ISSN 0081-2463
South African Journal of Psychology, 42(3), 2012, pp. 423-433

Red, Green, Blue, Red, Argh!
A missing shift in processing: the Stroop task does not affect facial recognition

Alicia Nortje and Colin Tredoux
Department of Psychology, University of Cape Town, South Africa
Alicia.nortje@gmail.com

The transfer-inappropriate processing shift is one explanation for the verbal overshadowing effect: a phenomenon that prevents accurate face recognition following a verbal description of that same face. Rather than examining a shift from configural to featural processing, this experiment investigated whether a shift from automatic to controlled processing could have similar deleterious effects on facial recognition. Automatic, and controlled, processing were induced using a computer-based Stroop task. Participants (*n* = 288) processed the font colour, or word reading conditions of the Stroop task; afterwards, their facial recognition was tested. Even though the Stroop effect did occur, the differences in facial recognition between participants who responded to the automatic condition and participants who responded to the controlled condition of the Stroop task did not differ significantly. The results suggest that the processing shift from configural to featural processing, which is induced by the Navon letters, would better explain the verbal overshadowing effect than the processing modes underlying the Stroop task.

Keywords: automatic processing; controlled processing; facial processing; Stroop task; transfer-inappropriate processing shift; verbal overshadowing effect

An eyewitness contributes to a police investigation in two important ways: (1) their ability to accurately describe a perpetrator assists the police with finding a suspect, and (2) when viewing the line-up, their identification of the suspect as the perpetrator, based on their memory of them, will lead to a trial. These two stages are integral to each other: without a good description, the police have no basis for finding the suspect; however, if the witness does not identify the suspect as the perpetrator, the suspect will walk free. Both of these stages rely on the eyewitness being able to rely on their memory of the perpetrator when writing a description and making an identification. However, research suggests that by providing a verbal description of the perpetrator, eyewitnesses are at risk of having impaired recognition of that same perpetrator. Put differently, description of the perpetrator, which is a prerequisite for finding a suspect, is produced at a cost: The eyewitness may have such impaired recognition of that perpetrator that if he is present in the line-up, the eyewitness may not accurately identify him – this is known as the verbal overshadowing effect (VOE) (Schooler & Engstler-Schooler, 1990).

Schooler (2002) proposes that this phenomenon arises from a transfer-inappropriate processing shift (TIPS): The verbal description, which normally emphasises the features of the perpetrator's face, induces a featural processing mode which is not conducive to facial recognition (Tanaka & Farah, 1993). While faces are processed both configurally and featurally, the former type – which emphasises the configuration of features and their spatial relations to one another (Rakover, 2002) – benefits facial processing the most (for example, Bartlett & Searcy, 1993; Sergent, 1984; Tanaka & Farah, 1993; Yin, 1969; Young, Hellawell, & Hay, 1987). This cognitive shift from configural processing to featural processing prevents optimal facial processing, thus impairing the eyewitness's ability to identify the perpetrator accurately.

The theory proposed by Schooler (2002) is a revision of the transfer appropriate processing shift theory (Morris, Bransford, & Franks, 1977). Morris and colleagues (1977) suggested that memory performance was better when information was processed in the same manner at encoding and retrieval – regardless of the depth of encoding as emphasised by the levels of processing theory (Craik & Tulving, 1975). In their research specifically, retrieval was better when participants encoded and retrieved words using the same type of cues (rhythmic or semantic) at both stages, than when using a different type of cue at retrieval than at encoding. Even though semantic cues required a deeper level of processing than rhythmic cues do, their results clearly demonstrated that retrieval was facilitated when the studied material was processed in the same code at both encoding and retrieval (Morris et al., 1977).

Similarly, since faces are better processed configurally, facial recognition will benefit from a previous task which induces such a configural processing mode, unlike the verbal description which induces featural processing. This implies that the nature of the task, rather than the verbal content of the task, causes the VOE. Therefore, this effect is not caused by verbal description tasks only, because *any* task that encourages a featural processing mode should impair recognition.

Macrae and Lewis (2002) demonstrated this by replacing the verbal description task with Navon letters (Navon, 1977): a set of stimuli consisting of a larger letter that is recreated out of smaller, strategically placed lowercase letters, which normally differ from the gestalt that they form (for example, a capital 'B' that is created out of lowercase 'm's). After watching the same stimuli video used by Schooler and Engstler-Schooler (1990), their participants were divided into three groups: featural, procedural, and control. Both the featural and the procedural groups viewed the same set of Navon letters, but their responses differed. The featural group named the lowercase letters in the Navon stimuli, whereas the procedural group named the gestalt, which these lowercase letters formed. The authors hypothesised that participants who completed the featural-processing task would perform more poorly than the participants in the procedural group who responded to the gestalt. The control group completed an unrelated task. The results supported their hypothesis: The procedural group performed best at the face recognition task and achieved 83% accuracy, while the control and featural groups had 60% and 30% accuracy, respectively. These results suggested that when the task, which preceded facial recognition, was processed in the same manner as faces were, facial recognition would be facilitated. This effect has been replicated: Brand (2004) was able to replicate these results in a web-based experiment. Perfect, Weston, Dennis, and Snell (2008) argue, however, that the Navon letters are not process-pure, since the Navon gestalt is always perceived to some degree – participants may have processed it unintentionally, even if they were required to process only the lowercase letters. With this in mind, Perfect et al. (2008) utilised two different sets of Navon letters: in one set, the lowercase letters were placed further from each other, thus making it difficult to process the gestalt, and in the other set, the standard Navon letters were used. The authors argued that this manipulation resulted in the former set of stimuli to be locally-biased: Since the lowercase letters are placed further apart, it is easier to process them individually, but more difficult to perceive the gestalt that they form. The standard Navon letters were treated as globally-biased, as it is easier to perceive the gestalt than the lowercase letters. It was hypothesised that when participants processed either of these sets, then, depending on the processing bias, that particular processing mode would be induced – and this should impair or benefit facial recognition respectively. Two experiments were reported: In the first experiment, participants responded to only one set of the adapted Navon letters, followed by a standard recognition task of an upright face; in the second experiment, the upright face was replaced with an inverted face. (Processing an inverted face disrupts configural processing (Tanaka & Farah, 1993; Yin, 1969), therefore this recognition task should benefit from featural processing.) In both experiments, participants were divided into four groups that differed according to which set of Navon letters were viewed, and the required response: They processed either the locally-biased or the globally-biased Navon letters, and their responses were either the local letter or the gestalt that was formed. For the first experiment, the authors expected participants who had answered with the gestalt letter to be faster and more accurate than participants who had responded to the local letter; for the second experiment, a reversal in performance was expected with locally-responding participants performing better and faster at recognising the inverted face. Results from the first experiment were surprising: Participants' recognition accuracy improved after matching stimuli bias and stimuli response conditions (i.e., global bias/global response, local bias/local response). Results from the second experiment showed that despite ceiling effects for recognition accuracy, participants were faster at recognising the inverted face following incongruent conditions, such as global bias/local response, local bias/global response. Within the transfer-inappropriate processing shift framework these results were unexpected, because participants should have benefited only from responses which were processed in the same mode that the face was, i.e. the global processing conditions. Yet they performed well when responding globally to a globally-biased letter, *and* when responding locally to a locally-biased letter.

To explain their results, Perfect et al. (2008) proposed a revision to the transfer-inappropriate processing shift theory, specifically, that a shift from automatic to controlled processing could better account for these results. (Schooler [2002] had also suggested that some results seen in earlier research on verbal overshadowing could be explained by an automatic/controlled processing shift, but at that stage, no research had explicitly tested this theory.) There is some justification for this argument: Novices tend to approach new tasks in a controlled manner, whereas experts will process such tasks automatically (Shiffrin & Schneider, 1977; for example, Beilock, Carr, MacMahon, & Starkes, 2002; Beilock, Bertenthal, McCoy, & Carr, 2004; Flegal & Anderson, 2008). As experts in facial processing, humans process faces automatically: fast, parallel, and unconscious (Shiffrin & Schneider, 1977) and facial recognition should be facilitated if the preceding task has the same, automatic nature. This was demonstrated in Perfect et al.'s (2008) first experiment: Participants were better at facial recognition when they responded to Navon letters in the automatic condition – i.e. when the response and the bias matched – whereas in the second experiment, participants were faster at recognising an inverted face (which does not benefit from configural processing) after inhibiting the automatic response to the Navon letters. Schooler (2002) had suggested this theory as an alternative explanation for the VOE a few years prior, because previous research had demonstrated that this effect disappeared if participants were encouraged to make identification decisions within five seconds (Schooler & Engstler-Schooler, 1990); was attenuated after repeated trials (Schooler & Engstler-Schooler, 1990); and was only apparent when participants had to recognise faces of the same ethnicity as their own, and not when participants had to recognise faces from a different ethnicity (Fallshore & Schooler, 1995). The first two examples may represent the following two characteristics of automaticity: fast decisions, and controlled tasks becoming automatic through repeated practise (Shiffrin & Schneider, 1977). The last example may demonstrate how we process faces of people, who are from a different ethnic group to ours, in a controlled manner. This controlled processing may be a result of our lack of expertise with such faces. If automaticity develops from expertise and practise, and we are experts at processing faces who belong to the same group as we do, then we should process such faces automatically; in contrast, our lack of expertise with other-ethnicity faces may result in those faces being processed in a controlled manner. However, through increased contact and exposure (which is similar to how expertise and automatic processing develops), other-ethnicity faces are processed more holistically (Michel, Caldara, & Rossion, 2006), and this may increase their vulnerability to a verbal description.

A few studies have subsequently investigated the effect of Navon letters on facial recognition. Processing the global condition of the Navon letters encouraged participants to process a subsequent face composite holistically (Gao, Flevaris, Robertson, & Bentin, 2011). Processing the Navon letters before encoding a face and before recognising a face affected recognition significantly (Lewis, Mills, Hills, & Weston, 2009). Participants had higher recognition accuracy when the type of Navon responses required before both stages were congruent (i.e. local responses were required at both stages) than when different Navon responses were required, thus supporting the transfer-appropriate processing shift theory.

Brand (2004) investigated the claim made by Perfect et al. (2008) by replacing the Navon letters with the Stroop task, since the Navon letters were influenced by two sets of processes: featural/global processes, and automatic/controlled processes. Without teasing apart these two sets of processes, it is unclear which is influencing the facial recognition accuracies and latencies. 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). During the standard administration of this task, colour words (for example, blue) are written in a different ink colour (for example, 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 difference in response latencies is the Stroop Effect. This effect is very reliable (Lezak et al., 2004; see MacLeod, 1991 and Macleod, 2005 for a review). Initially, it was widely accepted that the Stroop task demonstrated that certain tasks 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.

Brand utilised the same encoding and recognition stimuli that Finger and Pezdek (1999) had, and ran a web- and laboratory-based experiment. Since his experiment was run on computers, the administration of the Stroop task was adapted: Participants read one word at a time on a computer screen, and had an unlimited amount of time to type out their responses. Typing the font colour was considered a controlled processing task, whereas reading and typing the colour word was considered an automatic processing task. However, requiring participants to type out their responses may have lengthened the time taken to perform the Stroop task, and allowed for more time to process and inhibit the controlled processing condition of this task and correct their responses – thus preventing the Stroop effect from occurring. His results demonstrated that the Stroop condition did not significantly affect facial recognition or response latency, but without reporting the response latencies and accuracies for the Stroop task, it is unclear whether the experimental manipulation even occurred. Additionally, demographic information was not reported about the participants, so it is unclear whether they were of the same-ethnicity as the face which they were required to recognise. It is doubtful whether the experimental manipulation of the Stroop occurred, and it is unclear how to interpret these non-significant differences in facial recognition scores between groups. It could be that a shift from automatic to controlled processing may not affect facial processing, which explains why there were no differences between the groups; however, it is unclear whether that shift in processing even occurred.

Brand (2004) was right to use a Stroop task to investigate whether automatic or controlled processing would affect facial recognition. However, his method was incorrect. The response given to the Stroop should have been a single key press, instead of a lengthy typed response which diminished the Stroop effect. By typing out a response to the Stroop, participants had more time to inhibit their automatic responses and their reaction times would have increased, thus preventing the Stroop effect from occurring. Although Brand (2004) did randomise the order of the Stroop items, he did not control for the consecutive presentation of the same stimuli, thus facilitating the Stroop effect in some conditions, but not in others. Lastly, his participants had to complete the Stroop for 5 minutes, but he didn't report the number of Stroop items that they processed – so it is possible that his participants did not process the same number of Stroop items, thus either strengthening the processing shift, or weakening it.

The aim of this experiment was to expand the hypothesis presented in Brand (2004) and to investigate whether a task, which induces automatic or controlled processing, would cause results that resemble those of the verbal overshadowing effect. Facial recognition should be facilitated following a task that encouraged the same processing mode that is used to process faces. People are better at recognising faces from the same group as their own, and have developed an expertise for this. Automaticity arises from expertise. Therefore, it was hypothesised that recognition for in-group[1] faces would be better following an automatic-processing task than a controlled-processing task. To test the beneficial effects of a controlled-processing task, an out-group face was used instead of an inverted face.

METHOD

Design and setting

This experiment had a split-plot design with two between-subject factors: encoded face group (in-group, out-group), and processing type (automatic, controlled).

The experiment consisted of three trials, and each trial consisted of three stages: encoding, interval trial, and recognition. The within-subject factor was the encoded face (of which there were three in total), and the order that the participants saw the three faces was counterbalanced. The dependent variables were recognition accuracy, a total score ranging from 0–3 calculated by summing participants' line-up responses, and reaction times which were measured in milliseconds (ms).

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

Participants ($n = 288$) were recruited from the undergraduate student population (average age = 20.27 years) at the University of Cape Town. Half of the participants were white South Africans, and from the same group as the targets used in this experiment; the remaining participants belonged to the out-group. It was hypothesised that the in-group and the out-group differed in their group expertise, which was a reflection of own-group familiarity. The in-group had higher expertise with faces from their own group, and therefore, should process such faces automatically, whereas the out-group, due to their lack of expertise, should process them in a controlled manner.

Materials

Encoding stimuli. Three male targets were selected, and each acted in a film which was used as the encoding stimulus. In each film, a target steals money out of a bag, which he found in an empty classroom. The target's face and upper body were visible in the film, so participants could view each target's gait and body type/build. The sound was removed from the video. A pilot study was conducted to investigate whether the length of the films resulted in ceiling effects for facial recognitions. The videos were reduced from 48 seconds to 10 seconds, and in this latter condition, on average, the three targets were correctly identified by 64.1% of participants.

Recognition stimuli. A simultaneous line-up, with two different orders, was created for each target. These line-ups, which were the same size as those used by Perfect et al. (2008), contained seven foils and one target, and were constructed according to the guidelines by Malpass, Tredoux, and McQuiston-Surrett (2007). A modal description was created for each target, and six mock witnesses had to search through an electronic database of photographs of different faces and find suitable foils who matched each physical description best. Line-ups were built for each target and contained their seven foils who were most frequently chosen as matches. All the images were standardised for inter-ocular distance, and were in a frontal position. The images in the database had been photographed with different cameras in different locations and lighting, resulting in artefacts. Participants could have used these artefacts to try and figure out who the suspect could be, rather than remember who the suspect was. Therefore, the background in the photographs was replaced with a grey texture, a black cloak was drawn over their clothing, and an artificial shadow was added between each face and the background.

Participants were recruited as mock witnesses, and were instructed to choose the target from each line-up, with or without the modal description. This procedure allowed for line-up effective size and bias to be measured. Line-up bias measures how much the line-up is biased towards or away from the target, and should not differ from chance (1/number of line-up members; 1/8 = 12.5%). Line-up effective size is a measure of how many line-up members are considered appropriate alternative suspects for the target. Even though a line-up may comprise eight members, only three of those are considered alternatives to the suspect. This reduction in line-up effective size increases the risk of misidentification.

This mock witness procedure was repeated a number of times, and infrequently chosen foils were replaced with more suitable foils until line-up bias and line-up effective size were within an acceptable range.

Transfer inappropriate processing task: Stroop task. Both the Navon letters and the Stroop task require participants to demonstrate cognitive inhibition when processing the non-dominant perceptual dimension of either task (i.e. inhibit processing the gestalt letter or reading the word). However, unlike the Navon letters, which are governed by two sets of processes (automatic/controlled, featural/global), the Stroop task is governed by only one (automatic/controlled). Additionally, the Stroop task was suitable for this experiment, because an interval task was required which would force 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).

Using the guidelines suggested by MacLeod (2005), five distinguishable colours (red, blue, green, magenta, and yellow) were chosen from those available in E-Prime 1.1, and each colour was assigned a number key in the keyboard numpad. A small, white sticker, with the name of the colour written on it in black ink, was stuck on its respective key, so that participants did not have to remember which key was assigned to which colour.

The Stroop task consisted of two sets of Stroop stimuli, and each set contained 100 Stroop items. Each set differed according to the processing condition they would be used for. In the automatic condition, participants had to read the word colour written on the screen and press the key which matched that word. Each word was presented in its matching font colour, thus facilitating the ease and automaticity of the task. These 100 Stroop items were presented in a random order. In the controlled condition, participants had to press the key which corresponded to the font colour that the word was written in. Each word was written in a different font colour than the word itself, and each Stroop item was randomly presented, but neither font colour nor word was repeated consecutively.

Each Stroop item was presented on the screen until the participant responded with a keyboard press. This was followed by a black fixation point (*) and a blank screen, each presented for 250 ms. The Stroop condition was complete once the participant had responded to all 200 items

Procedure

Participants were run in groups, and each participant was assigned to a computer. The experiment was built using E.Prime 1.1., psychological software used for building, running and analysing psychological experiments (http://www.pstnet.com/eprime.cfm). All text was displayed in font type Arial with a font size of 18, and screen resolutions were set to 1024 × 768 pixels. The font colour was black (except for the Stroop conditions), and the background colour was always silver.

Participants began the experiment by watching the encoding video. Afterwards, they went through a familiarisation process with the Stroop task, and then completed both sets of the Stroop stimuli – the last set of stimuli determined which processing mode would affect facial recognition. Once they had responded to all 200 Stroop stimuli, participants were presented with a line-up that contained the target who acted in the previous video; participants were asked to identify whom they had seen in the video, but were warned that the target may or may not be present in the line-up. Participants made their decision by pressing the number, which was assigned to that line-up member, on the keyboard; if the participant thought the target was not present, they could press '0'.

After participants had made their decision, they were presented with a forced-choice condition, where they were forced to choose who the target was and could no longer indicate that the target was not present. After making their decision, they were asked a number of questions about their decision making.

This procedure was repeated three times, until each participant had viewed each video.

Ethical approval was granted from the Ethics Committee, and research participation commenced only after participants had read and signed the consent form. The consent form stated that the findings from this research would be reported in a journal article; however, all participants' data would be recoded with a randomised participant number, and would therefore be anonymous. Participants knew that they could leave the experiment if they felt uncomfortable or distressed. Once participation was completed, participants were debriefed and thanked.

RESULTS

Stroop accuracy and Stroop latency

The experimental manipulation was a Stroop task with two different conditions: automatic and controlled. Normally, a Stroop task would be administered one word at a time, and participants would be required to respond verbally. However, in this experiment a computer-based Stroop task was used. There is some risk with this adaptation, since research has shown that the size of the Stroop effect is not as large when induced through a computer-version of this task, most likely because keypresses are slower than verbal responses (Brand, 2004; MacLeod, 2005).

The Stroop accuracy data and response latency data were analysed with a paired-sample *t* test. The results show that the experimental manipulation was successful, and participants were significantly faster at the automatic condition of the Stroop (861.31 ms versus 1180.97 ms), $t(287) = -28.334$, $p < 0.001$, $r = .86$. Additionally, participants were more accurate at the Stroop items in the automatic condition than the controlled condition (97.66 % versus 88.89%), $t(287) = 7.302$, $p < 0.001$, $r = .40$

Combined recognition accuracies of all targets: Initial Choice and Forced Choice

The 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, ranging from 0 to 3, was calculated for each participant by summing their accuracy scores across all perpetrators. An average accuracy score was calculated for the Initial Choice condition and the Forced Choice condition and the data were analysed with an ANOVA.

From the descriptive statistics, it appeared that in-group participants performed better than the out-group participants (2.02 versus 1.84); however, these differences only approximated significance, $F(1,284) = 3.484$, $p = 0.063$, $\omega^2 = .0086$ (Table 1) There were no significant differences between processing type groups, nor was there a significant interaction between participant group and processing type.

In the forced choice condition, participants were instructed to either choose a face if they selected "not present" previously, or if they had chosen a face, they had to choose that same face. Two outcomes were expected: recognition would either remain the same or it would improve. Overall, recognition accuracy performance improved in the forced-choice condition, $t(287) = -9.670$, $p < 0.001$, $r = 0.497$ and in-group participants still performed better than the out-group participants (2.31 versus 2.18). Participant group and processing type did not interact significantly, and processing type did not significantly affect recognition accuracy.

Reaction times

Perfect et al. (2008) found significant differences between their participants, specifically, that their participants were significantly faster following the automatic condition of the Navon letters – suggesting that facial recognition was easier after the automatic condition.

In this experiment, participants did not differ significantly in their reaction times, and no significant main effects or interaction effects were present (Table 1).

Table 1. ANOVA results for reaction times following line-up decisions for the initial presentation of the line-ups

Initial choice	df	*F*	η	*p*
Processing type	1	.376	.001	.540
Participant group	1	1.541	.005	.215
Processing type × participant group	1	1.373	.005	.242
Error	284			

Reaction times re-analysed

When examining the recognition rates for each target, it was apparent that Target 3 was the most easily recognised, followed by Target 2 and then Target 1 (recognition accuracy was 86.1%, 56.3%, and 50.0%, respectively). Since it was easiest to recognise Target 3, the response latencies associated with his line-up would be faster, thus lowering the average response latency for each participant. The recognition rate and response latency for Target 3 was removed, and the data were re-analysed using an ANOVA. No significant differences were observed for the recognition accuracy. However, the response latencies did differ from those reported previously.

With Target 3's data removed, the average reaction time slowed, but there was still no significant main effect for participant group or processing type. The interaction between these two variables was marginally significant, $F(1,284) = 3.321, p = .069, \omega^2 < .001$. Surprisingly, it was in the opposite direction predicted: in-group participants performed faster after the controlled processing task, while out-group participants performed faster after the automatic processing task (Figure 1). This difference in response latencies did not come at a cost to accuracy, since in-group participants were not more accurate at facial recognition than out-group participants were.

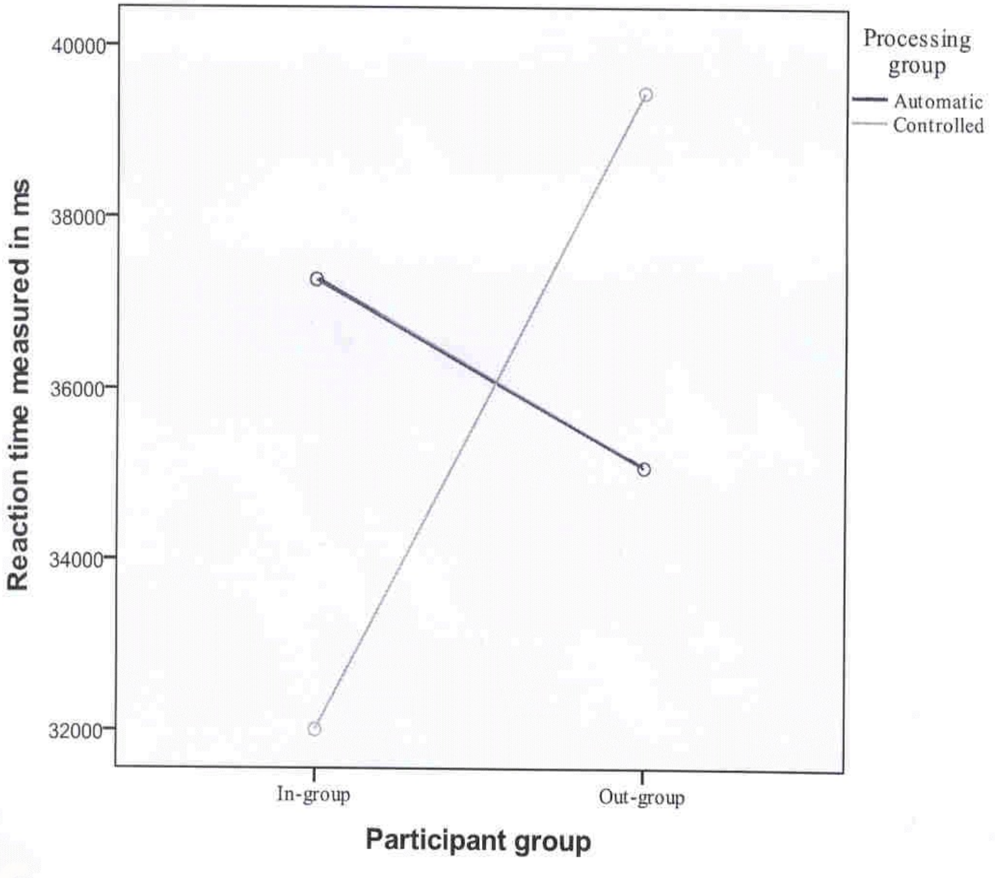

Figure 1. The interaction between participant group and processing group on time taken to make a decision. The interaction between participant group and processing group was nearing significant, but it was in the opposite direction predicted. In-group participants were faster at making line-up decisions after the controlled-processing tasks, whereas out-group participants were faster after the automatic processing task.

DISCUSSION

Overall, in this experiment, no evidence was found to support the hypothesis that automatic processing would benefit in-group participants' facial recognition, nor that out-group participants benefit from controlled-processing tasks. These non-significant results are similar to those presented by Brand (2004); however the manipulation of the Stroop task was effective in this experiment. Therefore, despite the experimental manipulation occurring, this did not affect participants' recognition rates, and it appears that the Stroop task did not produce similar results to those seen with the Navon letters. This suggests that the processing shift between automatic/controlled processing does not affect facial recognition, and that the results seen in experiments which used the Navon letters are better explained by a shift in global/local processing.

Brand (2004) was able to replicate the Navon letter effect on facial recognition with a web-based experiment, but he was unable to do so with a laboratory-based experiment, and he argued that this was a result of low statistical power in the second experiment. Unfortunately, the set of Navon stimuli used in the three different experiments (Brand, 2004; Macrae & Lewis, 2002; Perfect et al., 2008) differ from one another, and it is possible that this could account for the differing results found in the literature. Overall, from other studies, it appears that accuracy is higher following the global response to the globally-biased Navon letter, which is the standard Navon letter, than following the local response.

Based on the literature reviewed for this experiment, an important feature that linked the Stroop task and the Navon letters was the inhibition required by participants to suppress their response to the dominant perceptual domain of a stimulus in favour of the non-dominant domain. However, it is possible that the amount of inhibition required by both tasks is not the same: 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 same automatic quality. Unfortunately, Perfect and colleagues (2008) only presented response latency data for the second experiment, and therefore it is unclear how much inhibition was required of participants to allow them to respond against the automatic/dominant perceptual domain.

It is unclear why Perfect et al. (2008) obtained the results that they did. The global-processing group in Perfect's experiment performed worse than the global-processing group in Macrae and Lewis's (2002) study (0.65 and 0.83 accuracy, respectively), despite using the same encoding face and line-up. The only differences between these two experiments were the length of time spent responding to the Navon letters (participants in Perfect's study responded for 5 minutes, compared to 10 minutes for Macrae and Lewis's study, respectively), and possibly, the size of the stimuli. The size differed between the two sets of Navon letter stimuli that Perfect and his colleagues (2008) used: The globally-biased stimuli were printed on smaller sheets, and therefore the required attentional focus was narrower. This should have heightened the effect of the distracting elements of the stimuli (Chen, 2003), thus increasing the inhibition effort needed by participants who had to respond to the non-dominant domain. Unfortunately, this theory does not account for all the results presented by Perfect et al. (2008).

Limitations and future research

There are some limitations to this experiment. For example, including a target-absent line-up could yield interesting results and provide information about participants' decision-making. It has been suggested that the verbal overshadowing effect increases criterion response, so participants adopt a much stricter decision making strategy (Clare & Lewandowsky, 2004).This means that participants are more likely to reject the line-up, which is interpreted as verbal overshadowing, but with target-absent line-ups, that would be the correct decision. It would be interesting to compare accuracy rates between these two types of line-ups to see whether they would differ as a result of the Stroop task. With this in mind, future research in this area should include target-present and target-absent line-ups, as well as a forced-choice decision. This would then determine whether participants were incorrect, because they had a stricter criterion, or because they did not know who the perpetrator was.

While not significant, the in-group and out-group differed in their response latencies, and this difference was in the opposite direction hypothesis. In-group participants were faster following the controlled-processing task, whereas out-group participants were faster following the automatic-processing task. At this stage, it is unclear why this is. The other-group participants may have differed in their proficiency in English, and found the automatic condition so much easier than the controlled condition, that this inflated the difference in their response latencies. Future research should include target faces from both in-groups and out-groups in order to truly determine whether a cross-effect is present; if this pattern in response latencies is reversed, then this would suggest that it arose from a cross-group/in-group interaction, and not a procedural anomaly.

The aim of this experiment was to investigate what effect processing the Stroop task would have on facial recognition, specifically within the framework of automatic/controlled processing. Most psychologists are familiar with the Stroop task, and with theory surrounding automaticity; however, measuring and researching automaticity/controlled processing has proved to be a difficult task. Although, most psychologists can provide examples of automatic behaviour, it is not clear how to operationalize this concept when researching it (for example, read the suggestions given by Moors and De Houwer, 2006). It is contentious whether the Stroop task does induce automatic or controlled processing, and perhaps the videos were not processed in an automatic/controlled mode. Perhaps one way, of bettering this design, would be to replicate the procedure used by Lewis et al. (2009) and to include a set of Stroop tasks prior to encoding and prior to recognition to investigate whether recognition is best if the same processing mode is induced at both of these stages. However, this still does not address the problem of measuring this cognitive process.

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. It is possible that the levels of inhibition required by these two tasks are so different that they are not comparable, and perhaps what is more important, is the type of processing modes that the tasks induce. The Stroop task may not affect facial recognition in the same manner that the Navon letters do, because a shift from automatic to controlled processing is not as deleterious as a shift from procedural to featural processing is.

ACKNOWLEDGEMENT

This research was funded by grants from the National Research Foundation.

NOTE

1 While the terms 'in-group' and 'out-group' are traditionally used when addressing intergroup relations in Social Psychology, such as prejudice, these terms are also now commonly used in eyewitness and face recognition research to describe the recognition advantage that members of one group tend to show over others. These groups could be based on age, race, gender, ethnicity, or nationality, among other characteristics. See Sporer (2001) for a more detailed motivation and explanation regarding this usage.

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, direction 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.

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

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

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.

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

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

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.

Gao, Z., Flevaris, A. V., Robertson, L. C., & Bentin, S. (2011). Priming global and local processing on composite faces: Revisiting the processing-bias effect on face perception. *Attention, Perception & Psychophysics, 73*, 1477-1486.

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 ed.). New York, NY: Oxford University Press, Inc.

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). Washington, DC: American Psychological Association.

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

Malpass, R. S., Tredoux, C. G., & McQuiston-Surrett, D. (2007). 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.

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. *Psychological 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.

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.

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

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.

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

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. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, & Law, 7*, 36-97.

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

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

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.