Self-Generated Misinformation in Perpetrator-Describing Techniques

Eden Clingman

CLNEDE001

Department of Psychology

Faculty of Humanities

University of Cape Town

Supervisors: Colin Tredoux and Kate Kempen

Word Count:

Abstract: 283 words

Main Body: 9754 words

Abstract

Eyewitnesses to crimes often provide crucial input in police investigations. In cases where the perpetrator is unknown, eyewitnesses can be required to create a composite of the perpetrator, in order to aid police in the apprehension of a suspect. However, research suggests that composites are often inaccurate (Kovera, Penrod, Pappas, & Thill, 1997), and that composite creators may be more likely to misidentify the target in a later recognition task (Kempen, 2009; Wells, Charman, & Olson, 2005). The theory of self-generated misinformation has been proposed to explain these phenomena. This theory hypothesises that in creating a composite, one must select features one may not remember, and there is therefore a confabulation of the face features that may restrict access to the original memory trace. This experiment aimed to investigate if this theory of self-generated misinformation could explain why composites may be so inaccurate and may negatively affect recognition. Participants either created a composite (a FACES or an ID composite) or produced a verbal description (a free-recall or elaborate description). The results suggest that, compared to other groups, creating holistic (ID) composites led to fewer target identifications, with participants rejecting the line-up and making no identification. When forced to choose, these participants would incorrectly select foils. However, ID composites also appear to be the best quality product alongside free-recall verbal descriptions. This research proposes that as ID composites require their creators to commit to a large amount of detail, the holistic image may hamper their memory of the original face, decreasing their confidence, and thereby resulting in increased rejections of the line-up. However, as the following discussion is tentative theorising, more research must be conducted to provide evidence for the theory of self-generated misinformation.

Keywords: composites; identification; misidentification; descriptions; eyewitness memory; contamination; misinformation

PLAGIARISM DECLARATION

PLAGIARISM

This means that you present substantial portions or elements of another's work, ideas or data as your own, even if the original author is cited occasionally. A signed photocopy or other copy of the Declaration below must accompany every piece of work that you hand in.

DECLARATION

1. I know that Plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

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

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

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

NAME: **Eden Clingman**

STUDENT NUMBER: **CLNEDE001**

ASSIGNMENT: **Research Project**

Signature: Eden Date: 17 November 2016

Table of Contents

Self-Generated Misinformation in Perpetrator-Describing Techniques

Facial recognition in cases of criminal prosecution has proven to be problematic. Eyewitnesses to crimes are often required to provide testimony in cases where police have little or no physical evidence. When witnesses are unfamiliar with the perpetrator of a crime and there is no established suspect, they may be asked by the police to construct a facial composite to help locate and apprehend the perpetrator. This is therefore a crucial step in the police process. A facial composite is a visual representation of the perpetrator, which is formed by the assembling of individual facial features (Frowd et al., 2005). Composite images are widely used by the South African police (Schmidt & Tredoux, 2006). For example, in the case *Mavangwana v S* (2010), a composite was created by the complainant, who witnessed a murder, to help the police identify possible suspects. This composite led the police to include Mavangwana as a suspect, and he was later charged and convicted of murder and attempted murder. However, research has shown that composites are often inaccurate in their depiction of target faces (Kovera et al., 1997), and may also negatively affect the recognition ability of the composite producer (Wells et al., 2005). Understanding why composites have these issues could help in the development of better police practices.

Eyewitness Misidentification

Eyewitness testimony can be unreliable, as witnesses may misidentify the perpetrators of crimes they have witnessed. Since 1989, 333 individuals have been exonerated using DNA evidence in the United States alone, with 72% of these overturned convictions originally being made on the basis of an eyewitness identification. Of these cases of misidentification, 16% occurred after witnesses were required to create a facial composite (Innocence Project, n.d.a.).

There are two main reasons why composites may lead to eyewitness misidentifications. Research has shown that composites can often be poor likenesses of the faces they were meant to depict (Kovera et al., 1997). This means that if they do not correctly depict the perpetrator, composites may inadvertently resemble an innocent individual. Some research has also shown that building composites can harm recognition performance, making eyewitnesses less likely to pick the correct perpetrator from a line-up (Wells et al., 2005).

The use of composites has real-world effects on eyewitness misidentification. For example, in 1984, Larry Johnson was wrongfully convicted of rape and sentenced to life in prison, largely based on an eyewitness misidentification. The victim in this case created a facial composite with the police, where she described her attacker as clean-shaven. This composite helped the police put a photo array together, in which Johnson's image was included. The victim identified Johnson as her attacker, in spite of his moustache, and also identified him in a line-up after his arrest. He was eventually exonerated 18 years later on the basis of DNA evidence (Innocence Project, n.d.b.). What is clear from this case, is that the composite created by the eyewitness was not a good enough likeness of the real perpetrator, because, as Johnson resembled the composite, his image was included in the photo array. The creation of the composite may also have affected the victim's memory, as she identified Johnson in both the photo array and the line-up.

Facial Processing and Composite Software Systems

There is some debate in the literature about how faces are processed. Even though some research has found that a face is perceived and encoded featurally, which involves encoding separate individual parts without their context (Macho & Leder, 1998), much research suggests that a face is encoded a whole (holistically and configurally) (Searcy & Bartlett, 1996; Tanaka & Farah, 1993; Young, Hellawell, & Day, 1987). Today, there are two main types of composite software systems that are based on these differing ways of processing. The more widely used are the featural composite systems, such as E-FIT, Mac-a-Mug and FACES, which contain libraries of individual adjustable features that can be inserted into a face and manipulated by size, colour, and position (Frowd et al., 2005).

The newer, holistic composite systems – which include programs such as ID, EvoFIT and EFIT-V – create composites through the selection and 'breeding' of whole faces. Users are presented with a set of random, synthetically generated faces, from which they must choose those faces that best resemble the target face. The selected faces are then 'bred' together to form a new smaller set of faces which are better likenesses of the target. This 'evolutionary' process is repeated several times until the user is satisfied with one of the generated faces (Frowd et al., 2005).

Effects of Composite Production on Recognition

In the past, only manual systems, in which users place acetates or jigsaw pieces of facial features together to form a face (Frowd et al., 2005), and sketch artists were used, whereas today there are a wide array computerised systems available. The older, manual methods of creating composites were shown, in some cases, to have a facilitating effect on memory. Mauldin and

Laughery (1981) found that recognition of a face increased when the participants produced an Identikit composite than if they did not (90% of the composite producers correctly selected the target, compared to 60% of the control group). However, Comish (1987) and Davies, Ellis, and Shepherd (1978) found that participants who created a composite made more recognition errors than control participants.

Research that has been conducted on the effects of composite construction on memory, using the computerised systems, has been mixed, with some authors finding a facilitating effect, no effect, or a detrimental effect on memory. Some research has found that constructing composite images, using the mechanical and computerised methods, has no effect on memory (Davies et al., 1978; Maskow, Schmidt, Tredoux, & Nunez, 2007). A meta-analysis found that the more recent, computerised systems had a facilitating effect on memory, improving participants' recognition rates on subsequent line-ups (Meissner & Brigham, 2001). Such systems allow manipulation of every aspect of the face, with each feature having hundreds of options for selection (Frowd et al., 2005). Other research, however, has shown that the computerised methods hamper recognition. Wells et al. (2005) found that building a composite negatively affected participants' ability to accurately identify the target. The composite producers could only select the target 10% of the time, compared to 84% for the control group. From this, the authors concluded that composites had contaminated memory. This effect has also been established in other studies (Kempen, 2009; Yu & Geiselman, 1993), and Kempen (2009) found that participants who had made a composite were less likely to attempt an identification at all. A recent meta-analysis found that individuals who did not produce composites were 1.91 times more likely to correctly identify the target than those who did, and this had a weighted effect size of 0.65, suggesting that composite production could account for 65% of the variation in identification decisions (Tredoux et al., 2010). However, it must be noted that other research has failed to replicate Wells et al.'s (2005) finding that composite production significantly negatively affects memory (Dumbell, 2008; Maskow et al., 2007).

The literature is clearly equivocal regarding the effects of composite production on memory. However, as there is currently no alternative method of locating unfamiliar perpetrators, if composites do have a detrimental effect on memory, the reasons for this effect must be investigated.

Composite Quality

The general consensus in the literature is that computerised methods are capable of producing good likenesses of a face when the participant is working directly from a photograph (Davies, van der Willik, & Morrison, 2000; Frowd et al., 2011). However composite quality is poor when relying on recent memory (Brace, Pike, Allen, & Kemp, 2006; Frowd et al., 2005) and is even worse when the memory is several days old (Frowd et al., 2011). Considering that real-life situations often necessitate the creation of a composite at least several days after a crime, this may suggest that composites constructed in real-world crimes may often not resemble the perpetrator.

The literature suggests that, of the modern systems, holistic systems produce better likenesses. Early research by Frowd, Hancock, and Carson (2004) found that featural systems produced more accurate composites than holistic systems. However, a later study found this to be the case only when the composite was constructed with the target in view, and that holistic systems produced more accurate composites than featural systems when constructing from memory (Tredoux, Nunez, Oxtoby, & Prag, 2007). Frowd et al. (2007) found that composites created using the holistic EvoFIT were three times more accurate than those of the featural ProFIT.

Composite Interference Theories

The *transfer inappropriate processing shift* theory postulates reasons why composite production may have a negative effect on recognition, and why composites may be so inaccurate. This theory suggests that if the mode of processing at memory recall matches the mode of processing at encoding, retrieval of information is facilitated (Schooler, 2002). Faces are processed and memorised holistically – that is, they are dependent on representations that do not have an internal part structure (Tanaka & Farah, 1993), whereas constructing facial composites is a feature-based process (Lech & Johnston, 2011). Thus constructing a composite in a featural mode of processing may interfere with the holistic processing that naturally occurs between encoding and recognition, resulting in decreased recognition rates.

However, an alternative theory has been developed. A meta-analysis conducted by Meissner and Brigham (2001) found that giving more detailed facial descriptions, a cued process that requires a significant amount of information, produces a greater decline in recognition performance than having to describe only what one remembers (free recall). In applying this to composites, Wells et al. (2005) theorised that the detail required to produce a composite necessitates witnesses guessing what certain features of the target face look like, as they do not remember them, and all features of the face have to be provided. It has been shown that verbal descriptions can lead to more accurate identifications than composite images (Christie & Ellis, 1981; Lech & Johnston, 2011), since, in these experiments, composite images required the provision of much greater detail than verbal descriptions. The inclusion of features that witnesses are unsure about may take the form of self-generated misinformation, and Finger and Pezdek (1999) have suggested that this misinformation may be internalised as being part of the target face. This may create a memory that is similar to and indistinguishable from the target source (Ackil & Zaragoza, 1998). Much psychological research has been conducted on the misinformation effect, demonstrating the susceptibility human memory has to misleading information (Loftus, 2005). This theory of self-generated misinformation may therefore describe how using featural systems can result in poorer recognition rates than holistic systems, which do not necessitate the remembering and selection of numerous individual features.

However, no research testing this theory experimentally has yet been conducted. In the theory of self-generated misinformation, creating a composite is similar to providing an elaborate verbal description, as both necessitate witnesses describing features they may not remember. Comparing these two conditions allows for the disentanglement of the visual aspect of composite production. Thus, if one were to compare the recognition rates of composite producers and individuals who provide an elaborate verbal description of a face, to that of individuals who provide a free recall verbal description, this theory could be assessed.

Significance

Misleading composites misattributed to innocent individuals, and incorrect or missed line-up identifications as a possible result of composite creation, can lead to serious miscarriages of justice. Research that has been conducted on the quality of composites and their effect on recognition is generally, equivocal. The current literature therefore cannot inform changes to current police practice, which might be necessary given the high numbers of overturned convictions. If this research shows that the elaborate verbal description and composite groups perform similarly on subsequent recognition, this may provide evidence in support of the theory of self-generated misinformation. This research may show that one form of creating a description of a target (such as the elaborate verbal description or the featural composite system) performs significantly better on the tasks than the others, which may identify better methods that could be used to apprehend suspects.

This research could also help establish if composite construction simply makes producers more conservative, as shown by Kempen (2012). If this is the case, and producers are still able to make a correct decision after having rejected the line-up, it could also inform police practice, in ensuring that witnesses who produce composites can make an appropriate line-up decision.

In summary – research has shown that composites can rarely be matched to the targets they are intended to depict. Research conducted on the effects of composite production on memory is equivocal, and this, along with the theory of self-generated misinformation, needs to be tested further. Exploring these theories may help inform composite software and update police practices so as to improve society's ability to correctly apprehend and convict criminals.

Research Aims and Question

The aim of this research is to discover which of the different methods of creating composites and verbal descriptions leads to increased target identification, decreased rates of line-up rejections, increased confidence ratings, and creates the best quality product. In particular, this study aims to answer the question: Does the theory of self-generated misinformation explain why composites are inaccurate and may negatively affect recognition? This research hypothesises that on recognition the composite construction groups will perform similarly to the elaborate verbal description group, which will both perform worse than the free recall verbal description group; that the ID (holistic) group will have better recognition performance than the FACES (featural) group; and that the composite groups will be more conservative in their recognition decisions.

Method

Creation of Line-ups

To create fair and unbiased line-ups for the main experiment of this research, two principles were adhered to: neither the target nor any foil should stand out from the other line-up members, and the foils should be equally good alternatives to the suspect (Malpass, Tredoux, & McQuiston-Surrett, 2007). According to Malpass et al. (2007), there are three steps which should be taken to achieve this, and these made up the following procedure.

Participants

Undergraduate Psychology students from the University of Cape Town were recruited through convenience sampling. Participation for the first step was acquired from friends of the researcher, and for the second and third, was motivated through the use of SRPP points.

Materials

Target stimuli. Eight initial target stimuli were used. These were photographs of white female faces, chosen for their indistinctive appearance to the researcher, between the ages of 18-25, with their hair tied into a ponytail. Frontal and three-quarter views were used depending on the requirements of each step, and the photographs were in full colour. They were standardised in height, width and pixel resolution. These faces were chosen because the face database contains a larger proportion of white female faces than any other group, providing more opportunities for similar foils. Additionally, the quality of composites produced by the FACES and ID programs was assessed in this study. As ID cannot accurately depict long hair, to avoid an inherent bias in the quality ratings of the composites produced using it, females with their hair tied back were used in the line-ups to allow for short hair to be created in the frontal view of a composite.

Presentation and answering materials. Stages one and three took the form of surveys created using Qualtrics (an online survey creator). The second stage consisted of participants searching through a database of photographs in the computer laboratory.

Procedure

Description generation. Eighteen participants completed an online survey, where they were presented with eight different faces, shown for five seconds each. Participants were asked to provide a verbal description of each face. The faces were presented in a random order for each participant, and they were in the frontal view so as to present the best-case scenario for memory retention.

After all of the faces were described, participants rated these faces, which were in view, on a Likert-scale of 1-7 (in order to ensure a decision) on different traits in order to establish their distinctiveness. The order of the traits and target faces were randomised for each participant.

Modal, or average, descriptions for each of the faces were created, by assessing which descriptors were used most often to describe each of the faces.

Foil selection. Thirty participants were seated at individual computer stations, and completed a consent form. They were presented with each of the eight faces and modal verbal descriptions, and had to choose 12 faces from the face database that they deemed similar to each of the target faces and descriptions. Both the target faces and the verbal descriptions were used because the literature supports the use of each of these in determining appropriate foils for line-ups (Malpass et al., 2007).

The five most frequently chosen faces for each target face were used as the foils for each line-up. Each line-up thus consisted of six faces, which has been established as the general line-up size used in the United States (Malpass et al., 2007). Photoshop was used to create eight different line-ups, in which the photographs of the faces were in a three-quarter view. Three-quarter images were used in the line-ups so that, in the experiment, the participants would not recognise the photograph (which was presented in frontal view) because of its idiosyncrasies, but rather the face itself (Bruce, 1982). To establish uniformity across the images, distinctive marks and jewellery were edited out. The line-up members were given the same background and T-shirt, to avoid clothing recognition. There were two orders of each line-up, with the target and foils being placed in a random order for each.

Line-up rating. One hundred and thirty participants completed an online survey, in which they were presented with each line-up and associated modal verbal description, and asked to choose which face they believe best matched the description. This was done to ensure that all of the faces adequately matched the descriptions used to create them. The line-ups were presented in a random order for each participant. Effective size, "*E*", an estimation of how many possible suitable foils are present in the line-up (Malpass et al., 2007), was calculated to assess the fairness of the eight line-ups. The bias of each line-up was also calculated, which involved assessing whether each line-up member was chosen at frequencies significantly different to chance. High bias is represented by fewer faces being selected significantly, whereas low bias is represented by more faces being selected significantly (Malpass et al., 2007). Four line-ups (and corresponding targets) with the highest effective size scores and least bias were selected for use in the experiment. The effective sizes and bias present in the line-ups are presented in Table 1.

Table 1

Effective Size and Bias Representation in the Possible Target Line-ups

	E	Bias
Target 1	4.69*	1 face
Target 2	2.55	2 faces
Target 3	4.58*	1 face
Target 4	3.10	1 face
Target 5	3.99*	2 faces
Target 6	3.81	1 face
Target 7	4.15*	2 faces
Target 8	3.41	3 faces

Note. Asterisk denotes the faces that were selected for use in the experiment.

Experiment

This experiment attempted to establish the relative accuracy of composites and verbal descriptions created using different techniques, and the effect the creation of these composites and verbal descriptions had on subsequent target recognition.

Participants

One hundred and sixty undergraduate Psychology students from the University of Cape Town were recruited for this experiment, through convenience sampling. A sample size of 122 was calculated on G-Power, using a power of 0.80 and an effect size of Cramer's V = 0.30. This is a moderate effect size, and was chosen to reflect Cramer's V = 0.23 attained in the similar study conducted by Kempen (2012). However, as the results attained in Kempen (2012) were only noticeable with 38 participants in each group, it was decided that there should be 40 participants per group, yielding a sample size of 160.

Participation was motivated through the use of SRPP points.

Materials

Line-ups. The four line-ups previously created were used (Appendix H). Four targets were used to protect against stimuli effects should any of the faces be particularly distinctive.

Presentation and answering materials. The experiment was run on computers using E-Prime 2.0, which records keyboard responses and reaction times.

Composite construction software. The computerised software programs FACES 4.0 and ID were used to create the composites.

Procedure

Encoding. Participants were randomised to one of the four conditions and received a random target face and line-up order. They were seated at individual computer stations and completed a consent form. An E-Prime 2.0 slideshow was opened, in which instructions were shown, and a target face was displayed for five seconds. Participants then engaged in crossword puzzles as a distractor task for five minutes. Subsequently, they rated the face on several characteristics to facilitate deeper encoding, after which they had to give a free recall verbal description of the face. Participants then completed a confidence assessment, in which they rated on a Likert-type scale how well they believed they could remember the target face.

Free recall (control) condition. The participants watched an unrelated video and were told that they would be asked questions following the video.

Elaborate verbal description condition. Participants provided an elaborate verbal description of the target face, which necessitated their filling in, on paper, an exhaustive list of features describing the face, that correspond to the FACES categories. This description generation list (see Appendix K) was based on the one used by the South African Police Services.

Training. Participants in the composite conditions watched a tutorial on how to use either FACES or ID, and were helped in constructing an unrelated practice face.

FACES condition. Participants created a composite of the target face using FACES.

ID condition. Participants created a composite of the target face using ID.

Confidence. Participants in the elaborate verbal description, FACES and ID conditions were once again asked how confident they were that they would be able to recognise the target face if her photograph were placed among five photographs of reasonably similar women.

Recognition. Two days later, all participants were emailed a unique online survey created using Qualtrics. Each survey corresponded to each participant's condition, target face, and line-up order. In this, they were asked to select the target face from a line-up. Even though the targets were always present in these line-ups, the participants were informed that the target may not be present and had the option to make no selection. If a participant rejected a line-up as not containing the target, they were later presented with the line-up again, and were forced to make a selection. Confidence in the decision made was collected for each time the line-up was presented.

Debriefing. Participants were fully debriefed at the end of the survey and were given contact details if they had any questions.

Quality Assessment Phase

The quality of the composites and descriptions produced was assessed using an online survey created with Qualtrics, which was distributed to 780 UCT students through a campus-wide announcement. This served to ascertain how useful the constructions and descriptions would be in aiding identification. A composite or verbal description was presented alongside the target's corresponding line-up, and the participants were asked to select which photograph they believed it to represent. Each participant saw four types of generated information (a control simple description, an elaborate verbal description, a FACES composite, and an ID composite) for all four target faces, resulting in 16 items per participants. No participants rated the same item twice, and the items were randomly selected for presentation per participant.

Results

All of the statistical tests were conducted twice, once analysing the data that was produced when participants made their first line-up selection (labelled here as the *unforced* results), and once analysing the data that was produced when participants were then forced to select a face in the line-up (if they had previously decided that the target was not present – this data has been labelled as the *forced* results). Where appropriate, the vertical axes on the graphs presented have been scaled up to better depict the patterns present between the conditions.

Analysis of Line-up Identifications

Unforced identifications. The frequencies of correct responses (hits), incorrect responses (foil identifications), and rejections (stating that the target was not present) are shown in Table 2.

Table 2

Percentage Correct, Incorrect and Rejections of Unforced Line-up Decisions.

Condition	Decision: Correct (Hits)	Decision: Incorrect (Foil Identifications)	Decision: Rejection (No Identification)	Total
ID	30.00% (12)	7.50% (3)	62.50% (25)	100.0% (40)
FACES	40.00% (16)	12.50% (5)	47.50% (19)	100.0% (40)
Elaborate Verbal	47.50% (19)	7.50% (3)	45.00% (18)	100.0% (40)
Control	47.50% (19)	10.00% (4)	42.50% (17)	100.0% (40)

Note. Numbers in parentheses are frequency counts.

Each participant had to select the face from a line-up which they thought to be the original target face they saw at the encoding stage. Upon initial inspection, the values presented in Table 2 indicate that the elaborate verbal and control groups correctly identified the target the most, with each group achieving 47.50% accuracy. They also suggest that the FACES group made the most foil identifications, while the ID group made the most rejections, with participants making this decision 62.50% of the time. Figure 1 displays this pattern.

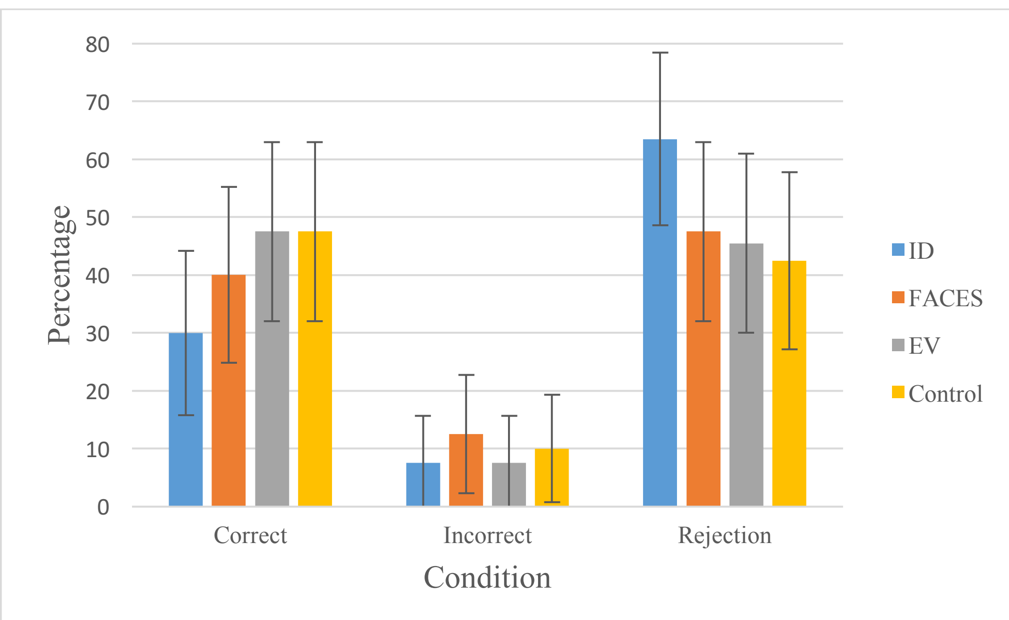

Figure 1. Percentage of unforced correct, incorrect, and rejection decisions per group in the line-up identification task. I-bars are 95% confidence intervals.

For purposes of analysis, rejections were coded as incorrect decisions. Table 3 displays the results of the chi-square tests conducted on this data, with the addition of a combined groups condition, in which the ID and FACES groups were combined together to form a composite group, and the elaborate verbal and control groups were combined together to form a description group.

Table 3

Results of Chi-Square Tests on Unforced Decisions.

Condition	χ^2 Value	Degrees of Freedom	Significance (1-tailed)
Combined groups	2.58	1	.054*
Across groups	3.40	3	.167
ID vs FACES	0.88	1	.174
FACES vs Elaborate Verbal	0.46	1	.250
ID vs Elaborate Verbal	2.58	1	.054*
FACES vs Control	0.46	1	.250
ID vs Control	2.58	1	.054*

Note. * denotes approaching significance at the $\alpha = .05$ level.

These chi-square goodness-of-fit tests demonstrate three differences that approach significance. The description groups performed better than the composite group, at rates approaching significance, with the description groups achieving 47.50% accuracy whereas the composite groups could only identify the target 25.00% of the time, $\chi^2(1, N=160) = 2.58$, $p = .054$. Specifically, since both the elaborate verbal and control groups achieved the same levels of accuracy (47.50%), both groups performed better than the ID group (30.00% accuracy) at levels approaching significance, $\chi^2(1, N=160) = 2.58$, $p = .054$ (for both). However, there was no significant difference in the selections between all of the groups, and there was also no significant difference in the selections between the ID and FACES groups; the FACES and elaborate verbal groups; and the FACES and control groups.

Since multiple comparisons were conducted, the chances of making a Type I error have increased. If a Bonferroni correction was applied, the new alpha level would be $\alpha = .007$, meaning that there would be no significant results. However, the Bonferroni is a conservative test, and is often interpreted with caution. As this is exploratory research investigating a new theory, these tentative results are still useful, and will interpreted with this in mind.

Forced identifications. The frequencies of correct responses (hits) and incorrect responses (misses, or foil identification) are shown in Table 4.

Table 4

Percentage Correct and Incorrect of Forced Line-up Decisions.

	Decision		
Condition	Correct (Hits)	Incorrect (Foil Identifications)	Total
ID	60.00% (24)	40.00% (16)	100.00% (40)
FACES	72.50% (29)	27.50% (11)	100.00% (40)
Elaborate Verbal	80.00% (32)	20.00% (8)	100.00% (40)
Control	67.50% (27)	32.50% (13)	100.00% (40)

Note. Numbers in parentheses are frequency counts.

When forced to make a decision, the values presented in Table 4 indicate that the elaborate verbal group obtained the highest number of hits, achieving 80.00% accuracy. The ID group appeared to perform the worst with only 60.00% accuracy. Figure 2 displays the percentage of correct and incorrect forced decisions per group.

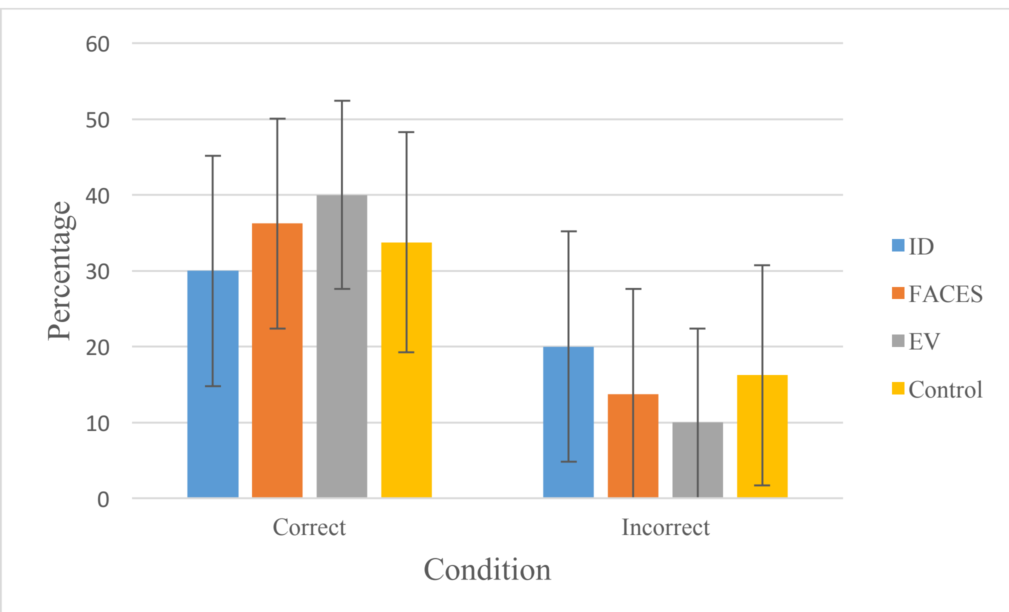

Figure 2. Percentage of forced correct and incorrect decisions per group in the line-up identification task. I-bars are 95% confidence intervals.

Table 5 displays the results of the chi-square tests run on this data.

Table 5

Results of Chi-square Tests on Forced Decisions.

Condition	χ^2 Value	Degrees of Freedom	Significance (1-tailed)
Combined groups	1.07	1	.150
Across groups	4.05	3	.128
ID vs FACES	1.40	1	.119
FACES vs Elaborate Verbal	0.62	1	.216
ID vs Elaborate Verbal	3.81	1	.025*
FACES vs Control	0.24	1	.313
ID vs Control	0.49	1	.242

Note. * denotes significance at the α = .05 level.

Just like when they were not forced to make a decision, the elaborate verbal group performed significantly better than the ID group (80.00% vs 60.00% accuracy respectively), $\chi^2(1, N=160) = 3.81$, $p = .025$. However, the other significant differences that were observed for unforced choices (between combined groups, and between the control and ID groups) were not evident for forced choices. In fact, no other significant differences between groups were observed for the forced decisions. However, with a Bonferroni-adjusted alpha of .007, the significant difference between the elaborate verbal and the ID group disappears. Again, this result will be interpreted tentatively.

Analysis of Reaction Times

Upon inspection, the reaction times for both the unforced and forced decisions were not normally distributed. As a result, the data was log-transformed to reduce skewness so that statistical analysis could be carried out.

Unforced reaction times. The amount of time taken for a participant to make a final unforced line-up decision was recorded by Qualtrics. The descriptive statistics of this raw (no transformation) data are presented in Table 6.

Table 6

Descriptive Statistics for Unforced Reaction Times.

Condition	*M*	Median	Minimum	Maximum	*N*
ID	18.30s (16.98s)	13.92s	0.97s	85.07s	40
FACES	17.78s (11.64s)	14.87s	0.98s	53.36s	40
EV	21.71s (14.88s)	20.20s	1.03s	50.14s	40
Control	21.40s (18.36s)	16.80s	0.02s	78.06s	40

Note. Standard deviations are presented in parentheses. EV refers to the elaborate verbal condition. Medians are presented as a better representation of central tendency as the distribution is skewed and may contain extreme reaction times.

The descriptive statistics suggest that the elaborate verbal group made their decision the slowest, with a median reaction time of 20.20s, whereas the ID group made their decision the fastest, with a median reaction time of 13.92s. A between-groups one-way ANOVA was conducted to examine if there were any significant between group differences for unforced reaction time across all the groups. Levene's test for homogeneity of variance was significant. The ANOVA was not statistically significant, $F(3, 156) = 0.51$, $p = .679$, partial $\eta^2 = .01$. There were also no significant differences between the individual groups themselves. Similarly, a between-groups ANOVA was conducted to examine unforced reaction times across combined groups (a composite and a description group). Once more, Levene's test for homogeneity of variance was significant, and the ANOVA itself was not statistically significant, $F(1, 158) = 0.09$, $p = .925$, partial $\eta^2 < .01$.

Forced reaction times. The amount of time taken for a participant to make a final forced line-up decision was recorded by Qualtrics. The descriptive statistics of this raw data are presented in Table 7.

Table 7

Descriptive Statistics for Forced Reaction Times

Condition	*M*	Median	Minimum	Maximum	*N*
ID	21.57s (18.42s)	14.10s	0.97s	85.07s	40
FACES	15.44s (13.21s)	12.19s	0.46s	60.28s	40
EV	21.92s (23.93s)	15.96s	1.03s	131.11s	40
Control	21.91s (22.41s)	14.30s	0.02s	98.74s	40

Note. Standard deviations are presented in parentheses. EV refers to the elaborate verbal condition. Medians are presented as a better representation of central tendency as the distribution is skewed and may contain extreme reaction times.

The descriptive statistics indicate that the elaborate verbal group had the longest median reaction time (15.96s) and the FACES group had the shortest (12.19s). A between-groups one-way ANOVA was conducted to examine if there were any significant between group differences for unforced reaction time across all the groups. Levene's test for homogeneity of variance was significant. Once more, the ANOVA was not statistically significant, $F(3, 156) = 0.68$, $p = .565$, partial $\eta^2 = .01$. There were also no significant differences between specific groups themselves. The between-groups ANOVA conducted to examine unforced reactions across combined groups, with a non-significant Levene's test score, was also not statistically significant, $F(1, 158) = 0.05$, $p = .824$, partial $\eta^2 < .01$.

These results suggest that the different groups did not take significantly different amounts of time to make either their unforced or forced line-up decisions.

Analysis of Confidence Scores

Participants were required to rate the confidence they had in their line-up decision out of 100. The mean confidence scores provided for the unforced and forced decisions are presented in Figure 3.

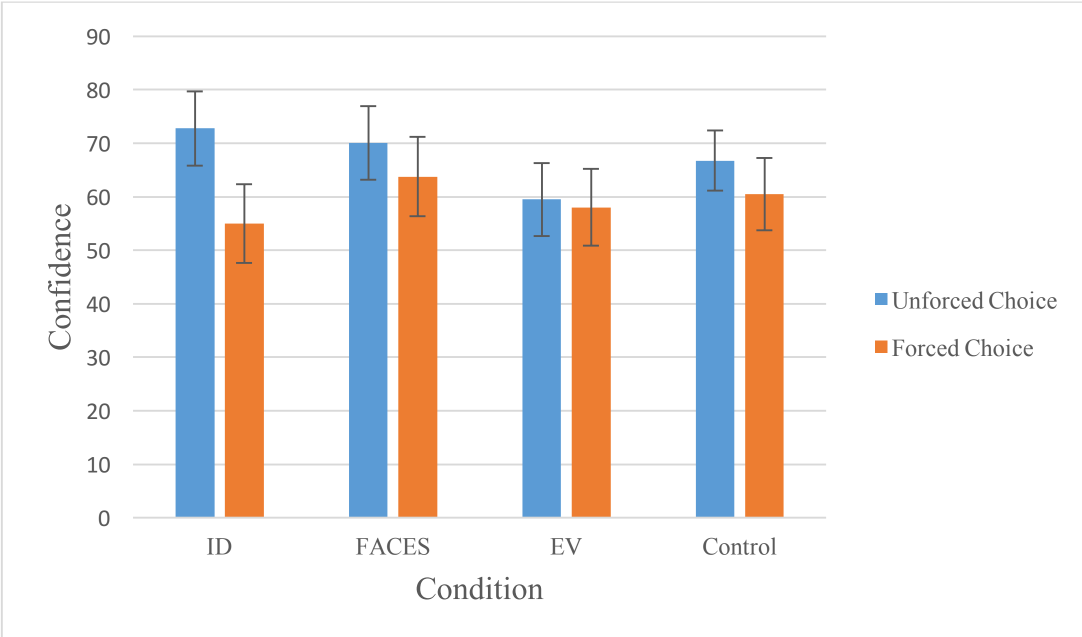

Figure 3. Mean confidence scores per group for unforced and forced decisions. I-bars are 95% confidence intervals.

Unforced confidence. The descriptive statistics of the confidence scores given for the unforced line-up decision have been provided in Table 8.

Table 8

Descriptive Statistics for Unforced Confidence Scores

Condition	*M*	Minimum	Maximum	*N*
ID	72.75 (22.42)	20.00	100.00	40
FACES	70.00 (22.19)	30.00	100.00	40
Elaborate Verbal	59.50 (21.95)	10.00	100.00	40
Control	66.75 (18.31)	20.00	100.00	40

Note. Standard deviations are presented in parentheses.

A between-groups one-way ANOVA was conducted to examine if any between group differences exist for unforced confidence across combined groups (a composite group and a description group). Inspection of boxplots of the distribution of the data indicated that the data was normally distributed with no outliers, and Levene's test for homogeneity of variance was not significant. The ANOVA was statistically significant $F(1, 158) = 5.99$, $p = .016$, partial $\eta^2 = .04$, suggesting that the composite group had significantly higher confidence scores ($M = 71.38$) than the description group (*M* = 63.13). However, the effect size is small, with participant condition explaining only 4% of the variance in confidence.

For a more detailed analysis, a between-groups one-way ANOVA was conducted on unforced confidence between all the groups. Once more, the data appeared normally distributed with no outliers, and Levene's test for homogeneity of variance was not significant. The ANOVA was statistically significant, $F(3, 156) = 2.89$, $p = .037$, partial $\eta^2 = .05$. This is a very small effect size, indicating that the participants' condition only explains 5% of the variation in their confidence scores. Notably, amongst the different groups, Tukey's post-hoc analysis indicates that the only significant difference was between the ID and the elaborate verbal groups, with the ID group having significantly higher confidence scores (*M* =72.75) than the elaborate verbal group (*M* = 59.50, *p* = .036).

Forced confidence. The descriptive statistics of the confidence ratings for forced line-up decisions have been provided in Table 9.

Table 9

Descriptive Statistics for Forced Confidence

Condition	*M*	Minimum	Maximum	*N*
ID	55.00 (23.85)	10.00	100.00	40
FACES	63.75 (23.82)	10.00	100.00	40
Elaborate Verbal	58.00 (23.22)	0.00	100.00	40
Control	60.50 (21.83)	20.00	100.00	40

Note. Standard deviations are presented in parentheses.

A between-groups one-way ANOVA was conducted to examine if any between group differences exist for forced confidence across all the groups. Inspection of boxplots of the distribution of the data indicated that the data within these four groups is normally distributed, and there are no outliers. Levene's test for homogeneity of variance was not significant. However, unlike for the unforced confidence, the ANOVA was not statistically significant, $F(3, 156) = 1.03$, $p = .383$, partial $\eta^2 = .02$. There were also no significant differences between specific groups themselves.

Another between-groups one-way ANOVA was conducted to examine if any between group differences exist for forced confidence across combined groups. Once more, the data is normally distributed, and Levene's test for homogeneity of variance was not significant. Once again, the ANOVA was not significant $F(1, 158) < 0.01$, $p = .973$, partial $\eta^2 < .01$. This means that when forced to choose, participants all had the same levels of confidence.

Relationship Between Confidence and Accuracy

Even though one might expect there to be a link between identification accuracy and confidence in identification judgements, research has shown that there is only a weak correlation between them (Sporer, Penrod, Read, & Cutler, 1995). To investigate whether this was the case for both forced and unforced line-up decisions, a point-biserial correlation was conducted between the unforced confidence of initial choosers (those who did not reject the first line-up) and the accuracy of these choices. Participants who were more accurate had higher levels of confidence after their line-up decision, $r = .27$, $p = .007$. However, the point-biserial correlation between forced confidence and accuracy is, while positive, very weak ($r = .07$) and not statistically significant, $p = .194$.

Investigating Confidence Modulation

Participants were asked to provide their confidence at three stages of the experiment: after encoding (pre-confidence), after having made a composite or provided an elaborate verbal description (mid-confidence), and after having made their forced line-up choice (post-confidence). Since the control group was not asked to provide a mid-confidence score, they were left out of the following analyses. For the purpose of these analyses, the post-confidence scores have been converted to a scale from 1-10 to match the scales of the pre- and mid-confidence stages. First, a one-way ANOVA was conducted on the pre-confidence scores for the different groups, to establish if there was a significant difference between them at this stage. A significant result here would indicate that there was some difference between the groups before any differing tasks were implemented. However, this ANOVA was not significant, $F(2, 117) = 1.02$, $p = .364$. As a result, a mixed-methods ANOVA was conducted on the three confidence scores obtained from the participants, to examine if there were any between group differences. The assumption for homogeneity of variance was upheld, but Mauchly's test of sphericity was significant ($p < .001$), and therefore the Greenhouse-Geisser results are presented where relevant.

There is a significant main effect for confidence period on confidence scores ($F(1.69, 193.86) = 3.83$, $p = .023$, partial $\eta^2 = .03$), but no significant main effect for participant condition ($F(1, 2) = 1.83$, $p < .166$, partial $\eta^2 = .03$) on confidence scores. There is a significant interaction between confidence period and participant condition on confidence scores, $F(3.37, 193.86) = 2.99$, $p = .027$, partial $\eta^2 = .05$. However, as this is a disordinal interaction, the main effects will not be interpreted in any further detail. Post-hoc tests indicate that participants in the elaborate verbal group ($M = 6.12$) had significantly higher confidence scores at the mid-confidence stage than the FACES ($M = 5.02$, $p = .011$) and ID ($M = 4.79$, $p = .002$) groups. The FACES group had significantly higher confidence scores at the post-confidence stage ($M = 6.37$) compared to the pre-confidence stage ($M = 5.27$, $p = .010$) and the mid-confidence stage ($M = 5.02$, $p = .002$). Contrastingly, the ID group had significantly lower confidence scores at the mid-confidence stage ($M = 4.79$) compared to the pre-confidence stage ($M = 5.47$, $p = .024$).

Figure 4 shows the patterns of confidence scores across these different stages for the different groups.

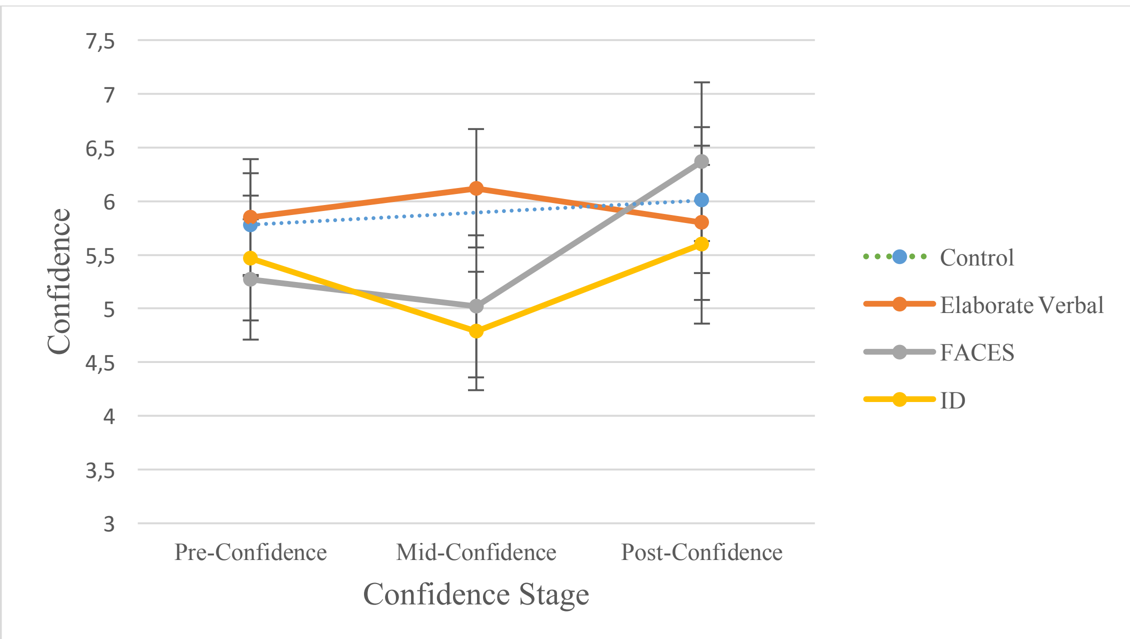

Figure 4. The pattern of confidence scores for the different groups at the different confidence stages. I-bars are 95% confidence intervals.

Description and Composite Quality Assessment

The quality of the composites (ID and FACES) and verbal descriptions (elaborate and simple descriptions) created in the experiment were assessed in a survey. The line-up decisions from this Qualtrics quality assessment survey were summed for each line-up member across all of the conditions and targets. The correct target identifications for these line-ups were then summed across the targets to obtain an overall proportion of the total selections. These percentages are presented in Figure 5.

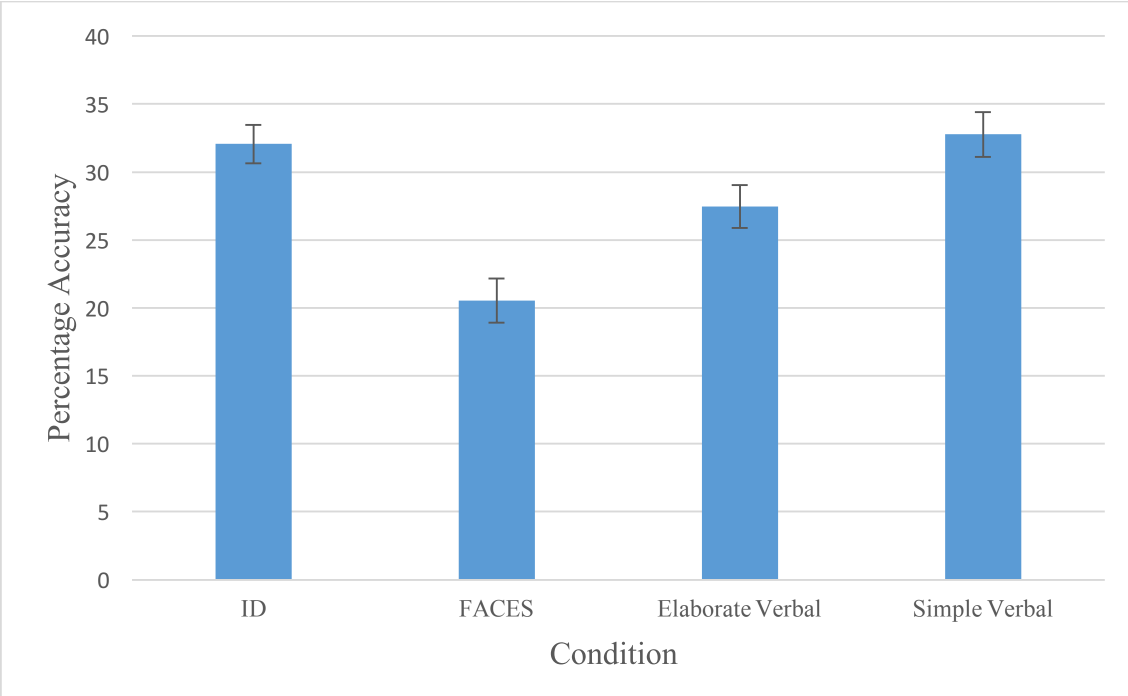

Figure 5. Percentage accuracy choices, per group, in the quality assessment task. I-bars are 95% confidence intervals.

Binomial probability distributions were used to compute the probabilities for each of the groups to investigate whether each differed significantly from chance levels, which would mean that someone informed by the image or description would perform significantly differently to someone who simply made a random selection with no prior knowledge. Chance, which is calculated as the total percentage divided by the number of possible choices, is 100/6, or 0.17. All of the groups differed significantly from chance, with $p < .001$, and therefore it appears that all of the information provided was useful in some regard in aiding identifications.

Chi-square goodness-of-fit tests were conducted on the quality assessment line-up decisions, assessing whether there was any difference amongst the accuracy of the decisions as relating to the information each group provided. There was a significant difference amongst all of the groups, $\chi^2(3, N=6240) = 145.92$, $p < .001$. Specifically, using the FACES images (20.54% accuracy) led to significantly fewer correct decisions than the ID images (32.05% accuracy, $\chi^2(1, N=6240) = 106.56$, $p < .001$), the elaborate verbal descriptions (27.47% accuracy, $\chi^2(1, N=6240) = 40.98$, $p < .001$), and the simple descriptions (32.76% accuracy, $\chi^2(1, N=6237) = 118.95$, $p <$

.001). Using the elaborate verbal descriptions (27.47% accuracy) also led to significantly fewer correct decisions than the ID images (32.05% accuracy, $\chi^2(1, N=6240) = 15.67$, $p < .001$) and the simple descriptions (32.76% accuracy, $\chi^2(1, N=6237) = 20.72$, $p < .001$). Overall, using the description information (60.23% accuracy) led to significantly more correct responses than the composite images (48.01% accuracy, $\chi^2(1, N=12477) = 22.39$, $p < .001$).

A Bonferroni correction was applied resulting in an alpha level of $\alpha = .007$. However, all of the analyses appeared to withstand this adjusted level, and all of the comparisons are therefore still significant.

Discussion

This experiment required four groups of participants to describe a target face with each group using one of four means of doing so – by creating a holistic ID composite, a featural FACES composite, an elaborate verbal description, or a simple free-recall verbal description. Participants then had to attempt to identify the target face from a line-up two days later. The quality of these products created by the participants was assessed by new participants who had to use the constructions to attempt to select the target face from a line-up. The following discussion is a tentative exploration of the results that will attempt to relate them to the theory being investigated.

The results suggest that creating composite images has some effect on memory – memory appears to be hampered once construction has taken place. For their unforced decision, the description groups were better than the combined composite groups at correctly identifying the target face from a line-up at rates approaching significance. As hypothesised, the composite groups were more conservative in their recognition responses than the description groups. Upon further examination, however, it was specifically the ID group that obtained fewer hits than the control and elaborate verbal groups – although these results were marginally significant. It was apparent that the ID group performed so poorly not because they consistently selected foils instead of the target, but because they were incorrectly rejecting the line-up. When participants who rejected the line-up were forced to make an identification, the ID group also performed significantly worse than the elaborate verbal group, indicating that they were incorrectly selecting foils at this stage. There were no significant differences between FACES and the description groups with regard to identifying the target.

These results differ from previous research conducted by Kempen (2012). In that study, even though the composite group also made many incorrect rejections, it was the FACES group that performed in this way, and, when forced to make an identification, these participants largely selected the target correctly. However, in this research it was the ID group that was unable to correctly identify the target when forced to make an identification. This is in line with the results produced by Wells et al. (2005), where participants who built composites of a target face performed poorly at subsequent identification and identified innocent line-up members. Even though Wells et al. (2005) only used the FACES program, and not a holistic program such as ID, a similar conclusion can be drawn that creating composites may contaminate memory.

Nonetheless, even though the ID group performed the worst at both the unforced and forced identifications, the ID composites were significantly more useful in aiding identifications than both the elaborate verbal descriptions and the FACES composites, thus suggesting that they are of better quality than these descriptions and composites. This supports other research that has shown that holistic composite systems produce better likenesses than featural systems (Frowd et al., 2004; Tredoux et al., 2007). Thus it may be the quality of the ID composites that has an effect on participant memory. While Holland, Otzen and Sporer (1994, as cited in Maskow et al., 2007) found that participants who constructed good quality composites were more likely to accurately select the target, perhaps better quality composites only have this effect up until a point, where their good quality may begin to interfere with memory. The pattern evident in Figure 1 supports this notion. It is apparent from this graph that as the construction or description detail in the method increases, the proportion of correct decisions decreases, and the amount of rejections increases. The most realistic and detailed product is an ID composite, followed by a FACES composite, an elaborate verbal description, and, finally, a simple verbal description. Therefore, the problem may lie in the realism of the product: As the likeness of the product to the target increases, the tendency to initially reject the line-up, and then to select foils, appears to increase. This is in agreement with previous research that found that verbal descriptions lead to more accurate identifications than composite images (Christie & Ellis, 1981; Lech & Johnston, 2011), as a result of the increased detail necessary in the creation of a composite image compared with that of a verbal description. Similarly, this could explain why the older, manual methods of creating composites were sometimes found to have a facilitating effect on memory (Mauldin & Laughery, 1981), as these methods do not produce very detailed images.

This research hypothesised that, according the theory of self-generated misinformation, the featural system would result in worse recognition rates than the holistic system, because FACES requires more detailed input on separate facial features than ID does. This has clearly not been the case. However, the theory may, in fact, still account for this result. The high degree of realism in ID composites could be leading to *holistic* confabulation. The program does not require the user to make explicit selections for all of the features, especially those that may not be remembered. This was the reason that the original hypothesis asserted that the ID group would perform better than the FACES group. Yet, in creating a composite, the user must still commit to the overall appearance of the final face that is created. As this image contains many details (because, for example, it is extremely realistic, and contains texture and true-to-life colours) the user must commit to all of these, even if they may not accurately represent the original face in all respects. As a result, the image as a whole may take the form of self-generated misinformation, which may create a memory that is similar to and indistinguishable from the target face. Additionally, as the composite is holistic, but may not look exactly like the original target face, it may hamper memory because it is in the same holistic mode of processing as encoding and recognition. When forced to make an identification, the ID group might therefore be selecting foils because this new memory of the realistic composite may actually resemble the foils themselves, as they were selected for use based on their likeness to the target.

Similarly, the fact that the FACES group did not perform significantly differently from the other groups at correctly recognising the target can also be accounted for by this theory. In this experiment, the FACES composites themselves were the poorest quality products created by the participants, as they were significantly less useful in aiding identification than the ID composites, the elaborate verbal descriptions, and the simple descriptions. This poor quality of featural composites has been replicated often in the literature (Brace et al., 2006; Frowd et al., 2005; Frowd et al., 2011). Perhaps, therefore, because the FACES composites are of such bad quality, the new memory of the composite would not have been similar to the original memory of the target face at all, and the participants may thus not have been confused.

The analysis of the confidence scores obtained during the experiment adds further to this theory. The confidence modulation analyses show that after constructing the composite, the confidence of the ID group decreases significantly compared with their confidence before having done so, and compared to the other groups. In contrast, the confidence of the elaborate verbal group increases at this stage. This confidence is reflected in the actual identifications made, with the ID group performing worse than the elaborate verbal group at rates approaching significance. The elaborate verbal group could thus be rejecting the line-up so rarely because these participants are more confident in their memory, whereas those in the ID group are not confident, and therefore reject the line-up. Hence, as the process of making a realistic composite (which is a difficult task) leads to confabulation, the participants lose confidence in their memory of the original face, a phenomenon which affects their subsequent line-up decisions. Notably, the confidence of the ID group increased significantly post decision-making, despite their increase in selection of foils. Given that research has shown that people are more likely to believe confident witnesses (Luus & Wells, 1994), this result suggests that in spite of their poor recognition performance, holistic composite creators may still be believed regarding their identification selections.

However, it must be noted that the elaborate verbal group did not noticeably perform as hypothesised (similarly to the FACES group, and worse than the control group on the identification task). Nevertheless, this might be because the task itself was not satisfactorily comparable to creating a composite. During the experiment, participants in this group did not spend as much time creating their elaborate verbal descriptions as those creating a composite. Therefore, as the participants did not have their elaborate written descriptions reinforced for as long a period as the composite groups, confabulation could not occur. Also, even though participants were asked to describe every feature used in a FACES composite, and were told to describe even those features they did not remember well, many simply provided vague descriptions for each of these, and these descriptions may not have been detailed enough to cause confabulation. The verbal overshadowing effect, a phenomenon which may describe why participants perform poorly at face recognition tasks after having verbalised and described features of a face (Fallshore & Schooler, 1995), has been discussed often in psychological literature. This effect theorises that as verbalisation necessitates participants to access specific features of the memory trace, and as faces are encoded holistically, this may impair the original holistic memory trace in later recognition (Meissner & Brigham, 2001). However, just as in this experiment where the control and elaborate verbal groups performed similarly with their unforced decision, several studies have failed to replicate this phenomenon (Meissner & Brigham, 2001). In fact, research conducted by Meissner, Sporer, and Schooler (2007) suggests that face descriptions and face recognition utilise distinctly different processes that are based in opposite hemispheres of the brain. Perhaps, therefore, verbalisation does not have the same effects on memory as the creation of a visual image might, resulting in the comparatively better performance of the elaborate verbal group.

This study has produced an interesting finding with regard to the quality of composite images. In the quality assessment stage, even though using both types of composite images resulted in correct identifications at rates greater than chance, the ID composites, which were more useful than the FACES composites, were still only as useful as the simple verbal descriptions produced by the control group. These descriptions were often extremely vague, sometimes listing only the gender and hair colour of the target. This result is concerning, considering the expenses involved in using composite software systems, for example in the training of police officers and the purchase of the software and the computers on which to run it. Composite creation is also a very time-consuming process, and it may be difficult for traumatised victims and witnesses to sit through a lengthy interview. Therefore, these results question the efficacy of using featural composite images, given that they appear to be no more helpful in isolating members of the population than a much simpler method.

These results would of course need to be replicated for a concrete suggestion to be made to the police. However, if the results were replicated, one still could not simply suggest that composite images should not be used without providing a meaningful and practicable alternative. Perhaps, however, holistic composites could still be considered useful. Since they appear to be better able to identify a suspect than the other methods currently available, the fact that they may hamper eyewitness memory may not be reason enough to constrain their use. If their possible effect on memory is taken into account, and the police utilise a different method, other than line-ups, of confirming identification or of modulating witness confidence after composite construction, then holistic composite images may still prove extremely useful. However, another alternative may be found in the research that has been conducted on improving police interviewing techniques, which aims to obtain better descriptions from witnesses. A lot of research has focussed on the development of the Cognitive Interview, for example, which utilises psychological principles on memory, such as employing open-ended questions and retrieval mnemonics, to improve the retrieval of correct details from witnesses, without increasing the number of incorrect details provided (Finger & Pezdek, 1999). The meta-analysis conducted by

Memon, Meissner, and Fraser (2010) reiterates these positive findings. Therefore, perhaps there are indeed ways to acquire accurate descriptions that could be used instead of composite images to help identify unknown suspects.

Limitations and Future Research

There were limitations in this study. Analysis of the reaction times obtained for both the forced and unforced line-up decisions produced non-significant results. However, it must be noted that in this experiment, this measure is perhaps not an accurate or useful one. Even though the participants were all asked to complete the identification surveys on a computer or laptop, the data accrued by Qualtrics indicates that some participants used cell phones or tablets. Smaller screens may have required participants to scroll through the page in order to see all of the images, increasing their reaction time. Similarly, the speed of the internet connection that participants used to complete the survey could not be controlled, and slower internet connections may have resulted in increased reaction times, given that the images may have taken longer to load. Future research should take this into account, and set out a longer period for data collection so that participants can be brought back in, in order to complete the survey in the laboratory, where screen size and internet speed can be kept constant. Future research should also endeavour to improve the elaborate verbal description condition. As previously discussed, the time allowed to participants to engage with their descriptions should be increased. Moreover, participants should be encouraged to provide more information in their descriptions, so that the condition can be more comparable to the conditions prevailing in composite creation.

Another limitation is the fact that this study only used target present line-ups, and not target absent line-ups. In practice, if an eyewitness were to identify someone in a suspect-present line-up other than the suspect, the police would know that that individual was not the perpetrator, as unrelated individuals are used as foils. As a result, these findings do not necessarily suggest that eyewitnesses who create descriptions of perpetrators and then misidentify them, may cause innocents going to jail. Target absent line-ups are used to assess whether, in a situation where the police have apprehended the wrong suspect, a designated innocent suspect would be chosen by an eyewitness. However, using target absent line-ups along with target present line-ups would have meant doubling the sample size, and this was deemed not feasible for this study. Future research on this topic should thus use both target present and target absent line-ups in order to ascertain more real-world effects of the results. As this would utilise a greater sample size, this could also result in more analyses having statistically significant results.

Conclusion

In closing, this study aimed to discover which of the different methods of creating composites and verbal descriptions leads to increased target identification, decreased rates of line-up rejections, increased confidence ratings, and the creation of the best quality product. Based on a review of the literature, the theory of self-generated misinformation was proposed as an explanation for why composite images are inaccurate and may lead to hampered recognition ability. The study therefore aimed to assess the theory of self-generated misinformation, something that has never been done before. This research hypothesised that during recognition, the composite construction groups would perform similarly to the elaborate verbal description group, which would both perform worse than the free recall verbal description group; that the ID (holistic) group would have better recognition performance than the FACES (featural) group; and that the composite groups would be more conservative in their recognition decisions.

The results of this study indicate that holistic composites lead to decreased target identifications, increased line-up rejections, decreased confidence prior to decision-making and increased confidence post-decision-making, yet nonetheless also create the best quality product alongside a simple free-recall verbal description. There is a mechanism that can account for these results - the theory of self-generated misinformation. ID composites require their creators to commit to a large amount of detail, even though they may not have selected all of the features themselves. As a result, the holistic image confabulates their memory of the original face. This decreases their confidence in this original memory, and results in increased rejections of the line-up. The FACES group did not perform as poorly as expected on recognition, but produced the product of the least quality. Understood within the theory of self-generated misinformation, creating an inaccurate FACES composite does not affect participants' confidence in their memory of the target face, and they are therefore able to perform similarly to the verbal groups on the recognition task.

This research hypothesised that the theory of self-generated misinformation would require the elaborate verbal description group to perform similarly on recognition as the featural composite group. However, the results did not find this to be the case. This may be a result of verbal and visual processing having different effects on memory, thereby making the conditions not directly comparable. Yet of the hypotheses proposed for this study, only one is supported by the results, in that the composite groups were shown to be more conservative in their recognition decisions. As a result, these results cannot be deemed definitive proof for this theory.

However, it must be noted that the results do reiterate an important finding – that the quality of composite images is poor. Even though the ID composites were one of the most useful types of products in aiding identifications, they were only as useful as simple verbal descriptions, which lack detail almost entirely. As a result, one must wonder whether the resources put into using composite images as a means of suspect identification might not be better utilised elsewhere. These results also indicate that holistic composite images may negatively affect recognition performance, by increasing rejections without subsequent correct identifications when choosing is forced. In practice, this means that using holistic composite systems could lead to witnesses incorrectly identifying innocent suspects who could then face lengthy custodial sentences.

Of course, this study takes an exploratory approach, as it attempts to hypothesise a theory that can account for the quality of composites and the effect on memory they may have. As a result, before any strong conclusions regarding its veracity can be drawn, more research should be conducted in order to assess the theory of self-generated misinformation.

References

Ackil, J. K., & Zaragoza, M. (1998). Memorial consequences of forced confabulation: Age differences in susceptibility to false memories. *Developmental Psychology, 34*(6), 1258-1272. doi:10.1037/0012-1649.34.6.1358

Brace, N. A., Pike, G. E., Allen, P., & Kemp, R. I. (2006). Identifying composites of famous faces: Investigating memory, language and system issues. *Psychology, Crime & Law, 12*(4), 351-366. doi:10.1080/10683160500151159

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

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

Comish, S. E. (1987). Recognition of facial stimuli following an intervening task involving the Identi-kit. *Journal of Applied Psychology, 72*(3), 488-491.

Davies, G., Ellis, H. D., & Shepherd, J. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology, 63*(2), 180-187. doi:10.1037/0021-9010.63.2.180

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*(1), 119-124. doi:10.1037/0021-9010.85.1.119

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

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

Finger, K., & Pezdek, K. (1999). The effect of the cognitive interview on face identification accuracy: Release from verbal overshadowing. *Journal of Applied Psychology, 84*(3), 340-348.

Frowd, C. D., Bruce, V., Ness, H., Bowie, L., Paterson, J., Thomson-Bogner, C., . . . Hancock, P. (2007). Parallel approaches to composite production: Interfaces that behave contrary to expectation. *Ergonomics, 50*(4), 562-585.

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 & Law, 11*(1), 33-52. doi:10.1080/10683160310001634313

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., . . . Hancock, P. J. B. (2011). The psychology of face construction: Giving evolution a helping hand. *Applied Cognitive Psychology, 25*(2), 195-203. doi:10.1002/acp.1662

Innocence Project. (n.d.a.). *Eyewitness Misidentification*. Retrieved March 18, 2016, from http://www.innocenceproject.org/causes-wrongful-conviction/eyewitness-misidentification

Innocence Project. (n.d.b.). *Larry Johnson*. Retrieved March 18, 2016, from http://www.innocenceproject.org/cases-false-imprisonment/larry-johnson

Kempen, K. (2009). *It's the thought that counts: Face composite production can hamper recognition performance.* (Unpublished Honours thesis). University of Cape Town, South Africa.

Kempen, K. (2012). *The effects of isolated featural and configural composite construction on recognition accuracy.* (Unpublished Masters thesis). University of Cape Town, South Africa.

Kovera, M. B., Penrod, S. D., Pappas, C., & Thill, D. L. (1997). Identification of computer-generated facial composites. *Journal of Applied Psychology, 82*(2), 235-246. doi:10.1037/0021-9010.82.2.235

Lech, A. M., & Johnston, R. (2011). The relative utility of verbal descriptions and facial composites in facial identifications. *International Journal of Bio-Science and Bio-Technology, 3*(3), 1-16.

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12*(4), 361-366.

Luus, C. A., & Wells, G. L. (1994). The malleability of eyewitness confidence: Co-witness and perseverance effects. *Journal of Applied Psychology, 79*(5), 714-723.

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.

Malpass, R. S., Tredoux, C. G., & McQuiston-Surrett, D. (2007). Lineup construction and lineup fairness. In R. Lindsay, D. 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*(3), 351-357.

Mavangwana v S 2010 (25321) JOL 1 (SCA).

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

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.

Memon, A., Meissner, C. A., & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law, 16*(4), 340-372.

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.

Schmidt, H., & Tredoux, C. (2006). Utilisation and usefulness of face composites in the South African Police Service - an evaluation study. *South African Journal of Criminal Justice, 19*, 303-314.

Schooler, J. W. (2002). Verbalization produces a transfer inappropriate processing shift. *Applied Cognitive Psychology, 16*(8), 989-997. doi:10.1002/acp.930

Sporer, S. L., Penrod, S. D., Read, J. D., & Cutler, R. L. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118*(3), 315-327.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology, 46A*(2), 225-245.

Tredoux, C. G., Kempen, K., Nortje, A., Christ, B., Hoogenhout, N., & Kaiser, A. (2010). *Do face composites contaminate memory? A meta-anlysis*. Unpublished manuscript, University of Cape Town.

Tredoux, C. G., Nunez, D. T., Oxtoby, O., & Prag, B. (2007). An evaluation of ID: An eigenface based construction system. *South African Computer Journal, 37*, 90-97.

Wells, G. L., Charman, S. D., & Olson, E. A. (2005). Building face composites can harm lineup identification performance. *Journal of Experimental Psychology Applied, 11*(3), 147-156. doi:10.1037/1076-898X.11.3.147

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 Behaviour, 20*, 280-292.

Appendix A

Ethical Considerations

This study received ethical approval from the Department of Psychology (Appendix B).

Consent, Voluntary Participation, and Confidentiality

All participants were asked to provide informed consent, after having been briefed about the tasks they were to complete, the confidentiality of their responses, the voluntary nature of their participation, and their ability to withdraw from the study at any point (Appendices C, D, E, F). Participation in this study did not carry any foreseeable risk for participants. The benefits for participants included participating in psychological research, attaining SRPP points required for their courses, and learning about an important police practice.

Appendix B

UCT Ethics Committee Email Indicating Approval

From: Rosalind Adams < rosalind.adams@uct.ac.za>
Date: Mon, Jun 6, 2016 at 12:14 PM
Subject: Ethics Approval
To: Eden Clingman <traibusch@gmail.com>

Morning All

Your Ethics application for your Proposal has been approved I will be sending your reference numbers in the next day or two

Regards
Ros

Appendix C

Participant Consent Form for Step Two of Line-up Creation

PARTICIPANT CONSENT FORM

Thank you for your interest 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 name and 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: _____________________________

Race: ____________________________

Length of time residing in South Africa: _______________________________

Course for which you are receiving SRPP (course code):_________________

For your time and participation, you will receive **60 minutes** (2 units) towards your SRPP requirement.

Please note that at any time during study if you feel uncomfortable or experience any distress, you are free to leave. All of your responses will be kept confidential.

THE STUDY: In this study, you will be exposed to various photographs of people. You will be given either a face description, or a face, and required to sort through the photographs to see which faces best match the face, or description. 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. Once you have completed the study, please refrain from telling other potential participants about the details of this experiment, as this will bias the results.

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

Participant Consent Form for Experiment

Fun with Faces: SRPP Face Rating Study

Dear Student,

Thank you for your interest in this study!

Purpose

This study is being conducted for my Honours research at the University of Cape Town. The purpose of this study is to see how participants perceive and rate different faces. Please note that more information will be given to you at the end of the experiment.

Procedure

If you agree to participate, you will attend two sessions of this experiment. The first session should take approximately one hour, and the second session (which will take place two days later) should take 30 minutes. You will be asked to view some faces, and then to rate the face on certain traits and characteristics. You will then view a film and be required to answer questions on the film you watched. The entire experiment will be presented on computer. Your student number and name is required for SRPP purposes, but will in no way be linked to the results of the study. Your responses will remain anonymous and your responses will be kept strictly confidential.

You will need to complete both sessions of the experiment in order to receive your SRPP points.

Possible Risks

There are no anticipated psychological or physical risks associated with this study. The experiment is computer-based, non-intrusive, and you will not be exposed to emotional stimuli. This experiment will only take up 1:30 hours of your time, for which you will be compensated by receiving 3 SRPP points.

Benefits

You will receive 3 SRPP points if you complete both sessions of the experiment. Please note that you will NOT receive any SRPP points if you do not attend BOTH sessions.

Voluntary Participation

Participation in this experiment is voluntary, and you can, at any stage, withdraw from the experiment. Please note, however, that your SRPP points will not be awarded unless both sessions are complete.

Confidentiality

Your name, details, and responses will be kept strictly confidential. This consent form will be locked away, and the computer files will be encrypted. Your consent form will not be linked to your responses. Only your demographic information (race, sex, age) will be used in the experiment. Your name is required for SRPP purposes, but will not be linked to your responses.

Contact

You will be debriefed at the end of the study and given more information. If you have further questions, you may contact the researcher, Eden Clingman, via email at: **clnede001@gmail.com**, or by SMS: **072 731 4450**. If you have any questions regarding the ethics of this experiment, or about your rights as a participant, you can contact the Research Ethics Committee, Department of Psychology, Cape Town on: **021 650 3417**.

I, as a participant, have read the above information and am aware of the possible benefits and risks in this experiment. I have no further questions at this point. I hereby give consent to voluntarily participate in this experiment, knowing that my responses will be collected but will in no way be linked back to my personal details. I have been offered a copy of this document for future queries.

______________________________ ______________________________

Name of Participant Student Number

______________________________ ______________________________

Course code for SRPP Points (e.g., PSY1004F) Date

______________________________ ______________________________

Signature of Participant Signature of Witness

Appendix E

Description Generation Survey for Line-up Construction

In this session, you will be asked to view a series of faces and to rate these faces on various characteristics and traits. These questions are a series of rating scales. It should take about 15 minutes to complete.

Your participation in this study is completely voluntary. You will be free to withdraw from this study at any time. The study is not expected to involve risks of physical or mental discomfort to you in any form and the results of the study will be presented in a way so that, if published, your identity will not be revealed, and confidentiality of data will be ensured by careful safekeeping.

Important note: to be eligible to participate in this research you must be over the age of 18.
Please fill in the information below:
Full name:
Age:
Gender:
Which ethnicity or ethnic group you identify with:
Do you agree to take part in this study? Yes/No

By clicking 'Next' you consent to take part in this study.

In this survey, you will be asked to provide a written description of ten female faces.
You will study each face for five seconds. After studying the face, the page will advance automatically to the next slide, where you can write your description. Please note that you cannot move backwards to a previous slide, so read each slide carefully, and pay attention to instructions.
The verbal face description is a *broad* physical description of the face that you would give a police officer to identify a suspect (e.g. blonde hair, brown eyes, round face). Other examples of

physical descriptions could be gender, ethnicity, and facial features. You do not need to describe clothing or jewellery. You can separate your description with commas.
After you have described all the faces, you will then be asked to rate each face on a variety of characteristics.
Please click 'Next' to advance to the next slide and to start describing the faces.

Characteristics Ratings

Please rate the above face on the following characteristic from 1 – 8:
Criminality: *Even though a person may be innocent, some faces are perceived as belonging to someone who is more likely to commit a crime, than other faces. To what extent does this face resemble a criminal?*
Note that '1' – not at all criminal; '8' – extremely criminal.
Wealth: *Regardless of how wealthy people really are, some faces look like they belong to people who are wealthier than others. How wealthy do you think this person is?*
Note that '1' – not at all wealthy; '8' – extremely wealthy.
Memorability: *For many reasons, some faces seem easier to remember than others. How easy do you think it will be for you to remember this face?*
Note that '1' – not easy to remember at all; '8' – extremely easy to remember.
Typicality: *Some faces are more average-looking than other faces. To what extent does this face look average?*
Note that '1' – not at all average-looking; '8' – extremely average looking
Attractiveness: *How attractive do you think this face is?*
Note that '1' – not at all attractive; '8' – extremely attractive
Familiarity: *Even though you may not have seen this person before, some faces seem more familiar to us than other faces. How familiar does this face seem to you?*
Note that '1' – not at all familiar; '8' – extremely familiar
Age: *How old do you estimate this person is? (Give an estimate in years)*

Appendix F

Line-up Rating Survey for Line-up Construction

This study is being conducted as part of an Honours degree by Eden Clingman at the University of Cape Town.

In this session, you will be asked to view a series of face arrays and select a face that best matches a written physical description. It should take about 5 minutes to complete.

Your participation in this study is completely voluntary. You will be free to withdraw from this study at any time. The study is not expected to involve risks of physical or mental discomfort to you in any form and the results of the study will be presented in a way so that, if published, your identity will not be revealed, and confidentiality of data will be ensured by careful safekeeping.

Important note: To be eligible to participate in this research you must be over the age of 18
Please fill in the information below:
Full name:
UCT student number:
The course code for which you would like your SRPP points applied:
Age:
Gender:
Which ethnicity or ethnic group you identify with:
Do you agree to take part in this study? Yes/No

By clicking ‘Next’ you consent to take part in this study.

In this survey, you will be given eight different face arrays, along with eight different verbal, physical person descriptions.

Your task is to read the description above the face array, and decide which one face you think best matches the description above the array. Above each face is a number from 1 to 6. Please select the number of the face you think matches the description. You must only choose one face.

After selecting a face and clicking 'Next', the page will advance automatically to the next slide, where you will be given the next description and array. Please note that you cannot move backwards to a previous slide, so read each slide carefully, and pay attention to the instructions.

The verbal face description is a broad and vague physical description of the face (e.g., blonde hair, brown eyes, round face). Please note that the descriptors and features in the description are not listed in order of importance (e.g., you do not need to match on hair first, and then on eye colour second). There is no right or wrong answer to which face you select.

Please click 'Next' to advance to the next slide to begin the face selection task.

Appendix G

Elaborate Verbal Description Sheet

Name: ____________________ **Student Number: ______________**

Please describe the following features of the face you were shown earlier. Please do your best to describe all of these in as much detail as possible, but features marked with an asterisk (*) must be completed:

Race*: ____

____

Gender*: ____

____

Hair*: ____

____

Head shape*: ____

____

Eyebrows*: ____

____

Eyes*: ____

____

Nose*: ____

____

Lips*: ____

____

Jaw shape/Chin*: ____
____

Skin tone*: ____
____

Ears*: ____
____

Build/Weight*: ____
____

Moustache: ____
____

Goatee/beard: ____
____

Forehead lines: ____
____

Eye lines: ____
____

Chin lines: ____
____

Glasses: ____
____

Moles/Freckles: ____

____

Scars/Tattoos/Marks: ____

____

Jewellery: ____

____

Tattoos: ____

____

Appendix H

Examples of Target Line-ups

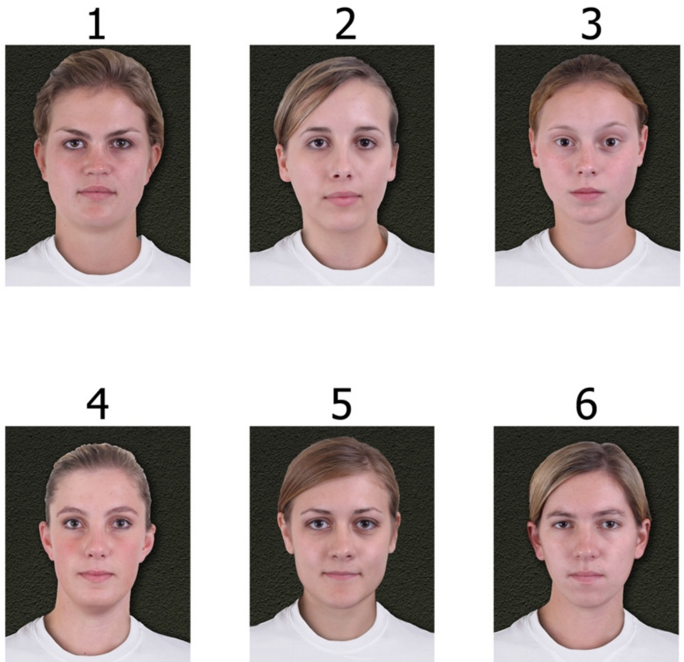

Figure 6. Line-up for target 1. Target is in position 2.

Figure 7. Line-up for target 2. Target is in position 3.

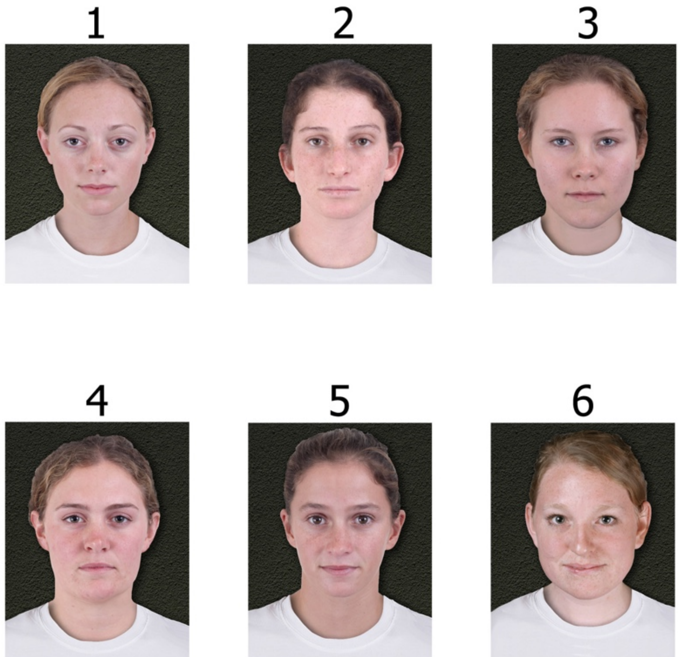

Figure 8. Line-up for target 3. Target is in position 4.

Figure 9. Line-up for target 4. Target is in position 3.