0.00% (0)	0.00% (0)	0.00% (0)
C (Multiple members)	0.00% (0)	5.56% (2)	2.85% (1)	0.00% (0)
D (Not confident)	8.33% (3)	8.33% (3)	22.86% (8)	11.11% (4)

Note. Numbers in parentheses are frequency counts.

The proportion of participants who indicated “N” (that they did make a selection) followed the same response pattern as the correct responses for the unforced decisions. Interestingly, the featural-assembly group had the highest proportion of “D” responses (they had a slight idea of who the person was, but were not very confident in their decision). The whole-face group had the second highest proportion of “D” responses. Both of these groups were required to configure the features of the composite into a whole face. The featural-assembly group differed significantly from the control and isolated-featural groups, $\chi^2(1, N = 71) = 2.86, p = 0.045$, Cramer’s $V = 0.20$. There was no significant difference between the featural-assembly and whole-face groups, $\chi^2(1, N = 71) = 1.74, p = 0.09$, Cramer’s $V = 0.16$. The isolated-featural group had the highest proportion of “A” responses (the original person was definitely not there). The isolated-featural group thought the target was not there at significantly higher rates than the control group, $\chi^2(1, N = 72) = 6.82, p = 0.004$, Cramer’s $V = 0.31$, and the whole-face group, $\chi^2(1, N = 72) = 5.14, p = 0.012$, Cramer’s $V = 0.27$. There was a marginally significant difference between the isolated-featural group and the featural-assembly group, $\chi^2(1, N = 71) = 2.46, p = 0.058$, Cramer’s $V = 0.19$. Thus, only being exposed to piecemeal features of a composite resulted in participants rejecting the lineup because they thought the target was not present. This is in contrast with the featural-assembly group who were less confident about their selection, resulting in higher proportions of lineup rejection.

Meta-Cognitive Questions

After the lineup task, participants were asked three questions which required them to rate: their level of confidence of their lineup decision, how difficult they thought the lineup task was, and how similar they thought the lineup members were to each other. They answered using the number keypad on the keyboard, on a scale from 0 (not at all confident/difficult/similar) to 9 (extremely confident/difficult/similar). Some participants saw the target present lineup first, and the target absent lineup second, and vice-versa. Thus, when calculating the values, the responses were calculated into two columns: target present 1st, and target present 2nd. This indicated whether participants saw the target present lineup first, or second. The means for each of these columns were calculated (see Table 3.8).

Table 3.8

Mean Responses for Meta-Cognitive Questions Across Individual Targets Following Lineup Task

	Target 1		Target 2		Target 3	
	TP	TA	TP	TA	TP	TA
Confidence	6.98 (1.63)	7.56 (2.06)	6.67 (2.26)	7.48 (1.73)	6.50 (1.74)	7.51 (1.65)
Difficulty	3.99 (2.32)	3.57 (2.42)	4.85 (2.47)	3.60 (2.54)	4.73 (2.31)	4.58 (2.16)
Similarity	4.20 (1.66)	3.88 (1.94)	5.56 (1.83)	4.69 (1.52)	5.44 (1.84)	4.94 (1.76)

Note. Numbers in parentheses are standard deviations.

The average responses between groups did not vary greatly. The confidence, difficulty, and similarity responses were all similar across targets. This could possibly suggest that some targets did not stand out more than others and that the lineups were all constructed equally.

Interestingly, Target 1, target absent, had the lowest rating of similarity between lineup members, and also had the highest confidence rating. The relationship between confidence and similarity across targets was moderately, negatively correlated, $r = -0.67$, $p < 0.01$. Thus, as similarity between the lineup members decreases, confidence increases. There

was a strong, negative correlation between confidence and difficulty, $r = -0.70, p < 0.01$. This suggests that as difficulty decreases, confidence increases. There was a strong, positive relationship between difficulty and similarity, $r = 0.87, p < 0.01$. Thus, as similarity between lineup members increases, difficulty also increases.

A point-biserial correlation was conducted between confidence and accuracy for target present lineups, and found a positive relationship between the two variables, $r = 0.20, p = 0.02$. This indicates that as accuracy increases from inaccurate to accurate, so confidence, too, increases. A point-biserial correlation conducted between confidence and accuracy for target absent lineups also found that as responses become accurate, confidence increases, $r = 0.21, p = 0.01$. Average confidence values were calculated for each condition, see Table 3.9.

Table 3.9

Descriptive Statistics for Confidence Values for All Conditions

Condition	Lineup Type	
	Target Present	Target Absent
Control	7.28 (1.67)	7.56 (2.03)
Isolated-Featural	7.00 (1.80)	7.42 (1.27)
Featural-Assembly	6.17 (1.82)	7.09 (2.27)
Whole-Face	6.39 (2.19)	7.50 (1.78)

Note. Numbers in parentheses are standard deviations.

A one-way ANOVA conducted on confidence between groups, for target present lineups, showed a significant difference between the means of these groups, $F(3, 142) = 2.68, p = 0.049, \eta^2 = 0.05$. A post hoc LSD test revealed a significant difference between the control and featural-assembly groups, $p = 0.02$, and between the control and whole-face groups, $p = 0.047$. This suggests that the featural-assembly and whole-face groups, which utilised configural assembly of features, rated their confidence significantly lower than the control group.

Reaction Time

The mean reaction times for correct and incorrect responses were calculated. A factorial ANOVA was conducted between condition and response on reaction time (see Table 3.10).

Table 3.10

Mean Reaction Times for Correct and Incorrect Responses Across Groups

Condition	Response	
	Accurate	Inaccurate
Control	11.16 (8.52)	20.77 (21.67)
Isolated-Featural	16.12 (12.24)	17.20 (10.54)
Featural-Assembly	18.11 (14.15)	20.80 (20.76)
Whole-Face	20.19 (13.54)	19.89 (13.18)

Note. Measured in seconds. Numbers in parentheses are standard deviations.

A factorial ANOVA for target present lineups found no interaction effect between condition and response on reaction time, $F(3, 143) = 0.58, p = 0.63, \eta^2 = 0.01$. There was no main effect for condition ($F(3, 143) = 0.59, p = 0.63, \eta^2 = 0.01$) or for response ($F(1, 143) = 1.59, p = 0.21, \eta^2 = 0.01$).

However, as there was no significant difference between the groups, multiple chi-square tests were conducted between accuracy (which is a dichotomous variable) and reaction time. This was done in order to determine the optimal reaction time for accuracy. No participants got the answer correct after 53 seconds, so this was used as the final reaction time. The quickest correct response occurred just under three seconds, so this was used as the upper limit for the first group. Fifty-one chi-squares were conducted and the final result plotted on a line graph, see Figure 3.6.

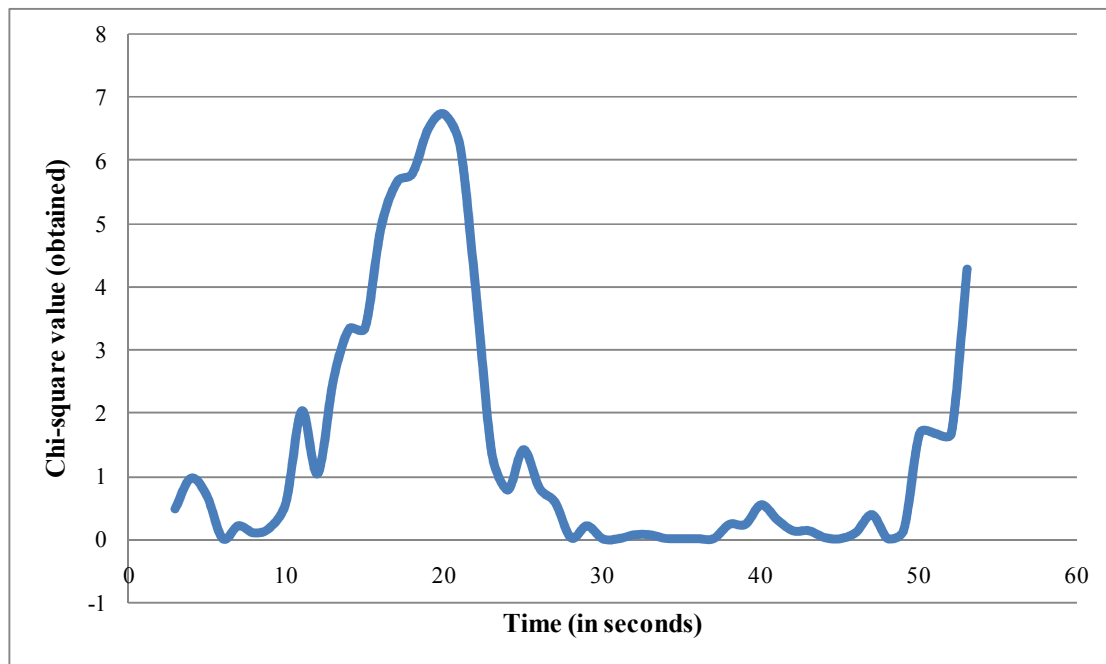


Figure 3.6. Chi-square values for reaction time and correct responses on the lineup task. The highest point indicates the reaction time for which accuracy was highest.

Figure 3.6 suggests that both accuracy and reaction time are optimised at 20 seconds. Thus, participants took their time when making an accurate decision. The graph decreases as the number of correct responses levels off, but incorrect responses increase. The graph increases at 50 seconds where the number of correct responses begins to increase again. A point-biserial correlation was conducted on accuracy and reaction time and found a weak, negative relationship, $r = -0.12$. This would suggest that as response changes from 0 (not accurate) to 1 (accurate), reaction time decreases. However, this result was non-significant, $p = 0.15$. Confidence and reaction time were significantly, negatively correlated, $r = -0.40$, $p < 0.01$. Thus, as confidence increases, reaction time decreases. Participants who were more confident about their decisions made faster responses.

Discussion

Initial analysis found an overall significant difference between the accurate identifications for all the conditions. However, when the results were further analysed, there was only a significant difference between the control condition and isolated-featural condition. Participants who were selecting only the piecemeal features during construction performed worse during the recognition task. The featural-assembly and whole-face groups performed roughly the same with regards to correct selections. These are both the groups that had some involvement in configuration of the composite features. Interestingly, both the isolated-featural and featural-assembly groups selected no fillers, although they had the highest number of incorrect answers. When the incorrect responses were analysed, it was found that 43 out of 46 incorrect answers were lineup rejections. The isolated-featural and featural-assembly groups rejected the lineups more frequently than the other conditions. Thus, it was not that the construction groups were incorrectly selecting a filler, but rather that they refused to select anyone from the lineup.

The forced decision results were analysed and found no significant difference between any of the groups. All conditions obtained near ceiling effects. Participants initially appeared to not know the answer, however, when forced to choose, they accurately selected the target. The isolated-featural, featural-assembly, and whole-face conditions all performed significantly better in the forced-choice analyses than in the unforced decisions. This supports the criterion shift theory (Clare & Lewandowsky, 2004) and also fits with findings in previous research (Kempen, 2009; Wells et al., 2005; Yu & Geiselman, 1993). Participants became more conservative in their selection and refused to make any identification at all. Unforced decisions for the target absent lineups also indicated near ceiling effects. This suggests that participants did know who the correct target in the target present lineup was, and were not merely guessing.

However, one caveat might be that the targets used at encoding and recognition were in the same pose (frontal position). This was necessary so that the features could be extracted for lineup purposes in Experiment 1. The images at recognition were in full colour, flipped horizontally, and had edited clothes and backgrounds, so as to prevent photograph recognition as opposed to target recognition. However, the results from the unforced-choice decisions did not reach ceiling effects; participants were still conservative in their responses. Thus, the same-pose photograph possibly had no overall effect on participants' responses. Furthermore the accuracy of selections was spread equally across all the targets, indicating that some targets were not biased over others. When the lineup bias was calculated, the data indicated two possible biased lineups. However, when the data was re-run without these lineups, the results remained the same, indicating that the possibly biased lineups still had no effect on the overall results.

Interestingly, with regards to the lineup rejection reasons, the featural-assembly and the whole-face construction groups had the highest proportion of "D (not confident)" responses (they had a slight idea of who the person was, but were not very confident in their decision). These were both groups that utilised configural and holistic aspects in their composites. This further supports the results obtained on the target absent lineups; participants were confident that the target was not there, and selected the target even under unforced conditions. However, when participants were faced with the target, they became conservative in their decision criteria. Also interestingly, the isolated-featural group scored the highest proportions for "A (not present)" responses (the original person was definitely not there). Thus, being exposed to only piecemeal images resulted in participants rejecting the lineup because they thought the target was not present, as opposed to configural construction of a face which made other participants more conservative.

Accuracy and confidence were positively correlated; thus, participants' confidence increased as their answers changed from incorrect to correct. Difficulty and similarity were also positively correlated, which indicates that as lineup similarity increased, so participants' perceived difficulty also increased. Confidence and difficulty were negatively correlated, thus, as difficulty decreased, participants' confidence increased. There was a significant difference in confidence ratings between the control and featural-assembly, and control and whole-face groups. Both of these construction groups rated their confidence significantly lower than the control groups. These two groups were allowed to configure the composite features. This corresponds to the lineup rejection reasons: The featural-assembly and whole-face groups had the highest proportion of "D" (not confident) responses. The proportions of overall rejection answers, although significant, were still low. The majority of lineup selection reasons was "N" (participants did make a selection). Thus, participants who *did* select were confident about their selection, which may have contributed to a higher confidence average across the conditions.

There was no significant difference between the type of condition and reaction time, indicating that all the groups spent the same amount of time deliberating on the lineup decision. However, confidence and reaction time were negatively correlated. Thus, as confidence increases, so reaction time decreases. The optimal reaction time which maximised accuracy was at approximately 20 seconds.

These results indicate some clear findings. Composite construction appears to hamper recognition in a lineup task. However, when these results are analysed, participants are rejecting the lineup. Indeed, roughly the same number of incorrect fillers was selected across all conditions. The isolated-featural group had the highest number of lineup rejections, followed by the featural-assembly group, then by the whole-face group. It is interesting to note that the featural-assembly and whole-face groups performed similarly throughout this

experiment. Both of these conditions had a configural aspect to them. This is in comparison with the isolated-featural group which had the highest number of lineup rejections. The isolated-featural group also indicated that their reason for rejection was that they thought the target was not present in the lineup.

This fits with the VOE effect in such a way that being forced to select the individual features may have masked or overshadowed access to the visual memory. Contrastingly, the fact that the featural-assembly group was allowed to assemble their features at the end of the first session seemed to eliminate this effect and resulted in their performance being similar to that of the whole-face condition. Both of these conditions indicated that they were not very confident about the target, thus they rejected the lineup. It appears as though this reasoning fits with the criterion shift theory, and this is further evidenced by the significant increase in accurate responses when the forced-choice answers were analysed. This supports Yu and Geiselman's (1993) finding that participants who constructed a composite were more likely to reject the lineup. However, they did not analyse forced-choice responses in their study, so they were unable to investigate whether or not participants were correctly selecting the target. These results partially fit with those from Wells et al. (2005) in which composite constructors also rejected the lineup at higher rates. However, when Wells et al. analysed their results, participants who constructed composites were incorrectly selecting lineup fillers. This does, however, fit with previous research (Kempen, 2009) in which forced-choice answers significantly improved lineup accuracy. However, it is not clear why this effect occurs. Since the effect was stronger with the isolated-featural group, this may indicate that the act of featural selection creates a mask which prevents access to the entire configural face. However, constructing the composite into a whole-face partially eliminates this VOE. Furthermore, composites constructed may not be of good quality. This misinformation present in the entire composite may be recoded and restored as a veridical representation of

the target. The composite may then match multiple members of the lineup, creating participant uncertainty. Thus, they would rather reject the lineup completely than make an incorrect decision. When the criterion threshold is lowered, however, participants are less stringent about their decision criteria, and select the correct target.

The VOE uses SDT to suggest that generating verbal descriptions (or constructing a composite) will cause participants to set their criterion threshold higher. Participants are therefore less likely to respond yes to anyone in the lineup. When the results are analysed, it appears as though this verbal description or construction has affected accuracy. However, when the incorrect choices are analysed, this high criterion threshold has also resulted in a decrease in false alarms. Participants are adopting a more stringent threshold and rejecting the lineup completely. These participants are more conservative and would rather “err on the side of caution”. However, when forced to make a selection, participants lower their threshold. This results in a greater number of hits, as seen in this experiment, without sacrificing other lineup members in the form of false alarms. As can be seen in the target absent lineups, participants are correctly rejecting the lineup. Thus, participants are aware of whom the correct target is; they are not merely lowering their criterion threshold and generating an increase in hits *and* false alarms. The shift from a high criterion threshold to a low criterion threshold appears to maximise hits without sacrificing false alarms. Composite construction does not result in a decrease in recognition accuracy, as postulated by Wells et al. (2005), but rather results in participants adopting a higher threshold below which they will not make a decision. In this way, it is safer to allow a guilty person to go free, rather than to incarcerate an innocent individual.

Research on the VOE by Meissner et al. (2001) induced participants to adopt either a high response criterion, or a low response criterion. In the high response criterion group, Meissner et al. warned participants to only include verbal descriptions that they were

absolutely sure about. In the low response criterion group, the authors told participants to include as much information as they could, even if it felt that they were guessing. Meissner et al. found that participants who increased their response criterion had an increase in description quality, and an increase in subsequent target identifications. Conversely, participants who shifted to a low response criterion performed worse on a subsequent recognition task. If we link composite construction to the production of a verbal description, participants may be including features that they are unsure about, to the point of guessing. This may be particularly applicable to the isolated-featural group; they were given a feature selection sheet. They may have felt that they were required to fill in as many features as possible. Thus, these participants may have felt more conservative in their lineup decision, and rejected the lineup. As discussed earlier, Meissner et al. found that participants who generated misinformation were more likely than the control group to perform poorly on the recognition task. Thus, participants in the construction conditions may have adopted a low response criterion for composite construction, resulting in a poor quality composite. Meissner et al. also found that participants who generated misinformation were unable to distinguish between the source of the misinformation they had generated during the intervening task, and information they had encoded from the target. This self-generated misinformation may have resulted in participants becoming more conservative; participants may have tried to match what they thought was a veridical representation of the target, to someone in the lineup. This may have led participants to shift to a more stringent criterion response, and reject the lineup completely.

Palmer, Brewer, and Weber (2010) also conducted a study in which some participants were exposed to two lineups, as in this experiment. They concluded that participants who viewed a second lineup were more likely to perform poorly after viewing the initial lineup. They suggest that witnesses may have found the initial lineup to be difficult, and thus adopted

a more stringent response threshold for the second lineup. They also suggest that participants may have initially expected the target to be present in the first lineup. When the target was not present (a target absent lineup), this may have affected participants' confidence about selecting correctly in the second lineup. In this experiment, the initial results of the target present lineup indicated that participants were adopting a higher criterion response threshold and rejecting the lineup. The initial results of the target absent lineup indicated that participants were correctly rejecting the lineup. Participants were either exposed to a target present/target absent lineup sequence, or target absent/target present sequence. In the case of the former sequence, participants may have responded that the target was not present, as the composite may have visually overshadowed the veridical representation of the target. According to Palmer et al. (2010), participants will therefore be less likely to respond in a subsequent lineup. In this sequence, the subsequent lineup was the target absent lineup. Participants may have felt less confident at this point, and rejected the lineup, which, in this case, is the correct answer. In the target absent/target present lineup sequence, participants may have begun the task expecting the target to be present. When the target was not present, participants may have become more conservative and less confident, and rejected the second lineup (target present), accounting for a decrement in target hits. However, this suggestion of subsequent lineups does not fit with the high recognition rates achieved by the control group. It can then be tentatively assumed that the reduction in accuracy is due to a stringent criterion threshold brought about by composite construction.

Meissner, Tredoux, Parker, and MacLin (2005) suggest that simultaneous lineups encourage familiarity-based processing. Furthermore, they state that shifts in criterion are associated with changes in familiarity processing. Thus, it could be suggested that constructing and re-encoding a composite allows for participants to internalise the misinformation and think of the composite as a veridical representation of the target. When

faced with the lineup, participants experience a reduction in familiarity, as the composite may not resemble the target. This could lead to a shift in response criterion, resulting in participants being less likely to make any selection on the lineup.

CHAPTER FOUR:

EXPERIMENT 2B

Method

This experiment aimed to investigate the quality of composites constructed by the featural-assembly and whole-face construction conditions. This was conducted by means of a matching task.

Participants

A total of 28 participants was used in this study. Participants were undergraduate Psychology students between the ages of 18 and 24 years ($M = 20.38$, $SD = 1.89$). Students participated in exchange for course credit. Demographic information indicated that 11 participants (39.29%) were male and 17 (60.71%) were female. As the stimuli used were of White, female faces, the race distribution of participants was also recorded. A total of 19 (67.86%) participants identified themselves as “White”, and nine (32.14%) participants identified themselves as “Black” or “Coloured”. Testing took place in a computer laboratory in the Psychology Department at the University of Cape Town.

Materials

Target stimuli and composites. The target faces used in this experiment were the three White, female targets used in Experiment 2A. The composites used in this experiment were the composites constructed by the whole-face construction group, and composites constructed by the featural-assembly group. Thus, there were 36 composites from each group. The isolated-featural composites were not used in this experiment as the composites consisted of only individual features. The target faces and composites were all in greyscale. They were all shown in frontal position and standardised to 5.77 cm in width and 7.43 cm in height, with a resolution of 350 x 450 pixels (8-bit).

Presentation and answering materials. All the instructions, target stimuli, and composites were presented using E-Prime 2.0. The entire task was computer-based. The slideshow was presented on a computer screen with a resolution of 1024 x 768 pixels. This attempted to minimise experimenter effects and to standardise the experimental procedure.

Slideshow and task construction. The slideshow contained detailed instruction slides, which the participant could read through at their own pace. The instructions stated that participants had to select a composite (from two possible selections) that they thought best resembled the target photograph. One of the randomly drawn selections was a featural-assembly composite, and the other selection was from the whole-face construction set. The slideshow displayed the target photograph in the middle of the screen, with one of the composites from each construction group on either side. The position of the featural-assembly and whole-face composites was randomised throughout the entire experiment. Participants had to select the composite they thought best matched the target photograph. Once they had made a selection, another two composites were randomly drawn from the featural-assembly and whole-face construction groups. As there were 72 composites, these were divided into their two respective groups, featural-assembly, and whole-face construction, totalling 36 composites in each set. As there were three targets, each target had 12 pairings. Once participants had selected amongst these 12 pairings, they moved on to the second target, and then the third target. The target sets were randomised throughout the experiment.

Procedure

All participants were seated at a computer in the laboratory. Each workstation contained a computer and consent form. Participants were required to fill in their demographic information and sign the consent form.

The slideshow was opened. Instruction slides notified the participants that they would be shown a photograph in the middle of the screen, with two composites on either side of the photograph. They were instructed to select the composite they thought best resembled the photograph in the middle. If they thought the composite to the left of the photograph resembled the target, they indicated “Q” on the keyboard. If they thought that the composite to the right of the photograph resembled the target, they indicated “P” on the keyboard (see Figure 4.1 for an example of the matching task). Participants were allowed to take as long as they needed to respond. Once they had made their selection, two new composites were randomly generated. Each composite was only shown once. Participants continued to make their selections for that target, until a new target face appeared, with a new set of composites. This continued for the third target. The “sets” of targets were randomised.

Once participants had completed the matching task, they were debriefed and told the purpose of the study.

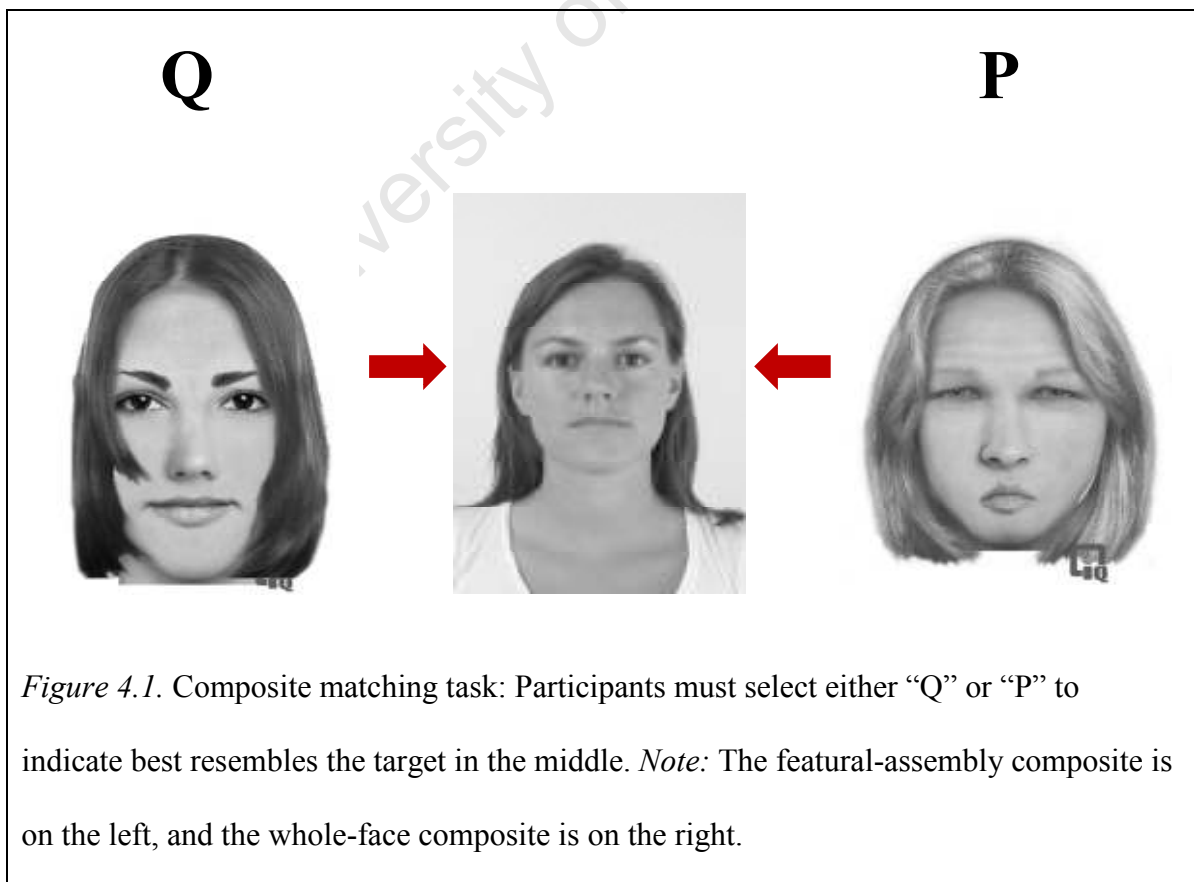


Figure 4.1. Composite matching task: Participants must select either “Q” or “P” to indicate best resembles the target in the middle. *Note:* The featural-assembly composite is on the left, and the whole-face composite is on the right.

Results

This experiment investigated whether participants would be more likely to match a featural-assembly composite or the whole-face composite to the target. This method chose to pair the composites off against each other to investigate if one construction technique resulted in more accurate composites than the other construction technique. Participants had to select a composite between a random pair (of a featural-assembly and whole-face composite). As there were 72 composites, there were 36 decisions to be made. Each composite was only displayed once and the pairings never remained the same between participants.

Composite Selections

Participants who selected a featural-assembly composite were scored with a “1” in the featural-assembly column, and participants who selected a whole-face composite were scored with a “1” in the whole-face column. The scores were then summed to give an indication of whether participants were selecting the featural-assembly composite or whole-face composite at higher frequencies. Figure 4.2 below displays the percentage selections between the composite types.

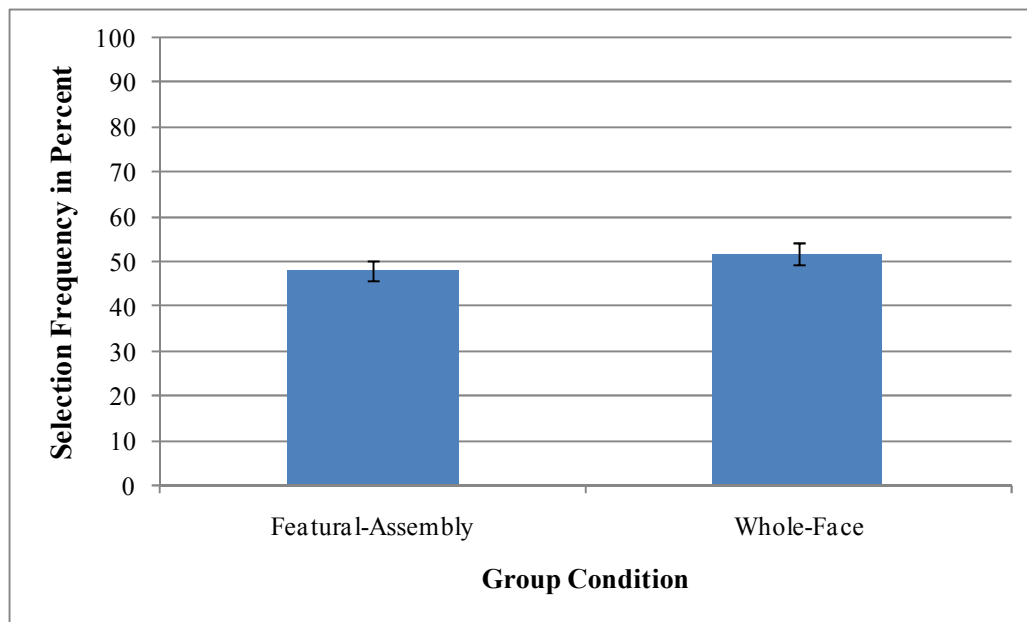


Figure 4.2. Percentage of composite selections, choices, per group, in the matching task. I-bars are 95% confidence intervals.

The featural-assembly composites were selected 48.11% of the time, while the whole-face composites were chosen 51.89% of the time. As can be seen in the graph, and with the error bars, there was no significant difference between the selection frequencies, $\chi^2(1, N = 2016) = 2.86, p = 0.09$, Cramer's $V = 0.04$. An odds-ratio was performed and indicated that the featural-assembly composites were 0.86 times less likely to be chosen than the whole-face composites. However, the odds-ratio value is very close to 1, and this may be an indication that both groups are equally likely to be selected at chance rates. As the composites were not rated further, it is unclear whether there is no difference between the types of composite construction, or if both composites are so poorly constructed that participants were selecting composites at rates equal to chance.

Reaction Times

The reaction times for each participant's response during selections was averaged per group (featural-assembly and whole-face) and then averaged across the total number of participants. The total average reaction times per group were compared to investigate whether

one group of composites was chosen at faster rates than the other group. Reaction times are given in seconds. There was no significant difference between reaction times for the featural-assembly composites ($M = 11.46$, $SD = 6.93$) and the whole-face composite ($M = 12.25$, $SD = 8.27$), $t(54) = -0.54$, $p = 0.59$, Cohen's $d = -0.10$. Participants selected the composites at equal speeds. Figure 4.3 below indicates the mean reaction time of selections per group (featural-assembly and whole-face).

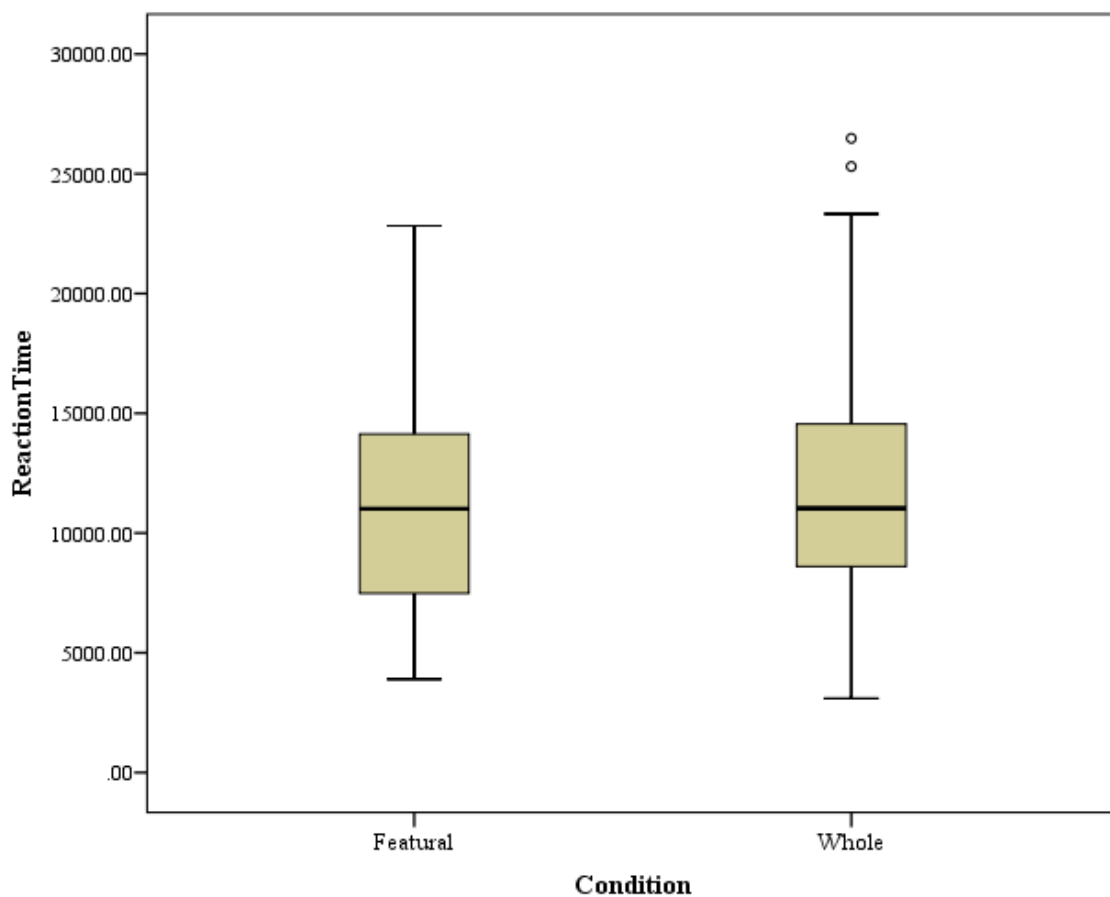


Figure 4.3. Mean reaction times per group, in the matching task. O-circles are outliers.

Thus, it is possible that the composites a) look very similar, or that b) both composites look nothing like the target. It would therefore be very interesting to test this in a later study.

Discussion

This experiment aimed to investigate which composite construction method resulted in a better matched composite. Two composites from each group were paired off against each other. Participants had to indicate which composite they thought matched the target. This experiment did not aim to investigate whether the composite could be matched to the target, but rather which construction technique fared better. Experiment 2A indicated no significant difference between the featural-assembly and whole-face conditions with regards to recognition accuracy. However, it was hypothesised that there may be a difference between the two types of composites generated.

The featural-assembly composites were chosen at lower rates than the whole-face composites (48.11% vs. 51.89%, respectively). However, the results indicated no significant difference between the two types of constructions. The featural-assembly composites were 0.86 times less likely to be matched to the target than the whole-face composites. This result, however, is interesting in that participants who selected the features using a piecemeal method produced a composite that was equal in quality to the whole-face composites. Thus, two deductions may be made: Selection of features during construction (which occurs with the manual and computerised composite systems) does not necessarily account for poor quality composites, as the standard whole-face composites were matched equally poorly. The second possibility is that feature selection accounts for *all* the misinformation displayed in the whole-face construction. However, many participants in the feature construction groups did not select isolated face lines and shading for their composites. This is possibly due to the fact that these are parts of a face that add more to the holistic representation of the composite. The whole-face group were able to select these and possibly make their composite look holistically more like the target. Furthermore, the featural-assembly group were exposed to configural aspects of the face when they were required to assemble the features into a whole-

face. Perhaps a better matching task would be to match the individual features between the two composite types.

The mean reaction selection times for both composite types were not significantly different. Thus, participants were selecting the different types of composites at the same pace. It was not easier to decide on one type of composite, rather, participants deliberated over which composite they thought best matched the target.

This experiment supports other literature which suggests that composites rarely depict the target they are intended to resemble (Davies & Valentine, 2007; Frowd et al., 2005). If chance is 50%, participants selected between the types of construction at equal frequencies. Thus, it is possible that both the featural-assembly and whole-face composites were so poorly constructed that participants resorted to guessing. Future research should have the composites rated on a scale, or matched to targets and fillers in the lineup to get an indication of composite quality for each condition on its own, and not in relation to the other construction type.

CHAPTER FIVE: GENERAL DISCUSSION

Limitations and Future Research

There were some limitations to this study. Firstly, for the recognition experiments (Experiment 1 and Experiment 2A), the photographs used at encoding and recognition should have been different poses. Research has suggested that it is advisable to use different photographs at encoding and recognition because the target needs to be recognised, and not merely the photograph (Bruce, 1982). It is unlikely that the perpetrator will look exactly the same as he did in the crime, and it is even more unlikely that the encoding event will be a static photograph. Bruce found that participants recognised the target quicker and more accurately if the picture remained unchanged. However, as previously discussed, it was not feasible to use a lineup of a different view apart from a frontal view, as the nature of the feature lineups were in a frontal position. However, the images in this study were manipulated at encoding and recognition. They were not identical images at the exposure and test phases. The horizontal view, clothes, colour, and background were all manipulated. One limitation is that the target could have been exposed in the three-quarter position. This may have reduced the accuracy and ceiling effects. However, the unforced-choice results did not indicate a ceiling effect. If the photograph pose was affecting recognition accuracy, conservative participants would have been more certain of their lineup selection. Furthermore, Wells et al. (2005) used same-pose photographs in their experiment, yet still managed to obtain 10% accuracy for composite constructors. There was nothing spurious about the lineup photographs in this experiment, as the results remained the same even when one set of possibly biased lineups were removed. Wells et al. did not report their effective size and lineup bias. Future research should encourage all authors to publish their lineup bias,

as this can influence the outcome of recognition studies. There was also nothing idiosyncratic or distinctive about the targets, as accuracy rates were evenly distributed across the targets.

The composite matching task could have been conducted differently. Each composite could have been displayed on its own with the entire target present lineup. Results may have indicated if composites matched other members of the lineup better than the target. This task gave no indication of the quality of the composite. However, as this paper intended to investigate another method of composite construction, the composite matching task was only implemented to see if one method resulted in higher quality composites. A total of 72 composites had to be rated. Participants may have become fatigued if required to rate each composite. Nevertheless, more descriptive data might have reflected if one composite was of poorer quality than the other.

The purpose of selecting features in isolation was to investigate the results of feature selection without being able to reference them in comparison to other features. For example, a nose might look bigger than it actually is because of a relatively small mouth. This task aimed to investigate if isolated features are less biased than features selected in a whole face. The composite matching task should have also focused on matching separate features. For example, participants in the whole-face construction condition could have written down the number of their individual features used on their composite. A direct feature-to-feature comparison could have been made. It may be that whole-face composites were matched to the target because participants were able to insert holistic-type lines and shading on the composite, which could have made them appear more human.

Experiment 2A did not make use of the featural lineups used in Experiment 1 as the results indicated that the task was too difficult. However, if participants in the isolated-featural condition were able to select their features, and were thus in a featural mode of processing, they may have performed better on the featural lineup task than on the whole face

lineup task. This could be investigated in a later study. It was assumed that they would have performed as poorly as the controls in Experiment 1. However, the controls in Experiment 1 encoded the face holistically, by rating the face on several traits. Thus, the controls might have been in a holistic processing mode, resulting in poor recognition rates in Experiment 1.

University of Cape Town

General Discussion

The first experiment in this paper investigated whether participants are able to accurately select isolated features (not embedded in a face) out of a multiple-choice lineup. The results were apparent; participants are able to select a target's feature at chance levels. These feature lineups resemble multiple channel lineups (Pryke, Lindsay, Dysart, & Dupuis, 2004). In this method, multiple lineups are presented to the witness, for example: a frontal photograph lineup, a sound clip of voices, or a body lineup. A correct identification is made if the witness is able to accurately identify the target in all of the lineups. In an ideal situation, participants should be able to identify all the features of the target. However, as evidenced in Experiment 1, this is not the case. The chance of participants accurately selecting from all six lineups is $(1/6)^6$. Participants are unable to correctly identify the target in any of the features. The results they obtained are tantamount to random guessing. This suggests that participants are unable to select isolated features that are not embedded in a face. This supports previous research conducted (Searcy & Barlett, 1996; Tanaka & Farah, 1993) which suggests that features cannot be identified in isolation and are greatly influenced by their relationship with other features. Furthermore, it is not a natural task to select disembodied features out of a lineup. Many of the features look too similar, possibly making the task too difficult for participants. During the encoding session, participants were exposed to the target face for five seconds, and then engaged in a holistic encoding task to ensure the face was properly encoded and deeply processed. Craik and Tulving (1975) suggest that tasks which induce a deeper mode of processing (semantic, as opposed to structural) will result in a deeper encoded stimulus. However, the results in Experiment 1 point more towards a parallel mode of processing. It may not necessarily be that objects that are semantically encoded are also structurally encoded. Thus, if a face is semantically encoded, it does not have to follow that it will also be satisfactorily structurally encoded. This further fits with the work of Tulving

(1972) (as cited in Groome, 2006) and the ESP. This suggests that objects are better retrieved if they are recalled in the same manner in which they were encoded. Thus, as the target face was encoded holistically, participants were unable to recall the individual features. This experiment should have implemented a group which received structural encoding, as opposed to holistic encoding. This would fit with the face recognition work of Winograd (1976) in which participants who semantically encoded a face were able to identify the target more often than participants who rated the face on structural characteristics. However, since Winograd did not implement a featural lineup, further investigation of this theory may be worthwhile. Experiment 1 indicated that if participants are unable to correctly select the target feature from a six-option choice, it might be a difficult task to construct a correct representation of the composite.

In Experiment 2A, although there appeared to be a significant difference between all the groups, only the control and isolated-featural groups were significantly different. There was no significant difference between the control and the whole-face groups. This does not fit with the Wells et al. (2005) finding, in which the control group obtained 84%, and the construction group achieved only 10%. This research implemented an encoding time of five seconds, compared to 180 seconds that Wells et al. used. It appeared (in this experiment) that selecting individual features did contaminate memory. However, it was not memory contamination in the way that Wells et al. explained it. They suggested that constructing a composite contaminates memory to such an extent that an incorrect filler is selected. However, this experiment showed that participants chose to reject the lineup rather than make an incorrect selection. Participants seemed to be more conservative and “err on the side of caution”. This was evident when participants gave their reason for lineup rejection. Both the featural-assembly and whole-face constructors indicated that they thought they knew who the target was, but they were not very confident in their decision. However, when they were

forced to choose, all three construction groups performed significantly better in recognition accuracy. Thus, participants *did* know who the target was; they had just implemented a stricter response criterion.

The reason for this is unclear. It is possible that a VOE effect might have resulted in participants utilising a higher criterion. However, the VOE typically applies to verbal descriptions of a face, as describing specific features can hamper access to the face memory. Meissner et al. (2001) suggested that participants who selected incorrect features were more likely to do worse in a recognition task. Indeed, Holland et al. (1994, as cited in Maskow et al., 2007) found that participants who constructed low quality composites were more prone to misidentifications than participants who constructed composites of a higher quality. Thus, constructing a poor quality composite may overshadow access to target memory. Participants who are forced to elaborate information (such as composite construction) frequently showed impaired recognition (Meissner et al., 2001). Ackil and Zaragoza (1998) note that self-generated misinformation is more likely to be internalised and misplaced. Information that is self-generated is likely to be monitored as coming from a different source. Thus, participants who construct a composite may internalise the image (with the misinformation). When it comes to the recognition task, participants retrieve that information and remember the composite and not the target. Participants may then attempt to match the mental composite to anyone in the lineup. The composite would therefore appear to induce a visual overshadowing effect (Kempen, this publication). Because access to the memory trace is blocked, participants may doubt their confidence about their selection. The control group took approximately 8.66 seconds to make an accurate decision, while the whole-face construction group took approximately 15.24 seconds to indicate their selection in the lineup. As only the unforced-choice reaction times were documented, participants who constructed a composite took roughly double the amount of time to make their decision and still rejected

the lineup. It may be that participants are trying to access the memory trace, but because access is prevented, participants are unsure if they know who the target is, and thus reject the lineup. As discussed in the introduction, we have two distinct streams of processing: the ventral stream and the dorsal stream. Evidence suggests that the ventral stream is concerned with recognising objects, while the dorsal stream is more suited to detecting change in the environment (Beck et al., 2005; Zeki, 2003). Participants may be slower at responding because they are not able to automatically recognise the target, using the ventral stream. It may be possible that they are trying to match the members of the lineup to their memory of the composite. They are thus trying to detect the changes and differences in both the target face and the composite. The discrepancy in reaction time may therefore be due to the different streams used for processing.

The aim of these studies was to investigate the underlying processes concerned with composite construction. It also aimed to find if composite production does have a negative effect on recognition, as postulated by Wells et al. (2005). One cannot explain the large detrimental effects found in the Wells et al. paper. Indeed, Meissner and Brigham (2001a) found that participants who constructed a composite were 1.56 times more likely to identify the target. However, a more recent meta-analysis found that a control group is 1.91 times more likely to select the target, when compared with the composite construction groups (Tredoux et al., 2010). The results of this paper initially pointed in the direction of the Wells et al. findings. However, when the forced-choice results were analysed, composite constructors (from all groups) performed no worse, and no better, than the control group. The results of the unforced-choice decisions *do* have some bearing, however. In a real life situation, law enforcement will not force a witness to make a lineup selection. As seen in the isolated-featural group, participants only obtained 52.78% accuracy. If composites are making witnesses this conservative, guilty suspects may be let free half of the time. This

paper aimed to investigate what processes could account for this decrease in recognition confidence.

It was initially hypothesised that the configural aspects of the construction may hamper memory. The hypothesis suggested that constructing the features into a whole face might result in the entire face being re-encoded and remembered. However, this hypothesis was not entirely confirmed or rejected. The isolated-featural group was more likely to reject the lineup which suggests that it may be the featural aspect of constructing a composite which results in a criterion shift. The featural-assembly and whole-face groups did not differ significantly from the control group; however, they did perform much lower. Moreover, both featural-assembly and whole-face groups performed similarly throughout the experiment. Their composites were also rated to be of equal quality to each other. Thus, selecting the features first, and then configuring them into a face, did not hamper the featural-assembly group. It is impossible to imagine an entirely configural composite system, but the aim of this experiment was to isolate the featural aspect of construction from the composite process. As discussed earlier, much literature has advocated that featural composite systems are not a natural method of composite construction. If we look at the initial analyses, it appears that isolated, featural construction makes participants more conservative regarding lineup selection. This is evidenced by the featural-assembly group who did not differ significantly to the control group. Being able to configure the features into a face appears to make participants less conservative. However, it may be that the whole-face composite constructed is of such poor quality that it cannot possibly compete with the memory of the original target. For example, Fallshore and Schooler (1995) found that the VOE did not occur when participants were required to describe an out-group face. In the same way that we use different processes to perceive and process faces that are distinctly different from our own, it may be that low quality composites do not elicit the VOE.

This paper was not able to replicate the Wells et al. (2005) finding of impaired recognition; however, it did support earlier findings (Kempen, 2009; Yu & Geiselman, 1993) that composite construction does appear to hamper recognition, but only in the sense that it makes participants shift their response criterion and become more conservative. This would need to be further investigated in future research as this widespread practice of composite construction has real life consequences. If composites are making witnesses misidentify the target (as Wells et al. suggest), then innocent individuals may be wrongfully incarcerated. If, however, composites are making witnesses more likely to refuse a selection, then guilty perpetrators may be set free.

University of Cape Town

Conclusion

In closing, these experiments aimed to investigate the processes of composite construction that may account for a decrement in recognition. If composite production *does* result in a decrease in lineup hits, this needed to be investigated as it has real-world consequences. If eyewitnesses are making false alarms, they could be sending innocent individuals to jail. The processes surrounding composite construction were investigated in these experiments. People are unable to accurately select features from a lineup, only being able to do so at levels equal to chance. This may suggest that people who are required to construct a composite may have difficulty selecting appropriate features for use in assembling a face. Previous literature has suggested that the featural nature of composite systems may be accounting for the decrease in target recognition. They suggest that parsing features together is an unnatural process, as faces are encoded using configural and holistic processes. As discussed in the review of the literature, composite system quality seems to be negatively correlated with recognition accuracy. As composite system quality increases, so recognition appears to decrease. Thus, composites that appear more “face-like” and “human-like” may be contaminating memory. Participants may encode this “new” composite representation as being a veridical depiction of the target. Thus, it was hypothesised that if faces are encoded configurally and holistically, being in the same mode of processing during construction might open participants up to configural and holistic contamination. It was hypothesised that featural processing of a face is far removed from natural, configural encoding; thus, featural construction should not hamper recognition. This has never been properly investigated as it is difficult to imagine *only* a featural composite system, or *only* a configural composite system. However, this experiment attempted to remove the featural construction from the composite production process. It was initially hypothesised that it may be the configural aspects of construction that hampers recognition, as viewing a whole complete composite may result in

that face being re-encoded as a veridical representation of the target. Other literature has shown that viewing a composite may result in decreased recognition (Kempen, 2009; Wells et al., 2005). This experiment showed that only the isolated-featural group differed significantly from the control group. This is opposite to what I expected to find. However, this decrease in hits was not coupled with an increase in false alarms (in which an innocent person may be incarcerated) but rather an increase in incorrect rejections (allowing guilty suspects to be set free). Participants who engaged in isolated-featural construction appeared to become more conservative. SDT suggests that participants are more uncertain about their decision and set a higher response criterion threshold. When they are forced to choose, the isolated-featural group achieves the same hit rate as the control group. It may be that participants in the isolated-featural group felt forced to confabulate and select extra features. This confabulation may result in misinformation which is then internalised and visually overshadows the memory for the original target. This could explain why participants exposed to a misleading composite also perform poorly – if they see the composite as an accurate attempt of the target, they may encode that as the original memory. When faced with the lineup, they are unable to monitor the sources of the misinformation, so instead reject the lineup.

The isolated-featural group did not perform as expected. Indeed, when the featural-assembly group were required to assemble their features into a full composite, the significant effect disappeared. Moreover, it appeared as though engaging in configural processing of a full face benefited recognition. However, the whole-face composite may have been of such poor quality that it offered no competing source of misinformation. The two featural composite construction groups (isolated-featural and featural-assembly) rejected the lineup at significantly higher levels compared to the control group. Interestingly, there was no significant difference in lineup rejection between the control and standard, whole-face

construction groups. Thus, piecemeal feature selection *does* appear to make participants more conservative in their lineup selection. However, participants in the featural-assembly and whole-face groups rated their confidence lower than the control group. Both construction groups also had the highest proportion of “not confident” responses as a reason for lineup rejection.

This leads to a peculiar fork in the road. The evidence indicates that engaging in configural spacing of features leads to participants becoming less confident. This supports the finding by Clare and Lewandowsky (2004) in which holistic verbalisation resulted in an increase in lineup rejections. On the other hand, the isolated-featural constructors had significantly fewer hits than the control group due to a criterion shift. But what is causing this criterion shift? If we presume that configural groups experience a criterion shift due to decreased confidence, what accounts for the criterion shift in the isolated-featural group? This group had the highest proportion of “definitely not present” lineup rejection responses. Thus, some innate property in featural construction must have made them believe that the target was not present. If featural and configural processing *are* distinct processes, then some other explanation (aside from decreased confidence) should account for the isolated-featural performance. It was suggested earlier that selecting individual features may lead to over-elaboration and confabulation of misremembered features. Without the presence of other helping features in context, participants may have to rely on many specific memories for each feature, as opposed to one memory for the entire face. After selecting all the individual features, participants may recode this information as belonging to another source. This may explain why participants thought the target was not present when they were required to give a reason for lineup rejection. Meissner et al. (2001) found that when participants generated fewer incorrect features, they were more likely to have fewer misidentifications. With composite construction being a difficult task, there is a risk of participants generating more

overall features which could result in an increased opportunity for those features to be incorrect. More particularly, this may be the case with isolated featural construction. This would result in higher misidentifications (or misses). Participants are unable to select a target feature from a lineup, performing only at chance rates. This may suggest that participants are guessing or confabulating features when it comes to composite construction. Finger and Pezdek (1999) instructed participants to remember every detail about the face and provide an extremely detailed description. Participants experienced a greater VOE than participants who engaged in free recall. Thus, when participants are constructing a standard, whole-face composite, they may rely on other features, and the spacing between features to compensate for the other features that they cannot remember. However, we can only speculate about these theories. This needs to be investigated in future studies. Previous research suggests that verbalisation is a left hemisphere operation, whereas face recognition is a predominantly right hemisphere function. Although feature differentiation and selection bears some resemblance to verbalisation, more research needs to be conducted to investigate whether a visual overshadowing effect (VIOE) occurs (Kempen, this publication). This appears to occur slightly during composite construction. Research has indicated that even being exposed to a composite can lead to a decrement in recognition (Brown, Deffenbacher, & Sturgill, 1977; Comish, 1987; Kempen, 2009; Wells et al., 2005). Visualisation of a composite leads to a partial decrease in recognition accuracy. The original memory trace may be blocked, or masked, by the whole face composite. This results in participants shifting their criterion threshold in a recognition task. This may also be a reason why more detailed composite systems experience a decrease in target recognition. These later, more detailed systems are, ironically, designed to eliminate the need for a third-party operator, and thus, verbalisation. However, being exposed to more detail requires participants to recall many more specific

details which they may not have encoded. Having to visualise many more features may lead to an over-elaboration and confabulation of features.

Future research should be conducted to further develop holistic composite systems. Although participants *did* know the correct answer when forced to choose, law enforcement would not force an eyewitness to make a decision. This research suggests that participants are more willing to err on the side of caution and set a guilty suspect free, rather than imprison an innocent individual.

University of Cape Town

References

- Ackil, J. A., & Zaragoza, M. S. (1998). Memorial consequences of forced confabulation: Age differences in susceptibility to false memories. *Developmental Psychology, 34*, 1358-1372.
- Aspley Limited. (1993). E-fit [Computer software]. Hatfield, U.K.: Aspley Limited.
- Baizer, J. S., Ungerleider, L. G., & Desimone, R. (1991). Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. *Journal of Neuroscience, 11*, 168-190.
- Beck, D. M., Muggleton, N., Walsh, V., & Lavie, N. (2005). Right parietal cortex plays a critical role in change blindness. *Cerebral Cortex, 15*, 1736-1741.
- Bartlett, J. C., & Memon, A. (2007). Eyewitness memory in young and old adults. In R. Lindsay, D. Ross, J. Read, and M. Toglia (Eds.), *Handbook of Eyewitness Psychology: Memory for People* (pp. 309-338). Mahwah, NJ: Lawrence Erlbaum.
- Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology, 25*, 281-316.
- 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. (1988). *Recognizing Faces*. Sussex, U.K.: Lawrence Erlbaum.
- Cabeza, R., & Kato, T. (2000). Features are also important: Contributions of featural and configural processing to face recognition. *Psychological Science, 11*, 429-433.
- Christie, D., & Ellis, H. (1981). Photofit constructions versus verbal descriptions of faces. *Journal of Applied Psychology, 66*, 358-363.

- Clare, J., & Lewandowsky, S. (2004). Verbalizing facial memory: Criterion effects in verbal overshadowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 739-755.
- Comish, S. (1987). Recognition of facial stimuli following an intervening task involving the Identi-Kit. *Journal of Applied Psychology*, *72*, 488-491.
- Cote, P. (2005). *FACES: The Ultimate Composite Picture (Version 4.0)* [Computer software]. Saint Hubert, Quebec, Canada: Interquest.
- Craik, F., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology, General*, *104*, 268-294.
- Davies, G., Ellis, H., & Shepherd, J. (1977). Cue saliency in faces as assessed by the Photofit technique. *Perception*, *6*, 263-269.
- 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*, 35-42.
- Davies, G. M., & Valentine, T. (2007). Facial composites: Forensic utility and psychological research. In R. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of Eyewitness Psychology Vol. 2* (pp. 59-86). Mahwah, New Jersey: Lawrence Erlbaum and Associates.
- 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*, *85*, 119-124.
- Deffenbacher, K. A. (1983). The influence of arousal on reliability of testimony. In S. M. A. Lloyd-Bostock and B. R. Clifford (Eds.), *Evaluating witness evidence* (pp. 235-251). Chichester: Wiley.

- Desimone, R., Albright, T. D., Gross, C., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *The Journal of Neuroscience*, *4*, 2051-2062.
- 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: Why descriptions impair face recognition. *Memory & Cognition*, *25*, 129-139.
- Duchaine, B. C., Dingle, K., Butterworth, E., & Nakayama, K. (2004). Normal greeble learning in a severe case of developmental prosopagnosia. *Neuron*, *43*, 469-473.
- Dumbell, E. (2008). *Face composite production effects on witness memory: Case not closed*. Unpublished Honours' thesis, University of Cape Town.
- 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*, 951-962.
- Ellis, H., Davies, G., & Shepherd, J. (1978). A critical examination of the Photo-fit system for recalling faces. *Ergonomics*, *21*, 297-301.
- Fallshore, M., & Schooler, J. (1995). Verbal vulnerability of perceptual expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1608-1623.
- Finger, K., & Pezdek, J. (1999). The effect of the Cognitive Interview on face identification accuracy: Release from verbal overshadowing. *Journal of Applied Psychology*, *84*, 340 - 348.
- Fisher, R. P., & Craik, F. I. M. (1977). Interaction between encoding and retrieval operations in cued recall. *Journal of Experimental Psychology: Human Learning and Memory*, *3*, 701-711.

- Frowd, C. D., Carson, D., Ness, H., Richardson, J., Morrison, L., McLanaghan, S., & Hancock, P. (2005). A forensically valid comparison of facial composite systems. *Psychology, Crime and Law, 11*, 33-52.
- Frowd, C. D., Hancock, P. J. B., & Carson, D. (2004). EvoFIT: A holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on Applied Perceptions, 1*, 19-39.
- Frowd, C. D., Pitchford, M., Bruce, V., Jackson, S., Hepton, G., Greenall, M., McIntyre, A., & Hancock, P. J. B. (2010). The psychology of face construction: giving evolution a helping hand. *Applied Cognitive Psychology, 25*, 195-203.
- Gibson, J. J. (1950). *The Perception of the Visual World*. Boston: MA: Houghton Mifflin.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15*, 20-25.
- Green, A. G., & Geiselman, R. E. (1989). Building composite facial images: Effects of feature saliency and delay of construction. *Journal of Applied Psychology, 74*, 714-721.
- Greenberg, S., & MacGregor-Hannah, M. (2010). Disruptive effect of holistic bias on processing of other-race faces following face categorization. *Perceptual and Motor Skills, 110*, 567-579.
- Gregory, R. (1966). *Eye and Brain*. London: Weidenfeld & Nicolson.
- Groome, D. (2006). *An Introduction to Cognitive Psychology: Processes and Disorders*. Sussex, U.K: Psychology Press.
- Hancock, P. J. B. (2000). Evolving faces from principal components. *Behavior Research Methods, Instruments and Computers, 32*, 327-333.
- Haxby, J. V., Hoffman, E.A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*, 223-233.

- Hayward, W. G., Rhodes, G., & Schwaninger, A. (2008). An own-race advantage for components as well as configurations in face recognition. *Cognition, 106*, 1017-1027.
- Innocence Project. (n.d.). *Facts on Post-Conviction DNA Exonerations*. Retrieved October 30, 2011, from <http://www.innocenceproject.org/Content/351.php>
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience, 17*, 4302-4311.
- Kempen, K. (2009). *It's the thought that counts: Face composite production can hamper recognition performance*. Unpublished Honours' thesis, University of Cape Town.
- 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.
- Lindsay, R. C. L., Wallbridge, H., & Drennan, D. (1987). Do the clothes make the man?: An exploration of the effect of lineup attire on eyewitness identification accuracy. *Canadian Journal of Behavioural Science, 19*, 463-478.
- Loftus, G. R., & Harley, E. M. (2005). Why is it easier to identify someone close than far away? *Psychonomic Bulletin and Review, 12*, 43-65.
- Loftus, E. F., & Greene, E. (1980). Warning: Even memory for faces may be contagious. *Law and Human Behavior, 4*, 323-334.
- Loftus, E., & Loftus, G. (1980). On the permanence of stored information in the human brain. *American Psychologist, 35*, 409-420.
- Loftus, E., & Palmer, J. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13*, 585-589.

- Macho, S., & Leder, H. (1998). Your eyes only? A test of interactive influence in the processing of facial features. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1486-1500.
- Macrae, C. N., & Lewis, H. L. (2002). Do I know you? Processing orientation and face recognition. *Psychological Science*, 13, 194-196.
- Malpass, R. S., Tredoux, C. G., & McQuiston-Surrett, D. (2007). Lineup construction and lineup fairness. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology: Memory for people* (pp. 155-178). Mahwah, NJ: Erlbaum.
- Maskow, N., Schmidt, H., Tredoux, C., & Nunez, D. (2007). *Face composite production does not affect eyewitness identification accuracy*. Unpublished manuscript, University of Cape Town.
- Mauldin, M. A., & Laughery, K. R. (1981). Composite production effects on subsequent facial recognition. *Journal of Applied Psychology*, 66, 351-357.
- Meissner, C., & Brigham, J. (2001a). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology*, 15, 603-616.
- Meissner, C., & Brigham, J. (2001b). 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. A., Brigham, J. C., & Kelley, C. M. (2001). The influence of retrieval processes in verbal overshadowing. *Memory and Cognition*, 29, 176-186.
- Meissner, C. A., Sporer, S. L., & Schooler, J. W. (2007). Person descriptions as eyewitness evidence. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology: Memory for people* (pp. 1-34). Mahwah, NJ: Erlbaum.

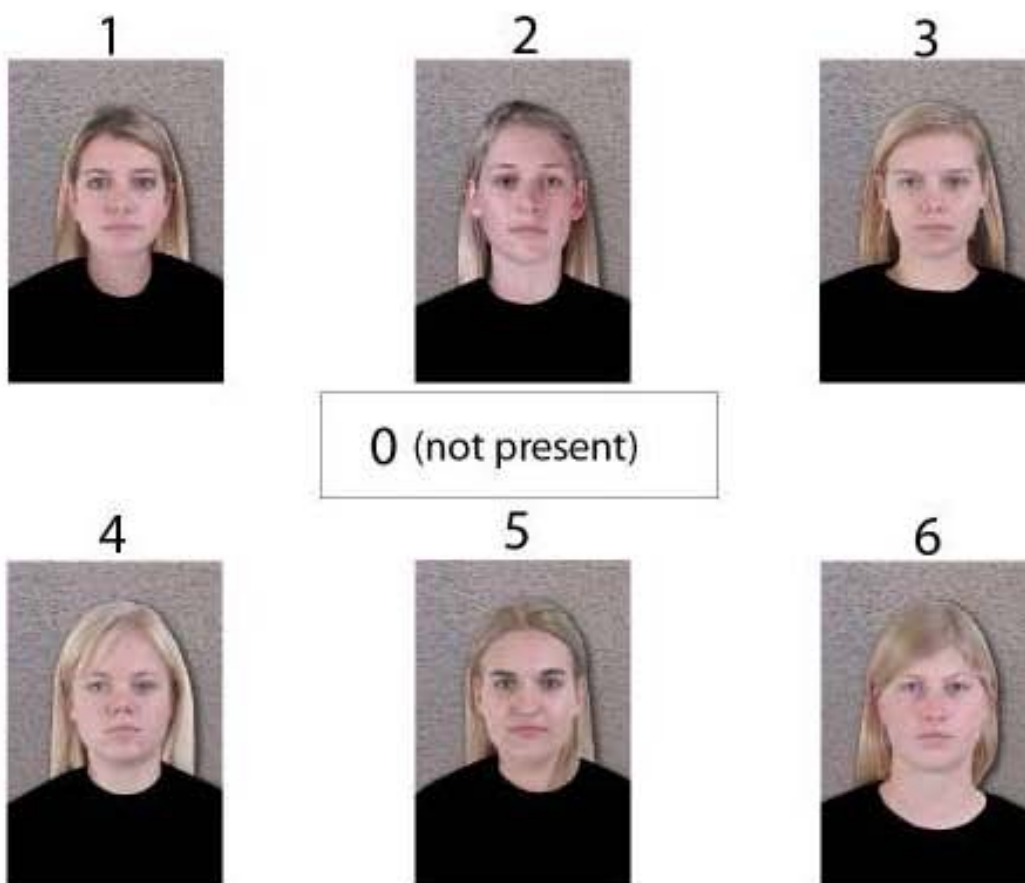
- Meissner, C. A., Tredoux, C. G., Parker, J. F., & MacLin, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory. *Memory and Cognition, 33*, 783-792.
- Michel, C., Caldara, R., & Rossion, B. (2006). Same-race faces are perceived more holistically than other-race faces. *Visual Cognition, 14*, 55-73.
- Milner, A. D., & Goodale, M. A. (1995). *The Visual Brain in Action*. Oxford: Oxford University Press.
- Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience, 9*, 555-604.
- Palmer, M. A., Brewer, N., & Weber, N. (2010). Postidentification feedback affects subsequent eyewitness identification performance. *Journal of Experimental Psychology: Applied, 16*, 387-398.
- Pryke, S., Lindsay, R. C. L., Dupuis, P. (2004). Multiple independent identification decisions: A method of calibrating eyewitness identifications. *Journal of Applied Psychology, 89*, 73-84.
- Rakover, S., & Teucher, B. (1997). Facial inversion effects: Parts and whole relationship. *Perception and Psychophysics, 59*, 752-761.
- Schacter, D. L. (1996). *Searching for memory: The brain, the mind, and the past*. New York: Basic Books.
- Schooler, J. W. (2002). Verbalization produces a transfer inappropriate processing shift. *Applied Cognitive Psychology, 16*, 989-997.
- Schooler, J. W. (2011). Unpublished results hide the decline effect. *Nature, 470*, 437-437.
- Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memory: Some things are better left unsaid. *Cognitive Psychology, 22*, 34-71.

- 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.
- Shaharazam, I. (1986). The Mac-a-Mug Pro manual [Computer software]. Milwaukee, WI: Shaharazam.
- Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, 2, 139-156.
- Shapley, R. (1995). Parallel neural pathways and visual function. In M. S. Gazzinga (Ed.), *The Cognitive Neurosciences* (pp. 315-324). Cambridge, MA: MIT Press.
- Shepherd, J. W., Gibling, F., & Ellis, H. D. (1991). The effects of distinctiveness, presentation time and delay on face recognition. *European Journal of Cognitive Psychology*, 3, 137-145.
- Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology*, 78, 22-33.
- Sporer, S. L. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law*, 7, 36-97.
- Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, 118, 315-327.
- Stanny, C. J., & Johnson, T. C. (2000). Effects of stress induced by a simulated shooting on recall by police and citizen witnesses. *American Journal of Psychology*, 113, 359-386.
- Stebly, N. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior*, 16, 412-424.

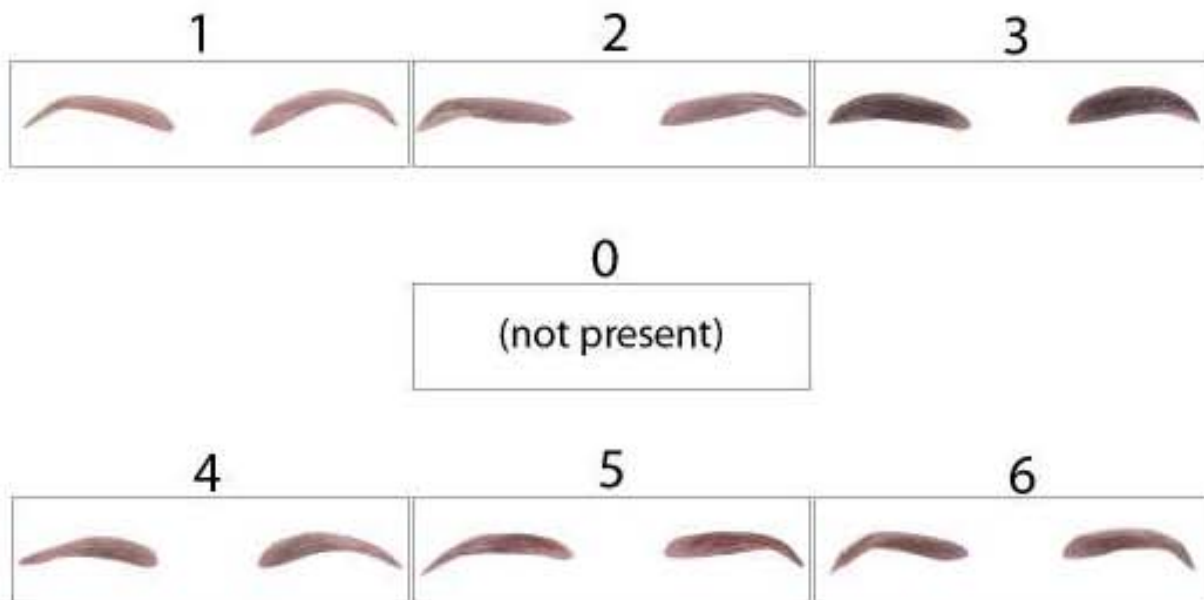
- Tanaka, J. W. & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology*, 46, 225-245.
- Tanaka, J. W., & Farah, M. J. (2003). The holistic representation of faces. In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes: Analytic and holistic processing* (pp. 53-74). Oxford: Oxford University Press.
- Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory and Cognition*, 25, 583-592.
- Taylor, K. T. (2001). *Forensic art and illustration*. Boca Raton, FL: CRC Press.
- Topp, L. (2006). *Exploring composite construction and its effects on eyewitness memory*. Unpublished Master's thesis, University of Texas, El Paso.
- Tredoux, C. G., Kempen, K., Nortje, A., Christ, B., Hoogenhout, N., & Kaiser, A. (2010). *Do face composites contaminate memory? A meta-analysis*. Unpublished manuscript, Department of Psychology, University of Cape Town, South Africa.
- Tredoux, C. G., Nunez, D. T., Oxtoby, O., & Prag, B. (2006). An evaluation of ID: An eigenface based construction system. *South African Computer Journal*, 37, 90-97.
- Tredoux, C. G., Rosenthal, Y., da Costa, L., & Nunez, D. T. (1999). *Face reconstruction using a configural, eigenface-based composite system*. Paper presented at the Third Biennial Meeting of the Society for Applied Research in Memory and Cognition, Boulder, CO.
- Tulving, E. (1983). *Elements of episodic memory*. Oxford: Oxford University Press.
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 352-373.
- Valentine, T., & Bruce, V. (1986). The effect of race, inversion and encoding activity upon face recognition. *Acta Psychologica*, 61, 259-273.

- Wells, G. L. (1978). Applied eyewitness-testimony research: System and estimator variables. *Journal of Personality and Social Psychology, 36*, 1546-1557.
- Wells, G. L., Charman, S. D., & Olson, E. A. (2005). Building face composites can harm lineup identification performance. *Journal of Experimental Psychology: Applied, 11*, 147-156.
- Winograd, E. (1976). Recognition memory for faces following nine different judgements. *Bulletin of the Psychonomic Society, 8*, 419-421.
- Wogalter, M. S., Laughery, K. R., & Thompson, B. G. (1986). *Eyewitness identification: Composite construction on subsequent recognition performance*. Unpublished manuscript. University of Richmond.
- Yarmey, A. D. (2004). Eyewitness recall and photo identification: a field experiment. *Psychology, Crime and Law, 10*, 53-68.
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*, 141-145.
- Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16*, 747-759.
- Yu, C., & Geiselman, R. (1993). Effects of constructing Identi-Kit composites on photospread identification performance. *Criminal Justice and Behavior, 20*, 280-292.
- Zeki, S. (2003). The disunity of consciousness. *Trends in Cognitive Sciences, 7*, 214-218.

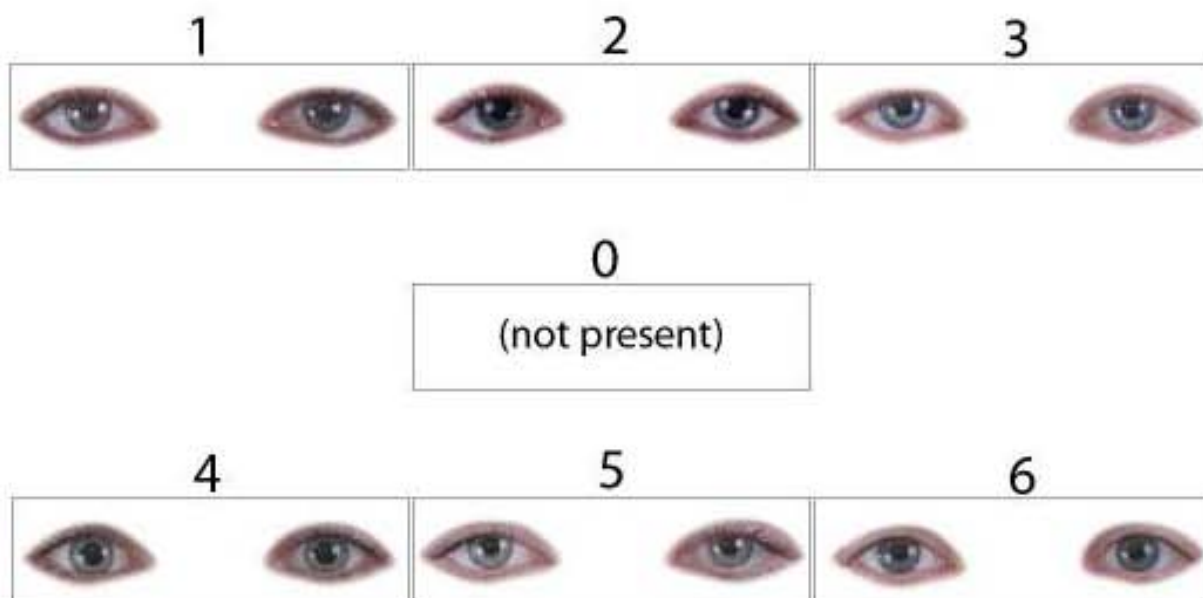
Appendix A
Whole face lineup, target 1 (position 3)



Appendix B
Feature eyebrow lineup, containing target 1 (position 2)



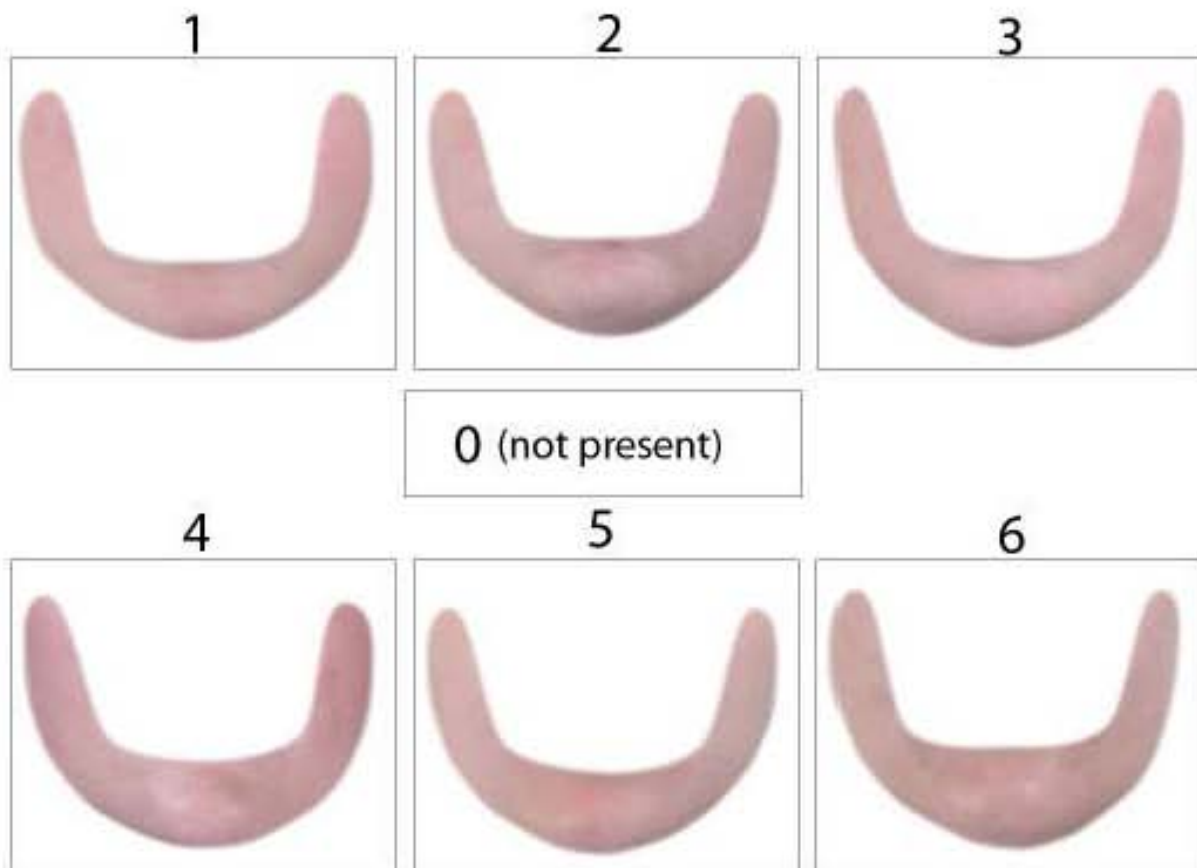
Feature eyes lineup, containing target 1 (position 6)



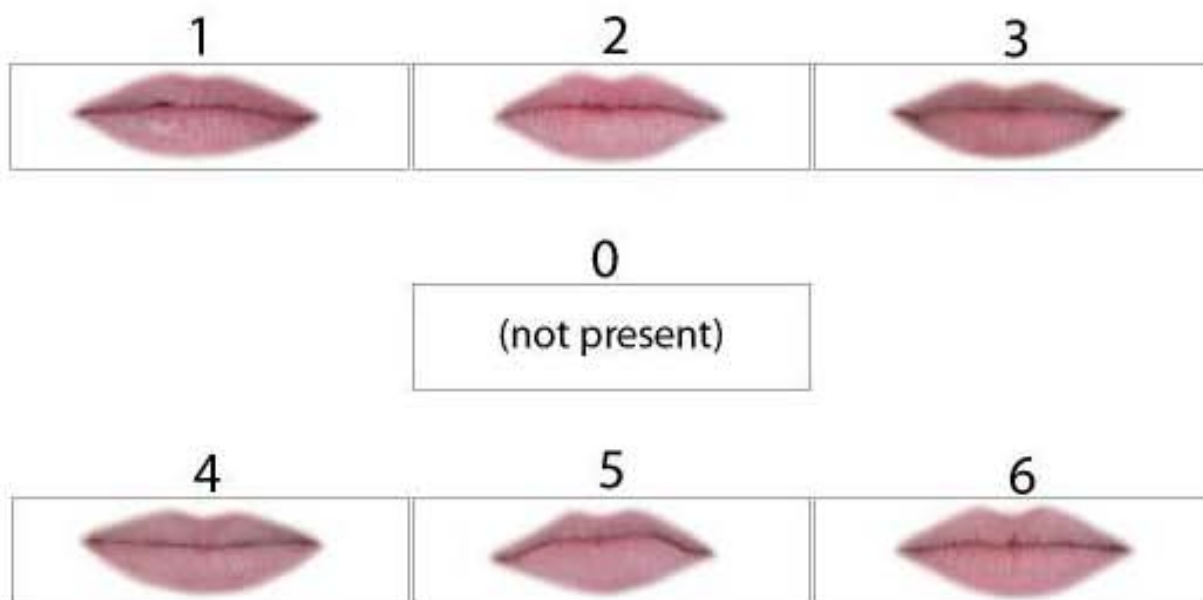
Feature hair lineup, containing target 1 (position 3)



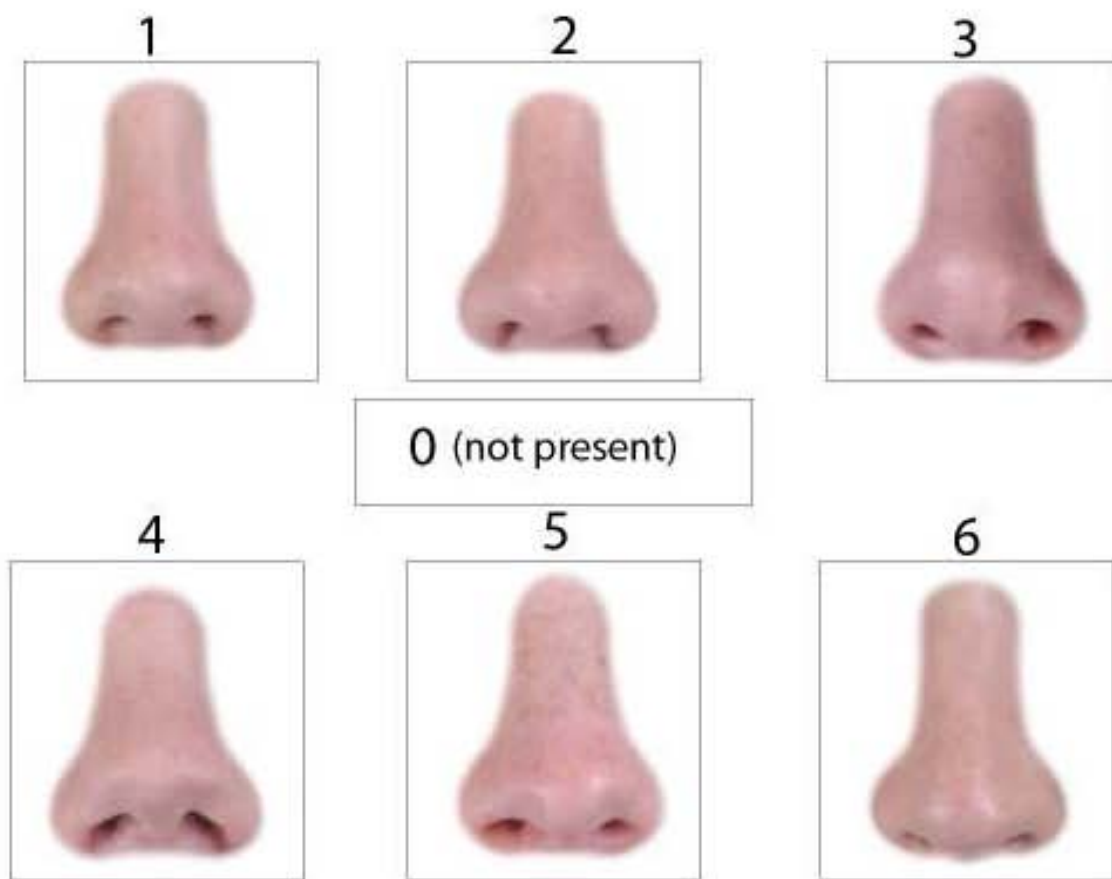
Feature jaw lineup, containing target 1 (position 5)



Feature mouth lineup, containing target 1 (position 4)



Feature nose lineup, containing target 1 (position 1)



Appendix C

Participant consent form

PARTICIPANT CONSENT FORM

REF #2009023

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.

Name: _____

Student No.: _____

Age: _____

Sex (please tick):

MALE

FEMALE

Race (please tick):

WHITE

BLACK

COLOURED

INDIAN

OTHER: _____

Country of birth: _____

Length of time residing in South Africa: _____

For your time and participation, you will receive **90 minutes** (3 units) towards your SRPP requirement. The first session will constitute **60 minutes** and the second session (two days from now) will take approximately **30 minutes**. If you fail to turn up for your second session, you will not be given your points from the first session. So **please** make a note of the date and time of your second session.

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: You will be required to rate faces on a variety of characteristics. During this study you may be required to complete various tasks. All these tasks are computer-based. Please take care when answering, and do so carefully, as your input and opinion is valued. 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 076 047 9726 (Kate).

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: _____

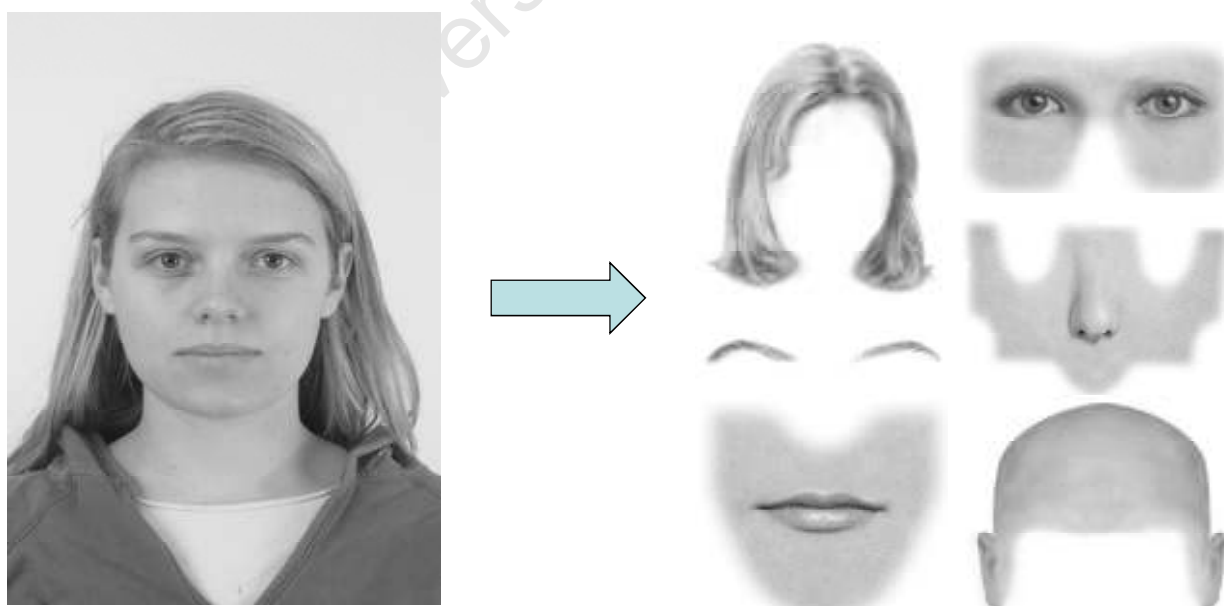
DATE: _____

Appendix D
Different methods of composite construction per group

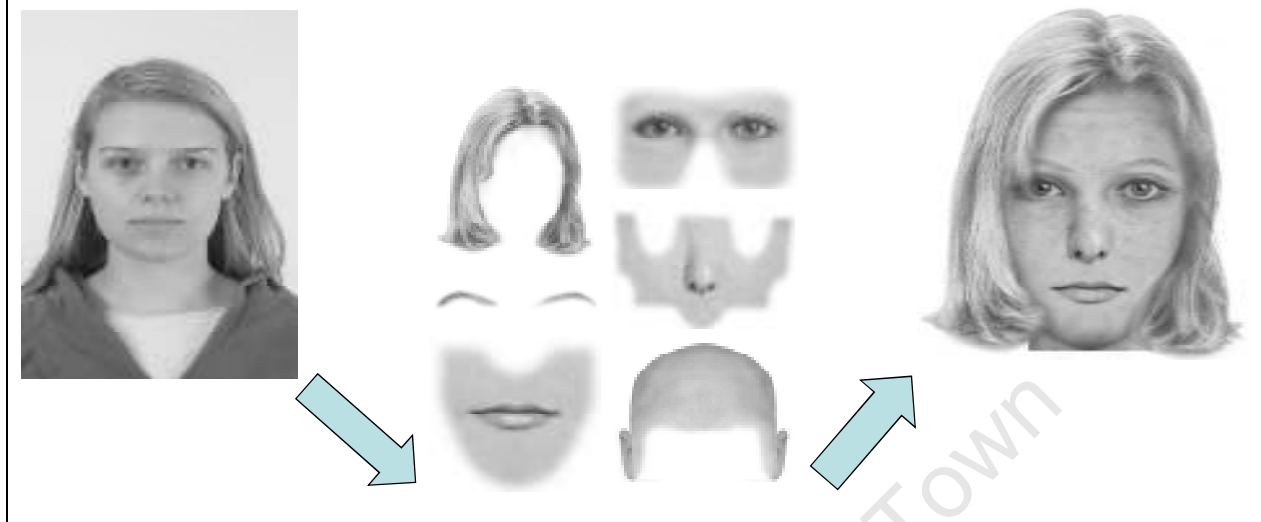
Whole-face construction: Participants construct a holistic, configural face



Isolated-featural construction: Participants select isolated features only – they never see those features in a whole face



Featural-assembly construction: Participants select isolated features – then spatially organise those features into a whole composite



University of Cape Town

Appendix E

Feature selection sheet, participants write the number of the specified feature in the space provided

NOTE: If no feature selection is made, please leave the section **blank.**

	Feature	Selection Number
	Hair	
	Head Shapes	
	Eyebrows	
	Eyes	
	Noses	
	Lips	
	Jaw Shapes	
	Moustaches	
	Beards	
	Goatees	
	Skin Tones	
	Forehead Lines	
	Eye Lines	
	Smile Lines	
	Mouth Lines	
	Chin Lines	
	Head Wears	
	Glasses	
	Moles	
	Scars	
	Piercings	
	Tattoos	

Appendix F

Screenshot of FACES 4.0 composite system, indicating the feature number above the relevant feature (feature – hair; number – 141)



Appendix G

Mockwitness match-to-description task, containing target 1 (position 1)

1



2



3



TASK: Which of these faces best matches the written description?
= female, blonde hair, blue eyes, rounded jaw, fair skin

4



5



6

