Face off: automatic versus controlled processing.
Does a shift in processing affect facial recognition?

Alicia Nortje

NRTALI003

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 the Humanities

University of Cape Town

2011

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 a few people who assisted and supported me through this project:

My supervisor, Professor Colin G. Tredoux, was infinitely patient, and provided me with much guidance and advice. His meticulous approach to detail and quality helped shape this thesis into a better project, and his dedication to our research team has made all of us better students.

My friends and family had to endure my continuous babble about my thesis, and we shared many tears and joys over countless cups of coffee. I am so grateful for their support.

Alexis van der Merwe, I carry your heart with me, and I am never without it – thank you for everything.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Working from the transfer-inappropriate processing shift (Schooler, 2002), this project aimed to investigate whether a shift from automatic to controlled processing would impair face recognition rates, much like the manipulated Navon letters do (Perfect, Weston, Dennis, & Snell, 2008), thus providing an alternative explanation for the mechanism underlying the verbal overshadowing effect (Schooler & Engstler-Schooler, 1990). Expertise leads to automaticity (Shiffrin & Schneider, 1977), and our expertise in processing faces of the same group as ours (Malpass & Kravitz, 1969) suggests that we may process such faces automatically; however since we have less familiarity with faces of a different group, those faces may be processed in a controlled processing mode. In order to induce automatic or controlled processing, a Stroop task was utilised as the word-reading and font-colour conditions were deemed automatic and controlled-processing tasks respectively. 288 participants were recruited, and were divided into four groups depending on their ethnicity (white or other) and processing type (automatic or controlled). All participants encoded a face (by watching a video), completed two Stroop tasks, and then had to identify that same encoded face from an 8-person line-up. This procedure was repeated three times, and three different faces were used in each trial. Much effort was made to ensure that line-ups were well constructed, thus allowing for better conclusions to be drawn about the results. Reaction times, recognition accuracy, and confidence levels were measured. Despite participants performing at different rates in the Stroop task (thus demonstrating that the experimental condition occurred), in general, my results demonstrated that neither participant ethnicity, the processing mode induced by the Stroop, nor an interaction between them had an effect of recognition rates. Thus, the data suggest that a shift from automatic to controlled processing

does not induce the same effects as those seen with the Navon letters (Perfect et al., 2008), and may not be a possible explanation for the verbal overshadowing effect.

Chapter One: Introduction

The verbal overshadowing effect (VOE) (Schooler & Engstler-Schooler, 1990) is a phenomenon where describing a face reduces later accurate recognition of that same face. This is a 'Catch-22 situation' in eyewitness testimony: a description is required to find the perpetrator; however, this act of providing such a description may impair later accurate recognition of that perpetrator from a line-up. Research suggests that the VOE does not necessarily result from the verbal description, but rather from the cognitive processing mode, specifically featural processing, that underlies this task (Macrae & Lewis, 2002). This mode is not conducive to facial processing, thus facial recognition is impaired (Tanaka & Farah, 1993). Since it is the induced processing mode that affects facial processing, any task that causes a similar processing mode should have the same deleterious effect on line-up recognition. This has already been established (Macrae & Lewis, 2002); much research has demonstrated that tasks which induce featural processing (which may arise from elaborate verbal descriptions) may negatively impact facial recognition, unlike holistic processing tasks that may facilitate it. Most of the experimental tasks used have solely induced either featural or holistic processing, for example the Navon letters (Navon, 1977); however, a manipulation of one such task, the Navon letters, produced results that suggested that there may be another type of processing mode that could affect facial recognition – automatic/controlled processing (Perfect et al., 2008). Since this manipulation may have induced two types of processing, it is not clear which affected facial recognition. To investigate this, an experimental task that *only* induced automatic/controlled processing was used in this experiment.

Verbal Overshadowing: Definition, Causes and Effects

VOE was first demonstrated in a series of six experiments by Schooler and Engstler-Schooler (1990). Their participants viewed a 30 second video of a bank robbery, and afterwards were assigned to two groups: either a control or description group. After viewing the video, the description group was given 5 minutes to describe the face of the robber; the control group completed an unrelated task (about which no details are given). Afterwards both groups were instructed to identify the robber from an 8-person line-up. The groups differed significantly in their performances: The control group correctly identified the perpetrator more often than the description group had (64% vs. 39%). A deeper investigation into the description group's errors revealed that these were similar to those made by the control group: Both groups had a higher proportion of misidentifications to misses (59% and 60% respectively), suggesting that the description group was not reluctant to make an identification, they were just poor at it.

Three theories have been suggested as explanations for this effect: recoding interference (Schooler & Engstler-Schooler, 1990); criterion response (Clare & Lewandowsky, 2004); and transfer-inappropriate processing shift (TIPS) (Schooler, 2002). The premise of recoding interference is that the verbal code that arises from describing the face overshadows the visual code of that same face, thereby preventing access to it at the time of recognition; this verbal code does not alter, or delete the visual code, but merely prevents access to it. This was supported by research where a verbal or visual code was solely created, such as when the encoded stimulus is a verbal statement (Schooler & Engstler-Schooler, 1990), or the verbalisation task is replaced with a visualization task (Schooler & Engstler-

Schooler, 1990). Additionally, the verbal code can be prevented from activating at recognition by lowering the recognition time to 5 seconds, thus resulting in best recognition (Schooler & Engstler-Schooler, 1990), or removed by presenting visual cues (Brandimonte, Schooler & Gabbino, 1997). However, since the recoding interference refers to the conflict between the verbal and visual codes of the *same* encoded object, it fails to explain why describing a different face to the one originally presented would negatively impact recognition accuracy (Dodson, Johnson & Schooler, 1997), or why describing a face would negatively impact recognition of an item from a different semantic category, such as a car (Brown & Lloyd-Jones, 2003). Additionally, the recoding theory does not explain why the VOE is not observed when the encoded and described faces are of a different ethnicity to the participant, yet is observed when they are of the same ethnicity (Fallshore & Schooler, 1995).

The second theory, criterion response, suggests that the VOE arises from a strict criterion response which participants adopt (Clare & Leandowsky, 2004). After a verbalisation task, participants were significantly more likely to indicate ‘not present’ when shown a line-up (if given this option); thus results indicated more misses with target-present line-ups, which manifested as verbal overshadowing; however, if presented with target-absent line-ups, these participants made more correct rejections. When participants were forced to choose a perpetrator from the target-present line-up, the VOE disappeared. These results have not been consistent, as they have not been replicated in all experiments that utilised a ‘not present’ option (for example, Fallshore & Schooler, 1995).

Schooler (2002) proposed that the verbal overshadowing effect arose from a transfer inappropriate processing shift (TIPS): a modified version of the theory of transfer appropriate processing (Morris, Bransford & Franks, 1977). Their research stemmed from the levels of processing research that stipulated that a deeper level of encoding would better aid with recognition, whereas a shallow level would harm recognition (Craik & Tulving, 1975). The transfer appropriate processing theory states that retrieval will be facilitated when an object is processed in the same manner at encoding and at retrieval; however, if the processing styles differ, then recognition will be impaired. Participants were provided with test stimuli which they encoded at either a semantic level (a typical example of deeper encoding) or a rhyming level (an example of a shallow encoding), and afterwards had to recognise the test stimuli in a semantic- or rhyming-recognition test. Results demonstrated that participants performed better on tests that matched the manner in which they encoded the stimuli: Semantically encoded stimuli were better recognised in the semantic-recognition test, while rhythmically encoded stimuli were better recognised in the rhyming-recognition test. These results were consistent across all three experiments, leading the researchers to conclude that the significance and effect of encoding is dependent on the type of recognition test used. Schooler (2002) reframed the verbal overshadowing effect within this paradigm, and hypothesised that the type of processing mode induced by the description tasks were not congruent with the type of processing needed at recognition, which resulted in poorer recognition of the encoded face.

In order to explore the TIPS, it is necessary to understand how faces are processed. There are a number of pivotal studies that have explored aspects of facial processing; from these, two different processing modes have been identified, configural and featural processing, and while faces are processed in both, configural processing appears to benefit facial recognition and processing more (Yin, 1969; Sergent, 1984; Young, Hellawell & Hey, 1987; Bartlett & Searcy, 1993; Tanaka & Farah, 1993). Configural processing emphasizes the configuration of the features and their spatial relationships to one another; featural processing processes each individual feature in isolation (Rakover, 2002). Research suggests that faces benefit from configural processing, because facial features are better recognised when they are placed within a face, rather than alone (Tanaka & Farah, 1993), and facial recognition is hampered when the face is inverted (which disrupts the configurations among the features, Yin, 1969; Tanaka & Farah, 1993). Brain imaging studies have supported these results: Faces elicit a greater response than common objects do from a brain region which is aptly named the fusiform face area (Kanwisher, McDermott & Chun, 1997), and when inverted faces are processed, a significantly decreased response is observed in the fusiform face area compared to the response elicited from upright faces (Kanwisher, Tong & Nakayama, 1998; Gauthier, Behrmann, & Tarr, 1999; Haxby et al, 1999; Rossion & Gauthier, 2002).

Studies using chimeric faces (which are “created’ by aligning together the top and bottom half of two different faces) have demonstrated a similar result (Young et al., 1987): Participants performed worse when they had to recognise only one half of an aligned face, but performed better when the halves were misaligned. By aligning the bottom and top halves, a gestalt of a real face is formed, making it difficult for participants to differentiate between the different halves as this requires featural processing of a stimulus that is normally processed configurally; however, when the halves are misaligned, this facial gestalt is not formed, and participants are better able to process each half in isolation (see Figure 1.1).

Figure 1.1. An aligned and misaligned chimeric face. Both left and right images are chimeric faces, and are constructed from the top and bottom halves of two different faces. When these two halves are aligned, like the image on the left, a gestalt is formed, and is processed configurally; however, when the halves are misaligned, like the image on the right, it is much easier to process each half separately.

The disruption of the configural relationships among facial features of an inverted face has been supported in studies utilising the Thatcher effect (Bartlett & Searcy, 1993): A grotesque Frankenstein-face that is created by inverting only the eyes and the mouths in an upright face (see Figure 1.2). This distortion is obvious when the face is upright; however, if the face is inverted, the distortion is not readily apparent.

Figure 1.2. An example of the Thatcher effect. The eyes and the mouth have been inverted and placed back in the image. When the image is inverted (left), then the grotesque appearance of the features is not obvious; however, once the face is reverted to its upright position (right), it is fairly obvious that the features have been inverted.

Faces are considered to be processed differently from objects, as facial features are better recognised when placed within a face than when viewed in isolation; however, parts of houses do not benefit from the whole, configural arrangement of the house (Tanaka & Farah, 1993). Interestingly, parts of inverted faces and scrambled faces do not benefit from the configural arrangement either. Neuropsychological evidence suggests that the mechanism underlying facial processing and object processing is different through a double dissociation between face and object processing; patients presenting with the neuropsychological syndrome Prosopagnosia fail to recognise familiar faces, however are able to recognise objects, whereas other neuropsychological evidence has recounted Patient CK whose facial recognition ability was spared at the cost of object recognition (Moscovitch, Winocur, Behrman, 1997, as cited in Kanwisher, 2000).

Within the facial processing literature, there seems to be contention about the nature of our facial processing ability – is it a domain-specific ability that is uniquely sensitive to faces (Kanwisher, 2000), or is it an expertise-specific ability capable of differentiating between exemplars of any category of expertise (Gauthier & Nelson, 2001)? Some research has demonstrated that expertise within a category can mimic the effects seen in the domain-specific literature, for example, dog experts were as impaired at recognising inverted images of dogs as with faces (Daimond & Carey, 1986), and both car and bird experts recruited the same brain regions that were normally associated with facial recognition (Gauthier, Skudlarski, Gore, & Anderson, 2000). This would suggest that our ability to process faces did not arise from a unique structure or need to do so, but rather from an expertise that has most likely developed from continued exposure to that particular category of objects, faces. However, McKone, Kanwisher and Duchaine (2007) have pointed out that (1) the dog expertise effect has not been replicated, and (2) the measured area for the FFA, as defined by Gauthieret al. (2000), was rather restrictive, and if it were altered to the measurements used by McKone et al., then the results would not be as significant. Within this preference for human faces, the FFA shows even greater activation for same-ethnicity faces than for other-ethnicity faces (Golby, Gabrieli, Chiao & Eberhardt, 2001), providing physiological evidence of the own race bias (Malpass & Kravitz, 1969). Twenty male participants (10 African American, and 10 European-American) were recruited, and had to view 42 photographs from each of the following four categories: African American male faces, European American male faces, antique radios and a fixation cross; each item was presented for 3.5 seconds, and participants were instructed to pay attention as they would have to recognise these objects later. Using a 3T MRI, functional imaging was performed on these participants to determine whether different brain regions were activated when viewing same- and other-ethnicity faces. Recognition results indicated that participants were significantly better at recognising own-ethnicity faces than other-ethnicity faces; specifically, the African-American (AA) participants performed significantly better at other-ethnicity recognition than the European-American (EA) participants. Two different criteria were used to define the FFA: a strict definition where the FFA was more narrowly defined (Kanwisher et al., 1997), and a broader definition (Gauthier et al., 2000). When the FFA was more stringently defined, activation was seen in only 9 participants (5 AA & 4 EA) as compared to 19 participants (10 AA and 9 EA) when using the flexible condition. Regardless of which FFA definition was used, significantly greater FFA activation was seen when participants viewed faces of the same-ethnicity compared to other-ethnicity faces. Within each FFA-definition groupings, all AA participants demonstrated greater activation when viewing same-ethnicity faces, whereas 75% of EA participants demonstrated greater FFA activation when viewing same-ethnicity faces. It could be argued that these results reflect the amount of social contact (or lack thereof) that each group had with each other, as suggested by the contact hypothesis (Brigham & Malpass, 1985), and in that case, it would then be expected that the African-American participants would have more social contact with an out-group than the European-American group would, and subsequently, they should have higher FFA activation for white faces (Eberhardt, 2005). Instead, the results demonstrated that every AA participant had greater FFA activation when viewing black faces, whereas only 75% of EA participants demonstrated this same-ethnicity effect.

In order to investigate whether the FFA was showing more activation for own ethnicity faces as a result of the ethnicity of the face, instead of the familiarity, Golarai, Gabrieli, Chiao & Eberhardt (2004, as cited in Eberhardt, 2005) showed White participants black or white faces that were either intact or whose features were scrambled inside the face area. White faces with scrambled features were used as stimuli, because it was presupposed as highly unlikely that anyone would have a high level of experience with such stimuli, thus possibly isolating stimuli familiarity from stimuli ethnicity. Golarai et al. (2004) replicated and extended the ethnicity effect reported by Golby et al. (2001) - there was higher activation in the FFA when White participants were exposed to same-ethnicity faces, than to other-ethnicity faces, *and* FFA activation was higher for intact and scrambled same-ethnicity faces than for intact other-ethnicity faces. These results suggest that the FFA is highly sensitive to faces, and even more so when viewing faces of the same ethnicity than faces of a different ethnicity; to borrow a succinct observation, there is an "expertise effect within an expertise effect" (Hayward, Rhodes & Schwaninger, 2008, p. 1018).

In summary, there is still much debate about the mechanism underlying our facial processing ability: specifically, is it an innate ability that is unique and highly sensitive to faces, or an expertise that arises from repeated exposure and which could occur with any category that we become experts of. It appears as though faces are processed configurally and featurally, and that configural processing improves processing and subsequent facial recognition. While research suggests that faces are processed differently to objects, it appears as though there is a sub-speciality where other-race faces are processed differently to own-race faces.

Linking Verbal Overshadowing through Automaticity to Facial Processing.

When participants describe a face, a featural processing mode is activated (Fallshore & Schooler, 1995), and this processing mode hampers subsequent facial recognition. However research suggests that not *all* verbal descriptions will impair facial recognition:

Elaborate verbal descriptions that focus on the physical qualities are more likely to impair recognition (Meissner & Brigham, 2001), while verbal descriptions which focus on the ‘personality’ assigned to that face may facilitate facial recognition (Wells & Turtle, 1987; MacLin, 2002). While both configural and featural processing are utilised in facial processing, featural processing alone may not facilitate facial recognition, hence elaborate verbal descriptions, which encourage a piecemeal, featural account of the face, may impair facial recognition, whereas the latter descriptions, which encourage configural processing, do not.

The negative impact of verbalisation on recognition extends beyond faces, and has been shown to impair other objects where participants had considerable expertise, such as wine-tasting (Melcher & Schooler, 1996), colour (Schooler & Engstler-Schooler, 1990), visual imagery (Brandimonte et al., 1997), mushroom identification (Melcher & Schooler, 2004), and insight problem solving (Schooler, Ohlsson & Brooks, 1993). Although these perceptual domains do not have configural qualities (for example, wine), they do have other similar qualities which require holistic/global processing to form the entire percept; similarly, procedural/local processing catch the ‘featural quality’ of the stimuli. Domains where one has high perceptual expertise, but low verbal expertise, are more vulnerable to the negative effects of featural processing style induced by verbalisation (Schooler & Melcher, 1996). These perceptual domains require holistic processing, and therefore would benefit from tasks that would induce such a processing style; however, when filler tasks induce a featural processing mode which is then carried over to the recognition stage of the experiment, recognition will be impaired. Much like facial processing, these mismatching modes result in verbal overshadowing. For example, both mushroom experts (Melcher & Schooler, 2004) and wine experts (Melcher & Schooler, 1996) performed worse on a recognition test after they verbally described the stimuli (mushrooms or wine respectively) that they had previously studied. Participants were trained as either perceptual or conceptual mushroom experts: in the former, participants learnt how to visually distinguish among different mushrooms species, and in the latter, they learnt the taxonomy and features of each species in order to distinguish among them (Melcher & Schooler, 2004). When the training was completed, participants were exposed to a mushroom of which they wrote a detailed description, and after this, they were instructed to choose the studied mushroom from 6 photographs of different types of mushrooms. Perceptually-trained participants had poor recognition after verbalisation, but conceptually-trained participants performed better after verbalisation. These results suggest that increased familiarity with a perceptual domain with abstract qualities, which are difficult to describe, results in increased vulnerability to the negative effects of a verbalisation task. However if the increased perceptual familiarity and expertise is matched with a similar level of verbal expertise, then recognition performance will not be as negatively affected – verbal overshadowing does not occur, because there is no shift in processing (Melcher & Schooler, 1996). This was demonstrated in research that specifically investigated a domain where participants had different levels of verbal and perceptual expertise: wine drinking (Melcher & Schooler, 1996). Participants were divided based on their training into three groups: trained wine experts, untrained wine drinkers and non-wine drinkers. All the participants had to taste two red wines, and after a describing both, they had to identify both wines from 5 plastic, unlabelled glasses: 4 were filled with other wines, and 1 contained the previously tasted wine. The untrained wine drinkers had the worst recognition following verbalisation, whereas non-wine drinkers benefited from the verbalisation and the experts' recognition scores were unaffected. The untrained wine drinkers had performed so poorly, because their high perceptual expertise was unmatched by their low verbal expertise in the wine domain, whereas non-wine drinkers and wine-drinkers did not have differing levels of verbal and perceptual expertise – both were low and high respectively.

If verbal overshadowing occurs as a result of a processing shift, then it is of no importance what the *content of the verbalisation is,* so long as this shift occurs. Therefore any task that requires featural processing should impair recognition of configural processing stimuli, and the reverse relationship should exist as well. When the verbalisation task was replaced with the Navon letters (Navon, 1977), which act as both a featural and configural processing task, such an effect was observed (Macrae & Lewis, 2002). Navon letters are capital letters which comprise of lowercase letters, for example, a group of aligned lower-case D's form a capital 'B', or lowercase 'X's that form an 'S' (see Figure 1.3). Participants were shown the original video that Schooler and Engstler-Schooler (1990) used in their VOE study, and then divided into three groups: featural, procedural and control. Both the featural and procedural group were shown the Navon letters: the featural group was asked to identify the lowercase letters, while the procedural group was asked to identify the large capital letter. By focusing on the lowercase letters, this group engaged in a featural processing task, and as expected, the proportion of participants who made correct identifications was lower than both the procedural group (who had engaged in the procedural task of identifying the large letter) and the control group (0.3 vs. 0.83 vs. 0.6 respectively). Not only did this study demonstrate that a featural processing task, which was unrelated to the encoded stimuli, could induce verbal overshadowing, but complemented previous findings that a procedural processing task (for example, listening to music, or completing a maze) prior to recognition would improve recognition accuracy (Finger, 2002).

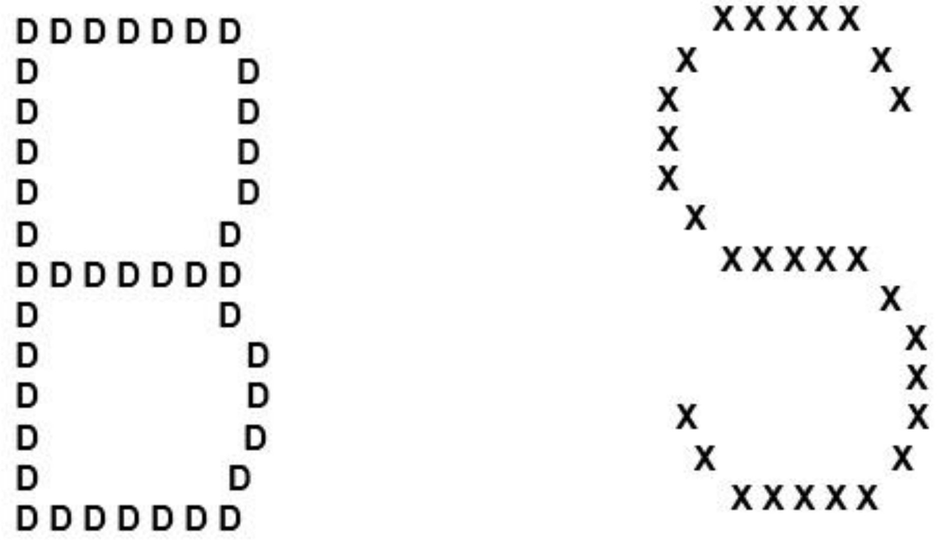


Figure 1.3. Two examples of Navon letters. Both images are constructed out of multiple lowercase letters and form a single capital letter, ‘B’ on the left, and ‘S’ on the right. Typically, participants are asked to respond by naming either the lowercase letters, or the large capital letter; each response utilises a different type of processing: featural and configural respectively.

Additional support for the TIPS theory was found when participants had to complete the same type of Navon letters task prior to both encoding and recognition; recognition was enhanced when both sets of Navon letters utilised global or local processing, but impaired when either set activated a different type of processing from the other (Lewis, Mills, Hills & Weston, 2009). 90 participants were recruited and placed in one of nine different groups; each group had a different combination of the three different Navon tasks prior to encoding (control, local and global) and three different Navon tasks prior to recognition (control, local and global).In order to ensure that the effect of the Navon task would perpetuate throughout the recognition and encoding stages, participants would complete Navon tasks that were strategically placed between each presentation of the 14 stimuli faces. Results suggested that the Navon effect was replicated, and recognition was indeed better following the Global-Navon letters at the recognition stage – but only for the Control and Global group. Those participants who had responded to the local-processing Navon letters during both encoding and recognition had better d’ scores: a signal detection measurement (Macmillan & Creelman, 1991) that investigates participants’ response bias to stimuli by comparing their hit rates to their false alarms rate, and calculating the difference between those computed z scores. Both sets of results suggest that it is not the processing style of the task that occurred prior to recognition that ultimately affects recognition, but rather the congruency between both processing types of the encoding and recognition tasks which facilitated recognition, and that the effect of the Navon letter was greater during the recognition stage than the encoding stage. The authors acknowledged that since the facial stimuli used during the encoding and recognition stage were the same, picture recognition, rather than memory retrieval may have been tested. Therefore, a second experiment was conducted where two different expressions of each face was used (28 X 2 = 56 photographs); one was shown at encoding, and the other at the recognition stage. The control groups were removed from this design, and only four experimental groups remained with 10 participants in each: local encoding-local recognition, local encoding-global recognition, global encoding-global recognition, and global encoding-local recognition. The results from these experiments mimicked those of the first experiment, with a significantly higher d’ when the processing modes were identical for both sets of Navon letters (the higher the d’ value, the better the participants were at detecting, in this example, the faces). Additionally, no significant effect was seen for the Navon task processing mode used prior to recognition, leading the authors to suggest that the typical Navon effect that is seen in prior experiments may be due to the incongruency between the processing mode that is typically used by participants when processing faces, without the influence of any tasks, and the processing mode that is induced by the Navon letter when it is used as an interval task.

Interestingly, research has shown that the Navon letters may not have activated the same processing styles as concentrating on the features or personality of a face (Weston, Perfect, Schooler & Dennis, 2007). Working from the premise of the transfer appropriate processing shift, it was hypothesized that recognition should be greater when stimuli were recognized and encoded using the same processing style, than when the recognition and encoding processing styles differed. Therefore, Navon letters tasks were introduced before the face encoding (as an encoding task), and again before the facial recognition (as a filler task); either set of Navon tasks encouraged either featural or configural processing. 32 Face stimuli were used, and consisted of own- and other-race faces. Results demonstrated that recognition was best following an interval task that encouraged configural processing, regardless of whether the encoded face was of the same ethnicity or a different ethnicity to the participant. When only the Navon letters encoding task (which occurred prior to encoding) were replaced with a configural/featural command, such as asking participants to study the personality or the features of the face, recognition of own-ethnicity faces benefited from the filler task which encouraged configural processing, while different-ethnicity faces benefited from the filler task which encouraged featural processing. When compared to the control group, improved accuracy was observed across *all* conditions for the filler task; specifically, configurally-encoded stimuli benefited from both global processing and verbalisation interval tasks, whereas locally-encoded stimuli benefited from both local processing and verbalisation interval tasks. Following these results, the experimental design was altered so the face stimuli were only of the same ethnicity – and again, accuracy scores benefited from a local-processing and global-processing filler task, regardless of the processing type of the encoding task. Within Schooler's TIPS theory, a filler task and recognition task should not benefit each other if they encourage different types of processing; therefore, the authors tentatively concluded that the Navon letters may not have induced the same configural/featural processing styles as those induced when participants focused on the personality/facial features.

Perfect et al. (2008) stated that the commonly used forms of Navon letters may not be appropriate ways of assessing the two processing styles, because the letters are biased towards global processing – they have global precedence: it is extremely easy to process the large letter, suggesting that even in a local processing condition, one could engage in some global processing (see Figure 1.4).

Figure 1.4. A globally-biased and locally-biased Navon letter. The image on the left is globally-biased: it is much easier to process the letter ‘F’ than the smaller ‘x’s. However, it is much easier to process the smaller ‘x’s when viewing the image on the right, as this image is locally-biased. Perfect et al. (2008) stated that most Navon letters, like the image on the left, are globally-biased, therefore global-processing always occurs to some extent.

In order to investigate whether the bias of the Navon letters was affecting facial recognition, two sets of adapted Navon letters (They were biased either locally or globally.) were used, and participants had to respond to either the small or large letter, which required local or global processing respectively. Therefore in a 2 (precedence of Navon letters) x 2 (featural vs. global response) design, participants were shown the same video that was used in Macrae & Lewis (2002), and were then divided into four groups: global precedence/global response; global precedence/featural response; local precedence/global response; and local precedence/featural response (Perfect et al., 2008), the results were unexpected: recognition was better after the global precedence/global response *and* local precedence/featural response. In a second experiment, the upright face was replaced with a face composite that was presented during the recognition stage as this relies on featural/local processing (Young et al., 1987), and therefore should benefit from a prior featural processing task. Participants were again divided into the same groups as above, and after each task were presented with the face composite. Ceiling effects were seen in the recognition data, but the response latencies showed that faster decisions were made after the global precedence/local processing and local precedence/global processing conditions. Therefore results from both experiments did not conform to the predictions of the TIPS theory. For experiment one, there was a significant interaction between Navon letter precedence and participant response type, specifically, recognition accuracy was higher when participants responded globally to the globally-biased letters, *and* locally to locally-biased letters – this second set of results is not accounted for within TIPS theory. The results from the second experiment did not conform either, as the recognition speed, which is interpreted as a measure of how easy it was to perform the task, was faster following the local precedence/global processing task and the global precedence/local processing task.

The authors suggested that the difference in recognition scores could be better accounted for by a shift from automatic to controlled processing; therefore, to reframe this suggestion within their experiments, a face is processed automatically and should benefit from a prior task that is also processed automatically – hence the global-precedence/global-response, and local-precedence/local-response conditions resulted in higher recognition accuracy. The composite task, which required controlled processing (because the inclination to analyse the face holistically had to be suppressed in order to shift attention to only one face half), benefited from prior tasks whose response required a different processing mode than that induced by the precedence, for example, the global precedence/local processing, and local precedence/global processing conditions. Therefore, it may be that the emphasis of the TIPS theory should not only be on the incongruency of configural and featural processing, but should include the incongruency of automatic and controlled processing.

This is not the first time that automatic and controlled processing have been suggested as a reason for verbal overshadowing (Schooler, 2002; Melcher & Schooler, 2004); and as facial-processing experts (whether through an innate ability or from a developed expertise), humans process face automatically. Initially, automatic and controlled processes were placed in opposition to each other; automatic processes were fast and unconscious, and were activated without taxing cognitive resources, whereas controlled processes were slow, required attention and usually required serial activation, because of constraints that were placed on cognitive resources (Shiffrin & Schneider, 1977). These categories were considered mutually exclusive with tasks falling within the boundaries of only one category, so much so that exclusion from one category by default meant inclusion in the other. This dualistic approach has been refined: these two processes are considered opposite ends of a spectrum, and mental processes are a combination of features from both categories (Cohen, Dunbar & McClelland, 1990), and the classification of mental processes as automatic or controlled is not static (MacLeod & Dunbar, 1988). For example, most research that investigated controlled and automatic procedures utilised the Stroop task: a task that uses stimuli that have two conflicting dimensions (the word meaning and the font colour), and participants are required to respond to only one while ignoring the other. Participants are quicker to respond to the written word, while ignoring the ink colour that the word is written in; however, when asked to say the ink colour, participants performed much slower. Researchers suggested that this is because the word meaning, which is easily read, is activated much faster than the font colour, and therefore in order to respond to only the font colour, participants need to suppress the lexical activation. This suppression of the 'automatic process' requires time and cognitive resources, which is reflected in the longer reaction time to say the ink colour condition of the Stroop task. Processing ink colours and words are not always controlled and automatic processes respectively and participants can be trained to respond to new stimuli as though it were automatic. The Stroop words were replaced with polygons and each polygon was given a name which participants learnt in a brief training phase. After the training, the ink colour of these polygons interfered with the naming of the polygons, suggesting that ink colour was automatically processed (whereas originally it was processed in a controlled manner), and after 20 days of extensive training, the polygon naming now interfered with the ink naming (MacLeod & Dunbar, 1988). These results demonstrate the difficulty of classifying tasks as merely automatic or controlled, because through training, a controlled task can become automatic.

Automatic tasks are characterized as fast; working in parallel with other tasks and therefore not affected by limitations of cognitive resources; activated unconsciously, and once activated they tend to run their full course (Shiffrin & Schneider, 1977). Automatic processing is difficult to learn, but once learnt, it is difficult to change. Therefore, over multiple trials it is possible to learn to process a task automatically (Shiffrin & Schneider, 1977). Alternatively, controlled tasks are very dependent on cognitive resources, as they are slower, require more attention and occur in serial. Therefore, performance speed and accuracy will decrease if controlled processing tasks have to compete with other cognitive tasks for mental resources. Controlled tasks are easily learnt, and easily modified (Shiffrin & Schneider, 1977). It is proposed that automatic processing works within a stimulus-response paradigm, where a stimulus, or multiple stimuli, activates a set response (Shiffrin & Schneider,1977), and the more training (expertise) that one has with this stimulus-response task, then the stronger the automatic processing becomes (Logan, 1988). For instance, new tasks, such as learning how to drive, are treated as controlled tasks, but they become automatic as the user (or driver in this example) gains more familiarity and expertise with it. Therefore, high expertise should benefit from automatic processing conditions (for example, reduced time to complete the task), and low expertise should benefit from controlled processing conditions (for example, more time to complete the task) (Jackson, Ashford & Norsworthy, 2006). Most experts struggle to perform well in their respective expert-tasks when they are required to reflect on that task's procedure, for example golf experts putt poorly when they have to attend to how they swing their clubs and ensure that they keep the club head straight, but perform better when they have to split their attention between two tasks (Beilock, Carr, MacMahon & Starkes, 2002), whereas novice golfers demonstrate the opposite pattern of performance: they performed better when required to reflect on a task's procedures than when their attention was divided (Beilock, Bertenthal, McCoy & Carr, 2004). This effect was not limited to reflecting while performing the task, because just describing their actions prior to putting hampered expert golfers' performance, while novice golfers were unaffected (Flegal & Anderson, 2008). If experts are forced to perform the task under a time constraint the effects of the verbalisation task are released. Expert putters performed better under a time constraint than when encouraged to be as accurate as possible and without time constraint (Beilock et al., 2004), and participants were better at recognising a face when they only had 5 seconds to make a decision – even after providing a verbal description of that face (Schooler & Engstler-Schooler, 1990) This negative effect of verbalising a sensorimotor domain extends to perceptual abilities as well (Melcher & Schooler, 1996; Melcher & Schooler, 2004).

There does seem to be some suggestion that verbalisation causes controlled processing, which then interferes with the automatic processing of faces. Although not tested explicitly, previous research does suggest this. For example, like experts perform better under timed conditions, when the recognition time for faces was reduced to 5 seconds, there was less interference from the prior verbalisation task and recognition was better (Schooler & Engstler-Schooler, 1990). Additionally the negative effects of verbalisation are attenuated over multiple trials – perhaps a similar pattern as that seen when controlled tasks become automatic due to training and exposure. This argument may extend to the ORB, because perhaps our expertise with same race faces endangers us to the negative effects of a controlled processing task (like a verbal description) which then presents as verbal overshadowing. In contrast, our lack of expertise with other race faces bolsters against this effect, and should benefit from a controlled processing task, which is why the VOE is not observed when recognising faces of a different race. However if contact and exposure to such faces are increased, then those faces are processed holistically (Michel, Caldara & Rossion, 2006), and through the development of this expertise, their vulnerability to verbal overshadowing is increased.

Brand (2004) conducted a web-based experiment that investigated whether the Stroop task could affect facial recognition and reaction time. His experiment followed from research which suggested that upright faces were recognised better when processed in the left visual field (and subsequently the right hemisphere), unlike inverted faces which were better recognised when presented in the right visual field (and subsequently the left hemisphere (Leehey, Carey, Diamond & Cahn, 1978 and van Kleeck, 1989, as cited in Brand, 2004). The Navon letters and Stroop task showed similar hemispheric preferences: the featural condition of the Navon letters and the word condition of the Stroop were processed faster in the left hemisphere, while the global condition of the Navon letters and the colour condition of the Stroop were processed better in the right hemisphere. Therefore, Brand (2004) aimed to induce the verbal overshadowing effect by replacing the Navon letters with the Stroop task as they seemed to engage the same hemispheric processing. Participants (n = 194) viewed a target face, positioned at 45 degrees in the photograph, for 2 seconds, and then completed a ten minute filler task – the first five minutes were dedicated to an unrelated sliding block puzzle, and a Stroop task occurred during the remaining 5 minutes. Participants were assigned to either a word-naming or colour-naming condition of the Stroop, with the response dictated by the group that they belonged in. Afterwards, participants had to recognise the target face from a target-present line-up that consisted of 6 frontal faces. Participants could choose a face, or indicate not present, and were required to rate their confidence in their choice.

Recognition accuracy and confidence ratings did not differ significantly between the two groups (70% versus 74%, $p > 0.05$ and 6.89 versus 6.95, $p > 0.05$ respectively), however this could be due to the failure of the experimental condition. Instead of relying on a key-press, participants were required to type their responses to the font- and word-Stroop conditions, and this may have afforded them more opportunity to produce the correct answer, and overcome the Stroop effect. The Stroop effect is affected by the response modality: less interference and more facilitation is observed when a keyboard-press response is required as opposed to an oral response (as cited in MacLeod, 1991), and this effect would be exacerbated by requiring participants to type out the entire response. Not surprisingly, this lengthened participants' reactions time (12 seconds versus 10 seconds for the colour and word condition respectively), but the difference between these two means was not significant. Therefore, it appears as though the experimental condition was not induced.

There are other metholodological problems with this experiment, namely that no demographic data was collected about participants, such as area of residence or ethnicity. The participants could have represented different ethnic or social groups, which may have differed from the target face, thus causing an ORB effect which lowered the entire recognition rate for each group. Brand used the same target face and line-up that Finger and Pezdek (1999) had used in their Experiment 3: the target face was that of a 31-year-old white male whose face was positioned at a 45^0 angle, and the line-up consisted of 6 faces (1 target and 5 foils) that were matched on a verbally similar description. However neither effective line-up size and bias, nor any ratings of the target face, such as distinctiveness, were reported. Effective line-up size and bias are measures of line-up fairness (Malpass, Tredoux & McQuiston-Surrett, 2007); effective size refers to the number of foils in the line-up who would be considered feasible perpetrators, for example, if the perpetrator has red hair, then the line-up effective size would be higher if many of the foils had red hair. Line-up bias refers to the likelihood that the suspected perpetrator is chosen out of a line-up, and can be biased towards or away from them. If the line-up bias is high, then choosing the perpetrator is an easy task, and could indicate that the perpetrator stands out, or the foils are not appropriate matches; if the bias is low, then this could indicate that it is too difficult to identify the perpetrator, and that the foils are too similar. Ideally, the bias should approximate chance (1/number of people in the line-up), thus each person is chosen equally. Without this information, it is impossible to know how well the line-up was constructed, which affects how easy or difficult it was to choose the perpetrator. This type of information is vital for eyewitness research, as it ensures that the stimuli used are fair and unbiased, thus allowing for better conclusions to be drawn from the tested hypotheses.

Aims

It appears from the literature that the VOE is not solely caused by a verbalisation task, but can be induced through the transfer-inappropriate processing shift. This shift could be from holistic processing, which is utilised when encoding faces, to featural processing, which is encouraged when describing the face or reading the small-case letters in Navon letter - or it could be a shift from automatic to controlled processing. There is some support that experts perform their respective expert-tasks in an automatic manner, and that their performance is negatively affected by controlled processing, much like how our expertise for facial processing is impaired by verbalisation. Even though there is some debate about whether facial processing is domain specific or an expertise effect, there is a definite improvement for processing faces of the same ethnicity than faces of people belonging to a different ethnicity; verbalisation only affects the former. Processing the featural condition of the Navon letters causes the same detrimental effects on facial recognition as verbalisation does, but the effect cannot be simply attributed to the featural component of the Navon letters. The Navon letters do not solely cause global and featural processing, but have an inhibitory effect which is reminiscent of automatic and controlled processing, thus mudding the conclusions drawn: was the effect seen because of a shift from global to featural, or from automatic to controlled?

The aim of this research project is to investigate whether facial recognition is affected by a shift from automatic to controlled processing. Additionally, the own race bias is also examined within this framework, as our expertise for recognition of faces of the same race as ours may benefit from a task that induces automatic processing, while recognition of other race faces should benefit from a controlled processing task. My hypotheses are as follows:

White participants should have higher recognition scores and perform faster following an automatic-processing task, whereas other participants should perform better and faster at recognition after a controlled-processing task.

Chapter Two: Method

Design and Setting

This experiment had a mixed repeated design: the between-subject factors were ethnicity (white, other) and processing type (automatic, controlled), and the within-subject factor was the order that the perpetrators (there are 3 different orders) are presented in (see Table 2.1. for the number of participants in each group). The dependent variables were Recognition Accuracy and Reaction Times.

Table 2.1.

The number of participants that were in each condition in the experiment.

		Processing type groups	
Participant ethnic groups	Total n	Automatic	Controlled
White	144	72	72
Other	144	72	72

The experiment was divided into three trials, and each trial comprised of the following three stages: encoding, interval task and recognition. Participants had to recognise a different perpetrator in each trial, and a total of three perpetrators were used in the entire experiment. Perpetrator order varied so that order and perpetrator identity were not confounded.

Randomisation schedules were created for the groups: White Male, White Female, Other Male, and Other Female. Each schedule comprised of blocks that contained 12 entries: all the possible combinations of experimental conditions, such as perpetrator order (3), target position (2) and processing type (2). The Target position varied in the 8-person line-up, and was either in one of two positions; the two assigned positions varied for each perpetrator so that participants didn't notice a pattern in the perpetrator position. In total, there were 2 blocks (24 participants) for White Male and Other Male, and 10 blocks (120 participants) for White Female and Other Female.

All participants were tested on the computers in the ACSENT Laboratory in the Department of Psychology, University of Cape Town. Ethical approval was granted from the Research Ethics Committee in the Department of Psychology.

Participants

The sample consisted of 288 participants (240 females), ranging between 18 and 27 years in age (M= 20.27; SD = 1.55), who were recruited from the undergraduate student body at the University of Cape Town. Participants were recruited through either the Department of Psychology's Student Research Participation Programme (SRPP) and received SRPP points, which are a compulsory requirement for their undergraduate degree, or they were recruited through advertisements and received R20 for their participation. Neither the advertisements that were used for the SRPP programme and on campus mentioned that there was a recognition test involved, as this could have primed participants to the recognition task. Instead, these advertisements stated that the aim of the experiment was to investigate facial perception. All participation was voluntary, and participants were reminded that they could leave voluntarily. Each participant had to complete a consent form (see Appendix A) that explained the aim and research procedure before participating.

The sample size, N = 272, was calculated using an adjusted power of 0.70 and the odds ratio (2.26) that was reported in Experiment One in Perfect et al. (2008).The sample for my experiment had a predetermined 5:1 female-to-male ratio (67% of students enrolled in Psychology at UCT are female), and therefore, it was necessary that my sample size was divisible by 6, thus the final sample size was 288.

Half of the sample size (144) marked themselves as ‘White’, and were born in South Africa; the remaining half had indicated that they were either Black, Coloured, Indian, or Other (for example, Asian), and they were grouped into the participant group ‘Other’.

I specifically wanted two groups of participants for my experiment: a low-expertise and a high-expertise group, because I had hypothesised that both groups would perform differently from each other following the automatic/controlled processing task (because high expertise with a task leads to automatic processing of that task). My participant groups differed in the amount of familiarity that they had with processing faces that were similar to my targets (who were white South Africans). The low-expertise group consisted of participants who had indicated that they were either black, Indian, coloured or ‘other’, and my high-expertise group consisted of participants who had indicated that they were white; “expertise” was a reflection of own-group familiarity, thus the high-expertise group would have higher familiarity with their own-group than the low-expertise group, and these differing levels of familiarity may encourage different modes of processing. The only exclusion criterion was that these participants, who had indicated that they were white, had to be South African, as our perpetrators were White South Africans; this exclusion criterion was not a necessity for participants who had indicated that they were black, Indian, coloured or ‘other’. This criterion was used for white South African participants, because research has shown that recognition accuracy can differ across what seems to be the ‘same’ ethnic group, and to such an extent that it mimics the ORB (Chiroro, Tredoux, Radaelli & Meissner, 2008). For example, South African participants were worse at recognising North American faces than South African faces, and white and black participants were better at recognising white and black South African faces respectively. Both of these groups of participants performed so poorly on North American faces that the recognition results suggested an ORB. Therefore, the authors’ conclusions were that faces that appear to be from the same ethnic group are not necessarily processed as such, and instead are treated as a (geographical) out-group, thus poorer recognition occurs. Therefore, in order to avoid such results from appearing in my research, my high-expertise group consisted of only White South Africans.

Materials

Selection of targets.

This experiment used three perpetrators based on their availability and willingness to participate for a small reward. Their participation required that they perform in a short video, and have their photograph taken.

Encoding stimuli: video construction.

Three short films were recorded with one perpetrator acting in each. The setting for the video was an empty classroom in the Department of Psychology. The video depicts the perpetrator committing an act of theft: he steals money out of a satchel that he finds in an empty classroom (please see Appendix B for a list of hyperlinks to an online storage location that contains all the research stimuli used, including the original videos, edited videos, and line-ups). Throughout the video, the perpetrator is facing the camera, but often he looks towards the door and the windows, as though checking whether anyone has seen him, and this allows the viewer to see his face from different angles.

The perpetrators were run through the script a number of times in order to desensitise them, and familiarise them with the script. Recording commenced once they felt more comfortable, and they were guided and cued through the script, for example, they were reminded to look at the door, or at the window. There were two types of videos recorded for each perpetrator: a ‘CLOSE’ where only the shoulders and head of each perpetrator could be seen once they had opened the satchel, and a ‘FAR’ video, where the entire torso and head of each perpetrator was visible. Once the recordings were completed, each perpetrator had their photograph taken, which would be used for the recognition stage of the experiment.

The videos were then edited, and two 48-second videos, one containing the footage from the close-up video and the other containing footage from the far-away video, were made for each perpetrator. The original sound in the video was removed, because one could hear the researcher cueing each perpetrator, and was replaced with the muffled sound of music playing through earphones.

Recognition stimuli: line-up construction.

A line-up was constructed for each perpetrator, with two orders: Perpetrator 1 was placed in position 3 or 6; Perpetrator 2 was placed in position 2 or 8; and Perpetrator 3 was placed in position 4 or 5. The two orders differed across the line-ups, because I was concerned that if I had used the same line-ups positions for each line-up, then the participants may realise this (since this is a repeated measures design) and may use this strategy when making a choice. These simultaneous line-ups were the same size as those used by Perfect et al. (2008), and contained 7 foils and 1 target. The line-ups were constructed according to the guidelines from Malpass et al. (2007), specifically a modal verbal description of each perpetrator was created, and used to find matching foils. Six, white females, who were recruited through the Department of Psychology’s Student Research Participation Programme (SRPP), were run as a group in the ACSENT laboratory. A Microsoft PowerPoint slideshow was opened for each participant, and the second slide contained the face of the first perpetrator. The participants were instructed to view the face, exit the slideshow, delete the photograph, and then return to the PowerPoint presentation. The next slide contained a distracter task that instructed participants to make as many three or more lettered words from a given word. Once participants had provided as many words as possible, they then moved on to the next slide that contained the following instructions '*We would now like you to describe the face you saw just before the work task. Your description could be used as the basis for identifying that person'*, and space was provided for their typed response. This procedure was repeated with the remaining two perpetrators, until a description of all three participants was obtained. A modal description was created for each perpetrator, and consisted of the characteristics that were mentioned the most often by the 6 participants (see Table 2.2 for the modal descriptions, and Appendix C for the full descriptions provided by each participant).

Table 2.2.

Modal descriptions for each perpetrator.

Perpetrator	Modal Description
1	Brown hair, brown eyes, messy/styled hair, and maybe blemished skin.
2	Long, dark brown hair; thick eyebrows; thin lips and a thin face
3	Longish hair; dark blonde or light brown hair, chubby and brown eyes

The descriptor 'blemished skin' for Target 1 was preceded by 'maybe' as it was only mentioned by 3 (50%) participants. There was some risk for including this descriptor, because instead of being treated as a characteristic that may or may not be present, it could be considered by the participants as an always-present descriptor.

After the modal descriptions were created, 6 white participants were recruited to select foils that matched these modal descriptions. Each modal description was given to these participants, one at a time, and they were instructed to search through a .pdf album (a digital database that contained standardised photographs of faces of white males who were undergraduate students at UCT) and to select seven faces that matched the description best. All these results were tabulated, and the faces that were selected by the most participants were the foils. These faces were then cut out of their respective photographs and superimposed on top of a coloured square used as a background – a different colour for each face. The rationale for using different coloured backgrounds was that the photographs had been taken under different conditions with different cameras, and therefore the quality and colouring differed among them. Rather than editing the photographs to minimise the differences among them, differently coloured background were inserted in order to make them so dissimilar that a relative comparison strategy is discouraged (Schmidt, 2010). All three line-ups are included in Appendix D.

Once, the line-ups were constructed, participants were recruited as ‘mock witnesses’, and were required to select a target from each line-up (with or without the modal description); this allowed the bias and effective size of each line-up to be calculated (Malpass et al., 2007). Line-up bias refers to how much the line-up is biased towards, or away from the perpetrator; without the visual knowledge of an actual witness, a mock witness should not choose the perpetrator more than chance (1/number of line-up members) if the line-up is well constructed and all the foils are appropriately selected. A line-up can be biased in a number of ways, such as an artefact of the photograph or distinctiveness of the target, and line-up bias is calculated by dividing the number of times the perpetrator was chosen by the number of mock witnesses. Even though a line-up may have a set number of members, not all of these are considered suitable options for the eyewitness, thus reducing the effective size of the line-up. These line-up members are dismissed by eyewitnesses, and are not taken into account when deciding the identity of the perpetrator, thus increasing the risk of false identification. Four mock witness experiments were run: three with the modal description and one without, and slight changes were made to either the line-ups, the descriptions or the instructions to the participants (see Table 2.3 for the results and changes). All of the undergraduate students recruited were matched on ethnicity to the perpetrators.

The line-ups that were created for the fourth mock witness experiment were used in the main experiment. Some of the less suitable faces, such as those that were never chosen, were removed, and replaced with faces that better matched the modal description and physical similarity of the perpetrators. The coloured backgrounds were replaced with textured grey backgrounds, and an artificial shadow was added between each face and the background. The perpetrators' clothing was erased, and a black T-shirt was drawn onto the image. All the faces were standardised for inter-ocular distance, and were in a frontal position (please see Appendix D for the old line-ups and Appendix E for the standardised line-ups). The instructions to participants were as follows, "*I'm going to show you three line-ups, and you will be required to pick a target from each line-up. These line-ups are meant to simulate a real line-up, and you should assume that they were constructed by the police. Who is the suspect that matches the following description best?*" Only the description for Line-up 1 was unaltered; 'thin lips' was removed from the modal description for Line-up 2 and 'chubby' was removed from the modal description for Line-up 3, as both were the most pronounced in our perpetrators. The biases for all three line-ups were optimal; however the effective size for Line-up 2 was rather small with only 2.72 members considered suitable foils. Despite the poor effective size for Line-up 2, these three-line-ups were used for the main experiment, and there were no better suited photographs in the photograph databases that could be used to replace the less appropriate foils in Line-up 2.

Table 2.3.

All the mock witness experiments and their associated effective sizes, biases and instructions.

Mockwitness experiment		Line-up 1	Line-up 2	Line-up 3	Instructions	Changes
1: non-verbal	n	133	133	133	You have been asked by the police to help identify a suspect from a line-up. A line-up consists of photographs of similar looking men. In each line-up, a suspect is present.	
	Effective size	5.68	4.87	5.15		
	Bias	14.29%	3 %	15.54 %		
2: verbal	n	9	9	9	You have been asked by the police to help identify a suspect from a line-up. A line-up consists of photographs of similar looking men. In each line-up, a suspect is present. Please choose the perpetrator that fits the following description best.	The modal verbal description of each perpetrator was provided to participants.
	Effective size	1.53	1	5.4		
	Bias	22.2%	100%	22.2%		
3: verbal	n	102	102	102	I have a bunch of faces here, and I need you to pick which face matches the description best.	Words such as ‘police’, ‘perpetrator’ and ‘line-up’ were not included in the instructions with the intention of preventing participants from choosing a face based on some form of criminality. Some foils were replaced. The descriptor ‘thin lips’ was removed from Perpetrator 2’s modal description.
	Effective size	4.15	5.97	2.69		
	Bias	40%	22%	55%		
4: verbal	n	145	112	87	I’m going to show you three line-ups, and you will be required to pick a target from each line-up. These line-ups are meant to simulate a real line-up, and you should assume that they were constructed by the police. Who is the suspect that matches the following description best?	Faces that were never chosen in the previous paradigms were removed. The colour backgrounds were removed and replaced with a textured grey background. Photographs were standardised for intraocular distance. ‘Chubby’ was removed from Target 3’s modal description.
	Effective size	5.78	2.72	4.82		
	Bias	17.24%	15.18%	10.34%		

Video type and video length.

Three pilot experiments were conducted to determine (1) which video of the two types of videos created (‘CLOSE’ or ‘FAR’) would be more appropriate for the experiment, and (2) what the optimal length of that video should be in order to avoid ceiling effects, and to allow for effect of the Stroop task to be properly demonstrated. Participants were assigned to one of two groups, which used either the ‘CLOSE’ or ‘FAR’ video as encoding stimuli, and the procedure that was followed was the same as that used in the main experiment. The length of the videos used in the first pilot experiment was 48 seconds (this was the original length of the entire video), but they were reduced to 19 seconds for the second experiment, and again reduced to 10 seconds. The reason for this was that there were apparent ceiling effects for all three perpetrators, and reducing the video to 19 seconds removed *only* the ceiling effects for Perpetrator 1 and Perpetrator 2, therefore these videos had to be shortened even more (see Table 2.4). After the three pilot experiments were run, the final decision was to use the 10 second ‘FAR’ video; this may seem counterintuitive as recognition performance across the three pilot studies was lower when the ‘CLOSE” video was used (which was the aim of reducing the video length), but the average recognition performance (across all three perpetrators) for the ‘FAR’ video was 64.1 %, whereas the average recognition performance for the ‘CLOSE’ video was only 35.9%. Using the ‘FAR’ video meant that were was more range for participants to perform better and worse, thus better testing the effect of the Stroop letters.

Table 2.4.

Recognition performance in each video length pilot experiments.

	Video type	Line-up 1	Line-up 2	Line-up 3	Average
Video Length 48 seconds	Far	50%	71.4%	78.6	66.67%
	n= 14	(7)	(10)	(11)	
	Close	85.7%	85.7%	100%	90.47%
	n=14	(12)	(12)	(14)	
Video length 19 seconds	Far	44.4%	55.5%	77.8%	59.2%
	n=9	(4)	(5)	(7)	
	Close	55.5%	88.9%	100	81.47%
	n=9	(5)	(8)	(9)	
Video length 10 seconds	Far	76.9%	30.8%	84.6%	64.1%
	n=13	(10)	(4)	(11)	
	Close	61.5%	15.3%	30.8%	35.87%
	n=13	(8)	(2)	(4)	

Note. The number of participants that correctly identified the perpetrator in each line-up is presented in parentheses.

TIPS task: Stroop task

Often the Navon letters are considered to have a Stroop-like quality: the necessary inhibition of the dominant dimension of the stimulus in order to process the other non-dominant dimension of that same stimulus. Perfect et al. (2008) manipulated the Navon letters on an automatic/controlled continuum, so that these letters were either biased globally or locally.

This bias is very similar to how automatic processes are defined, because both occur automatically, and without the need of conscious activation, and in order to not process this automatic domain, one would have to inhibit one's initial response. The Navon letters are, however, not guided by only one processing mode, as they clearly encourage global/local processing (which disrupts facial processing) as well (Lewis et al., 2009; Macrae & Lewis, 2002). Both Perfect et al. (2008) and Schooler (2002) have suggested that this inhibitory quality of the Navon letters may be the reason why these letters negatively affect facial recognition, specifically that processing the dominant dimension encourages automatic processing, which is a necessary and favoured processing mode of faces. Despite being described as Stroop-like, Navon letters and the Stroop task do not cause an inhibition in the same type of processing, specifically the Stroop task is used extensively in research on automatic and controlled processes, whereas the Navon letters are extensively used in research on global and local processes. In order to investigate whether the results seen in the study conducted by Perfect et al (2008) resulted from automatic/controlled processing, or an inhibition induced through these two types of processing, the Navon letter task must be replaced with a task that solely investigates automatic/controlled processing. However, defining a task as either automatic or controlled is extremely difficult, because neither of these two types of processes are clearly conceptualised, and without such definitions, it is not possible to measure a task as automatic or controlled.

The Stroop task (Stroop, 1935) has been used extensively in cognitive and clinical psychological research to measure attention and inhibition (Lezak, Howieson, & Loring, 2004; Strauss, Sherman, & Spreen, 2006) and automaticity (Posner & Snyder, 1975). In the traditional presentation of this task, colour words, such as green, are written in different ink colours, such as red or green, and participants have to identify the ink colour while ignoring the colour word. Responses are slower when the word colour and the ink colour differ, and this is known as the Stoop effect. Initially, it was widely accepted that the Stroop task demonstrated that certain processes were either automatic (for example, reading) or controlled (for example, processing the font colour); however this notion has been re-examined, and it is now accepted that processes lie on a continuum of automatic/controlled processing. Regardless of the nature of the particular tasks used in the Stroop, the effect seen with reading and colour is very reliable (Lezak et al., 2004; see MacLeod, 1991 and Macleod, 2005 for a review), thus explaining its popularity and common usage. For this experiment, the filler task requires a certain amount of inhibition from the participant in order to override the ‘automatic’ response in favour of the ‘controlled’ response, as suggested by Perfect et al. (2008), and the reliability of the faster response when processing words than when processing colours, suggests that the Stroop task may be an appropriate test to use.

For this experiment, the Stroop task was created using the guidelines set out by MacLeod (2005), who recommended that 3-5 easily distinguishable colours should be chosen. I chose five colours: red, blue, green, pink (this is named magenta in E-Prime), and yellow, and each colour was allocated to the 8, 6, 2, 4, or 5 keys respectively in the Numpad section on the keyboard. Each key had a small, white sticker on it with the colour name written on it in black ink (the key press may have been facilitated if the letters were written in the same ink colour as the colour itself), so that participants could check which keys to press and did not have to memorise the key placement. E-Prime has built-in functionality that provides numerous options, such as font colour, font type, font size and so forth, to the researcher when building experiments; however there are only 16 font colours available to choose from. From this limited selection, I chose the 5 colours mentioned above, because they were bright, and there was little risk that they belonged to another colour category that had low hue discrimination (such as cyan, purple, maroon or olive).

The Stroop task used in this experiment was scripted in a programming language called E-Basic, which is a collection of libraries written in VBSscript. Participants had to respond to two sets of Stroop stimuli, and each set contained 100 Stroop stimuli. In the automatic condition (or Stroop word condition), participants had to read the word colour written on the screen and press the key corresponding to that word. Each word was presented in its matching font colour, for example, ‘blue’ was presented in ‘blue’ and the corresponding key was ‘8’. The decision to present each word in its matching font colour was to allow for the responses to be extremely easy and allow for facilitation. Together all five stimuli words constituted a cycle, and words were randomly selected from the cycle until all the words had been displayed, thus beginning a new cycle; there were 20 cycles in total. In the controlled condition (the Stroop font condition), participants had to press the key allocated to the font colour of the word displayed on the screen. Each stimulus word was written in a different colour to the word written, for example, ‘blue’ was never presented in its same font colour blue. Additionally, the Stroop word that followed the currently displayed word was never the same written word, nor displayed in the same font colour as that word – this was to discourage facilitation. This was done through a randomisation script, which was used by E-Prime, which sorted all the Stroop items when the experiment began. The Stroop stimuli were displayed on the screen until the participant had responded using one of the 5 keys from the Numpad; a single black fixation point ‘*’ appeared on the screen for 250ms after participants had made their choice, and then followed by a blank screen presented for 250ms. The Stroop condition was complete once the participant had responded to all 200 items.

Procedure

Participants were run in groups of up to 6 members, and each participant was allocated to a computer. The experiment only began once all participants had signed the consent forms; these forms provided information regarding each participant’s gender and ethnicity so that they could be assigned to their respective experimental groups. The researcher would open the experiment file for the participant, type in their unique participant code, and then begin the experiment. All participants were instructed to read the experiment instructions carefully, and to keep their hands away from the keyboard if only reading, thus preventing them from accidentally pressing a key. Each participant was given a pair of earphones, because there was some sound in the video.

The entire experiment was built in E-Prime 1.1., psychological software used for building, running and analysing of psychological experiments (Psychology Software Tools, Pittsburgh, PA). All text was displayed in font type Arial with a font size of 18, and screen resolutions were set to 1024 x 768 pixels. The font colour was black (except for the Stroop conditions), and the background colour was always silver. This background colour was chosen, because it was important that the colours of the Stroop task were easy to distinguish from the background, and it could be difficult to read certain colours (for example, yellow) against a white background. The video files were run through E-Prime, using a script that had a command line that opened Classic Media Player: media playing software provided with Windows. Spacebar was the allocated key for when participants wanted to move to the next slide/screen, and the screen display would not change unless participants pressed spacebar.

During the first stage of the experiment, participants were instructed to view the 10-second encoding video, and pay attention. Afterwards, the fill task, which was the Stroop task, began. The interval task consisted of three stages: a familiarisation stage, and then two Stroop tasks whose order varied depending on which processing type group (automatic, controlled) the participant was in. During the familiarisation stage, the participant was instructed to look at the keys and their respective stickers on the Numpad of the keyboard. These instructions were then followed by an image of the Numpad, as not all the participants were familiar with computers. They were instructed to press the colours in the following order: blue, red, green, pink and yellow, and followed by spacebar to view the next slide. After they had familiarised themselves with the keys, they were guided through a practice trial Stroop task where they were presented with 15 trials of an array of four plus-signs presented in a randomly selected colour from the 5 designated font colours (The instructions were as follows *"Good job! Now we're going to practise a bit more. Press the correct key (8, 6, 2, 4 or 5) that corresponds with the colour of the symbols that will be displayed on the screen. The displays will not change until you press the correct key. Ready ? <Press spacebar to continue>"*). The plus-signs would remain on the screen until the participant had pressed the colour that they were written in. Once a key was pressed, the plus-symbols were removed, immediately followed by a single black '+' and a blank screen (each presented for 250ms), and then the next Stroop Practise item. Once the participant had responded to all 15 items, the Stroop conditions would begin. The order in which the Stroop conditions were presented to the participants depended on which processing type group they were in: if participants were in the automatic group, the Stroop-font condition was presented first, followed by the Stroop-word condition, but if they were in the controlled group, then the Stroop-word condition was presented first, which was followed by the Stroop-font condition. Once each participant had completed both Stroop condition, the recognition stage would commence. During the recognition stage, participants were given the following instructions, "*You will now view a line-up. Please indicate which target you saw previously in the video. This target may or may not be present in the line-up. Make your choice by pushing the number key that matches the number under the face of the target. You may also press "0" if you do not think that the target is present. <Please press SPACE BAR to continue>*", and then were presented with a line-up image that remained on the screen until participants made a decision by pressing a key (which assigned to one of the faces in the line-up, or the not-present option). After this, a forced-choice decision was required from the participants where they were instructed to assume that the perpetrator was present in the line-up, and had to choose who they thought he might be. If they had already made a decision in the initial presentation of the line-up, they were asked to make the same decision (the instructions for this condition read as follows, "*You will now view the same line-up again. If you did choose a target from this line-up in the slide before this, please choose the same target. If you did not choose a target, please assume that the target is present in this line-up, and make your choice. Make your choice by pushing the number key that matches the number under the face of the target. <Please press SPACE BAR to continue>.*").

After participants had completed the forced–choice condition, they were presented with a few metacognitive questions that investigated how they had perceived the line-up, and how they had made their decisions. For example, these questions asked about their confidence in their choice, how similar they perceived the line-up members to be, and how they made their decisions (more information about these questions will be provided). Examining the metacognitive aspects of eyewitness memory sheds light on other factors that may have attributed to the participants' decision making (Sporer, 1993). Confidence has been investigated quite extensively (see the meta-analysis for a review: Sporer, Penrod, Read, & Cutler, 1995), because eyewitness testimony may lead to a positive or negative identification of the suspected perpetrator, which would ultimately shape the investigation (for example, would a new perpetrator be sought, or would the case move forward with the positively identified suspect as the suspected perpetrator); and much weight is assigned to the amount of confidence that eyewitness have in their decisions (Wells & Lindsey, 1983, as cited in Sporer, 1993). Many studies have found that confidence is not strongly correlated with accuracy; however, when choosers (participants who made a choice) and non-choosers (participants who did not make a choice) were separated, a meta-analysis revealed the confidence-accuracy (CA) correlation to be .29, and when choosers were divided into those who correctly chose the perpetrators, and those who chose foils, the CA correlations were .41 in an unweighted analysis (or .37 in a weighted analysis) and .12 respectively (Sporer et. al., 1995). This data reveals that confidence is a better predictor of accuracy among choosers than previously thought. There are other variables that affect confidence, for example instruction type (biased or unbiased, Brewer & Wells, 2006) and perpetrator distinctiveness (Brigham, 1990).

In order to measure participants' confidence in their responses, participants were asked to rate their confidence about the decision regarding their line-up choice on a Likert scale ranging from 0% to 100%, with anchors for 0% and 100%. The question read as follows: "*"I'm _________ % confident that I chose the correct person from the Line-up". Please type in either (without the % sign): 0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%*

0% means that you are not at all confident. 100% means that you are extremely confident. Press <ENTER> when you are happy with your decision."

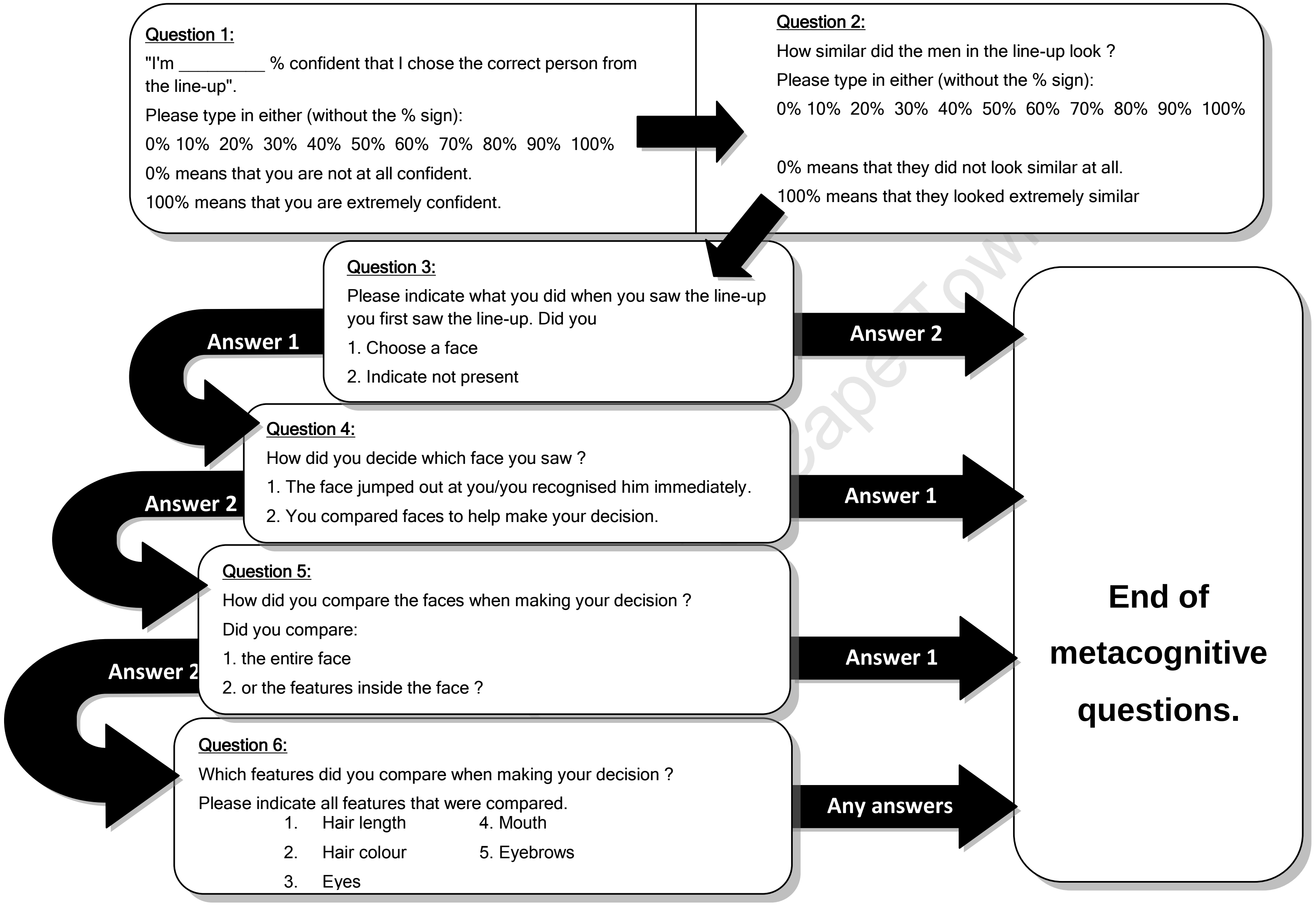
After this, participants were required to answer how similar they thought the men in the line-up looked. This question had a similar format to the above question, with participants rating similarity on a scale from 0% to 100%, with anchors for 0% and 100%. Although the effective size and bias had been previously measured for the line-ups, it was important to know how similar participants perceived the line-up members to be, as (1) it was possible that white participants would perceive the line-ups members as more dissimilar than the other participants (who had indicated that they were not white), and (2) that if participants perceived line-ups members as more dissimilar, then it may be easier to choose the perpetrator.

The next set of questions examined how participants decided who the perpetrator was. Participants were guided through these questions based on the answers that they gave; depending on their answers, participants either proceeded to the next question, or they skipped the remaining questions and the metacognitive investigation ended here (see Figure 2.1.) This set of questions began by asking whether participants had initially chosen a face from the line-up, or indicated not-present; if they indicated not-present, then there was no need to ask them the remaining questions as these specifically dealt with the strategies that lead to their perpetrator selection. Instead, they only had to answer one question, which was when they rejected the line-up, was it because they did not recognise the perpetrator, or because they were unsure if he was present. If the participants had chosen a face, they had to answer the following questions:

1. How did they decide to choose the face: did the face jump out at them (i.e. the pop out effect) (Neissner, 1964; Treisman & Gelade, 1980), or had they compared faces. If the face jumped out at them, this could indicate that there was a distinct quality about this face (or artefact about the image), or some quality about the line-up (such as low effective size) which aided recognition.
2. If participants had answered that they had compared faces, they had to indicate how they had compared them. Specifically, participants had to answer if they had relied on a holistic comparison strategy, or a featural comparison strategy. It was expected that white participants may rely on a holistic comparison strategy, whereas the remaining participants who rely on a featural processing strategy due to the other race effect/own race bias. Additionally, I wanted to investigate whether either of the Stroop conditions would encourage a particular type of processing strategy used by the participants.

3. If participants had indicated that they compared the faces featurally, they were then asked to list which features they had compared.

This procedure was repeated three times, and after the last trial participants were thanked, and debriefed.



Question 1:

"I'm __________ % confident that I chose the correct person from the line-up".

Please type in either (without the % sign):

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

0% means that you are not at all confident.

100% means that you are extremely confident.

Question 2:

How similar did the men in the line-up look ?

Please type in either (without the % sign):

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

0% means that they did not look similar at all.

100% means that they looked extremely similar

Question 3:

Please indicate what you did when you saw the line-up you first saw the line-up. Did you

1. Choose a face
2. Indicate not present

Question 4:

How did you decide which face you saw ?

1. The face jumped out at you/you recognised him immediately.
2. You compared faces to help make your decision.

Question 5:

How did you compare the faces when making your decision ?

Did you compare:

1. the entire face
2. or the features inside the face ?

Question 6:

Which features did you compare when making your decision ?

Please indicate all features that were compared.

1. Hair length
2. Hair colour
3. Eyes
4. Mouth
5. Eyebrows

Figure 2.1. A Diagram that demonstrates how the participants navigated through the metacognitive questions based on their answers. All participants had to answer Questions 1 and 2, but from Question 3 onwards, their answers would dictate whether they proceeded to the next question or not.

Chapter Three: Data Analysis

Stroop Times and Stroop Accuracy

The experimental manipulation for this experiment was a Stroop task with two different conditions: automatic and controlled. Participants were expected to perform the automatic condition faster, because it would be easier to process the written word, and their performance should be slower in the controlled condition where they are required to process the ink colour. All participants went through both conditions of the Stroop, but in different orders. Normally a Stroop task would be administered one word at a time to participants, and research has shown that a Stroop effect is not always induced through a computer-version of this task. A paired-sample t-test was run to investigate where this experimental condition was successful – specifically, whether participants responded faster to the written word (the automatic condition) than they did to the ink colour (the controlled condition).

On average, participants performed the automatic condition of the Stroop task significantly faster (M = 861.31 ms, *SD* = 184.14) than they did the controlled condition of the Stroop task (M = 1180.97 ms, *SD = 289.30*), *t*(287) = -28.334, *p* < 0.001, *r* = .86 (see Table 3.1.). This demonstrates that the manipulation condition was achieved.

Table 3.1.

Descriptive Data: Average Stroop Accuracy and Reaction times for both groups of participants

	Stroop Condition				
	Automatic	Controlled	df	t	p
Stroop Accuracy	97.66	88.89	287	7.302	< 0.001
	(3.19)	(20.50)			
Stroop Reaction Time	861.31	1180.97	287	-28.334	< 0.001
	(184.14)	(289.30)			

Note: Standard deviations are reported in parentheses. Stroop accuracy is reported as the average percentage of correct Stroop items. Stroop reaction time is reported in ms.

Additionally, participants performed significantly better in the automatic condition of the Stroop task (M = 97.66, SD = 3.19) than they did in the controlled condition (M = 88.89, SD = 20.50), t(287) = 7.302, $p < 0.001$, r = .40. However, when looking at the average reaction times, the results suggest that participants did not sacrifice speed for accuracy in the controlled condition, as they were performing poorer at a slower rate.

Total Recognition Accuracies of All Three Perpetrators: Initial Choice and Forced Choice

Accuracy for each Perpetrator was measured as a binary variable: 0 indicated an incorrect answer and 1 indicated a correct answer. A total accuracy score was calculated for each participant by summing the accuracy scores across the three Perpetrators, thus giving a total between 0 and 3. From this, a mean accuracy score could be calculated for both the initial choice, where participants were asked to choose a Perpetrator while the ‘Not present’ option was available, and for the forced choice, where participants were forced to choose a face, and could not indicate ‘not present’, and despite not being a true continuous variable, this data could be analysed through an ANOVA. Therefore the dependent variable was the total accuracy score calculated for each participant, the independent variables were ethnicity and processing type of the Stroop task. The descriptive data is presented in the Table 3.2.

Table 3.2.

Descriptive Data: Recognition accuracy scores for the Initial Presentation of the line-ups

Stroop Condition	Participant Ethnicity			
	White		Other	
	Initial Choice	Forced Choice	Initial Choice	Forced Choice
Automatic	1.96	2.38	1.79	2.13
	(.863)	(.721)	(.768)	(.821)
Controlled	2.08	2.25	1.89	2.24
	(.801)	(.783)	(.848)	(.778)
Average	2.02	2.31	1.84	2.18
	(.832)	(.752)	(.808)	(.799)

Note: Standard deviations are reported in parentheses.

Initial presentation of the line-up.

The descriptive data suggests that white participants performed better than other participants following the first presentation of the line-ups (M = 2.02 vs. M = 1.84). White participants were better than other participants after both the automatic and controlled condition of the Stroop, and both groups of participants performed better after the controlled condition. This result was not predicted by my research hypothesis, as I had originally hypothesised that other participants would perform better after the controlled condition, whereas white participants would perform better after the automatic condition.

The data does suggest that there may be an ORB effect occurring (see Figure 3.1), because white participants were recognising the white perpetrators better than other participants were. However an ANOVA found these results to be non-significant, F(1,284) = 3.484, p = 0.063, $\omega^2 = .0086$. Without another group of perpetrators and homogenous participant group that consisted of members that are, for example, black, it can be only speculated that the results *may* suggest an ORB.

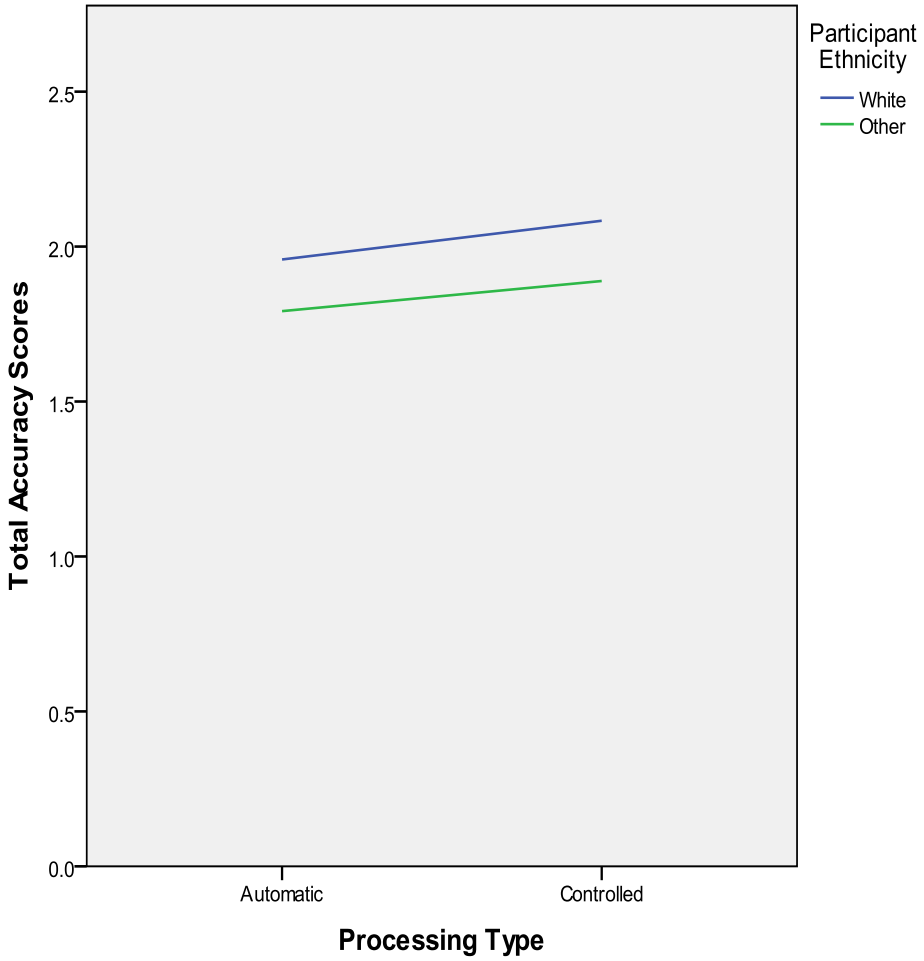


Figure 3.1: The combined recognition accuracies across processing groups for white and other participants for the initial presentation of the line-ups. The blue line represents white participants, and the green line represents other participants. The white participant group performed better than the other participant group.

The differences in performance following the controlled and automatic conditions were also not significant, as there was no significant main effect of processing type, F(1,284) = 1.319, $p = .252$, $\omega^2 = .0011$. There was no significant interaction effect between ethnicity and processing type, F(1,284) = .021, $p = .886$, $\omega^2 = -.0034$.

Forced choice.

Participants' recognition scores improved with the forced choice condition, F(1,284) = 95.153, $p < 0.0001$. White participants still performed better than other participants did (M=2.31 vs. M = 2.18), however the trend towards significance (as seen in the accuracy scores between white and other participants after the initial presentation of the line-up) disappeared as there was no significant difference between these two groups, F(1,284) = 2.079, $p = .150$, $\omega^2 = .0037$. Participants were instructed to choose the same face that they had chosen in the initial choice, if they had chosen one; otherwise, if they had indicated not present, they had to choose a face. Therefore participants could not perform worse in the forced choice condition than they had in the initial choice condition, because choosing not present or the wrong face were both marked as incorrect, and awarded 0 points. Therefore, the other participants' recognition accuracy must have improved in the forced choice, in order to lessen the gap between the two groups, suggesting that they may have had a stricter criterion in the initial choice.

White participants performed a bit better after the automatic condition than the controlled condition (M = 2.38 vs. M=2.25), and other participants performed better after the controlled condition (M= 2.24 vs M=2.1.3), which is what our hypothesis had predicted. An ANOVA yielded a non-significant main effect for processing type, F(1,284) = .006, $p = .940$, $\omega^2 = -.003$ and despite the pattern seen in the descriptive data, there was no significant interaction between ethnicity of participant and processing type, F(1,284) = 1.664, p = .198, ω^2 = .0023.

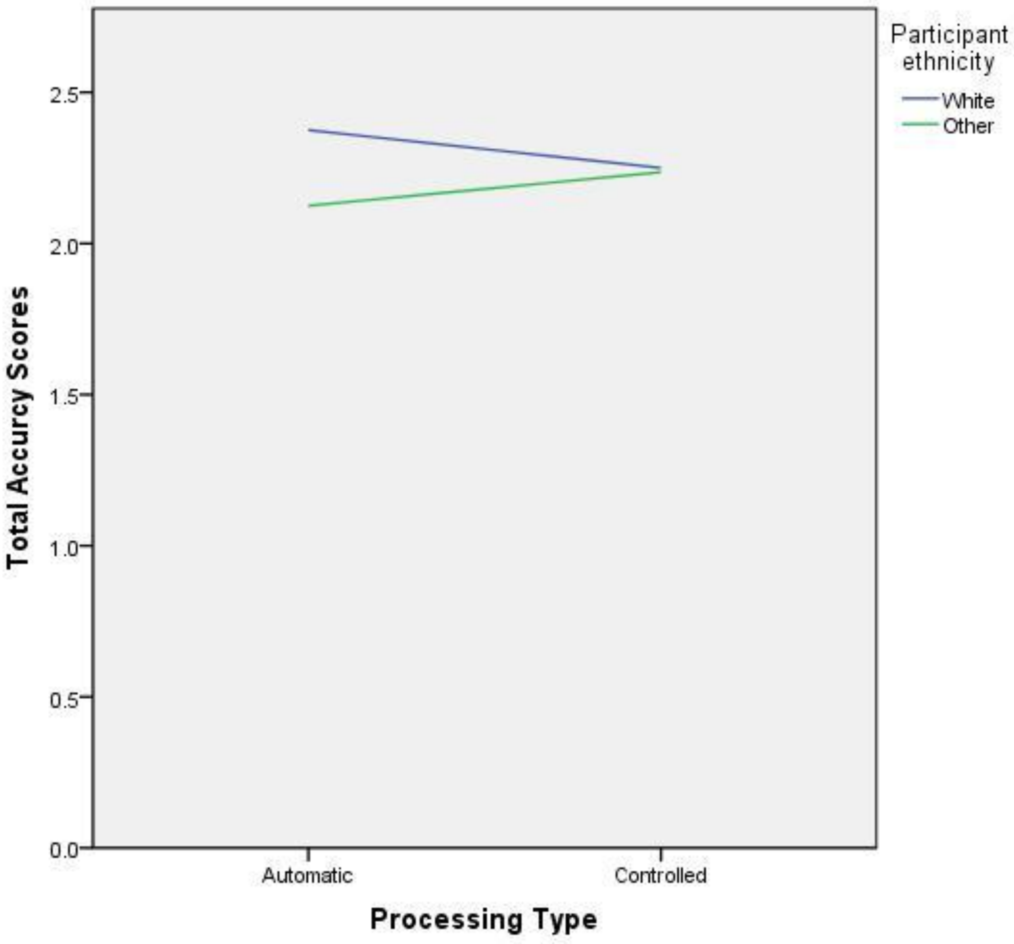


Figure 3.2: The combined recognition accuracies across processing groups for white and other participants for the forced-choice presentation of the line-ups. Participants are forced to choose who they suspect is the perpetrator, and can no longer indicate not-present. The difference between the accuracies scores of white participants and other participants has lessened to such an extent that it is no longer apporiximating significance (p = .150).

Reaction Times

Perfect et al. (2008) found significant results among the reaction times of their participants, specifically that participants in the congruent processing condition were faster at recognising the perpetrator than the participants in the incongruent processing condition. In my experiment, reaction times were only collected for the initial presentation of the line-up, because in the forced-choice presentation participants were instructed to assume that the Perpetrator was present, and could not indicate that he was not present; if they had chosen a face in the initial presentation of the line-up, they were instructed to choose that same face in the forced choice. Therefore the reaction time in the forced choice condition is not a reflection of the difficulty of the task, nor the time taken to make a decision.

In my experiment, the descriptive data (please see Table 3.3.) shows that participants responded slightly slower after the automatic condition (M = 16287.94, SD = 9397.18) than the controlled condition (M = 15632.85, SD = 8761.61, and that white participants responded slightly faster (M = 15296.84, SD = 8818.42) than other participants (16623.95, SD = 9307.94). Both differences were small, and an ANOVA showed that there were no significant main effects for processing type, F (1, 284) = .376, p = .540, ω^2 = -.0022 or participant ethnicity, F (1, 284) = 1.541, p = .215, ω^2 = -.0035. White participants appeared to respond faster after the automatic processing task than they did after the controlled processing task, and the reverse pattern was seen with other participants who responded quicker after the controlled processing task than after the automatic processing task. However, the interaction between ethnicity, reaction time and processing type condition was not significant, F(1, 284) = 1.373, p = .242, ω^2 = -.0035 (Table 3.3).

Table 3.3 .

Descriptive data: Average reaction times in ms across processing-type conditions for white and other participants.

	Automatic Condition	Controlled Condition	Average
White	14998.08	15595.60	15296.84
	(7734.11)	(9829.89)	(8818.42)
Other	17577.79	15670.11	16623.95
	(10708.57)	(7614.68)	(9307.94)
Average	16287.94	15632.85	
	(9397.18)	(8761.61)	

Note: Standard deviations are reported in parentheses.

Previous research has shown that decisions made under a time pressure were on average more accurate than those that were not. Dunning and Perretta (2002) suggested that all decisions made between 10 and 12 seconds were better precursors to accuracy, and Schooler & Engstler-Schooler (1990) demonstrated that the negative effect of verbalisation was lifted if the decision was made under 5 seconds. I partitioned my participants' reaction times into two groups: the first group contained all decisions made between 0 and 12 seconds (12000ms), and the second group contained all decisions made afterwards (please see Table 3.4).

Table 3.4.

Descriptive data for accuracy scores before and after 12 seconds.

	n	Average accuracy
Decisions made in less than 12 seconds	115 (39.9%)	2.28
		(.756)
Decisions made after 12 seconds	173 (60.1%)	1.70
		(.786)

Note: Standard deviations are reported in parentheses in the average accuracy column.

An ANOVA was run on the average accuracy scores with ethnicity, processing type and the banded reaction times as between subject factors. There was a significant difference between the accuracies of decisions made in fewer than 12 seconds, and decisions made after 12 seconds, $F(1,280) = 36.055$, $p < 0.001$, with the more accurate decisions occurring before 12 seconds than afterwards ($M = 2.28$ vs. $M = 1.70$). Even though more participants took longer to make decisions, those who had made decisions in less than 12 seconds were accurate more often.

Accuracy per perpetrator after the initial presentation

There were three perpetrators, and each participant was required to try to recognise all three. Although this data has been grouped together as a Total Accuracy score for each participant, I decided to investigate the accuracy scores for each perpetrator, because their effective line-up size differed among them, and could affect recognition, which was suggested in the pilot experiment that explored which of the two videos I should use.

When presented with each line-up initially, participants could choose a face (incorrect or correct), or they could indicate that the perpetrator was not present. Participants' responses could be grouped as one of the following: an incorrect rejection of the line-up when they indicated that the perpetrator was not present; a miss when they chose a foil; and a hit when they correctly chose the perpetrator. The responses given for each perpetrator were coded according to which of the three groups of answers they belonged to, and are presented in Table 3.5. Perpetrator 3 was recognised by most participants (86.1%), followed by Perpetrator 2 (57.3%) and then Perpetrator 1 (50.0%); these results suggest that there were ceiling effects for the recognition rates for Perpetrator 3. A deeper investigation of the errors made by participants revealed that 101 participants (70% of those that gave an incorrect answer) incorrectly rejected the line-up that contained Perpetrator 1, whereas 75 participants (61% of those that gave an incorrect answer) incorrectly chose a foil, and ‘missed’ Perpetrator 2. When participants were forced to choose a perpetrator from each line-up, and could no longer indicate not present, similar recognition rates and misses were obtained for Perpetrator 1 and Perpetrator 2. For Perpetrator 1, 193 participants correctly identified him, while 95 participants chose the incorrect person; compared to Perpetrator 2, 194 participants correctly identified him, and 94 participants made incorrect decisions. Therefore it appears as though participants had adopted a stricter response criterion when initially viewing the line-up containing Perpetrator 1, than when viewing the line-ups for Perpetrator 2 and Perpetrator 3. The recognition rates for Perpetrator 3 seemed to have ceiling effects, as 248 participants (86.1%) correctly identified him in the initial choice and this improved to 90.3 % in the forced choice condition.

Table 3.5.

Descriptive data: Proportion of participants in each group that chose the correct perpetrator

	Initial choice			Forced choice		
	Perpetrator 1	Perpetrator 2	Perpetrator 3	Perpetrator 1	Perpetrator 2	Perpetrator 3
Line-up rejection	101	48	26	*	*	*
	(35.1%)	(16.7%)	(9.0%)			
Miss	43	75	14	95	94	28
	(14.9%)	(26.0%)	(4.9%)	(33.0%)	(32.6%)	(9.7%)
Hit	144	165	248	193	194	260
	(50.0%)	(57.3%)	(86.1%)	(67.0%)	(67.4%)	(90.3%)
Total	288	288	288	288	288	288
	(100%)	(100%)	(100%)	(100%)	(100%)	(100%)

Note: Percentages are reported in parentheses.

It is interesting that the recognition accuracy for Perpetrator 3 seemed to have ceiling effects, because he did not have the highest bias, nor the lowest effective size (see Table 3.6.). The post-hoc study that explored the perceptual dimensions of each perpetrator (and will be discussed later) showed that white participants perceived Perpetrator 2 as the most distinctive, and the only between groups differences was that whit participants rated the picture quality of Perpetrator 2's image higher. However, despite this, there is some quality about Perpetrator 3 that makes him easier to remember. I suspect that this may be the medium of the encoding stimuli: the video. The studies that investigated the effective size and bias of each line-up utilised a copy of the line-up that was printed on a piece of paper, and the post-hoc facial rating study utilised the same image of the perpetrator that was used in the line-up. The recognition accuracy of Perpetrator 3 dropped to 30% when the close-up video was used, but was 80% for the far video. This leads me to suspect that the additional visual information in the far video that is available to participants encourages deeper encoding of Perpetrator 3, and may provide some cue that distinguishes him from the foils in the line-up.

Table 3.6.

The effective size, bias, similarity and average recognition score calculated for each perpetrator.

	Initial choice		
	Perpetrator 1	Perpetrator 2	Perpetrator 3
Effective size	5.78	2.72	4.82
Bias	17.24%	15.18%	10.34%
Similarity	41.15	54.73	39.29
	(24.65)	(22.19)	(25.52)
Average recognition	.50	.57	.86

Note: Standard deviations are reported in parentheses.

The proportion of participants who had correctly indentified each perpetrator was calculated, and is reported in Table 3.7. It appears as though all three perpetrators were better recognised by participants after the controlled condition of the Stroop, and Perpetrator 3 was accurately recognised most often.

Table 3.7.

Descriptive data: Proportion of participants in each group that chose the correct perpetrator

	Perpetrator 1		Perpetrator 2		Perpetrator 3	
	Automatic Condition	Controlled Condition	Automatic Condition	Controlled Condition	Automatic Condition	Controlled Condition
White participants	.56 (.500)	.56 (.500)	.51 (.503)	.65 (.479)	.89 (.316)	.88 (.333)
Other participants	.38 (.488)	.50 (.504)	.58 (.496)	.53 (.503)	.83 (.375)	.86 (.348)
Average	.47 (.501)	.53 (.501)	.55 (.499)	.59 (.493)	.86 (.347)	.87 (.340)

Note: Standard deviations are reported in parentheses.

The recognition rates for the three perpetrators were analysed with a repeated-measures ANOVA, and the between subject factors were ethnicity and processing condition. While ANOVA is not normally used to analyse binary data, as this violates ANOVA's assumption of normality, other researchers such as Sporer (Reinhard & Sporer, 2010) often do analyse such data with this statistical test, and cite Harris (1994, p428 – 432) as justification. Mauchley's test of sphericity indicated that the assumption of sphericity had been violated for PerpID, χ^2 (2) = 18.07, $p < 0.001$, and therefore the corrected degrees of freedom are used as calculated by Greenhouse-Geisser correction ($\varepsilon = .942$). The recognition accuracy for each of the three perpetrators were significantly different from one another, F(1.88, 534.92) = 56.488, $p < 0.001$, but there was no significant main effect for ethnicity of participant, F(1.88, 534.92) = .938, p = 387, nor processing type, F(1.88, 534.92) = .293, p = .733. Although not significant, the interaction between participant ethnicity, perpetrator and processing type did tend towards significant, F(1.884,534.92) = 2.552, p = .082.

Forced choice: Accuracy per Perpetrator

After initially being presented with each line-up and asked to choose a perpetrator or indicate not present, participants were presented each line-up again in a Forced-choice paradigm. This time they had to make a choice from the line-up, and were not allowed to indicate not present. This forced-choice data was analysed with a repeated-measures ANOVA.

The assumption of sphericity was violated, χ^2 (2) = .947, p < .001, therefore the corrected value for degrees of freedom was used as calculated by the Greenhouse-Geisser correction (ε = .949). The three perpetrators were recognised at significantly different rates from one another, F(1.90, 539.19) = 31.50, p < .001. This confirms the trend seen in the descriptive data: Perpetrator 3 is best recognised, followed by Perpetrator 2 and then Perpetrator 1. There was no significant main effect for ethnicity, F(1.90, 539.19) .691, p =.494, nor processing type, F(1.90, 539.19) = .520, p = .585, but the interaction between participant ethnicity and processing type on recognition accuracy was significant, F (1.90, 539.19) = 5.349, p = .006.

Table 3.8.

Descriptive data: proportion of participants in each group that chose the correct Perpetrator in the forced choice condition.

	Perpetrator 1		Perpetrator 2		Perpetrator 3	
	Automatic Condition	Controlled Condition	Automatic Condition	Controlled Condition	Automatic Condition	Controlled Condition
White participants	.78 (.419)	.64 (.484)	.67 (.475)	.74 (.444)	.93 (.256)	.88 (.333)
Other participants	.54 (.451)	.72 (.451)	.67 (.475)	.62 (.488)	.92 (.278)	.89 (.316)
Average	.66 (.475)	.68 (.468)	.67 (.473)	.68 (.468)	.92 (.267)	.88 (.324)

Note: Standard deviations are reported in parentheses.

Analyses rerun without Perpetrator 3 data

Our descriptive data of the recognition rates of Perpetrator 3 suggested that there were ceiling effects for Perpetrator 3. All the data relating to Perpetrator 3 was removed, and the analyses of the Total Accuracy scores and Reaction times were rerun with only the data from Perpetrator 1 and Perpetrator 2.

Total Recognition Accuracy for Perpetrator 1 and Perpetrator 2

Since accuracy data for Perpetrator 3 was removed, a new total accuracy score was calculated for each participants, and ranged from 0 – 2. An ANOVA was rerun to investigate whether there any differences between the recognition scores of participants' decisions when initially presented with the line-ups for Perpetrator 1 and Perpetrator 2 (see Table 3.9). On average, white participants appeared to perform better on the recognition scores than the other participants did (M = 1.14 vs. M = .99), and both groups performed better after the controlled condition. This difference in recognition performance between white and other participants was not significant, F (1,284) = 1.531, p = .83, ω^2 = .007 nor was the difference between processing groups, F(1,284) = 1.55, p =.214, ω^2 = .002. The interaction between processing type and participants ethnicity was not significant, F(1,284) = .172, p =.679, ω^2 = . -.003.

Table 3.9.

Descriptive data: recognition accuracy scores for the Initial Presentation of the Line-up for only Perpetrator 1 and Perpetrator 2.

Stroop Condition	Ethnicity			
	White		Other	
	Initial Choice	Forced Choice	Initial Choice	Forced Choice
Automatic	1.07	1.44	.96	1.21
	(.757)	(.669)	(.659)	(.711)
Controlled	1.21	1.38	1.03	1.35
	(.711)	(.659)	(.712)	(.715)
Average	1.14	1.41	.99	1.28
	(.735)	(.663)	(.684)	(.714)

Note: Standard deviations are reported in parentheses.

It was possible that line-up accuracy was affected by its respective line-up size and bias, as a small line-up effective size and high bias would make of correctly identify the perpetrator easier. In order to remove the effect that the line-up effective size and bias had on the accuracy scores of its matching line-up, a simple regression analysis was run. The aim was to partial out the effect of each line-up's effective size and bias, and calculate the residuals. Therefore the predictors were effective size and bias, and the DV was Perpetrator accuracy; however, the regression model was not significant, F(1,574) = 3.079, p = 0.08, and line-up effective size only accounted for 0.5% of variability in recognition accuracy scores (R = .078); line-up bias was an excluded variable. Therefore, due to the lack of a significant effect of line-up effective size on target accuracy, this analysis was abandoned.

There were no significant differences between the recognition rates in the Forced choice condition, nor were there any interaction effects.

Reaction times

Since Perpetrator 3 had ceiling effects for his recognition, it is possible that the time that it took for participants to choose him could be much shorter as it was easier to do so. With Perpetrator 3's reaction times removed, an ANOVA was rerun on the reaction times.

These results differed from those presented previously. Without the data for Perpetrator 3, the average recognition time slowed considerably. White participants made a decision faster than other participants (M = 34637.05, SD = 20004.36 vs. M = 37264.63, SD = 24729.30), and decisions were made faster following the controlled conditions, than the automatic conditions (Table 3.10); however, neither of these main effects were significant. Unlike what our hypothesis predicted, white participants performed slower after the automatic condition than the controlled condition, and this pattern was reversed for other participants. The interaction between ethnicity and processing approximated significance,

$F(1,284) = 3.321$, $p = .069$, $\omega^2 < .001$, suggesting that white participants performed faster following the controlled condition, and other participants were faster after the automatic conditions (please see Figure 3.3.).

Table 3.10.

Reaction times in ms across processing-type conditions for white and other participants.

	Automatic Condition	Controlled Condition	Average
White	37265.61	32008.49	34637.04
	(20147.01)	(19648.67)	(20004.37)
Other	35074.07	39455.19	37264.63
	(19993.82)	(28674.11)	(24729.30)
Average	36169.84	35731.84	
	(20030.47)	(24776.46)	

Note: Standard deviations are reported in parentheses.

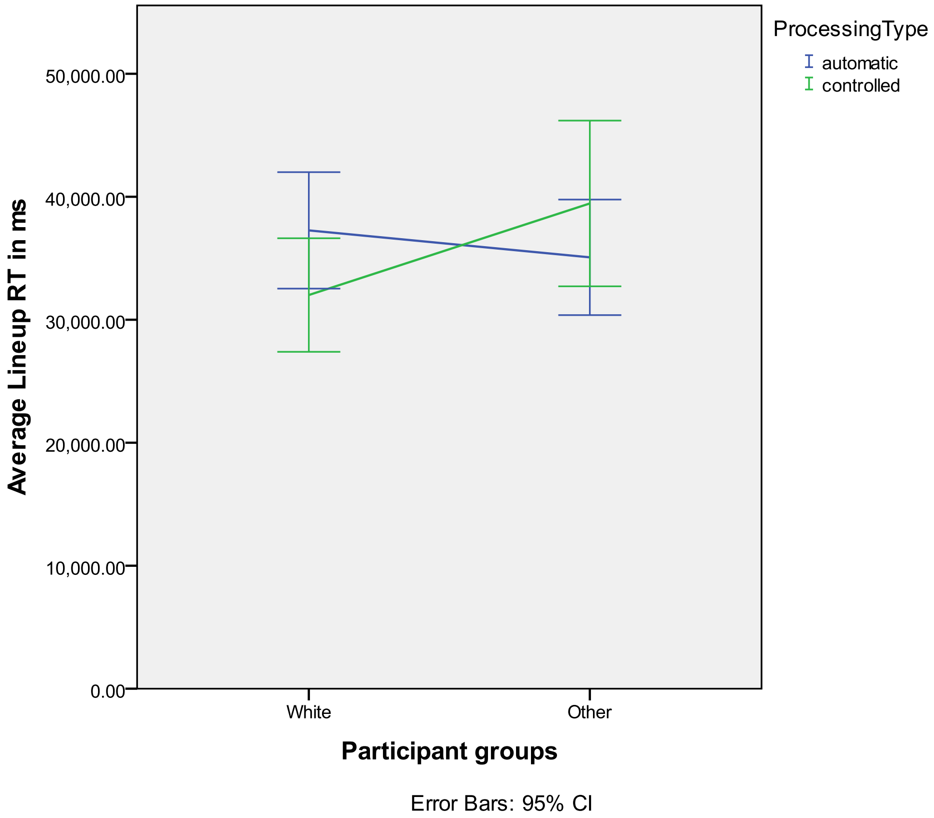


Figure 3.3. The average reaction times of both participant groups after automatic and controlled processing conditions. It appears that there is a weak interaction between participant ethnicity and processing type for white and other participants.

On average, identification of Perpetrator 1 and Perpetrator 2 was made quicker after the controlled condition of the Stroop task; however identification of Perpetrator 3 was quicker after the automatic condition of the Stroop task (see Table 3.11.). These results were not significantly different, F(1.853, 526.26), p =.685.

Table 3.11.

Descriptive data: Reaction times in ms for each Perpetrator

	Perpetrator 1	Perpetrator 2

					Perpetrator 3	
	Automatic Condition	Controlled Condition	Automatic Condition	Controlled Condition	Automatic Condition	Controlled Condition
White participants	14636.11 (8066.734)	16966.04 (11978.158)	19871.17 (15904.349)	18173.54 (12387.605)	10486.97 (8258.561)	11647.21 (12995.827)
Other participants	19449.83 (17518.510)	14501.43 (8469.380)	20449.90 (22709.807)	20128.10 (16558.721)	12833.64 (8440.788)	12380.79 (13020.42)
Average	17042.97 (13802.83)	15733.74 (10410.56)	20160.53 (19538.13)	19150.82 (14606.40)	11660.31 (8440.79)	12014.00 (12967.49)

Note: Standard deviations are reported in parentheses.

The descriptive data also suggests that white participants made quicker decisions when having to identify Perpetrator 1 and Perpetrator 3 after the automatic condition, while other participants were quicker making decisions for all three perpetrators after the controlled condition of the Stroop. An ANOVA found these differences to be non-significant, F(1.853, 526.26), p = 979. Additionally, there was no significant interaction effect among ethnicity, processing type and perpetrator, F(1.853, 526.26), p = .118.

On average, participants reacted slowest to the line-up for Perpetrator 2 (M= 19655.68, SD= 18225.928), followed by Perpetrator 1 (M=1638.35, SD= 12221.226), and fastest to Perpetrator 3 (M=11837.15, SD = 10923.248). These difference were significant, F(1.853, 526.26) = 27.86, p = 0.0001.

Mauchly's Test demonstrated that the assumption of sphericity was violated, χ^2 (2) = .921, p = 0.0001, and therefore the correct degrees of freedom as indicated by the Greenhouse-Geisser correction were used, ε = .927. The reaction time of the three perpetrators differed significantly from each other, F (1.85, 526.26) = 27.86, p = 0.0001, but these differences occurred without the influence of participant ethnicity, F (1.85, 526.26) = .016, p = .979, and without any effect of processing type, F (1.85, 526.26) = .355, p = .685.

Recognition rates of perpetrators were unaffected by the insignificant interaction between participants ethnicity and processing type, F (1.85, 526.26) = 2.181, p = .118.

When the reaction times for Perpetrator 3 were removed, and an ANOVA was rerun, the results did not differ much. The reaction times were significantly different between Perpetrator 1 and Perpetrator 2, F(1,284) = 7.989, p = .005; however, the interaction between reaction time, ethnicity and processing type only tended towards significant, F(1,284) = 3.503, p = .062. None of the other interactions were significant.

Metacognitive Data

Three of the metacognitive questions that participants were asked were analysed; these questions were: (1) their confidence about their line-up choice; (2) how they decided which face to choose; and (3) which comparison strategy, featural or holistic, they utilised.

Confidence rating.

Each participant had to rate how confident they were about their line-up choice; these three confidence scores were averaged for each participant and analysed with an ANOVA. The mean confidence reported by white participants (M = 73.70, SD = 16.36) was not significantly different from that reported by other participants (M = 73.65, SD = 17.20), F (1, 284) = .002, p = .962, $\omega^2 = -.003$ (see Table 3.12).

Table 3.12.

Descriptive data: average confidence reported by participants.

	White participants		Other participants	
	Automatic Condition	Controlled Condition	Automatic Condition	Controlled Condition
Confidence	73.45	73.95	75.64	71.56
	(16.77)	(16.05)	(16.21)	(19.63)

Average	73.70	73.65
	(16.36)	(17.20)

Note: Standard deviations are reported in parentheses.

There was no significant difference between average confidence ratings between processing type groups, F (1, 284) = .779, p = .962, ω^2 = -.001 nor was there an interaction effect between ethnicity and processing type, F (1, 284) = 1.269, p = .261, ω^2 = .001.

The confidence levels for each perpetrator were analysed with a repeated-measures ANOVA. Mauchly's test showed that the assumption of sphericity was violated, X2 (2) = .960, p < 0.05, therefore the corrected degrees of freedom calculated with Greenhouse-Geisser are reported. Between perpetrators, it seemed that participants were more confident about the recognition decisions for Perpetrator 3 (M= 85.88, SD = 22.03) than for Perpetrator 2 (M = 69.75, SD = 25.92) or Perpetrator 1 (M= 65.33, SD = 29.62).

These confidence levels were significantly different from each other, F(1.924,546.401) = 59.115, p < .0001; however, there were no significant main effects of ethnicity of participants, F(1.924, 546.401) = 1.491, p = .226, processing type, F(1.924, 546.401) = .696, p = .494, nor was there a significant interaction between perpetrator, ethnicity of participants and processing type, although it did tend towards significance, F(1.924, 546.401) = 2.685, 0 = .071 (see Table 3.13 for descriptive data).

Table 3.13.

Descriptive data: confidence levels for each perpetrator across both ethnic groups and processing-type groups.

	Perpetrator 1		Perpetrator 2		Perpetrator 3	
	Automatic	Controlled	Automatic	Controlled	Automatic	Controlled

	Condition	Condition	Condition	Condition	Condition	Condition
white	66.11 (29.58)	66.32 (26.26)	65.21 (25.75)	70.42 (27.34)	89.03 (14.77)	85.08 (23.46)
other	63.75 (31.06)	65.14 (31.84)	76.67 (24.72)	66.67 (24.78)	86.50 (22.75)	82.88 (23.31)
Average	64.93 (30.244)	65.73 (29.09)	70.94 (25.80)	68.54 (26.07)	87.76 (19.15)	83.98 (24.49)

Note: Standard deviations are reported in parentheses.

When Perpetrator 3 is removed from the analysis, the confidence ratings still differ from each other significantly, F(1,284) = 4.103, p =0.044: participants reported significantly higher confidence about their decision for Perpetrator 2 (M = 69.74, M = 25.91) than for Perpetrator 1 (M = 65.33, SD = 29.62). Confidence was unaffected by ethnicity, F(1,284) = 1.669, p = .197, or processing type, F(1,284) = .538, p .464. The interaction between confidence ratings, ethnicity and processing type tends towards significance, F(1,284) = 3.542, p = 0.061.

Average confidence, without the data for Perpetrator 3, was reanalysed through an ANOVA. Participant ethnicity and processing type had no effect on average confidence, and the interaction between them was insignificant.

Decision making strategy and processing strategy.

If participants had chosen the ‘not present’ option when presented with the line-up they did not answer the questions about decision making strategy, or processing strategy. As a result, the number of participants who did answer these questions varied across perpetrators: 40 participants (13.9%), 26 participants (9%), and 22 participants (7.6%) had chosen the ‘not present’ option for Perpetrator 1, 2 and 3 respectively.

For the decision making strategy, participants could either respond that the face had popped-out at them, or they had utilised a comparison strategy where they compared the faces in the line-up to one another. Just over half of the participants (152 and 154 participants respectively) compared faces for Perpetrator 1 and Perpetrator 2; however 200 participants experienced a pop-out effect for Perpetrator 3.

Table 3.14.

Frequency table listing the number of participants who utilised each decision making strategy for each perpetrator.

	Perpetrator 1 n=248	Perpetrator 2 n= 262	Perpetrator 3 n=266
Pop-out effect	96 (33.3%)	108 (37.5%)	200 (69.4%)
Comparison strategy	152 (52.8%)	154 (53.5%)	66 (22.9%)
Not present	40 (13.9%)	26 (9%)	22 (7.6%)

Note: Sample percentages are reported in parentheses.

The majority of those participants who relied on a comparison strategy answered that they used featural comparison when choosing a face out of the line-ups for Perpetrator 1 (55.4%) and Perpetrator 2 (60.3%). Therefore, rather than comparing the entire face of each male in the line-up to one another, these participants compared individual features. However, when presented with Perpetrator 3, 53.1% of participants relied on holistic processing.

Table 3.15

Frequency table listing the number of participants who utilised either holistic or featural processing when a perpetrator was chosen.

	Perpetrator 1 n=184	Perpetrator 2 n= 179	Perpetrator 3 n=147
Holistic processing	82 (28.5%)	71 (24.7%)	78 (27.1%)
Featural processing	102 (35.4%)	108 (37.5%)	69 (24.0%)
Participants not required to answer	104 (36.1%)	109 (37.8%)	141 (49%)

Note: Sample percentages of the total sample (n=288) are reported in parentheses.

Correlations

In order to investigate the correlation between confidence and accuracy, all data entries, where participants had rejected the line-up, were removed, as this would have no effect on an investigation in the real world. Therefore, only the confidence levels of choosers were used. Additionally, due to the ceiling effects seen in the accuracy rates of Perpetrator 3, his data was also removed.

A point-biserial correlation coefficient was calculated to investigate the correlation between accuracy an confidence, as one of my variables (accuracy) was a discrete dichotomy; confidence and line-up accuracy were positively correlated, $r_{pb} = .347$, $p < 0.01$, and accounts for 12.04% of variability in accuracy scores (please see Table 3.16).

Confidence was negatively correlated with reaction time, r – 0.345, p < .01, and accounted for 11.9% of the total variability in reaction time. Therefore, quicker decisions are correlated with higher confidence. However, since reaction time is significantly correlated with both confidence and accuracy, I recalculated the correlation between confidence and accuracy while controlling for reaction time. Confidence and accuracy were still significantly correlated, r =.262, p < 0.01; however confidence now only accounted for 6.9% of variability.

Lastly, a correlation was calculated between similarity and confidence; these two variables were not significantly correlated with each other, r = -.028, p = .563.

Table 3.16.

Correlations between confidence levels, accuracy, reaction times and similarity of choosers.

	Confidence	Accuracy	Reaction times	Similarity
Confidence	1			
Accuracy	.347* p < 0.001	1		
Reaction time	-.348* p < 0.001	-.329* p < 0.001	1	
Similarity	-.028 .563	-.078 .109	.031 .530	1

Note: Reported figures are Pearson's r, p, except for correlations with Accuracy which are point-biserial correlations.
* Correlation is significant at p < 0.01

Face Rating study

A post-hoc study was conducted to gather information about the perceptual dimensions of the target faces and their foils, because it was suspected that these could be possible confounding variables that were facilitating recognition. Initially, when the perpetrators were chosen, Perpetrator 1 and Perpetrator 3 were identified as “Typical” faces, and Perpetrator 2 was identified as “Distinctive”; however, there was no empirical data to support this. In order to investigate whether the faces used in the experiment, especially the Perpetrators’, differed in how they perceived, participants were required to rate each of the faces used in the main experiment on 6 perceptual dimensions: Attractiveness; Criminality; Distinctiveness; Neutrality of Facial Expression; Picture Quality and Typicality. All the dimensions were measured from ‘not at all’ to ‘much’, on a Likert scale from 1 to 9 respectively, and responses were collected from keyboard presses. The experiment was built in the E-Prime 1.0 software. The order that the faces were presented in, and the dimension that they were rated on, were randomly selected; once all the faces were rated, the next dimension was randomly selected until all the faces had been rated on all the dimensions. The display size of each face was the same that it was when placed in a line-up, as this was how participants in the main experiment would have perceived and rated the face when making decisions. A definition of each dimension was included in the instructions for the participant (please see Table 3.17 for a list of all the definitions).

24 participants were recruited who matched the demographic profile of the participants of the main study. Half of these participants (2 male, 10 female) had indicated that they were white, and the other 12 (2 male, 10 female) had indicated that they were black, coloured, Indian or Asian. The experiment took place in the ACSENT Research laboratory in groups of up to 5 participants. The entire rating took about 10 minutes.

A repeated-measures ANOVA was run on this data, and yielded the following data: white participants rated the three perpetrators significantly different on the attractiveness measure, F(2,22) = 6.775, p = .005, and distinctiveness measure, F(2,22) = 3.826, p = .037. Pairwise comparisons demonstrate that Perpetrator 1 was rated as more attractive than Perpetrator 2, p = .032, and Perpetrator 3, p = .002; Perpetrator 2 and Perpetrator 3 did not differ significantly on their attractive ratings. Perpetrator 2 was rated as significantly more distinctive than Perpetrator 1, p = .033, who was considered significantly less distinctive than Perpetrator 3, p = .068. Perpetrator 3 and Perpetrator 2 were not rated significantly different from each on the on the distinctiveness scale.

Other participants did not rate any of the three perpetrators as significantly different from one another on any of the perceptual dimensions; participant ethnicity did not have a significant effect on any of the perceptual dimensions.

Table 3.17.

The mean rating of each perceptual dimension for each perpetrator.

	White participants			Other participants		
	Perpetrator 1	Perpetrator 2	Perpetrator 3	Perpetrator 1	Perpetrator 2	Perpetrator 3
Attractiveness	3.83^{a*b*}	2.58^{a*}	2.33^{b*}	2.83	2.25	2.08
	(2.04)	(1.68)	(1.30)	(1.64)	(1.54)	(1.08)
Criminality	5.00	5.17	6.08	4.84	6.42	5.50
	(2.00)	(2.52)	(1.83)	(2.17)	(2.35)	(2.93)
Distinctiveness	3.83^{a*b*}	6.00^{a*}	5.08^{b*}	4.17	4.08	3.42
	(1.59)	(1.86)	(1.68)	(2.08)	(3.03)	(2.27)
Facial Expression	7.00	6.17	7.83	7.08	5.92	6.75
	(2.56)	(2.25)	(1.19)	(2.34)	(2.78)	(2.45)
Picture Quality	5.75	5.67	5.08	4.42	3.67	4.00
	(1.54)	(1.23)	(1.62)	(2.35)	(1.83)	(1.54)
Typicality	5.42	5.17	4.83	5.33	4.17	4.00
	(2.27)	(2.52)	(2.44)	(2.61)	(2.25)	(2.37)

Note. Standard deviations are in parentheses. ^{a*} and ^{b*} refers to significant within group differences, $p < 0.05$.

Chapter Four: Discussion

Summary of Results

Overall, there was no support for the hypothesis that white participants (or high-expertise participants) would be better at the recognition task after automatic processing, and that other participants (low-expertise participants) would have better recognition after a controlled-processing task; there was no interaction effect among the three variables: ethnicity of participant, processing type of the Stroop task, and recognition accuracy. These results are similar to what Brand (2004) found in his experiment, although he included only two variables: processing type and recognition accuracy. However, unlike Brand (2004), the manipulation of the Stroop task was effective in this experiment. Therefore, despite performing at different rates in the two Stroop conditions, participants' recognition rates were unaffected by this experimental manipulation.

There were two sets of results in this study: one set that contained Perpetrator 3, and the other that did not. Within the first set of results, the processing task, and the interaction between that and participant ethnicity did not significantly affect the recognition results (after the initial presentation of the line-up), while the main effect of ethnicity on recognition results only tended towards significance. In the forced choice condition, all participants performed better; however, the main effects and interactions remained non-significant. Reaction time, measured only during the initial presentation of the line-up, yielded similar results: no significant main effects of participant ethnicity, nor processing type; the interaction between participant ethnicity and processing type was also non-significant. When reaction time was partitioned into two groups depending on whether it was made in fewer than or more than 12 seconds, it became clear that most of the accurate decisions were made in fewer than 12 seconds.

The accuracy data was analysed for each perpetrator, and it became apparent that Perpetrator 3 was recognised significantly better than the other two perpetrators. This has important implications, as this could affect reaction time (for example, he may have been easier to recognise, and therefore the decision was made quicker), and it could affect confidence (for example, since he is easier to recognise, participants may be more confident in their decision when they selected him.). Interestingly, the line-up containing Perpetrator 3 did not have the highest bias, nor the lowest line-up effect size of the three perpetrators – in fact, the line-up containing Perpetrator 2 had this, yet Perpetrator 2's recognition results were more similar to Perpetrator 1 than to Perpetrator 3. Therefore, Perpetrator 3's data was removed, and the data was reanalysed.

The recognition rates following the initial choice and the forced choice were unaffected by participant ethnicity, processing type, and the non-significant interaction between the two. Neither participant ethnicity nor processing type significantly affected reaction time; however the interaction between processing type and ethnicity did tend towards significance – but in the opposite direction to that hypothesised! White participants made slower decisions following the automatic processing task, whereas other participants made faster decisions following the automatic processing task. Confidence remained unaffected by participant ethnicity and processing type, but confidence among choosers was significantly correlated with accuracy and reaction time.

Conclusion

Overall, it appears as though the Stroop task did not produce similar results to those seen with the Navon letter tasks, suggesting that the processing shift between automatic/controlled processing does not cause the difference in recognition results; rather the results seen in the Navon letters task resulted from a shift in global/local processing. In order to make this statement, I am assuming that the Stroop task is the best way to measure automatic and controlled processes, while it could be possible that it is not. How exactly would one know if a task is automatic or if it is controlled? It has been recommended that in order to properly investigate whether processes are automatic or controlled, researchers should rather investigate each defining feature of automaticity (Moors & De Houwer, 2006); in some cases, this may not be necessary, and the research question should provide guidance about which particular feature of automaticity is the most important for the investigation. Unfortunately, prior research has not established whether the Navon letters have automatic/controlled processing qualities, and it’s not extensively used in that research domain. For this particular experiment, what was most important was inhibition, as this is what is similar between the Navon letters and Stroop, since both require an inhibition of one response in order to give another. Of course, it is possible that the amount of inhibition required by both tasks is not the same, and perhaps a way to compare inhibition “effort” between the two tasks would be to compare the differences in reaction times for both conditions. It is possible that these tasks are not comparable in terms of inhibitory effect - the Stroop examines the automatic, over-learned reading response and how much cognitive inhibition is required to inhibit this in order to respond to the font colour, whereas neither dimension in the Navon letters seem to have this automatic quality. Although Perfect and colleagues (2008) used their sets of Navon letters in order to induce this automaticity (through the bias/precedence), without reported reactions time it is not possible to know how much inhibition (if any) was required by participants to respond to the non-biased dimension (for example, the local response to globally-biased letters).

With this in mind, it is necessary to re-evaluate why Perfect et al. (2008) saw the results that they did, specifically that recognition accuracy was worse when local-responses and global responses were given to a globally biased letter, and locally-biased letter respectively. Perfect et al. (2008) suggested that an inhibition effect caused these results: this inhibition primed participants to use controlled processing which, in turn, negatively affected their facial recognition scores. Perfect et al. utilised the same video and line-up that Macrae & Lewis (2002) had, and the second set of authors had seen similar results where participants' recognition accuracy was better following global-responses; however, Macrae & Lewis only used the traditionally-used Navon letters, which were globally-biased, whereas Perfect et al. used two sets of Navon letters (as already established). Neither Perfect et al., nor I had included a control group; unlike Macrae & Lewis who had. Interestingly, the recognition accuracy following their globally-biased Navon letters were .83, .60 and .30 for their global-response, control and local-response groups respectively (Macrae & Lewis, 2002); whereas the recognition accuracy from the participants in the globally-biased group in the experiment by Perfect et al., did not have the same high accuracy achieved by the global-response group in Macrae & Lewis, instead the global-response and local-response groups obtained .65 and .30 accuracy respectively (personal correspondence with the authors). Since both groups of authors used the same encoding and recognition stimuli, the difference in recognition accuracy may be a result of the difference (1) in the reported duration of the Navon task: Macrae & Lewis had their participants respond to 100 Navon stimuli in 10 minutes, whereas Perfect et al., had their participants perform the Navon letter task for 2 and a half minutes (30 letters were presented for 5 seconds each); or (2) in the size of Navon stimuli used. Macrae & Lewis used only one set of Navon letters which were printed on cards, whereas Perfect et al. had two sets of Navon letters: globally-biased letters which were 85mm x 55mm, and locally-biased letters which were 115mm x 95mm (it is unclear if the cards were printed on paper, or if they were presented on a computer screen). Therefore, in the globally-biased condition, the smaller letters were placed very close to one another to form the large letter (see the example of the figure printed in Perfect et al., 2008), thus the attentional focus for participants would be very small (because the small letters and the large letter that they formed are confined to a much smaller space), whereas the attentional focus was much larger for the locally-biased letters. This smaller attentional focus may have resulted in a greater Navon letter effect, because the effects of distractor tasks are greater when attentional load is smaller, and when both conflicting domains belong to same stimuli (Chen, 2003). This argument would mean that the distracting effect would be greater for participants who had to respond locally, as they would have a narrower attentional focus, thus their recognition performance should be the worst – and compared to the global-response group it was! This argument breaks down when one considers the locally-biased letters, because the stimuli was larger, and the ‘features’ (or the smaller letters) were widely spaced, thus attentional focus is larger, yet participants performed worse at the recognition task after responding globally.

However by distorting the two sets of stimuli so that the one is globally-biased, and the other is locally-biased, it is possible that the sets are no longer comparable. In the smaller globally-biased letter it is still readily apparent what the global and local letters are, unlike in the locally-biased letter where the identity of the global-letter is not as clearly accessible. For the locally-biased letters, the stimuli are larger, and the features are more widely spaced apart, thus it is easier to focus on the local features, while making it more difficult to focus on the global letter, but this does not necessarily mean that inhibition is still required in order to process the global-letter. It does not appear as though as much inhibition is required for the locally-biased letter, because the global and local letters are not competing as much as they are in the globally-biased condition. Perhaps this explains why the difference in recognition scores between participants who responded globally or locally to the locally-biased letters is not as large as it was for those who had responded to the globally-biased letters in Perfect et al. (2008). Unfortunately, it does not explain why recognition accuracy following the global-response for the locally-biased letter was worse.

Despite the overall hypothesis being not being supported, some interesting data did emerge from my experiment, such as the improvement in accuracy scores following the forced choice condition. Both groups of participants improved their recognition scores after the forced choice condition, and this improvement suggests that participants utilised a stricter response criterion causing them to favour the ‘not present’ option in the initial presentation of the line-ups. Initially, white participants performed better than other participants had, but this gap was lessened after the forced choice presentation. As discussed in the results section, it appears as though other participants had adopted a stricter criterion than white participants had following the initial presentation. In the initial presentation of the line-up, a miss and an incorrect line-up rejection were both scored ‘0’; however in the forced-choice condition, participants only scored ‘0’ if they chose the wrong face, because they could no longer answer ‘not present’. Therefore, it was not possible for participants to perform worse in the forced choice, as participants were instructed that if they had chosen a face initially, they were to choose that same face in the forced choice condition. This means that in order for the gap between white and other participants to lessen so much that the effect of ethnicity on recognition scores was no longer tending towards significance, other participants, specifically those within this group who had indicated ‘not present’, must have improved their recognition scores the most, because they had adopted the stricter criterion of the two groups. Curiously, confidence did not significantly differ between these two groups, and both participants reported high confidence (>70%); this suggests that other participants’ hesitancy to make an answer was not due to a lack of confidence.

While not significant, it appeared as though the effect from an interaction between ethnicity of participant and processing type on reaction time was approximating significance (once perpetrator 3 was removed) – however this significance was in the opposite direction hypothesised! White participants performed faster after the controlled condition, and other participants performed faster after the automatic condition. This unexpected pattern in reaction times could be a result of the instructions given. Previous research has shown that experts benefit from strictly timed conditions, where were explicitly explained to them; however, in my experiment, participants were not encouraged to make the decision as quickly as possible, but only to look at the line-up and choose a face. It is possible that the instructions given to my participants did not encourage a swift decision-making, but rather controlled processing. Since white participants should be experts at recognising faces of the same ethnicity, they should benefit from automatic processing tasks, but if the instructions given to them did not encourage this type of processing (unlike those used in other expertise literature that emphasise the speed of the decision), then controlled processing may be induced, thus causing them to make a decision slower after the automatic task of the Stroop, and faster after the controlled task of the Stroop. There is another possible explanation for the result that was seen. Govorun and Payne (2002) used the Stroop task in an experiment that investigated whether it would facilitate or prevent automatic bias. They hypothesised that processing any task, which tapped cognitive resources and caused cognitive load, prior to a weapon-identification task would lead to an increased reliance on automatic stereotyping. Participants were required to either complete the automatic or controlled condition of the Stroop, and afterwards they had to identify items as either hand guns or hand tools. Each item was preceded with a Black or White face. Participants were more likely to demonstrate racial stereotyping in the weapon identification task after performing the controlled condition of the Stroop task, for example, identifying the item as hand gun if it followed Black face, because the effect of inhibiting this condition of the Stroop task strains cognitive resources, therefore making it more difficult to inhibit stereotypes. This effect was more prominent in participants who required a large amount of control over their automatic responses, and therefore needed a large amount of cognitive resources to ensure this control. The authors explain that the difference seen between the stereotypical answers and correct answers in the weapon identification task was a function of automatic bias, A, multiplied by the probability that control will fail, 1 – C. Therefore, if the cognitive reserve is high, then there is a low probability that this control will fail, and this when multiplied with the automatic bias results in a low probability of stereotypical responses occurring; however, if the cognitive reserve is low, then there is a higher probability of stereotypical responses occurring. Their concluding argument was that the cognitive loading of the Stroop task affected the amount of cognitive control that participants had available to override the stereotype, and not their automatic bias towards an answer. Therefore, if processing the controlled condition of the Stroop task diminishes cognitive control, and encourages a reliance on automatic processing, then white participants would be more likely to utilise an automatic processing style when recognising faces, while other participants are unaffected by the cognitive load.

The difference in recognition accuracy among the three perpetrators was a very interesting finding, specifically that Perpetrator 3 was so well recognised, despite not having the lowest line-up effective size, nor the highest bias, and he was not rated as the most distinctive perpetrator. Yet, despite this, his accuracy suggested ceiling effects. The data from the pilot experiment, which investigated which version of the video stimuli to use, did suggest that recognition was better for Perpetrator 3 after the FAR video (84.6%) than it was after the CLOSE (30.8%) video; data for the perceptual dimension study was gathered from participants rating the photograph, which was used in the line-up containing Perpetrator 3, and not from the video that used in the encoding stage of the main experiment. It is possible that the additional visual data that was available in the video facilitated recognition, and that this data were not captured in his photograph. However two meta-analyses have found that (1) videos used as encoding stimuli do not predict verbal overshadowing (Meissner & Brigham, 2001), and (2) inconclusive results regarding the effect of using videos as encoding stimuli on facial recognition (Shapiro & Penrod, 1986). Shapiro and Penrod (1986) did hypothesise that recognition would be facilitated if videos, rather than photographs, were used as encoding stimuli. However, later research has offered inconsistent conclusions: participants were 70% accurate when presented with a still image (which was taken from a video) of a face, and had to decide whether that face was present in 10-person line-up (Bruce et al., 1999, as cited in Hancock, Bruce, & Burton, 2000), yet in other research, recognition rates of 76% were obtained when participants had to recognise faces that they had encoded from a video, while a recognition rate of only 46 % was obtained when participants had to encode still images of those faces from the same video (Pike, Kemp, Towell & Phillips, 1997). Further research needs to investigate the effect of encoding stimuli on line-up recognition, in particular how different types of videos affect recognition, and the amount of additional perceptual information that is available to participants and may facilitate their recognition. Also there may be a concern about how line-ups are constructed, because line-ups are normally constructed from photographs, and despite being fair, as measured by line-up effective size and bias, these line-ups may still not control for all the perceptual information which is available from videos.

Limitations and Recommendations for Future Research

There were a number of limitations with this study, and future research should seek to avoid or correct these in the future. The first flaw in this research design is that I had only White perpetrators. It would be beneficial to include another group of perpetrators that belong to a different ethnic group, such as Black South African, so that an ORB can be investigated properly. The results from this study tentatively suggested that an ORB occurred: white participants recognised the white perpetrators better than the other perpetrators did; however without Black perpetrators, and a homogenous participant group that is matched to the Black perpetrators, it can only be *speculated* that an ORB occurred.

Another limitation is that I did not use target-absent line-ups. Without such line-ups, I could not calculate any signal detection measures - which would have provided more information about the types of decisions that participants made. For example, my participants seemed reluctant to choose a perpetrator when initially given the line-up, however when they were forced to choose, they performed relatively well. If I had a target-absent line-up, I could investigate what this pattern would have looked liked – would they have performed better, because their reluctance to make a choice appears as though they are correctly rejecting the line-up?

Another methodological concern is how to measure whether a type of processing has been induced in participants in order to ensure that the experimental condition has worked. For example, my participants performed at significantly different rates in both conditions of the Stroop (i.e., they were slower following the font condition, and faster after the word condition), whereas Brand (2004) was unable to induce such results. Despite this, both of our experiments saw no effect of processing type on recognition accuracy. This could be because (1) the Stroop task does not induce the type of processing that we hypothesised, (2) the type of processing, which it does induce, did not carry over to the recognition stage of the experiment; or (3) the TIPS did not occur, because the manner in which the faces are processed initially is not incompatible with the processing type induced by the Stroop task. A way to investigate this could be to adopt the design used by Lewis et al. (2009), and replace the Navon tasks, which are used prior to encoding and to recognition, with one condition of the Stroop task, thus allowing us to investigate the TIPS with the Stroop tasks. Although this design is still limited by not having a way of measuring how the faces are processed after the Stroop words, it does provide opportunity to possibly influence how faces are processed in both the encoding and recognition stage, thus allowing for the TIPS to be better investigated.

What complicates this further is the difficulty in operationalising mental processes so that they can be measured and observed accurately. Unfortunately, automatic and controlled processing is predominantly studied in the expertise literature that focuses on motor-perceptual skills. Even though the Stroop is used in clinical practise and in research to demonstrate inhibition and cognitive control, it is not clear whether the processes which are being inhibited by the font colour condition of the Stroop task are necessarily automatic. Perhaps researchers could be more tentative when claiming that processes, or the effects seen, are a result of automatic/controlled processing, and as Moors and De Houwer (2006) suggest, rather focus on the exact quality of automatic/controlled processing that is being displayed. Although previous researchers have suggested that the Navon letters require inhibition, none have reported the differences in the processing times of these letters, thus not providing empirical evidence of this inhibition. Perhaps the reason why the Stroop task does not replicate the Navon letter effect is because of the obvious: the effect is not related to inhibition, and the Stroop task does not induce global/local processing.

References

Bartlett, J. C. & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology,* 25, 281 – 316.

Beilock, S. L., Carr, T. H., MacMahon, C., & Starkes, J. L. (2002). When paying attention becomes counterproductive: impact of divided versus skill-focused attention on novice and experience performance of sensorimotor skills. *Journal of Experimental Psychology: Applied, 8,* 6-16.

Beilock, S. L., Bertenthal, B. I., McCoy, A. M. & Carr, T. H. (2004). Haste does not always waste: Expertise, direct of attention, and speed versus accuracy in performing sensorimotor skills. *Psychonomic Bulletin & Review,* 11, 373 – 379.

Brand, A. (2004). A web experiment based enquiry into the verbal overshadowing effect (Unpublished doctoral dissertation). Cardiff University, Cardiff, Wales.

Brandimonte, M. A., Schooler, J. W. & Gabbino, P. (1997). Attenuating verbal overshadowing through colour retrieval cues. *Journal of Experimental Psychology*, 23, 915 - 931.

Brigham, J.C. (1990). Target person distinctiveness and attractiveness as moderator variables in the confidence-accuracy relationship in eyewitness identifications. *Basic and Applied Social Psychology, 11,* 101-115.

Brigham, J. C. & Malpass, R.S. (1985). The role of experience and contact in the recognition of faces of own-race and other-race persons. *Journal of Social Issues, 41,* 139-155.

Brown, C. & Lloyd-Jones, T. J. (2003). Verbal overshadowing of multiple face and car recognition: Effects of within – versus across-category verbal descriptions. *Applied Cognitive Psychology,* 17, 183 – 201.

Brewer, N. & Wells, G. (2006). The confidence-accuracy relationship in eyewitness identifications: Effects of lineup instructions, foil similarity and target-absent base rates. *Journal of Experimental Psychology: Applied, 12,* 11-30.

Chen, Z. (2003). Attentional focus, processing load, and Stroop interference. *Perception & Psychophysics, 65,* 888-900.

Chiroro, P. M., Tredoux, C. G., Radaelli, S., & Meissner, C. A. (2008). Recognising faces across continents: The effect of within-race variations on the own-race bias in face recognition. *Psychonomic Bulletin & Review, 15,* 1089-1092.

Clare, J. & Lewandowsky, S. (2004). Verbalizing facial memory: criterion effects in verbal overshadowing. Journal *of Experimental Psychology, 20,* 739-755.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: a parallel distribued processign account of the Stroop effect. *Psychological Review, 97,* 332-361.

Craik, F.I.M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology, 104,* 268-294.

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychological Science, 15,* 806-813.

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.

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.

Eberhardt, J. L. (2005). Imaging race. *American Psychologist, 60,* 181-190.

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

Finger, K. (2002). Mazes and music: Using perceptual processing to release verbal overshadowing. *Applied Cognitive Psychology, 16,* 887 – 869.

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.

Flegal, K. E., & Anderson, M.C. (2008). Overthinking skilled motor performance: or why those who teach can't do. *Psychological Bulletin and Review, 15,* 927-932.

Gauthier, I., Behrann, M., & Tarr, M. J. (1999). Can face recognition really be dissociated from object recognition? *Journal of Cognitive Neuroscience, 11,* 349 – 370.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. Nature and Neuroscience, 3, 191-197.

Gauthier, I., & Nelson, C. A. (2001). The development of face expertise. *Current Opinion in Neurobiology, 11,* 219-224.

Golby, A. J., Gabrieli, J. D. E., Chiao, J. Y., & Eberhardt, J. L. (2001). Differential responses in the fusiform region to same-race and other-race faces. *Nature Neuroscience, 4,* 845-850.

Govorun, O., & Payne, B. K. (2005). Ego-depletion and prejudice: separating automatic and controlled components. Social cognition, 24, 111-136.

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

Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., & Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. *Neuron, 22,* 189-199.

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.

Jackson, R. C., Ashford, K. J. & Norsworthy, G. (2006). Attentional focus, dispositional reinvestment and skilled motor performance under pressure. *Journal of Sport and Exercise Psychology, 28*, 49 – 68.

Kanwisher, N. (2000). Domain specificity. *Nature Neuroscience, 3*, 759 – 763.

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.

Kanwisher, N., Tong, F., & Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. *Cognition, 68*, B1-B11.

Lewis, M. B., Mills, C., Hills, P. J., & Weston, N. (2009). Navon letters affect face learning and face retrieval. Experimental Psychology, 56, 258 – 264.

Lezak, M. D., Howieson, D.B., & Loring, D. W. (2004). Neuropsychological Assessment (4th edition). New York: Oxford University Press, Inc.

Logan, G.D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492-527.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological bulletin, 109,* 163- 203.

MacLeod, C. M. (2005). The Stroop task in cognitive research. In A. Wenzel & D. C. Rubin (Eds.) Cognitive methods and their application to clinical research (pp. 17-40). Washiongton, DC: American Psychological Association.

MacLeod, C. M. & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 126-135.

MacLin, M., K. (2002). The effects of exemplar and prototype descriptors on verbal overshadowing. *Applied Cognitive Psychology*, *16*, 929-936.

Macmillan, N. A. & Creelman, C. D. (1991). Detection theory: A user's guide. New York: Cambridge University Press.

Macrae, C. N., & Lewis, H. L. (2002). Do I know you? Processing orientation and face recognition. *Psychological Science*, *13*, 194-196.

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology, 13,* 330-334.

Malpass, R.S., Tredoux, C. G., & McQuiston-Surrett, D. (2007). Line-up construction and line-up fairness. In R. Lindsay, D. Ross, J. D. Read, & M. P. Toglia (Eds.), Handbook of eyewitness psychology: Volume 2. Memory for people. Mahwah, NJ: Erlbaum.

McKone, E., Kanwisher, N., Duchaine, B.C. (2007). Can generic expertise explain general processing for faces ? *Trends in Cognitive Sciences, 11,* 8 – 15.

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

Melcher, J. M. & Schooler, J. W. (1996). The misremembrance of wines past: Verbal overshadowing and perceptual expertise differentially mediate verbal overshadowing of taste memory. *Journal of Memory and Language, 35,* 231 – 245.

Melcher, J. M., & Schooler, J. W. (2004). Perceptual and conceptual training mediate the verbal overshadowing effect in an unfamiliar domain. *Memory & Cognition, 32,* 618 – 631.

Michel, C., Caldara, R., & Rossion, B. (2006). Same-race faces are perceived more holistically than other-race faces. *Visual Cognition, 14,* 55 – 73.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psyhcological Bulletin, 132,* 297-326.

Morris,C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behaviour, 16,* 519-533.

Navon, D. (1977). Forest before trees: the precedence of global features in visual perception. *Cognitive Psychology, 9,* 335 – 383.

Neissner, U. (1964). Visual Search. Scientific American, 210, 94-102.

Perfect, T. J., Weston, N. J., Dennis, I., & Snell, A. (2008). The effects of precedence on Navon-induced processing bias in face recognition. *The Quarterly Journal of Experimental Psychology, 61,* 1479-1486.

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

Posner, M. I., & Snyder, C. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition* (pp. 55-85). Hillsdale, NJ:Erlbaum.

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

Reinhard, M. A., & Sporer, S. L. (2010). Content versus source cue information as a basis for credibility judgements: the impact of task involvement. *Social Pychology, 41,* 93-104.

Rossion, B., & Gauthier, I. (2002). How does the brain process upright and inverted faces? Behavioural *and Cognitive Neuroscience Reviews, 1,* 63-75.

Schooler, J. (2002). Verbalisation produces a transfer inappropriate processing shift. *Applied Cognitive Psychology, 16,* 989 – 997.

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

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General, 122,* 166-183.

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

Schmidt, H. C. (2010). Determinants of estimated face composite quality. (Unpublished doctoral dissertation). University of Cape Town, Cape Town, South Africa.

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

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84,* 127 – 190.

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence and decision times in simultaneous and sequential line-ups. *Journal of Applied Psychology, 78,* 23-33.

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

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). A compendium of neuropsychological tests (3rd edition). New York: Oxford University Press, Inc.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643-662.

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

Treisman, A.M. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97-136.

Wells, G. L. & Turtle, J. W. (1987). Eyewitness testimony research: current knowledge and emergent controversies. *Canadian Journal of Behavioural Science, 19,* 363 – 388.

Weston, N. J., Perfect, T. J., Schooler, J. W., & Dennis, I. (2007). Navon processing and verbalisation: A holistic/featural distinction. *European Journal of Cognitive Psychology, 20,* 587 – 611.

Yin, R., (1969). Looking at upside-down faces. *Journal of Experimental Psychology: Learning, Memory and Cognition, 81,* 141 -145.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16,* 747 – 759.

Appendix A

Consent form

SRPP VIDEO CLIP EXPERIMENT:

PARTCIPANT CONSENT FORM

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

Student No: ____________________________________

Age: _______________________

Sex: _______________________ (M, F).

Race (please tick): ☐ WHITE ☐ BLACK ☐ COLOURED ☐ INDIAN ☐ OTHER

If ‘Other’, please specify: _______________________________

Psychology course code: ________________________________

Is South Africa your country of birth ? _____________________ (Y, N)

If ‘No’, please specify your country of birth: ________________

If ‘No’, how many years have you been living in South Africa ? ________ (years)

Are you dominantly left-handed or right-handed ? ___________

For your time and participation, you will receive 30 minutes (1 unit) towards your SRPP requirement.

Please note that at any time during the 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 gladly assist you.

THE STUDY: This study investigates how human beings perceive faces. You will be required to watch a video clip, and then complete some cognitive tasks. All the instructions are included within the experiment. If you have any questions about the experiment, the researcher will be happy to answer them once you have completed the experiment.

By signing this consent 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 descriptive above, and filled in the required information. I am aware that any personal information will not be distributed.

SIGNATURE: _______________________________________

DATE: ____

Appendix B

Dropbox links and lists of materials

Stimuli videos:

Perpetrator 1 – Close video – original length: http://db.tt/0htla8M

Perpetrator 1 – Far video – original length: http://db.tt/de4pyKC

Perpetrator 1 – Close video – 10 seconds duration: http://db.tt/gOY3JBo

Perpetrator 1 – Far video – 10 seconds duration: http://db.tt/58Pwf7q

Perpetrator 2 – Close video – original length: http://db.tt/5rP92JN

Perpetrator 2 – Far video – original length: http://db.tt/YjVSh3J

Perpetrator 2 – Close video – 10 seconds duration: http://db.tt/jIdduAk

Perpetrator 2 – Far video – 10 seconds duration: http://db.tt/LuwidKW

Perpetrator 3 – Close video – original length: http://db.tt/h20Ib92

Perpetrator 3 – Far video – original length: http://db.tt/2PV1Pvj

Perpetrator 3 – Close video – 10 seconds duration: http://db.tt/yIC7X43

Perpetrator 3 – Far video – 10 seconds duration: http://db.tt/qYVBCb4

Line-up images:

Mock witness procedure 1 & 2: Perpetrator 1: http://db.tt/ZWiMKP4

Mock witness procedure 1 & 2: perpetrator 2: http://db.tt/t3iV1l8

Mock witness procedure 1 & 2: perpetrator 3: http://db.tt/A7eH6TQ

Mock witness procedure 3: perpetrator 1: http://db.tt/EkTTZk3

Mock witness procedure 3: perpetrator 2: http://db.tt/0QNY5Nx

Mock witness procedure 3: perpetrator 3: http://db.tt/yUleCVo

Mock witness procedure 4: perpetrator 1: http://db.tt/b2vrFvz

Mock witness procedure 4: perpetrator 2: http://db.tt/3PuF3SR

Mock witness procedure 4: perpetrator 3: http://db.tt/Syoisnq

Appendix C

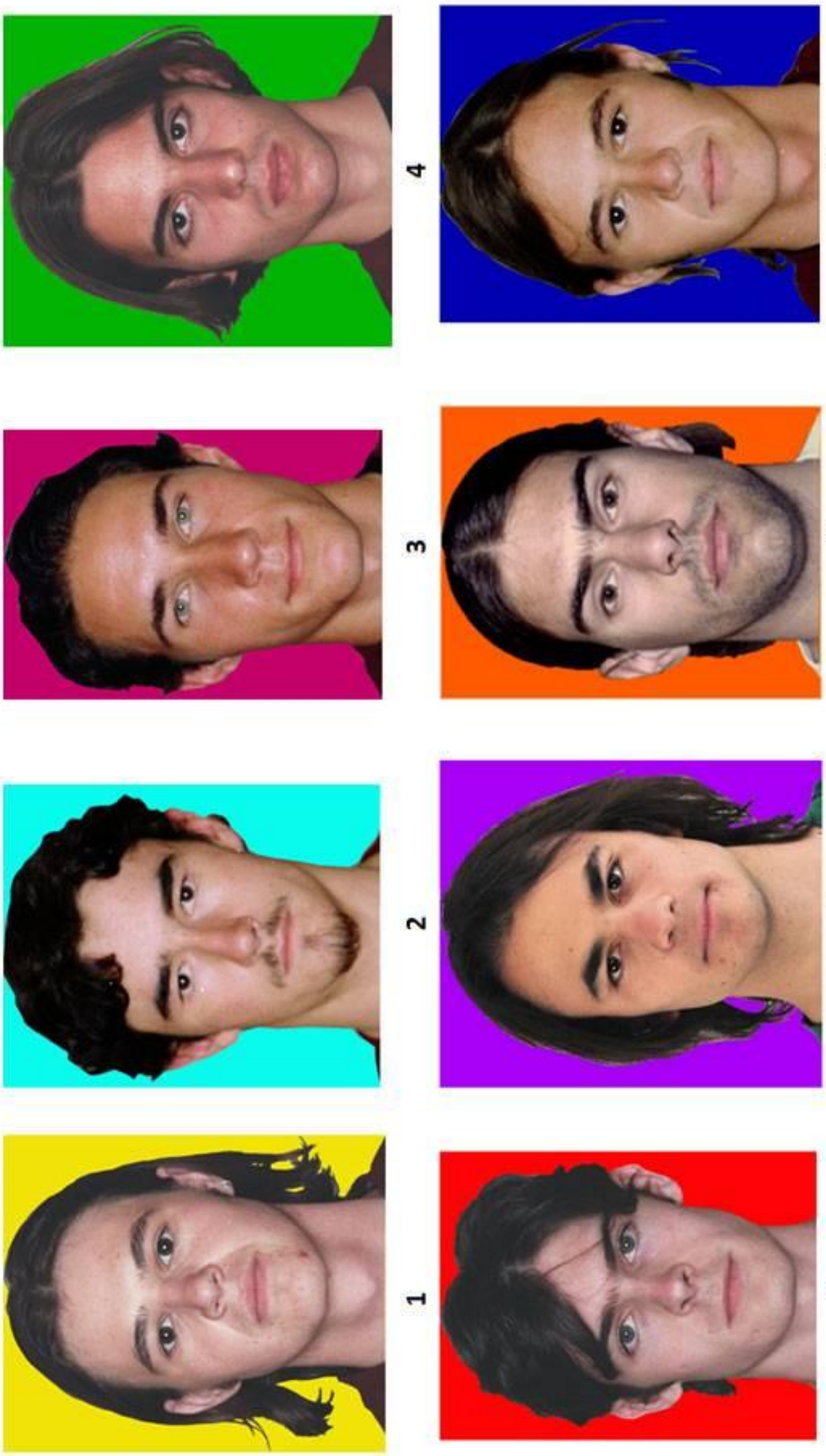
All descriptions for each perpetrator

Perpetrator	Participant	Descriptions	**Modal descriptions**
1	1	A white guy with a mohawk hairstyle. He has black hair. He looks like a jock. He was not smiling and had pimples on his face. He was wearing a white t-shirt.	**Brown hair** **Brown eyes** **Messy/Styled hair** **Maybe blemished skin**
	2	Untidy hair, lethargic, white shirt, white guy, hair colour- light brown, brown eyes	
	3	light brown, relatively greasy hair. White male. Pinkish skin. Relatively blemished. About age 22-26	
	4	A white males face. He had styled brown hair and brown eyes. He had a bit of facial hair	
	5	dark blonde medium length hair, dark eyes, white male, early 20’s, pointy nose, left eye slightly squinted inwards.	
	6	White, male, hair styled up, greasy skin, pink lips, dirty blonde hair	
2	1	A pale, white guy, he had long brown hair. He was wearing a green top. He had thin lips, he is a thin guy. He had a round nose	**Long dark brown hair** **Thick eyebrows** **Thin lips** **Thin face**
	2	Hair dark brown and long until the neck, thin, dry lips, thick eyebrows, white guy, tiny mole on forehead,green shirt	
	3	straight, thinning, black hair, quite long. Receding hairline. Relatively large forehead. Large black eyebrows. Beauty spots around the mouth.	
	4	A males face. Tanned skin. Black hair and dark brown eyes. Hair is thin and long. His face is narrow.	
	5	white male, early to mid twenties, dark shoulder length hair, dark eyes, slightly thick dark eyebrows, long face,	

		pointy chin	
	6	White, male, black, thin, shoulder-length hair. Thin face, pale.	
3	1	A white guy with longish messy blond hair. He had a green t-shirt on. He seemed a bit chubby.	**Longish Hair** **Dark blonde/Light brown hair** **Chubby** **Brown eyes**
	2	Green shirt, brown hair, white guy, chubby in the face	
	3	white male. Relatively long hair, dark blonde/light brown with blonde undertones, slight wave. Looks somewhat dirty. Eyes relatively quite close together, brown eyes. Pinched mouth. Pale freckles	
	4	A white males face. He had blondish hair with hazel eye colour. His hair was quite long, reached his ears. He has a bit of facial hair. He also had freckles around his nose and cheeks.	
	5	white male, early to mid twenties, rounded face, dark brown eyes, dark blonde hair, medium length hair	
	6	Chubby face, long hair (covering the ears) with the top of the hair combed over. White, male, dirty blonde hair	

Appendix D

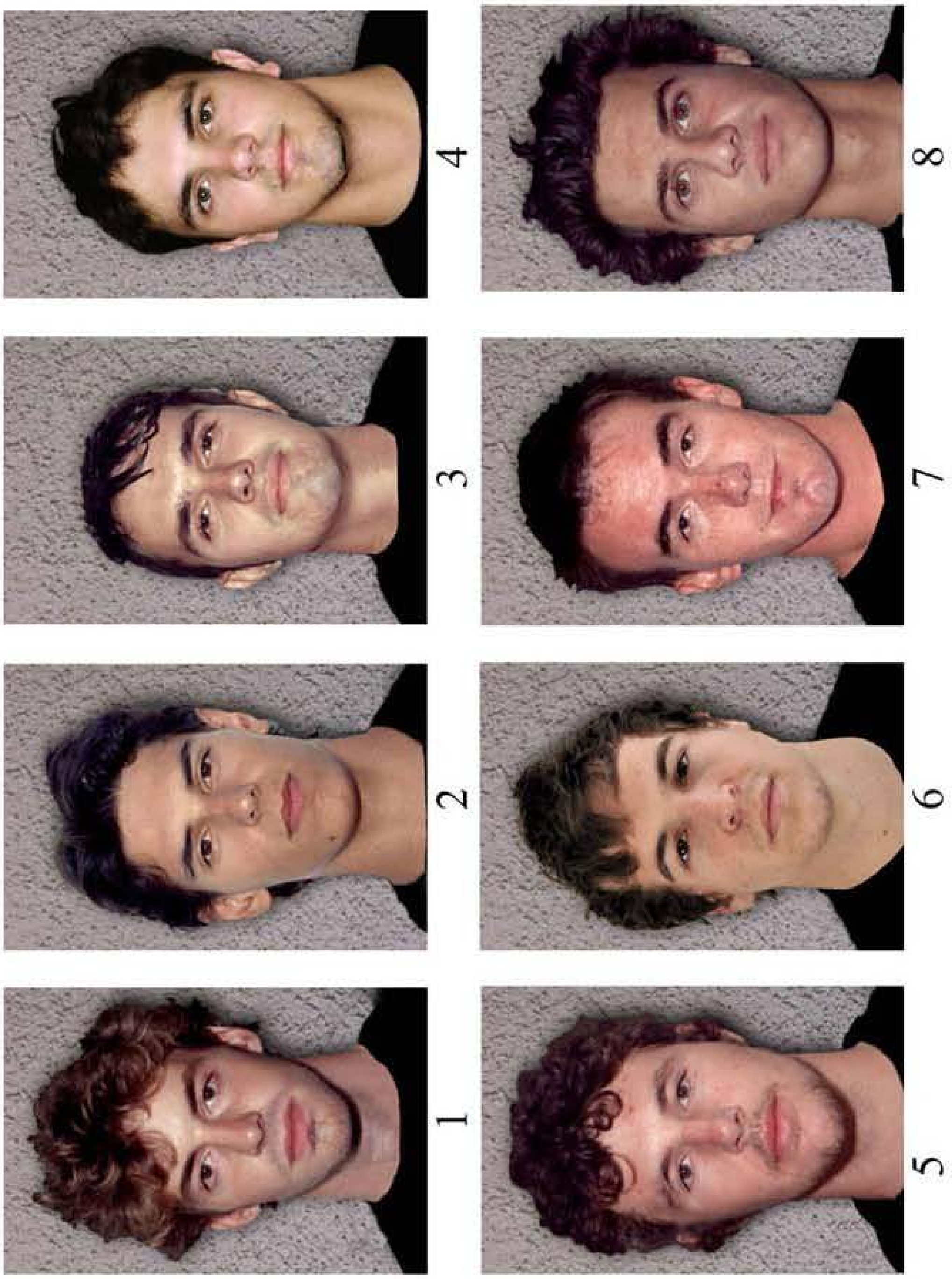
Line-up containing perpetrator 1 (position 3)



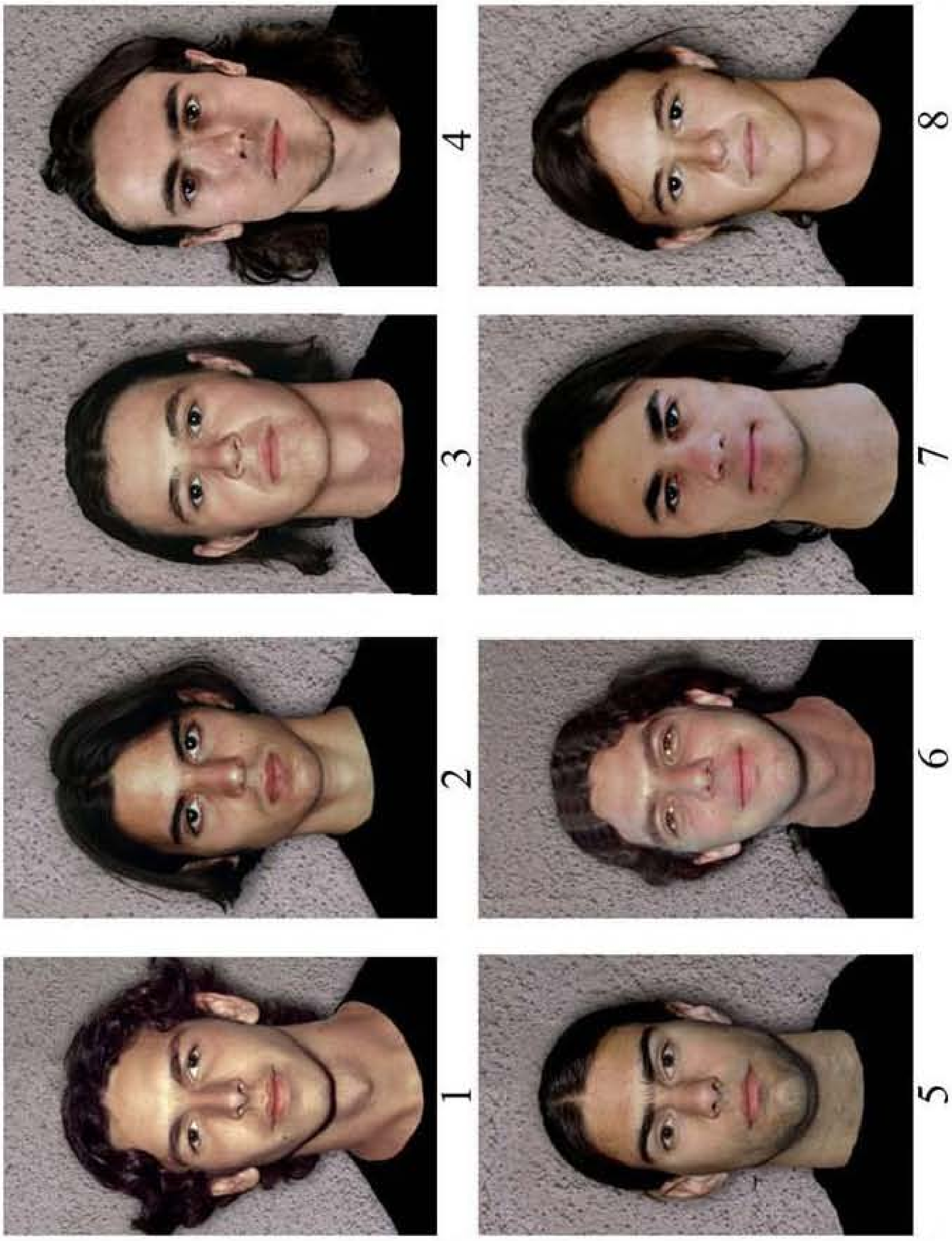
Line-up containing perpetrator 2 (position 6)

Line-up containing perpetrator 3 (position 4)

Appendix E



Line-up containing perpetrator 1 (position 6)



Line-up containing perpetrator 2 (position 7)

Line-up containing perpetrator 3 (position 5)