You will do better if I watch: Anonymity, identifiability and audience effects in a stereotype threat situation

Jared Daryn Forbes

Submitted in partial fulfilment of the requirements of the degree of Master of Social Science (Research Psychology) in the School of Psychology, University of KwaZulu-Natal, Pietermaritzburg.

2009

As the candidate's Supervisor I have approved this thesis for submission

DECLARATION

I, Jared Daryn Forbes declare that

(i)   The research reported in this dissertation, except where otherwise indicated, is my original work.

(ii)   This dissertation/thesis has not been submitted for any degree or examination at any other university

(iii)   This dissertation/thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

(iv)   This dissertation/thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

(a)   their words have been re-written but the general information attributed to them has been referenced;

(b)   where their exact words have been used, their writing has been placed inside quotations marks, and referenced.

(v)   Where I have reproduced a publication of which I am an author, co-author or editor, I have indicated in detail which part of the publication was actually written by myself alone and have fully referenced such publications

(vi)   This dissertation/thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

Signed: ……………………

Acknowledgements

I would like to thank my supervisor, Mike Quayle for all his help and support on the project. His assistance in both my honours year as well as these last two years while completing this project has been invaluable. He has always been there and willing to help and guide me.

I would also like to thank my participants for agreeing to take part in the study, without them I would have little to write.

Table of Contents

Abstract

The current study examined stereotype threat or lift (STL) in terms of various elements of social identity theory. STL occurs when a negative stereotype (or positive stereotype) about a group leads to a decrease (or increase) in performance on a task that the group identifies with. The primary focus was the relationship between STL and identifiability, whereby identifiability refers to whether one views one's self as an individual or as an anonymous part of a social group. The study examined STL in relation to humanities and science students' ability to recognise patterns using two short forms of the Raven's Advanced Progressive Matrices (APM) which was developed. The students completed matrices under two conditions; anonymity and visibility to an audience (in-group, out-group and experimenter). When visible, participants performed significantly better than when anonymous, regardless of the STL condition. When examining in-group identification, participants with high in-group identification experienced traditional STL effects while participants with low in-group identification experienced a reversal in effects.

Introduction

Stereotype threat or lift (STL) occurs when a negative (or positive) stereotype about a group, leads to a decrease (or increase) in performance on a task that the group values (Steele, 2003). It can affect any individual who is a member of a relevant group that is stereotyped with respect to a task or task-relevant domain. For example, previous research has shown that this effect negatively impacts on the academic performance of black students and positively affects the academic performance of white and Asian students when group stereotypes become relevant in academic domains (Aronson, Lustina, Good, Keough & Steele, 1999). The effect has also been shown to negatively affect a wide variety of groups in a wide variety of contexts, including white sports players, the aged, women in mathematics and science, affirmative action candidates and many others (see Walton & Cohen, 2003). However, explanations for the effect are so far not entirely convincing.

Although Steele and his colleagues (Steele, Spencer & Aronson, 2002 ; Aronson *et al.*, 1999) initially focused on the negative effects on task performance of negative stereotypes, Walton and Cohen's (2003) meta-analysis also revealed the phenomenon of stereotype lift - whereby groups that were positively stereotyped did better on a test of intellect when an out-group was negatively stereotyped. Stereotype threat and lift have both been established as an empirical effect (Walton & Cohen, 2003) and seem to work hand in hand in the given context, whereby the stereotype threat of one group causes the stereotype lift in the other group. Therefore the effect will be referred to as stereotype threat or lift (STL).

Recently it has been suggested that one can understand stereotype threat as a social identity theory phenomenon (Haslam, Salvatore, Kessler & Reicher, 2008). Social identity theory suggests that individuals evaluate themselves in terms of the groups they belong to and how positively these groups are represented (Oakes, Haslam & Turner, 1994). effect. Viewing stereotype threat through the lens of social identity theory may provide new avenues for understanding which individuals are vulnerable to stereotype threat as well as how much of an impact stereotype threat has on individuals in group situations.

The current study is part of an integrated program of research investigating several aspects of social identity theory in relation to stereotype threat. The main focus of the current study is identifiability. Identifiability refers to whether or not the individual is seen as an anonymous part of their group or whether they perceive themselves to be individually identifiable. However, other aspects of social identity theory were included for comparison across studies these are: intergroup permeability, intergroup conflict, group legitimacy, and group stability. Recent studies have shown identifiability to be highly relevant to identity processes, and thus potentially relevant to stereotype threat. For example a study by Postmes and Spears (2002) found that the activation of the stereotype led to gender-stereotypic behaviour only when the members of the groups were anonymous.

Two exploratory studies within the present programme of research induced stereotype threat in participants in conditions of high or low identifiability. The researchers found that under stereotype threat conditions the participants that have high identifiability perform better on a given task than participants with low identifiability. However, when stereotype threat is not activated the participants with low identifiability perform better on a given task than participants with high identifiability (Forbes, 2007). In the second study participants in the high identifiability stereotype lift condition performed best and participants in the low identifiability stereotype lift condition performed worst (Quayle & Reicher, 2007). Therefore, it is probable identifiability is an important component of STL. The current study further explored the effect of identifiability within a stereotype threat context.

Literature Review

Stereotype threat or lift

Stereotype threat or lift (STL) occurs when a negative (or positive) stereotype about a domain the individual or group identifies highly with is made relevant to performance (Steele, 1997; Walton & Cohen, 2003) and modifies task performance for members of stereotyped groups.

STL originated with Claude Steele at Stanford University, as a social-psychological explanation for the poor performance of African-American students at University. This research started in 1995, but since then the body of work on stereotype threat has grown immensely. The context in which STL has been detected also varies. STL has been demonstrated to affect various groups including, African-American students, other students from minority groups, white athletes, and women in mathematics and engineering (see Walton & Cohen, 2003).

Since STL is a 'situational modifier' or a 'threat in the air' according to Steele (1997) one cannot say with certainty in a given situation that any performance is influenced by STL. STL cannot be directly measured through interviews or survey methods. Therefore STL can only be inferred within an experimental context in which conditions are created which meet the criteria for STL, and task performance is measured and compared across threat and lift conditions. STL is inferred if the performance of the participant matches the nature of the group stereotype activated at the time they complete the task. For example, if the participants' social group is viewed negatively and their performance decreases compared to participants whose social group is viewed positively then one could infer that this was due to STL, so long as confounding variables are accounted for.

In order to create an STL condition there must be the following factors: firstly, the individual or group must have domain identification, which is the extent to which the domain is viewed as important to the identity of the group or individual (Steele, 1997; Smith & White, 2001); secondly, there must be self-definition, in other words, the task must be related to a defining feature of an individual's identity; and lastly, the stereotype must (a) be perceived as a generally known stereotype and (b) have self-relevance (Steele, 1997). Meta analyses (e.g. Walton and Cohen, 2003) demonstrate that STL is a well established empirical effect, d=.24 for lift and d=.29 for threat.

Importantly, the stereotype does not have to be *believed* to induce stereotype threat. It should just be in a domain in which the participants care about performing well during the given task. Therefore, stereotype threat may also be thought of in terms of performance expectancies in which either success or failure is determined as important by the participant. More specifically, that success or failure at a task would have a particular meaning or connotation (Hyde & Kling, 2001 cited in Smith, 2004). There has also been interest in stereotype threat from a motivational standpoint (Wheeler & Petty, 2001). Stereotype threat is also reliant on how others view individual performance in regards to the stereotypes around the groups. In this way if others expect you to perform in a certain way and, you do perform in that way, then this may lead to a self-fulfilling prophecy (Jussim *et al.*, cited in Smith, 2004).

It has been suggested that stereotype threat, is just a form of 'performance expectancy'. However, research into performance expectancy can change according to the participant and the duration of the study, that there may be different expectancies on different participants at different stages of the research whereas stereotype threat is concerned with the participants' value which they place on that task, as well as the outcome they expect from the task (Smith, 2004). It has been shown that the STL effect is most clear when the participant identifies highly with the domain of the stereotyped task (Aronson *et al.*, 1999; Steele, 1997; Smith & White, 2001 cited in Smith, 2004) or/and identifies highly with their social group. (Schmader, 2002 cited in Smith, 2004). Therefore STL is not as simple as performing as you expect yourself to perform, as would be expected if it were related to simple performance expectancies. In order to understand and study STL this can be best understood as indicated below.

"The basic paradigm for demonstrating stereotype-threat effects is such that group members are made aware (or not) of a domain-relevant group stereotype, and are then asked to take a (usually standardized) test in that domain" (Smith, 2004, p.181).

In order to induce the STL effect the stereotypes have been manipulated in numerous ways. The most common way in which the STL is represented is that the participants are presented with a stimulus or vignette invoking stereotypes that, on the task they are required to perform, the out-group outperforms the in-group or visa-versa (Aronson *et al.*, cited in Smith, 2004). The other way in which the stereotype can be manipulated is where the participant is simply

made aware of their stereotyped characteristic and made to report it, such as race or socio economic status (Croziet & Claire, 1998 cited in Smith, 2004).

The most frequently studied domains of performance within the STL literature are sex differences in mathematics and science performance, as well as the differences in general academic performance between white and black students (Walton & Cohen, 2003). Traditionally women would be told that they as women are not as good at maths as men, and in fact that men are superior in their mathematic ability. After this they would then be given a maths test and their results would be compared with the men's results. These results would also be compared with a control group where both the men and women would not have been made aware of the stereotype that men are better at mathematics than women. What has usually been found is that the experimental groups would perform in the stereotyped manner while the control groups would show little systematic differences in performance between the groups. The same sort of scenario has been repeated for white and black students and their academic ability.

Effort. Effort has been linked to potential explanations for STL. There are two such explanations for the role that effort may play in STL. The first explanation is that the participants react to the stereotypes almost like a learned helplessness situation. Because the participants are told through the stereotypes that they are worse at a particular task than another social group they simply will put less effort into the task than the participants who are told they are better at the task (Smith, 2004). A second explanation of effort is that the threatened group has an "I'll show them" experience and that there attempt at trying too hard leads to a decrease in performance on the task (Smith, 2004, p.183). Ways in which effort was tested have been numerous. Several studies have examined time taken on the task or number of items answered however there have been no significant findings from these (Aronson *et al*., 1999, Study 1; Leyens *et al*., 2000; Spencer *et al*., 1999, Study 2, cited in Smith, 2004). Other studies examined the number of items attempted in the task however again there were no significant findings (Ambady *et al*., 2001, Study 1, Study 2; Shih *et al*., 1999, Study 1, Steele and Aronson, 1995, Study 1, Study 2 cited in Smith, 2004). Self-reported effort has also been examined. Of the nine studies cited by Smith (2004) eight found no significant findings (Aronson *et al.,* 1999, Study 2; Gonzales *et al*., 2002; Keller, 2002; Keller and Dauenheimer, 2003; Smith and White, 2002, Study 1 and Study2; Steele and Aronson, 1995, Study 2; Stone *et al*., 1999, Study 1, Study2). Therefore although there are intuitively promising explanations for how effort may mediate the STL effect this has not been demonstrated in the research experiments.

Self-handicapping. "The concept of self-handicapping was described by Leary and Shepperd (1986) as an individual's tendency to protect the self by either actually behaving or simply claiming to have behaved in a way that would provide a believable external attribution or reason for failure on an important task" (Cited in Smith, 2004, p.184). It has been suggested that self-handicapping plays a role in the negative effects of STL. It is also then follows that participants that are threatened by the stereotype would then have more self-handicaps then those participants that are not threatened by STL. Studies using self-reported measures of self-handicapping reporting such measures as lack of sleep, lack of focus etc. have found no significant results of self-handicapping mediating STL (Croizet and Claire, 1998; Keller, and Dauenheimer, 2003; cited in Smith, 2004). However, a study by Keller (2002) demonstrated that reporting more self-handicaps did in fact mediate STL on the performance of the task (cited in Smith, 2004).

Anxiety. A further way in which STL effects are thought to be explained in task performance is that of anxiety. It is suggested that the STL creates a level of anxiety for the participants and in turn this disrupts there performance on the task (Smith, 2004). Early studies examining anxiety and STL found no effects these studies examined cognitive interference (Steele and Aronson 1995, Study 1; Gonzales *et al*., 2002; Keller and Dauenheimer, 2003; McKown and Weinstein, 2003 cited in Smith, 2004). Other studies using Spielberger *et al*.'s (1970) state-trait anxiety instrument (STAI) (cited in Smith, 2004) have also shown no significant results however Spencer et al. (1999, Study 3 cited in Smith, 2004) did find an experimental effect for anxiety but no mediator effects. In two studies Ford, Ferguson, Brooks and Hagadone (2004) examined the effect of humour in a stereotype threat situation. The authors were able to demonstrate in their second study that women under stereotype threat who had a coping sense of humour fared better on a maths test than those did not. There was also no difference in performance on the maths test in the control group with regards to a coping sense of humour (Ford *et al*., 2004). "Mediation analyses suggest that in the stereotype threat condition, state anxiety mediated the relationship between coping sense of humour and test performance. Women higher in coping sense of humour performed better because they felt less anxiety while taking the test". (Ford *et al*., 2004, p.643).

Osborne (2001), wanted to examine whether or not anxiety played a role in explaining the test performance differences of participants who differed in terms of gender and race. Osborne (2001) used a sample from a high school data base. From this a number of analyses were performed. What was found was that anxiety (self-reported measure) was significant in explaining the differences in performance on the test between whites and African Americans and whites and Latinos.

Osborne (2006), wanting to explain anxiety more holistically in terms of its relationship with STL sought to use some physiological measures when assessing anxiety. Males and female university students were sampled and placed into either a high or low stereotype threat conditions and given a maths test to complete. Osborne found that females who had been negatively stereotyped, in other words, high stereotype threat condition exhibited more signs of anxiety from the physiological measures (skin conductance, surface skin temperature and diagnostic blood pressure) (Osborne, 2006).

O'Brien and Crandall (2003) examined the effects of arousal on stereotype threat. The authors predicted that participants under stereotype threat would perform poorly on difficult tasks and that there performance would improve on easier tasks when they are compared with participants who are not faced with the stereotype. Within stereotype threat condition women performed better on the easy test compared to the control group. When examining the difficult test the women performed worse on the test compared to the control group. Within the control group for the women there was hardly any difference in their test scores for the easy and difficult tests. There were no significant differences for men when comparing the stereotype threat men to the men in the control group. The difference in test scores for between men and women in the stereotype threat condition was more than double than the difference between men and women in the control group. Other measures of anxiety have also not shown significant results (Smith, 2004). Therefore although on the face of it anxiety seems like a plausible mediator for the STL effects on performance, more research is needed in this regard.

Evaluation apprehension. Evaluation apprehension "negatively affects an individual's performance by creating concern for how other people are evaluating him or her and feelings of self-consciousness" (Smith, 2004, p.187). In order to test evaluation apprehension Spencer et al., (1999) created their own self-report measure, which was also used by O'Brien and Crandall (2003) however, there were no significant effects for this relationship (cited in Smith, 2004). Therefore there is no evidence as yet of evaluation apprehension as a mediator of STL effects.

Performance Confidence. STL effects of performance have also tried to be understood in terms of performance confidence. In this way it is suggested that STL may impact on the perceived confidence in a task and in this manner influence the behaviour of participants. Various operationalisations of performance confidence have been used (Steele & Aronson, 1995; Stone, 2002 cited in Smith, 2004). Other studies have used self-reported measures of performance once they have completed the task, however, these studies did not find significant results (Aronson *et al.*, 1999, Study 2; Keller, 2002; Kray *et al.*, 2001, Study 4; Shih *et al.*, 1999, Study 1 cited in Smith, 2004).

Test perceptions. Perceptions of the test have also been suggested as a mediator of STL effects on performance specifically, if the test is viewed as unfair towards a certain social group. Studies by Steele and Aronson (1995) as well as Stone (2002) (cited in Smith, 2004) have shown no significant results between how the test is viewed and if it mediates STL performance.

Self-directed emotion. Measures of emotion have also been introduced as possible mediators for the relationship between STL effects and on performance on a task. Self-esteem has been one such measure that is used. However, there were no significant findings in this relationship (Stone, 2002; Oswald & Harvey, 2001 cited in Smith, 2004). Depression was also used as a measure but again there were no significant findings for this relationship (Hausdorff *et al.*, 1999 cited in Smith, 2004).

Empirical evidence for mediators and moderators is which have been discussed above is mixed. "Indeed, a single explanatory process analysis might not only be incomplete but also account for why stereotype-threat mediation tests results are mixed. If the stereotype-threat–performance mediating chain of process variables is long and interconnected, single

mediation tests could be statistically unpowerful and thus difficult to document" (Judd and Kenny, 1981 cited in Smith, 2004).

Ways of reducing STL effects

Due to the negative effects of stereotype threat, particularly with regard to academic performance there has been increasing amounts of research to try and reduce the possible negative effects of STL.

Redefining the task-stereotype link. One such way to nullify or reduce the negative effects of STL is to redefine the link between the task and the stereotype. Showing that the stereotype is not relevant to the task is one way in which the negative effect can be reduced (Spencer *et al*., 2002, cited in Smith, 2004).

Increasing perceptions of malleability. A further way in which to reduce the negative effects is being able to view the stereotype as something that can be changed, for example educating students that they are always learning and therefore improving, that their abilities are not fixed. Another way is the use of role models who have been successful in spite of the stereotype (Aronson *et al*., 2002 cited in Smith, 2004).

Blurring category boundaries. The study by Rosenthal and Crisp (2006), examined if interventions to reduce intergroup bias can also be used in a stereotype threat context. The authors wanted to demonstrate whether or not blurring the boundaries between the groups would lead to a reduction in the level of stereotype threat experienced. Mediating effects of stereotype threat include anxiety, working memory capacity, and self-handicapping. Moderators include: emotional responses, changing perceptions of the situation, composition of the group, individual differences, identification with the group and stigma-consciousness (Rosenthal & Crisp, 2006). It is assumed that the participants need to be made aware of relevant differences between the in-group and the out-group, in order for the researcher to be able to successfully observe the effects of stereotype threat. Taking this a step further with regard to downward social comparisons, a person would not be able to know the perceived strengths and weaknesses of a group if they could not compare them to another group (Rosenthal & Crisp, 2006).

Within the study, the well tested stereotype of women in math was used. In order to blur the boundaries between the groups they asked participants, which were all female, to list characteristics which are common in both the in-group and out-group (Rosenthal & Crisp, 2006). Participants who listed the common characteristics between the groups scored better on the math test than participants in the baseline condition or the condition where differences were made apparent (Rosenthal & Crisp, 2006). Even in the final experiment when the participants were told their results were to be compared to men still performed better than the baseline and difference condition when they had to list common characteristics before the task (Rosenthal & Crisp, 2006).

Short-term responses to STL

Steele *et al.,* (2002) propose four short-term responses to STL. These are domain avoidance, self-handicapping, counter-stereotypic behaviour, and disengagement. Domain avoidance refers to the fact that the participants may just simply avoid the domain in which they have been stereotyped. The second response that of self-handicapping is creating reasons or explanations to explain the performance i.e. not enough sleep or not focused when taking the test (Steele *et al.,* 2002). Counter-stereotypic behaviour refers to when participants might try to disprove the stereotype in their performance thereby leading members of the social group to categorize them differently (Steele *et al.,* 2002). Trying to be categorised differently would lead you to identify less with your social group or a different social group, in this way counter-stereotypic behaviour links to permeability of groups. Permeability of groups refers to the extent that members of groups feel they are able to move between groups which is a feature of SIT which is discussed later. Lastly, disengagement refers to the participants decreased views of their ability with regard to their performance within a stereotyped task (Steele *et al.,* 2002).

Long-term responses to STL

The biggest concern for the negative affects of STL on participants is the of domain disidentification. Because the stereotype about the performance rests in a particular domain the participants may avoid the domain altogether (Steele *et al*., 2002). This may lead to less and less members of a particular social group within that domain therefore entrenching the "truth" of that stereotype further and further in that domain. Taking the example of women in maths, because of the negative stereotype about women's poor performance in maths, less and less women may try to take up math related subjects or careers. Leading on from this, the lack of women in these positions then only reflects the negative implications of the stereotype.

Criticisms of STL

Within the literature there are two main criticisms STL. "Shapiro and Neuberg (2007) suggested that stereotype threat actually means different things to different researchers because they often use stereotype threat as an umbrella concept without articulating the sub-concepts that define it" (Cited in Derks *et al*., 2008, p.167). The argument here is that many researchers are doing research which they term STL however, if it was inspected closer, and there concepts around what it is they were researching one would find it may not be related to STL.

The second criticism is that some researcher's argue that stereotype threat is just a case of priming. Priming is where a concept or idea is introduced before an experiment or task, in order to examine the impact of this prime on a person's thinking or behaviour. An example may be shown pictures of flowers before given a task, or shown a series of words relating to a particular personality trait or characteristic. Some experimental research on priming also uses stereotypes. Both priming and stereotype threat can lead to the same outcomes which is a decrease in performance. Thus it has been viewed as the same phenomena (Dijksterhuis & Bargh, 2001; Gladwell, 2005; Oswald & Harvey, 2000, cited in Stapel & Marx, 2006). Other researchers argue it is more than this, Marx, Brown and Steele (1999) cited in Marx and Stapel (2006) argue that there must be a situational pressure which is posed by the probability of being viewed or treated differently because of a negative stereotype aimed at one's group. This in turn may induce anxiety or concern around the consequences if they underperform

that is conforming to the stereotype. "Stereotype priming can affect anyone, whereas stereotype threat by definition, only occurs for those who are targeted by the relevant stereotype" (Marx & Stapel, 2006, p.244). Stereotype threat can be distinguished from priming as stereotype threat demonstrates a high level of concern for conforming to the negative stereotype. Stereotype threat also targets groups and thereby makes accessible a relevant social self (Marx & Stapel, 2006).

Limits of STL studies

STL studies have generally tried to identify possible mechanisms in order to explain what exactly is happening in the performance of a task when the social groups are stereotyped (Smith, 2004). Due to the nature of the experiments one has to control for a number of variables and tries to isolate a single variable within the study which could be used to account for the change in performance under stereotyping. However, no one variable being studied within STL has demonstrated that it may be a mediator of this effect (Smith, 2004). Smith (2004) states that the two most potential mediators from the research may be that of anxiety and performance confidence, but suggests the effects of other mediators related to STL are still unclear. It is surprising that although Smith, (2004) reviews Schmader's (2002) paper and notes the promising findings around gender identification but does not suggest that it or social identity may play a role in STL effects.

STL is a phenomenon experienced on a group level. Hence it is relevant as a function to group membership and is mediated through category group membership (Steele, Spencer & Aronson, 2002). A critical review of the stereotype threat literature leads (Haslam *et al*., 2008) to argue that stereotype threat cannot be viewed in individual cognitive terms. Viewing stereotype threat from a cognitive framework does not address the concept fully, it fails to address various aspects of group dynamics, intergroup relations and other possible explanations for STL therefore the cognitive view point is inadequate. Haslam *et al*. (2008) argues that stereotype threat can be better explained in terms of self and identity. As STL occurs on a group level, it has recently been argued that social identity theory may have conceptual resources and frameworks that could be fruitfully applied to the STL phenomenon (Haslam *et al*., 2008).

"Social identity theory (SIT) assumes that people are motivated to evaluate themselves positively and that insofar as group membership becomes significant to their self-definition they will be motivated to evaluate that group positively" (Oakes *et al*., 1994, p.82). Within SIT social categorization is important as a point of reference for the individual and their place in society, in order for the person to make sense as to how they 'fit in' in the world or their immediate space. "Categorization models of bias reduction do so by encouraging a weakening of the 'us' versus 'them' prerequisite distinction. One cannot positively favour 'us' as opposed to 'them' if we are not perceived as psychologically different from them" (Rosenthal & Crisp, 2006, p.503). It is argued that identity is derived from both individual and group resources. In other words identity is assumed to be comprised of elements of your social identity, the groups of people you associate with and the norms and values that you represent within those groups as well as your personal identity, which are characteristics that are unique to you as an individual (Oakes, Haslam & Turner, 1994). Tajfel suggests that identity is experienced on a continuum from personal identity to social identity but explains that in many circumstances individuals deal with each other as members of defined social categories (Tajfel, 1978). Tajfel argues that within social situations an individual's sense of group membership will increase when they have clarity of awareness, in other words, an

appreciation that they are part of a group; an understanding of positive and negative associations with membership; and an emotional investment in group membership (ibid).

Stereotype threat is viewed as quite mysterious but can be seen as one type of social identity threat (Derks *et al*., 2008). This is where a situation arises in which people are fearful that their social group, which they feel apart of will be viewed as unfavourable by others. This is where the group is a group which they identify with and draw some of their identity from (Tajfel & Turner, 1986, cited in Derks *et al*., 2008). This conception is much more than situational as suggested by stereotype threat as it includes all cases of possible marginalisation from other people and social groups as well as environmental cues (Derks *et al*., 2008).

In-group Identification and STL

A further element that may mediate the STL-performance relationship is that of in-group identification (Schmader, 2002 cited in Smith, 2004). In-group identification is an important variable within the SIT literature and therefore may help our understanding of STL. A group of studies by Spears, Doosje and Ellemers (1997) examined the effect of threats to the status of the group with regard to self-stereotyping. Status of the group with relation to self-stereotyping was defined as how similar people felt towards other in-group members. A further variable which was introduced was the participants were either classified in terms of high or low in-group identification. The main hypothesis of the study was the level of in-group identification would alter the way the participants responded to self-perceived or public perceived threats when the status of the group is targeted resulting in self-stereotyping. The results were that self stereotyping was less for low identifiers and was increased for high identifiers.

Within social identity theory it has been shown that in order to be viewed positively people will often differentiate their social group from a different or competing social group for example an in-group from an out-group. This a done through a downward social comparison in order to maintain a positive view of themselves ,however, simply categorising people into groups will not have this effect. "Social identity theorists have also argued that a degree of social identification is a prerequisite for group behaviour" (Spears *et al*., 1997, p.539). It could be argued that this inconsistency exists because SIT maintains that discrimination between groups happens when the identity of that group is threatened.   Social identity theory assumes that a group wants to be viewed in a positive light therefore downward social comparison or discrimination is used in order to create distinctions between the group as the positive standing of the one group rests on the negative comparison of the other group.

"One of the differences between real and artificial groups may lie in the fact that whereas artificial groups may give group judgements that merely reflect the experimentally imposed status differences, real groups may be more motivated to challenge the existing ranking of the groups" (Tajfel & Turner, 1986, cited in Spears *et al*., 1997, p539). Therefore in constructing our experiments it is useful to use pre-existing social groups rather than a minimal group paradigm as the pre-existing groups may give us more of an accurate reflection of group life particularly when we are relating them to other SIT variables.

The results of the four studies carried out by Spears and his colleagues (Spears *et al*., 1997) demonstrate that there was a main effect for in-group identification on social identity threats with regard to self-stereotyping. In this way, self-stereotyping was more frequent for high identifiers than for low identifiers. The hypotheses that low identifiers would distance themselves from the group when under threat and high identifiers would stick to the group when under threat found some support in three of the four studies (Spears *et al*., 1997). "Low identifiers in effectively distancing themselves from the group or its central tendency can be seen as relatively individualistic identity protection strategy. High identifiers on the other hand, is [sic] more collectivistic and loyal to the group" (Spears *et al*., 1997, p.550).

In two studies by Aronson *et al.* (1999) the authors examined the stereotypes in mathematics between Asian students and White students, invoking stereotype threat by telling white participants that Asian students out perform White students in maths. In the second study the authors identified the role of being either highly or moderately "math identified" and the relationship between STL (Aronson *et al.,* 1999). What was found was that when the students were more highly math identified the stereotype threat effects were more salient. The moderately identified group displayed the opposite pattern in performance. Therefore when the stereotype was not mentioned high math identified students performed better than moderately identified students, however when the stereotype was mentioned there was a reversal in effects whereby the high math identified students performed significantly worse than the moderately identified math students (Aronson *et al.,* 1999).

Although the authors (Aronson *et al.,* 1999) say that the math identification measure would be a measure within a domain and hence would be a measure of domain identification, it could be argued that this could be a display of in-group identification as the highly identified math students are therefore highly identified with their social group which is being negatively stereotyped. Whether it is domain identification or in-group identification the highly identified participants would have more "pressure" on themselves to succeed or care more about the outcomes either for their domain or social group.

Schmader (2002) examined the effect of social identity theory on stereotype threat. More specifically Schmader examined the possible role of in-group identification i.e. gender identification on men and women's math performance in a stereotype threat situation. What

was demonstrated was that women who were highly identified with their gender performed almost as well as the men when the stereotype was not made salient. However, when the stereotype was made salient low identified males and females performed almost the same within the test. The highly identified women performed significantly worse than the highly identified men (Schmader, 2002). Hence "women showed poorer performance compared to men on a stereotype relevant task when their social identity was linked to their test performance, but only if they considered gender to be an important part of their self-definition" (Schmader, 2002, p.199).

This pattern of performance has parallels within the literature on SIT. Ouwerkerk *et al*. (2000) hypothesized that when in-group status is low there would be more of an effort by the group to improve their social identity than when their status is high. The authors argue that evaluating preferences between the in-group and out-group is of little value to their research as it does not pose a threat or hold an individual accountable. Therefore, by merely displaying in-group favouritism it is not enough to demonstrate the individual's willingness to improve the social standing of their group. Thus the authors investigated the "individual effort that the group members are willing to exert to actually change the status quo." (Ouwerkerk *et al*., 2000, p.1551). "As predicted, stronger identification enhanced individual effort on behalf of the in-group when people's social identity was threatened, whereas no such effect was obtained when the current standing of one's group was favourable" (Ouwerkerk *et al.,* 2000, p.1557).

In-group identification is important to SIT and STL as the more you are invested in your group you are more likely to act on behalf of your group or in group normative ways. On the other hand if you are less invested in your group then you are less likely to perform for your group or in group normative ways. Two other SIT variables which influence in-group identification and therefore influence the effects of STL are permeability and identifiability which will be discussed below.

Permeability of groups

In the study by Ellemers et al., (1988) two experiments were performed in order to examine certain factors within social identification. When upward mobility is not an option then participants with high individual ability identify less with their group (Ellemers et al., 1988).

People would ideally like their group to be viewed in a positive frame of reference. People may try to do this in two ways, the one way is to try and improve their individual standing and the second way is the improvement of the group (Barreto & Ellemers, 2000). The more permeable the group boundaries the greater the possibility for individual mobility across groups. "Individual group members will try to leave their low status group in order to gain membership of a group with more positively valued characteristics, (i.e. higher status group). In other words, individual mobility will be the dominant strategy towards identity enhancement. However, individual locomotion across groups is by definition only feasible where group boundaries are permeable" (Ellemers et al., 1988, p.498). Even though Ellemers et al., (1988) found that permeability of group boundaries affected in-group identification of low status groups, it is not clear from the literature that it will only affect low status groups. Within the assumption of SIT that one must seek to be positively evaluated, allowing groups to be permeable or having knowledge that groups are permeable, being able to move to a higher status group, even at an individual level may lead to greater movement between groups in order to improve one's social identity. However, where group boundaries are impermeable, where it is almost impossible to change one's group membership (e.g. race or gender) and therefore the individual cannot move between groups they will be more likely to have higher in-group identification (Tajfel, 1975, 1978 cited in Ellemers et al., 1988). When a situation is likely to arise with low status group with impermeable boundaries, wanting positive evaluation in terms of SIT, it would have to be sought in terms of overall appreciation for the group and therefore in-group identification becomes more apparent. However, it should be noted that Tafel (1982) also suggests negative ethnocentrism whereby the in-group members actually favour the out-group rather than displaying an in-group bias (or an appreciation for their own group). Therefore for a group with low group status in-group identification would be more likely to be high if the group was seen to have impermeable boundaries while in-group identification would be more likely to be lower when the group boundaries were permeable as the group member could more easily leave the

group. Another SIT variable which may influence in-group identification is that of identifiability which is discussed below.

Identifiability

This specific study will focus on identifiability, which is the extent to which targets of STL, experience themselves as identifiable individuals or as anonymous members of a group. Identifiability has already been shown to be relevant to stereotype threat in two exploratory studies (Quayle and Reicher, 2007; Forbes, 2007). However, the nature of the interactions is far from clear. Haslam *et al*. (2008) argue that when a social identity is in conflict with individual motivations, it may lead to stereotype threat because who they are as individuals is at odds with how they are seen within their group (Haslam *et al*., 2008). Alternatively, anonymity may reduce the extent to which individuals are seen as separate from the group and anonymity may therefore increase targets' social identity as individuals and are less likely to be distinguished from the group. An equally likely explanation is that being identifiable within a group, may bring into question the group norms which in turn may leave the individual feeling personally responsible (Klein, 2003). Whether or not the person is seen within a group or as an individual may explain how much they feel in control over a situation or its outcome (Reicher & Levine, 1994). The current research aims to explore the mechanisms by which identifiability may intersect with STL and thereby resolve some of these questions.

Klein, Spears and Reicher's (2007) article on how SIDE relates to visibility towards an audience as well as how that may influence identity performance is important when considering the present study. Of particular interest is the examination of visibility. "Recent work on "respect" from the in-group also addresses this important sense of acceptance and group belonging" (e.g., Ellemers, Doosje, & Spears, 2004; H. J. Smith & Tyler, 1997 cited in Klein *et al*., 2007, p.32). Therefore it could be argued that when watched by the in-group it is important to perform according to stereotyped ways in order to earn the respect of the group, particularly if group identification is important.

"When they are individually identifiable, identity performance can help the social identity of specific individuals to be recognized and acknowledged by the out-group" (Klein *et al.*, 2007, p.41). In this way in may be important to perform well when being viewed by the out-group specifically when the in-group has been negatively stereotyped.

This may suggest that within an STL context where there is both positive and negative stereotypes about the group and they are visible to both the in-group and the out-group that several scenarios are possible. It is more likely that when there is a belief about being watched by the in-group that the participants will perform in stereotypical ways in order to gain in-group approval. Secondly, in terms of the out-group, it is more likely that participants would perform better when viewed by the out-group, as their social identity may be under more criticism, or that they have "more to prove" to the out-group regardless of the STL condition.

Within their study Barreto & Ellemers (2000) in the two experiments it was demonstrated that participants that are highly identified with the group will follow the norms set out by the group regardless of whether they are accountable or anonymous when completing the task. "It was also reported that although low identifiers do not pursue group goals when anonymous, accountability to the in-group may present a temporary motivation to do so" (Barreto & Ellemers, 2000, p.903).

In a study by Barreto and Ellemers (2000) through two experiments they examined how "group members' choices to work on individual or on group status improvement was examined as a function of degree of in-group identification (low, high) and accountability of responses (anonymous, accountable to the in-group)" (Barreto & Ellemers, 2000, p.891). The authors suggest that the amount which a group member identifies with their group would greatly influence whether members of the group act as individuals during a task or as a cohesive group (Barreto & Ellemers, 2000). It is preferred not to be part of the low status group therefore members of these groups usually try to improve how their group is viewed. Therefore there are clearly two types of action which emerge from this, either individual action or collective action (Barreto & Ellemers, 2000). It is argued that "high identified group members are likely to be concerned with the welfare of the group regardless of the

circumstances, whereas low identifiers are only persuaded to do so out of self-presentational concerns" (Barreto & Ellemers, 2000, p.891).

Ambady, Paik, Steele, Owen-Smith, and Mitchell (2004) wanted to examine whether individuation (disclosure of personal information, this included listing their favourite, book, movie and food, as well as listing three positive and negative personality characteristic about themselves) prior to performance would eliminate the negative effects of stereotype activation. This was done in two studies.

The authors suggest from reviewing other studies that individuation may be a possible means of reducing stereotype threat. However, the author argues that there has been virtually no research in the area. When individuals are stereotyped they are both a target and a perceiver. They are targets because the stereotype is directed at them. However, they are perceivers as well as they are able to recognise the stereotype. Wilder cited in (Ambady, 2004) suggests that the "disclosure of personal information accentuating one's unique qualities encourages a more multifaceted view that may distinguish a person from his or her in-group and counteract the stereotyping" (p.402).The author suggests that "if the salience of group identity is replaced by the salience of individual identity, the risk associated with negative stereotype activation might be attenuated and performance altered to reflect more accurately the unique capabilities of the individual rather than the stereotypes of the group" (Ambady *et al.,* 2004, p.402).

The study examined white females on a quantitative test. The stereotype was that women were inferior on the task. In both studies half of the females performed a computer task, which primed them with the stereotype. The other half performed a task which was similar but did not prime them with a particular identity. Within each of the groups half of the participants answered an individuation questionnaire while the other half of the participants answered a neutral questionnaire. The dependent variable was performance on the quantitative test. This was examined across the four groups (Ambady *et al*., 2004).

The first study found that women who were primed and individuated performed well compared to those who were primed and non-individuated. They performed as well as un-primed non-individuated participants.

In the second of the studies, on the individuation questionnaire they had to list one more negative trait than positive trait. Negative traits were listed after positive. Participants had the study run individually by a white female experimenter. Similar results were found to study one. Individuation allows you the option to distance yourself from the stereotype (Ambady *et al*., 2004).

Two exploratory studies within the present programme of research induced stereotype threat in participants in conditions of high or low identifiability. The researchers found that under stereotype threat conditions the participants that have high identifiability perform better on a given task than participants with low identifiability. However, when stereotype threat is not activated the participants with low identifiability perform better on a given task than participants with high identifiability (Forbes, 2007). In the second study participants in the high identifiability stereotype lift condition performed best and participants in the low identifiability stereotype lift condition performed worst (Quayle & Reicher, 2007). Therefore, it is probable identifiability is an important component of STL. The current study further explored the effect of identifiability within a stereotype threat context.

Influences of audience

A few researchers have suggested that the way in which the group members act is dependent on the audience's expectations of them, they may therefore change their behaviour in order to meet these expectations and ensure social desirability (Deutsch & Gerard, 1995; Kelley, 1952, cited in Barreto & Ellemers, 2000). Others argue that an audience will not affect group members' behaviour and that if the group membership is viewed as being most important, then the group members will act in accordance with the groups social identity needs (Barreto & Ellemers, 2000).

Some theorists maintain that "people's behaviour in the presence of an audience is strongly determined by a fundamental desire to avoid censure and seek positive evaluation by others" (Baumeister, 1982; Leary & Kowalski, 1990; Schlenker, 1980 cited in Barreto & Ellemers, 2000, p.892). Support of this has been shown when participants may behave in a way the group expects when visible to the group, but would ordinarily not act in that way if they were anonymous (Barreto & Ellemers, 2000).

"The Social Identity model of Deindividuation Effects (SIDE) has been developed with the aim of accounting for the effects of anonymity and group audiences on group members' responses" (Barreto & Ellemers, 2000, p.892). The model is based on the principles that people have deep seated motivations for behaving and being viewed within a particular group and also the audiences which the members are exposed to may influence how they respond in the group. There are two ways in which the SIDE model accounts for social influence. There is the cognitive and the strategic. The cognitive component refers to the importance of an assigned identity to a participant and how the visibility is varied between the participants. The visibility of other participants may lead to a greater realisation of individual differences between people and in turn could decrease group normative behaviour. (Barreto & Ellemers, 2000). The strategic component refers to the person's behaviour within a given context, this behaviour should be appropriate regarding the audience they are visible to and how they need to present themselves towards that audience (Barreto & Ellemers, 2000).

Deindividuation refers to a lessening of self-awareness leading the individual to feel more like an anonymous member of the group this is caused be darkness, clothing or immersion in a group (Reicher & Levine, 1994). The SIDE model suggests that it is not a lessening of self-awareness but rather a shift from personal awareness to group awareness. Anonymity, arousal and external focus are all concepts which have been linked to the study of deindividuation, however the processes are unclear. Currently the focus of the research is on anonymity particularly with regard to social identity theory. The authors wanted to test the hypothesis that if you are anonymous within an important social group this would lead to defining yourself and others in terms of that group and which would lead to greater group behaviour as suggested by social identity theory. Below, a study by Lea *et al.,* (2001), examined for some of these effects.

"In this study, visual anonymity was found to increase self-categorization in line with the social identity formulation of deindividuation as involving a shift of self-focus from personal to group-based aspects of the self (rather than a reduction in self-awareness)" (Lea *et al.,* 2001,p.534). An example of how visibility may operate within social groups is outlined below.

"Increasing the visibility of in-group members may increase their ability to support one another when faced with sanctions from a powerful out-group. Under these conditions, visibility may have a strategic effect on group attraction (and the expression of group norms) (Lea *et al.*, 2001, p.535).

Klein, (2003) makes the assumption that when a person is capable of expressing prejudices is identifiable as an individual, that person would choose not to. The person would like to be viewed as tolerant where the norms of that society are not to be prejudiced. However, Klein argues that this might not always be the case as various social identities may play a role, or need to be protected when expressing attitudes towards out-group members. By expressing these attitudes they are presenting a particular social identity. Identifiability also has a role to play as whether you are anonymous part of the group or personally identifiable may influence which social identity you represent and how you communicate this to an audience.

"In line with this view, Noel, Wann, and Branscombe (1995) have shown that individuals were particularly motivated to be perceived by other in-group members as espousing the in-group's norms. In order to achieve this goal, they expressed more in-group bias when identifiable to other group members than when their responses were anonymous" (Klein, 2003, p.253).

In this study Klein, (2003) examined Greek students' identities, as Greeks have their Greek identity but would also like to associate with their European identity. Prejudice towards Turks is viewed in line with Greek identity but not European identity. Klein examined being personally identifiable, identification with Europe and audience effects (Western European or Greek) on the expression of prejudice towards Turks. Klein hypothesized that when viewed by the European audiences there would be less prejudice towards Turks as being tolerant is viewed as positive within a European identity. The results supported this hypothesis. However, one surprising result from the study was when identifiable to Greek audiences European identification was positively associated with prejudice towards Turks. "This finding was unexpected. It might be understood in the framework of the conflictual relationship between the rest of the EU and Greece" (Klein, 2003, p.259)

Five studies by Jetten, Hornsey, and Advares-Yorno (2006), examined perceptions of conformity within the group between junior and senior members. Junior members were defined as second year students; senior was defined as third year student. The in-group was defined as a psychology student conducting the study, while the out-group was defined as a political science student conducting the study. Within study three junior members of the group referred to themselves' as more conformist when compared with senior members of the group when visible to the in-group audience; however this effect was not repeated for the out-group audience. This is important as it displays that someone will manage their identity differently whether it is in the form of an action of describing a personality characteristic depending on who is the recipient of the information that is whether they are an in-group or out-group member.

Social facilitation theory

Social facilitation theory, which is outside of the scope of STL and SIT, may be another way of explaining the audience effects on task performance. Social facilitation theory "involves the observation of behaviour when it occurs in the presence of passive spectators" (Zajonc, 1965.p.269). Social facilitation came about to explain the social effects, in our case audience effects on task performance (Bond & Titus, 1983). Within learning tasks Zajonc suggested that the presence of an audience would enhance the individual's dominant response. Zajonc goes on to suggest that these responses may be helped by drive, arousal and activation. The audience, even though passive may increase an individual's drive and arousal.

Importance of this social phenomenon

STL is an important social phenomenon because it partially explains how genuine intergroup differences in performance can be produced and perpetuated in society in the absence of any actual (e.g. biological, physical or cultural) differences between group members. Understanding the mechanisms and modifiers of stereotype threat is therefore important in order to understand the difficulties that are faced by minority groups. It is hoped that a better understanding of stereotype threat will allow people to better protect themselves against the negative effects of stereotype threat, or suggest structural variables that could be improved thereby helping minority group members to achieve well in areas where they are traditionally negatively stereotyped (Steele, 2003).

Aims and Rationale

Identifiability has already been shown to impact on the experience of stereotype threat (Quayle & Reicher, 2007; Forbes, 2007), although the results were difficult to interpret.

The present study will be an extension of previous exploratory studies (Quayle & Reicher, 2007; Forbes, 2007), showing the effect of identifiability on social identity theory in which the researcher will investigate the interactions between identifiability, identity and STL which will be explored in more detail.

If identifiability has some effect on the experience of the STL effect, as preliminary evidence suggests it does, there are at least two possible explanations as to why this may be so. One explanation is that identifiability causes a shift in identity in terms of group identity or individual identity. It is hypothesized that when a member of a group is highly identifiable this will cause a shift from their group identity to an individual identity. On the other hand when a member of a group has low identifiability, or remains anonymous within that group, then that group member is more likely to adopt their group identity.

The second explanation is that identifiability influences motivational factors such as accountability and social loafing. Social loafing occurs when there is low accountability for actions or productivity within a group and the group member is able to do less work at the expense of the group as they are viewed as an anonymous part of that group. If the group member is highly identifiable and therefore individually accountable for the performance of their group, this would diminish social loafing. When there is low identifiability it can be argued that the group member is seen as anonymous within the group and therefore there is a greater possibility for lowered accountability and increased social loafing.

The research project aims to apply a social identity theory framework to the empirically well-established phenomenon of stereotype threat. Specifically, the study aims to explore whether the extent to which a target of STL is viewed as an individual or as an anonymous part of a group and thereby impacts on their task performance (i.e. contributes to STL).

a. Does identifiability reliably influence the experience of STL?

b. Is there any difference between in-group, out-group or external visibility with respect to the effect of identifiability on STL?

Additional analysis was performed to examine whether or not other SIT variables influence the experience of STL.

Methodology

Research Design

The design was a mixed within/between subjects repeated measures. The within subjects factors were the scores for the Raven's APM pre and post-test, this meant that each participant completed both the Raven's APM pre and post-test. The between subjects factors are STL condition (threat, boost) and the audience manipulation (in-group, out-group or experimenter). This meant that each participant was either exposed to a threat or boost situation and then they were exposed to one of the three audience manipulations. This would allow the researcher to be able to identify how each participant scores on the Raven's APM while under threat or boost while under the anonymous condition (Raven's pre-test) and then how the participants scored in when visible to an audience (Raven's post-test). Identifiability is implied through the experimental design as at the Raven's pre-test the participants are in the anonymous condition while at the Raven's post-test they are in the identifiable condition.

The dependent variable for the study was the scores for the Raven's (APM). Both the pre-test and post-test results were used. However, it must be noted that there are limitations with what the researcher can say about the data when using the repeated measures design. For some of the results only the Raven's APM post-test were used. This is because of the unbalanced between subject factors, which means that the repeated measures design can only be used when examining the effects of anonymity and audience as there are the manipulations for the pre and post test. It cannot be used to explain any STL effect as there is no post-test manipulation. However it is assumed that the initial STL manipulation carries its effects through the whole design.

The first independent variable for the study was the stereotype condition (threat, boost). The second independent variable for the study was the audience that the participant thought they were visible to (in-group, out-group, or experimenter).

There were six possible conditions that the participants could be exposed to. The first of these conditions was that the participant is exposed to the stereotype condition (threat). All participants were exposed to a description of the experiment that invoked either stereotype threat or lift. They then completed the APM pre-test under anonymous conditions- alone in the testing room. Following this, they completed the APM post-test while being filmed and under the impression that the footage would be viewed and discussed by the in-group, the out-group or a group of experimenters.

The stereotype threat or lift manipulation was positioned outside of the repeated measures design and was placed prior to the Raven's APM pre-test as the main aim of the research was the role of identifiability. In this way, by activating the stereotype before the pre-test the researcher was able to examine the effects of STL at both the Raven's APM pre-test and post-test conditions. Therefore the presence or absence of STL can be detected as a between-groups difference, with the usual experimental assumption that any differences between conditions will be either random (as a result of random assignment of participants to conditions) or resulting from systematic differences related to the manipulation.

Manipulations

In order to test the hypotheses of the research, several manipulations had to be created, in order to be able to experimentally account for the differences (if any) in the participants' scores for the Raven's APM.

Status Condition (STL)

In order for stereotype threat to be present, two stereotyped groups had to be created or assigned. The research settled on the contrasting groups of science and humanities students. These groups were used for the following reasons. Firstly, the researcher avoided using groups that had important social meaning such as race or gender. This was done because if one used a group with important social meaning the negative effects from the experiment may carry over to other contexts and be hard to shake off. Secondly, university students, particularly humanities students would create a convenience sample.

The status condition for threat was manipulated in the following ways: firstly this was done by the wording in the booklet, this contained a cover sheet explaining the research as presented in the extract below, and it also contained the initial group membership tasks, the Ravens APM, and the Social Identity Theory Inventory. By manipulating the wording was hoped to create the threat effect by explaining that humanities students were worse at the Raven's (APM) than science students. A except from the threat booklet is included below.

[This study compares the performance of Science and Humanities students on a test of academic ability called Raven's Advanced Progressive Matrices.

This test has been found to be a very accurate and reliable measure of intelligence and academic ability and humanities students consistently perform badly. At the same time Science students perform very well.

Research has found that both Science and Humanities students are skilled at finding complex patterns, but the types of patterns that they are skilled at finding are substantially different.

Science students are skilled at finding logical patterns in datasets whereas Humanities students are skilled at finding intuitive patterns in symbolic figures.

Raven's Advanced Progressive Matrices requires logical data analysis and is therefore more suited to the skills of Science students whereas Humanities students are substantially disadvantaged in the task.

Before completing the Raven's Advanced Progressive Matrices please complete the following two tasks. The first is a test of logical pattern recognition and ideally suits the skills of Science students. The second is designed to test symbolic pattern-recognition skills and ideally suits the skills of Humanities students]

To induce a lift effect in the boost group the following phrasing was used:

[This study compares the performance of Humanities and Science students on a test of academic ability called Raven's Advanced Progressive Matrices.

This test has been found to be a very accurate and reliable measure of intelligence and academic ability and Science students consistently perform badly. At the same time Humanities students perform very well.

Research has found that both Humanities and Science students are skilled at finding complex patterns, but the types of patterns that they are skilled at finding are substantially different.

Humanities students are skilled at finding symbolic patterns in symbolic figures whereas Science students are skilled at finding logical patterns in datasets.

Raven's Advanced Progressive Matrices requires symbolic pattern-recognition skills and is therefore more suited to the skills of Humanities students whereas Science students are substantially disadvantaged in the task.

Before completing the Raven's Advanced Progressive Matrices please complete the following two tasks. The first is a test of logical pattern recognition and ideally suits the skills of Science students. The second is designed to test symbolic pattern-recognition skills and ideally suits the skills of Humanities students.]

Secondly it was done through the initial group membership tasks. In order to enhance participants' sense of group membership, and to ensure that they are mindful of the relevant group stereotypes, they were asked to complete two tasks prior to completing the experimental measure. One of the tasks required participants to find objects in an Escher print and was presented as a "symbolic pattern recognition task" while the second required participants to find numbers in a grid and was presented as a "numerical pattern recognition task." These descriptions are congruent with expectations of humanities and science disciplines respectively. The task that was congruent with the participant's own group was easy, while the task associated with the out-group was impossible and therefore experienced by participants as very difficult. Therefore in this study using humanities students it was very easy to complete the Escher print task by finding the relevant objects. However, in the number search task in a grid some sequence of numbers given to participants to find were not included in the grid thereby making the task impossible. Since participants' experience of completing the tasks matched the descriptions of the two group stereotypes this manipulation ensured that their group membership was relevant in the experimental context and that the relevant stereotypes are validated.

Audience Manipulation

The audience variable was manipulated in the following ways: firstly all participants would complete the Raven's APM pre-test alone in the testing room. Since they were not asked to identify themselves on materials, they were specifically informed that the test would be

anonymous and the camera lens was covered by a lens cap and pointed away from them, it was expected that this would be experienced as anonymous.

In the post-test condition the camera was rolling (while pointed at their desk with the lens cap removed) and participants were told that their performance on the subtest will be watched and discussed by either a group of humanities students (in-group), science students (out-group) or researchers (powerful others). Footage from the camera clearly showing their paper and pencil as they completed the task was shown on a small monitor in full view near the camera. In reality, even for the identifiable group, the camera did not produce any video footage.

Sample

The sample consisted of 60 humanities students who were on the Pietermaritzburg campus. However of the 60 participants only 53 were included as some students were actually registered for other degrees. Humanities students were only sampled as it was only the outcome of the humanities students' results which were of interest to the researcher. However, for the stereotype threat to be effective it had to be believable that science students were part of the sample as well. In reality no other students were allowed to take part.

Measures

Development of the Raven's APM subtests

The Raven's APM was selected as the dependent measure as it was viewed as appropriate for the study for the following reasons: Firstly the Raven's APM is as an ambiguous task that could be presented to participants as either a 'symbolic pattern- recognition task' or a 'logical pattern-recognition task' to invoke threat or lift conditions. Secondly, the Raven's APM has been used in previous studies of stereotype threat therefore it makes it more valid to re-use the Raven's APM as the STL effects have been shown previously. Thirdly the Raven's APM is quite a difficult task which may then make the stereotype threat more likely.

Within the Raven's APM there are two sets of matrices, the first set of matrices are practice items which are used to get participants used to answering the matrices, there are 12 in total. The second set of matrices of which there are 36 vary in difficulty from easiest to hardest. Using the full Raven's APM would require 45 minutes but because of the additional scales and questionnaires used in the current study and seeing as the study was a repeated measures study it was decided to divide the full APM into two equivalent subtests of the Raven's APM for the repeated measures design. In this way the participants would have enough time to complete the study and the researcher can have the repeated design effect.

During a literature search two short forms of the Raven's APM were found (Bors & Stokes, 1998; Arthur & Day, 1994). However they were not comparable in terms of difficulty and some items were included in both thereby making it impossible to include both in a repeated measures design. It was therefore necessary to construct two suitable short forms of the APM.

A through literature search for item analysis of the Raven's APM was done in order to collate a table of item difficulty across multiple studies. The proportion of participants getting each item correct was averaged across the available studies to arrive at item difficulty scores. This list enabled the researcher to randomly create two subtests of equal difficulty. The studies were found through various search engines including, Academic search Premier, Psych Info, Psych articles, Google Scholar and through the University library. The studies which were included were studies that listed an item analysis of the Raven's APM or item difficulty. In other words studies were included that enabled the researcher to be able to examine the different items of the Raven's APM and there difficulty. The following eight studies were included after the literature search (Alderton & Larson, 1990; Arthur & Day, 1994; Forbes, 1964; Rushton, Skuy & Fridjhon, 2003; Rushton, Skuy, & Bons, 2004; Salthouse, 1993; Unsworth & Engle, 2005; Vigneau & Bors, 2005). The table of collated data had different N values for different items as not all the studies included in the table used all 36 items from the Raven's APM in their original studies.

It was decided to keep the Raven's APM subtests at twelve items as this was the same length as the other short forms of the Raven's APM which had been located within the literature. These short forms also used the first two items within the A set of matrices in order to present the participants with some practice items therefore items one and two of set A of Raven's APM were included in the final scale as practice items. From the combined sample of Raven's APM items, the items were first sorted by difficulty. The challenge was to identify two subsets of items with equivalent difficulty that had suitable sensitivity for the sample of undergraduate university students. Therefore the six easiest and six most difficult items were excluded because previous studies had shown that either too many or too few participants got them right for them to be of much use. This left the researcher with 24 items of moderate difficulty with which to create the two 12 item subtests. To do so, the first two most difficult items were assigned to 'block 1', followed by the next two most difficult items to 'block 2', this was repeated until each item was assigned a 'block' of difficulty. Thirdly using the random number generator in Microsoft Excel 2003 a random number was generated for each item. Then the items were sorted by 'block' and then by random number, so that the blocks are sorted by order of difficulty, but the items within each block are in a random order. Lastly the first item in each block is then assigned to subscale 1 and the second item is assigned to subscale 2. This is repeated until each scale has twelve items. For the first Raven's subtest the

scale mean was 61.9 with a standard deviation of 19.78. For the second Raven's subtest the scale mean was 61.2, with a standard deviation of 19.81.

TABLE 1: Summary of combined samples of the Raven's (APM) and the percentage of correct responses (Sources: Alderton & Larson, 1990; Arthur & Day, 1994; Forbes, 1964; Rushton, Skuy & Fridjhon, 2003; Rushton, Skuy, & Bons, 2004; Salthouse, 1993; Unsworth & Engle, 2005; Vigneau & Bors, 2005.)

Raven's (APM) item	n correct total	N Total	% correct responses
1	5176	5528	94
2	5082	5528	92
3	5578	6172	90
4	4779	5528	86
5	4810	5528	87
6	4856	5528	88
7	4707	5528	85
8	4445	5528	80
9	4587	5528	83
10	5011	6172	81
11	4506	5528	82
12	4917	6172	80
13	3648	5528	66
14	4299	5528	78
15	4529	6172	73
16	4225	6172	68
17	3836	5528	69
18	3713	6172	60
19	3586	5528	65
20	3250	5528	59
21	3308	6172	54
22	2507	5528	45
23	2534	5307	48
24	2379	5307	45
25	2222	5307	42
26	2048	5307	39
27	1782	5307	34
28	1675	5951	28
29	1323	5307	25
30	1760	5951	30
31	1538	5951	26
32	929	5307	18
33	1336	5307	25
34	1086	5951	18
35	938	5307	18
36	308	5307	6

Manipulation checks for status manipulation

Stereotype threat/lift check

The stereotype threat/lift measure examines the extent to which the participants believe in the status manipulation that they read before beginning the dependent measures. Five items were developed these were: ("Based on the available information, [BOOST GROUP] as a group have a good reputation with respect to the Raven's Advanced Progressive Matrices?"), ("*Based on the available information, [THREAT GROUP] as a group have a good reputation with respect to Raven's Advanced Progressive Matrices?"), ("Based on the available information, Raven's Advanced Progressive Matrices is more suited to the skills of [BOOST GROUP] than [THREAT GROUP]?"), ("*Based on the available information, more [THREAT GROUP MEMBERS] are likely to do better than [BOOST GROUP MEMBERS]?"), ("Based on the available information, the worst performing participant will probably be a [THREAT GROUP MEMBER]?"). The items marked with an asterisk are reverse-coded. The responses to the scale were on a seven-point Likert scale. A score of one indicated strongly disagree ranging to a score of seven which indicated strongly agree. The measure was piloted with 23 psychology honours students and the Cronbach's alpha was 0.696.

Stereotype Agreement Manipulation Check

The stereotype agreement manipulation check examines the extent to which the participants reported agreement with the STL manipulation. It was completed after the Raven's APM and thereby giving a better sense of the efficacy of the manipulation. Six items were developed these were: ("How do you think the typical [BOOST GROUP MEMBER] would have experienced [the task] in terms of DIFFICULTY??"), ("How do you think the typical [BOOST GROUP MEMBER] would have experienced [the task] in terms of ENJOYMENT?"), ("How do you think the typical [BOOST GROUP MEMBER] would have PERFORMED on [the task]?"), ("*How do you think the typical [THREAT GROUP MEMBER] would have experienced [the task] in terms of DIFFICULTY?"), ("*How do you think the typical [THREAT GROUP MEMBER] would have experienced [the task] in terms of ENJOYMENT?"), ("*How do you think the typical [THREAT GROUP MEMBER] would have PERFORMED on [the task]?"). The responses to the manipulation check were on a four-point Likert scale ranging from "very badly" to "very well". Except for the two items

which addressed enjoyment and the anchors for this scale were very unpleasant to very enjoyable. The items marked with an asterisk are reverse-coded. The measure was piloted with 23 psychology honours students and the Cronbach's alpha was 0.978.

Audience manipulation check

The audience manipulation check examines the extent to which the participants believe that their results will be viewed by either the in-group, out-group or experimenter audiences. In order to assess these six items were developed these were: ("how likely is it that a group of researchers will see this questionnaire?"), ("how likely is it that a group of humanities students will see this questionnaire?"), ("how likely is it that a group of science students will see this questionnaire?"), ("how likely is it that a group of researchers will see this video?"), ("how likely is it that a group of humanities students will see this video?"), ("how likely is it that a group of science students will see this video?"). The responses were on a sliding scale with one representing very unlikely to seven which represented very likely.

The Social Identity Theory Inventory

The manipulation checks and all the scales within the Social Identity Inventory were developed by Michael Quayle, unless indicated otherwise.

Identifiability

The identifiability scale measures the extent to which you feel anonymous and part of your social group or personally identifiable. The scale has two parts. A group identifiability subscale and an individual identifiability subscale. The group identifiability subscale assesses the extent to which you feel part of your group, whereas the individual identifiability subscale assesses the extent to which you are personally identifiable. The responses to the scale were on a seven-point Likert scale. A score of one indicated strongly disagree ranging to a score of seven which indicated strongly agree.

Items two, four and five assessed group identifiability they were: ("I think my results reflect more on my group than on me as an individual?"), ("I am being seen more as a group member than as an individual?")("I felt like my group was in the spotlight while I was doing the task?")

Items one, three and six assessed individual identifiability they were:

("I felt identifiable as an individual while I was doing the task?"), ("People who see these test results will be able to recognise me in other contexts?"), ("I felt like I was personally in the spotlight while I was doing the task?").

In-group Identification

The in-group identification scale measures the extent to which participants feel closely identified with their social group. As most of the participants were presumed to be second language English speakers the scale was kept simple. This was done by taking six simple items from various in-group identification scales that were related. The items were taken from the following sources: Ellemers, Kortekaas and Ouwerkerk (1999), adapted with respect to Crisp & Beck (2005); Doosje, Ellemers, and Spears (1995); Costarelli (2007); Luhtanen and Crocker (1992); Adapted from Verkuyten and Nekuee (1999); Schubert and Otten (2002). The responses of five of the items were on a seven-point Likert scale from "strongly disagree" (1) to "strongly agree" (7). The final item was a graphical representation of closeness to the in-group from Schubert and Otten (2002) consisting of seven representations of varying closeness which were coded from one to seven. A score of seven for the last item represented the strongest in-group identification. The measure was piloted with 23 psychology honours students and the Cronbach's alpha was 0.853

Anxiety/Pressure/Tension

The anxiety/pressure/tension subscale examines the extent to which participants perceived themselves to be anxious, feeling pressured and/or tense while completing the Raven's (APM). Four items were taken from Deci and Ryan's (2005) intrinsic motivation inventory. The responses to the scale were on a seven-point Likert scale. A score of one indicated strongly disagree ranging to a score of seven which indicated strongly agree. The measure was piloted with 23 psychology honours students and the Cronbach's alpha was 0.898

Other measures were included only for comparability with other studies in the programme of research and will not be discussed in this report. These included:

Differentiation, Group Esteem Scale, Public Self Esteem Scale, Legitimacy, Permeability Stability, Intergroup Conflict & Hostility, Intrinsic Motivation Inventory (Enjoyment Subscale, Perceived Competence Subscale, Effort/Importance Subscale) Threat, and Category Salience

Demographic information

All the participants were required to report their, race, gender, nationality, and because the study concentrated on humanities students, the researcher needed to be sure they indeed were humanities students, they had to report their degree they had registered for as well as how they would describe their degree.

Procedure

The research participants were asked to volunteer to be apart of the study through the medium of posters which were put up around campus. They would have to email their name student number and a suitable time which they were able to take part in the research. Because the experiment and measures used were time consuming, a research assistant was employed to carry out the data collection the research participants were also compensated for their time and received R30. Hiring the research assistant also minimised the probability of experimenter effects which could have arisen if the researcher had undertaken the data collection.

Once a time had been confirmed with a participant, they would meet at the testing venue. The study was explained as fully as possible without making the participants aware of the true aims of the study. The experiment proceeded in the following way: the participants were given an informed consent form (attached as appendix B). The informed consent forms were signed and handed back to the researcher which the researcher kept. The participants were told about the study in as much detail as possible. It was necessary to use deception. Several levels of deception occurred. The first is that of blinding participants to the real purposes of the study with a bogus study description in order to avoid introducing confounding demand characteristics. The second was more problematic and involves (falsely) informing participants that the group to which they belong is either positively or negatively stereotyped regarding 'symbolic pattern recognition' or 'logical pattern recognition'. However, the notion that humanities students or science students are more or less 'symbolically intelligent' or 'logically intelligent' is not likely to be particularly stressful or damaging (especially compared to stereotype threat studies that deal with much more important characteristics such as race and intelligence). The third deception involved participants being deceived into believing that their performance will be discussed by in-group, out-group members or a group

of researchers. This was necessary to investigate the effect of audience on identifiability. The participants were informed that the study is voluntary and that they are free to leave, without having to give a reason. The participants were told that they are not forced to complete the questionnaire or items on the measures. The participants were told that they will not be prejudiced or discriminated in any way for not participating in the study, or not completing the study.

Once they were content with the terms of the study the informed consent was signed. After this the participants completed the two group membership tasks (the Escher print search task and the number grid search task). They had 90 seconds to complete each task. This task was designed to reinforce the stereotype associated with the group membership. The research assistant left the room to try to ensure anonymity.

After this was done they completed the manipulation check scale followed by the first Raven's APM sub-test. To examine the extent to which they agreed with the manipulation. The status manipulation has been described previously. Then the participants completed the identifiability questionnaire.

In the first sub-test the participant was always anonymous whereas in the second sub-test they were subject to a different identifiability condition. This was either the in-group, out-group or experimenter condition. The researcher carefully explained that the task would be recorded and the recording would be scrutinized by the relevant group. After this the video camera was switched on and the research assistant again left the room at which point the participants completed the second Raven's APM sub-test.

Each Raven's sub-test was given a time limit of fifteen minutes and the subtests were always in a random order to avoid confounds.   To ensure this a schedule of conditions was created using Microsoft excel. This is where the schedule keeps sets or cycles of six conditions together to ensure a balanced design at different sample sizes. Within each cycle the order of the six conditions was randomised. Then (independently) the order of the presentation of the two Raven's APM subtests were randomised for each participant in the schedule. Thus in the final schedule the order of presentation for condition is randomly ordered and the order of presentation of the subtests is randomly ordered.

This procedure ensured that differences in performance due to small differences in the short forms of the Raven's APM and sequence effects were randomly distributed across conditions.

Once the second sub-test was completed the research assistant was called by the participant. The research assistant then switched off the camera, and left the room once again to allow the participant to complete the battery of social identity questionnaires. Once all the questionnaires were completed the research assistant was called again from the other room. The participant was then debriefed on the true nature of the study and why deception had to be used within the study. After this, an additional form was signed by the participant stating that they had been debriefed and was still willing for their results to be retained in the study. They were paid R30 as a token of appreciation and asked to sign a separate form demonstrating that they had received the money.

Results

Introduction

Within the original sample of 60 participants seven were excluded because they were not registered as humanities students and therefore would not be subject to the status manipulation. However the full sample was included in the reliability analysis of the social identity measures as it was thought that their status condition would not impact on these results.

Demographics

Of the 53 participants in the final sample: 40 (75.5%) were female and 13 (24.5%) were male; 50 (94.3%) were Black African, 2 (3.8%) were white and 1 (1.9%) described him/herself as 'other'; 47 (88.7%) were South African, and 6 (11.3%) were from outside South Africa.

When students were required to respond in an open-ended fashion to which degree they were registered for 34 (64.2%) were registered for a social science degree, 12 (22.6%) participants were registered for an arts degree, 6 (11.3%) participants were registered for postgraduate degrees and 1 (1.9%) was registered for a humanities degree.

They were then asked make a forced choice to describe their degree 32 (60.4%) participants described their degree as a humanities degree, 18 (34.0%) described their degree as a mixed degree, 1 (1.9%) described their degree as an arts degree, 1 (1.9%) described their degree as a law degree and finally 1 (1.9%) described their degree as a commerce degree.

Scale Reliability

When reviewing the results first the scale construction and the reliability were checked for each of the scales and the two sub-tests of the Raven's APM. Cronbach's alpha was used in order to determine the internal consistency of the scales; this measures the correlation between each item and every other item. Cronbach's alpha ranges from zero to one. Zero would indicate that there is no internal consistency and one would indicate maximum internal consistency. A general rule of thumb is for most questionnaire scales is that an alpha is more than 0.75 is considered appropriate and therefore reliable (Durrheim, 1999). However, Cortina (1993) states that Cronbach's alpha is also a function of the number of items in a scale and must be interpreted with the number of items in mind. "Cortina demonstrated that a six-item scale with an average item correlation of .30 has a value of α of .72. Keeping the average correlation the same, but increasing the number of items to 12 and 18 increased α to .84 and .88, respectively" (Streiner, 2003, p. 101). "However, the purpose of a test or scale is also a factor in deciding whether the reliability is adequate. Aiken (1982) argues that if the scale is to be used to compare groups of people then a reliability of 0.65 is sufficient" (Finchilescu, 2002, p.216). Therefore, in this modest between-groups study a reliability of 0.65 was considered acceptable for short scales. Although alpha for the three manipulation checks (mentioned below) fell below 0.65, these were still used within the current study as they were essential to the design and could not be omitted. The identifiability scale was also included as it was essential to support our research question although it marginally falls below the defined acceptability of 0.65. The identifiability scale is different as it is both a scale on its own as well as a manipulation check for the identifiability condition.

STL manipulation check

The Cronbach's alpha is 0.370 for the five items, which is unacceptable, however this can be improved to 0.541 if item one is dropped from the scale leaving the corrected scale at four items. The number of participants for the scale was 53.

Stereotype agreement manipulation check

The Cronbach's alpha is 0.422 for the stereotype agreement scale for the six items, which is unacceptable. However if item two is dropped reliability improves to 0.499 leaving the corrected scale at five items.

Audience manipulation check

The Cronbach's alpha is 0.430 for the audience manipulation scale for the six items, which is unacceptable. If item one is dropped reliability improves to 0.452, which is still unacceptable leaving the corrected scale at five items.

Identifiability

The Cronbach's alpha for the identifiability scale is 0.633 for the six items. While it is not within the desired level of 0.65 it is still acceptable, because the scale relates directly to the research question. The number of participants for the scale was 53. It was decided to use the 53 participants as the identifiability scale is closely linked with the STL manipulation and therefore the researcher needed to ensure that only humanities students were sampled.

> *identifiability (individual).* When the scale was spilt into the individual and group identifiability, the individual identifiability sub-scale Cronbach's alpha was 0.606 for the three items which is acceptable.

> *identifiability (group).* The Cronbach's alpha for the group identifiability sub-scale was 0.647 for the three items which is acceptable.

Ingroup Identification

The Cronbach's alpha for the group identification scale is 0.763 for the 6 items. This Cronbach's alpha is greater than .7 and therefore can be considered reliable. The number of participants for the scale was 58.

Intrinsic motivation inventory

> *Anxiety/pressure/tension subscale.* The Cronbach's alpha 0.626 for the four items. However if the fourth item is dropped the reliability improves to 0.664 leaving the corrected scale at three items. The number of participants for the scale was 58.

Raven's (APM) subtest 1

The Cronbach's alpha for the Raven's APM subtest 1 is 0.641 for the twelve items. When including the practice items the alpha is 0.629. The number of participants used for the scale is 53. The number of participants is lower than the previous scales as the Raven's (APM) is a specific manipulation of the dependent measure and only registered humanities students could be sampled for this analysis.

Raven's APM subtest 2

The Cronbach's alpha for the Raven's APM subtest 2 is 0.671 for the twelve items. When including the practice items the alpha is 0.678. The number of participants used for the scale is 53. The number of participants is lower than the previous scales as the Raven's (APM) is a specific manipulation of the dependent measure and only registered humanities students could be sampled for this analysis.

Other measures were included only for comparability with other studies in the programme of research and will not be discussed in this report. These included:

Differentiation, Group Esteem Scale, Public Self Esteem Scale, Legitimacy, Permeability Stability, Intergroup Conflict & Hostility, Intrinsic Motivation Inventory (Enjoyment Subscale, Perceived Competence Subscale, Effort/Importance Subscale) Threat, and Category Salience.

These other measures are not discussed as they do not relate directly to the research question of the current study and were included only for comparability with other studies in the programme of research and will be analyzed at a later date.

TABLE 2: item analysis of new short-forms of APM compared to original results of combined samples from studies.

		Current Study				Combined Sample		
Scale	Item	Percent Correct	N Correct	N Total	Item	Percent Correct	N Correct	N Total
2	6	64.2	34	53	6	88	4856	5528
1	7	61.3	33	53	7	85	4707	5528
2	8	56.6	30	53	8	80	4917	6172
2	9	52.8	28	53	9	83	4445	5528
1	10	28.3	15	53	10	81	5011	6172
2	11	45.3	24	53	11	82	4587	5528
2	12	35.8	19	53	12	80	4506	5528
2	13	20.8	11	53	13	66	3836	5528
1	14	34	18	53	14	78	3648	5528
2	15	20.8	11	53	15	73	4299	5528
1	16	39.6	21	53	16	68	4225	6172
1	17	37.7	20	53	17	69	4529	6172
1	18	20.8	11	53	18	60	3586	5528
1	19	24.5	13	53	19	65	3713	6172
1	20	28.3	15	53	20	59	3250	5528
2	21	17	9	53	21	54	3308	6172
1	22	26.4	14	53	22	45	2534	5307
1	23	22.6	12	53	23	48	2507	5528
2	24	11.3	6	53	24	45	2379	5307
1	25	18.9	10	53	25	42	2222	5307
2	26	17	9	53	26	39	2048	5307
2	27	5.7	3	53	27	34	1782	5307
2	28	24.5	13	53	28	28	1675	5951
1	29	20.8	11	53	29	25	1323	5307

As discussed previously, the final two scales of the Raven's (APM) subtests were created by ranking items 6 to 29 by difficulty by aggregating data from previously published studies (N between 5307and 6172 for different items). The results of the current study (N = 53) are compared to these original data in table 1, above. The similarities can be seen within the table both item six and seven in both samples are the easiest items to correctly complete with 65.4% and 61.5% of the participants correctly answering those items in the current study. This is compared to 88% and 85% in the combined study. The two most difficult items in the study were items 24 and 27 with 11.5% and 5.8% of the sample answering them correctly.

Manipulation checks

In order to assess the validity of the experimental measures and manipulations several manipulation checks were conducted.

Audience

A bivariate correlation was conducted in order to examine the different visibility manipulations against the audience the participant was told would be able to view their results (i.e. in-group, out-group or experimenter). There were six items which were assessed: The participants had to report how likely it would be for the following to see the questionnaire: a group of humanities students; a group of science students; or a group of researchers. They then were asked to rate how likely it would be for the following to see the video: a group of humanities students; a group of science students; or a group of researchers. When examining the first item ("how likely is it that a group of researchers will see this questionnaire?") This was negatively correlated with an participants visibility to in-group $r(53)=-0.299$, $p=0.030$ and positively correlated with visibility to out-group $r(53)=0.106$, $p=0.450$ with the greatest positive correlation with visibility to experimenter $r(53)=0.196$, $p=0.160$ which is to be expected with our current manipulation even though the only significant correlation is the negative correlation, however the pattern is in line with the researcher's expectations for the manipulation. When examining the second question, ("how likely is it that a group of humanities students will see this questionnaire?") It is positively correlated with the visibility to in-group $r(53)=0.101$, $p=0.471$ and visibility to out-group $r(53)=0.85$, $p=0.545$ and negatively correlated with the visibility to experimenter, $r(53)= -0.189$, $p=0.175$ although the correlations are not significant, the highest positive correlation is with the visibility to in-group which is what is expected. When examining the third item,

("how likely is it that a group of science students will see this questionnaire?") This is not correlated with the visibility to in-group $r(53)=-0.006$, $p=0.965$, negatively correlated with the visibility to experimenter $r(53)=-0.159$, $p=0.255$ but positively correlated with the visibility to out-group $r(53)=0.163$, $p=0.245$ again although this is not significant it does follow the desired pattern. The fourth item ("how likely is it that a group of researchers will see this video?") was negatively correlated with the visibility to out-group $r(53)=-0.49$, $p=0.729$ and positively correlated with the visibility to in-group $r(53)=0.006$, $p=0.965$ and greatest positive correlation with the visibility to experimenter $r(53)=0.43$, $p=0.759$ again these were not significant but followed the desired pattern. The fifth item ("how likely is it that a group of humanities students will see this video?") was negatively correlated with the visibility to out-group $r(53)=-0.0.16$, $p=0.910$ and visibility to experimenter $r(53)=-0.232$, $p=0.95$ and positively correlated with the visibility to in-group $r(53)=0.244$, $p=0.78$. The sixth item ("how likely is it that a group of science students will see this video?"). This was negatively correlated with the visibility to in-group $r(53)=-0.008$, $p=0.954$ and visibility to experimenter $r(53)=-0.187$, $p=0.181$ but it is positively correlated with the visibility to out-group $r(53)=0.192$, $p=0.168$.

Although only one correlation was significant, all the correlations follow the desired pattern with regard to the visibility conditions therefore it can be argued that the audience manipulation was successful.

A series of one-way ANOVAs were conducted. The independent variable was the audience (i.e. in-group, out-group and experimenter). The dependent variable was the visibility manipulation checks. There were no significant differences between the audience and the likelihood manipulation checks. However, when conducting a post hoc test (LSD) there was a significant difference between the in-group and experimenter audience when answering question 5- The humanities group are likely to see this video. Confidence intervals lower bound= -3.515, upper bound= =0.143, mean difference =-1.765.

Status

This manipulation check assessed the participant's comprehension of the status manipulation and extent to which they were aware of the stereotype. An independent samples t-test was conducted to compare the threat and boost groups. The independent variable was the STL condition (threat, boost). The dependent variable was the STL manipulation check. A composite score was created selecting four of the five items. The Levene's test was not significant indicating that the assumption of homogeneity of variance was not violated. There was a significant difference between the two groups and their scores on the STL manipulation check. (d.f=51, F=2.371, t=-2.123, p=0.039). The mean score for the threat group was 14.074 (S.D.=4.323) while the mean score for the boost group was 16.346 (S.D.=3.393). Since the higher scores on the STL manipulation check indicate agreement with the stereotype that 'humanities students are better than science students at the Raven's APM,' and the boost group have been led to believe that humanities students are superior in the STL task (i.e. the Raven's APM) while the threat group has been led to believe that they are worse at the STL task when compared to science students, these results indicate that the manipulation was successful although the effect size was small.

Stereotype Agreement

This manipulation assessed the extent to which the participants agreed with the stereotype whereas, the previous manipulation assessed the participants' comprehension of the status manipulation, in other words how well they have understood how the stereotype has been explained. An independent samples t-test was conducted comparing the two STL groups and the stereotype manipulation checks. The independent variable was the STL condition (two groups). The dependent variable was the stereotype agreement manipulation checks. Following this a composite score for the scale was created, dropping item two in order to improve the reliability of the scale. The Levene's test was significant therefore equality of variances could not be assumed. There were no significant differences between the two groups and their scores on the composite scale. This demonstrates that there was no difference in stereotype agreement between the groups. This could be due to the low reliability of the measure (0.422), but it could also be due to the fact that simply being made aware of the stereotype, is different from agreeing with the stereotype. It could be argued that the design is weak because the groups do not show agreement with the stereotype. However,

it is generally agreed that this agreement is not required for STL to occur: the mere awareness of the stereotype is enough to create the STL effect (Aronson *et al.,* 1999; Steele, 2003).

> *" Stereotype threat… refers to the strictly situational threat of negative stereotypes, the threat that does not depend on cuing an internalized anxiety or expectancy. It is cued by the mere recognition that a negative group stereotype could apply to oneself in a given situation." (Steele, 1997, p.617).*

Identifiability

As mentioned previously this scale serves both as a manipulation check for the identifiability condition, as well as being able to assess the participants level of identity whether it is more individual or group based. This was done using several paired samples t-tests. Firstly, the results of the identifiability scale during the anonymity condition were compared with the results of the identifiability scale during the identifiability condition. The mean score for the identifiability scale during the anonymity condition was (M=24.51). While the mean score for the identifiability scale during the visibility condition was (M=26.49). There was a significant difference between the scores over the two conditions (t= -2.117, d.f.=52, p<0.039). In this way it could be argued that our identifiability manipulation was successful in that when they anonymity condition (alone in the testing room) their identifiability score was lower than when they were in the identifiability condition (believed that the video camera was filming). When examining the different levels of identity (group or individual) the following was found. When examining individual identifiability there was a significant difference between the anonymity condition and the identifiability condition (t= -2.004, d.f.=52, p= 0.050). The mean score for the individual identifiability subscale during the anonymity condition was (M=12.660). While the mean score for the individual identifiability subscale during the identifiability condition was (M=13.962). There was no significant difference on the group identifiability subscale between the anonymity and identifiability conditions. When examining the individual identifiability subscale with the group identifiability subscale at the identifiability condition there was a significant difference (t= 2.258, d.f.=52, p= 0.028). The mean score for the individual identifiability subscale during the identifiability condition was (M=13.962). While the mean score for the group identifiability scale was (M=12.528). Lastly, when examining the individual identifiability subscale at identifiability condition the with the group identifiability subscale at anonymity condition the there was a significant difference (t= 3.00, d.f.=52, p= 0.004). The mean score for the individual identifiability subscale during the identifiability condition was (M=13.962).

While the mean score for the group identifiability scale at the anonymity condition was (M=11.849). Hence, when the participant is in the anonymity condition they are more likely to feel like part of their social group, while when they are in the identifiability condition they are more likely to assume an individual identity or feel they are being evaluated as an individual. Due to the low reliability of the subscales no further analysis was done.

Excluding practice effects as an explanation for pre and post test differences
However, one problem that had been encountered is that practice effects had not been taken into consideration. Therefore it was needed to demonstrate that these results were not due to the practice effects but due to the audience manipulation.

Hence a second sample was used in order to ascertain if the results of the first sample could be attributed to practice effects. In order to assess for the practice effects a paired-samples t test was used. This compared the results of the Raven's pre-test scores with the Raven's post-test scores. The second sample was only used to for the investigation of practice effects, as the Ravens APM measures were inserted in the wrong order – appended to the end of the materials. The second sample could not be used for further analysis as all the SIT inventory questions were answered before the Raven's APM. However, the second sample was still used to asses for practice effects as the Raven's APM scales were randomised. Therefore the second sample could be used to examine if there was a significant difference between the two Raven's APM scales.

The results of the Paired samples t-test showed that there was no significant difference between the scores on the two Raven's scales. (d.f. =59;t=1.353; p=0.181). Therefore it can be assumed that the results of the first sample were not likely due to practice effects.

The main findings are grouped into three sections. The first finding from the study is that STL status alone has no impact on the performance of a stereotyped task. This may suggest that in order to achieve STL effects more is required than a simple separating into groups with stereotypes about performance (threat, boost). Second, the participants scores on the Raven's APM increased when they believed they were being viewed by an audience compared with when believing they were anonymous i.e. completing the Raven's APM alone in a room. It did not matter which STL condition the participant was in (i.e. threat or boost) -- merely believing that you were visible to an audience increased participants' performance. Thirdly, there were reversal effects of the STL condition performance on two levels: identification to the in-group and anxiety. When in-group identification was high there was the traditional STL effect which was statistically significant. However when identification was low there was a reversal of effects, namely the threat condition outperformed the boost condition. This same effect was mirrored when comparing participants with high and low reported anxiety.

Status

Firstly the researcher examined whether or not status i.e. STL condition (threat, boost) influenced the results of the stereotyped task (the Raven's APM). In order to do this a repeated measures ANOVA was conducted. As this allowed the researcher to investigate the status condition for the two groups across the Raven's APM pre/post test. The independent variables were STL status (threat, boost) and Audience (in-group, out-group or experimenter). The dependent variable was the score on the Raven's APM pre/post-test. Box's M tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups $F(15, 11150.353)=0.639$, $p=0.846$. The Box's M test is not significant, therefore we fail to reject the null hypothesis and conclude that the observed covariance matrices of the dependent variables are equal across groups. Next Mauchly's test of sphericity was used. The null hypothesis is that variances of the differences between conditions are equal. Because our dependent variable is only on two levels, sphericity is not an issue and can be assumed. Next the Levene's test was conducted to assess the equality of error variances. At the Raven's pre-test $F(5,47)=0.359$, $p=0.874$, and at the Raven's post-test

F(5, 47)=2.296, p=0.06, therefore both these tests are not significant and therefore equality of variances can be assumed. Lastly the Shapiro-Wilk test of normality was used this is because within the STL condition (boost, threat) there are fewer than 50 cases. At pre-test and threat the results are as follows d.f.=27, W= 0.899, p=0.13, at pre-test and boost, d.f.=26, W=0.946, p=0.184, at post-test and threat d.f.=26, F=0.170, p=0.35, and at post-test and boost d.f.=26, W=0.173, p=0.366. Although in the threat conditions the results are significant, this is the only assumption which is violated, which is only in the threat condition. However the ANOVA is a robust procedure and will therefore still be used.

There was no significant main effects for the status (F(1,47)= 1.643, p= 0.206; η^2=0.034). There were no other significant results. Therefore this shows that Status alone has no impact. From the results below it can be shown that merely being in a STL condition (threat, boost) did not significantly influence the results on the stereotyped task.

TABLE 3: Descriptive statistics of sample within the six conditions

Descriptive Statistics

	STL	Audience	Mean	Std. Deviation	N
SCALE.RavenPretest	Threat	Ingroup	3.1111	2.31541	9
		Outgroup	3.7000	2.05751	10
		Experimenter	3.8750	2.79987	8
		Total	3.5556	2.30940	27
	Boost	Ingroup	3.2222	2.04803	9
		Outgroup	3.2500	2.31455	8
		Experimenter	3.1111	1.96497	9
		Total	3.1923	2.02028	26
	Total	Ingroup	3.1667	2.12132	18
		Outgroup	3.5000	2.12132	18
		Experimenter	3.4706	2.34834	17
		Total	3.3774	2.15929	53
SCALE.RavenPostTest	Threat	Ingroup	3.7778	2.63523	9
		Outgroup	4.1000	3.41402	10
		Experimenter	3.6250	2.92465	8
		Total	3.8519	2.91816	27
	Boost	Ingroup	3.7778	1.98606	9
		Outgroup	4.5000	3.25137	8
		Experimenter	4.1111	1.96497	9
		Total	4.1154	2.35503	26
	Total	Ingroup	3.7778	2.26367	18
		Outgroup	4.2778	3.25044	18
		Experimenter	3.8824	2.39485	17
		Total	3.9811	2.63476	53

Being watched makes a difference to your performance

In order to investigate whether completing the Raven's APM anonymously or while visible to an audience influenced participants' scores a repeated measures ANOVA was conducted.

Within the repeated measures ANOVA mentioned above the dependent measure was the score on the Raven's APM pre and post-test. The independent measures were the audience condition (in-group, out-group, and experimenter) and the stereotype condition (boost, threat) A significant effect was found on the performance of the Raven's scales. It suggests that on the participants perform better on the post-test than on the pre-test scale. ($F(1,47)= 5.449$, $p= 0.024$; $\eta^2=0.104$). This further suggests that the difference in the Raven's scales may be attributed to the audience manipulation because the status manipulation occurred before the pre-test while the audience manipulation occurred between he pre- and post-test. Since practice effects were ruled out on an equivalent sample, the best explanation for the significant pre/post difference is the audience manipulation. Therefore the participants performed better when they felt as if they are being watched than when they are anonymous and completing the test alone. The mean score on the Raven's APM pre-test was (M=3.378, S.E.=0.310). The mean score on the Raven's APM post-test was (M=3.982, S.E.=0.380).

Reversal effects of STL condition.

These occur on two separate levels: identification with the in-group, and anxiety. First the researcher will examine identification with the in-group further.

From the reviewed literature it was thought that in-group identification may be an important variable when examining for STL. Therefore a median-split at 50 percent was created to distinguish between participants with high in-group identification and those participants reporting low in-group identification. The median split was created using the in-group identification scale which was summed and scored. The histogram below demonstrates the distribution of the sample and which were grouped as high identification with the in-group and which were grouped as low identification with the in-group. This shows that the sample of low and high identifiers is roughly normally distributed and therefore it is unlikely that being categorised in this way will skew the results.
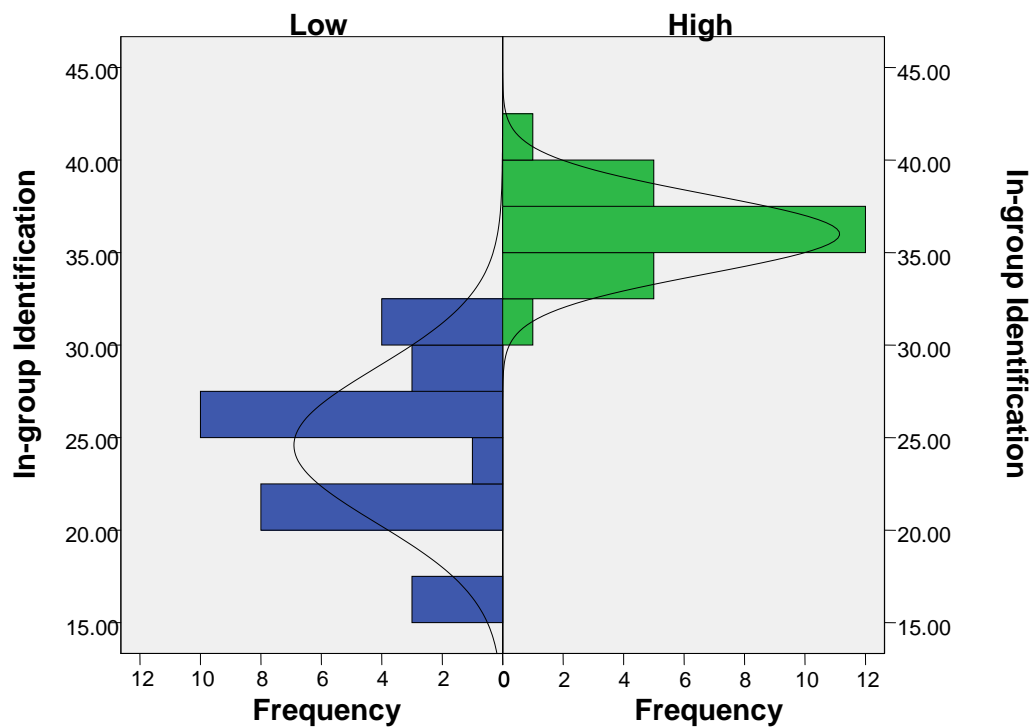

Figure 1: Histogram displaying the distribution of participants across the high and low in-group identification categories

High in-group identification

In order to investigate within a high in-group identification sample if there was a difference in scores on the Raven's APM and whether or not this was influenced by status and audience the participant was visible to The researcher first filtered for high in-group identification before conducting a repeated measures ANOVA. The dependent measure was the Raven's pre-test/post-test. The independent measures were audience (in-group, out-group, and experimenter) and status (lift, threat). Box's M tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups $F(12, 589.975)=$ 1.282, p=0.225. The Box's M test is not significant, therefore we fail to reject the null hypothesis and conclude that the observed covariance matrices of the dependent variables are equal across groups. Next Mauchly's test of sphericity was used. The null hypothesis is that variances of the differences between conditions are equal. Because our dependent variable is only on two levels, sphericity is not an issue and can be assumed. The Levene's test was significant for the Raven's APM pre and post-test therefore equality of variances cannot be assumed ($F(5, 18) = 4.750$, p = 0.006. $F(5, 18) = 2.905$, p = 0.043). Lastly the Shapiro-Wilk test of normality was used this is because within the STL condition (boost, threat) there are fewer than 50 cases. At pre-test and threat the results are as follows d.f.=11, W= 0.718, p=0.001, at pre-test and boost, d.f.=13, W=0.940, p=0.459, at post-test and threat d.f.=11, W=0.940, p=0.518, and at post-test and boost d.f.=13, W=0.963, p=0.794. Only one of the conditions is significant at pre-test threat, therefore normality cannot be assumed. However because the ANOVA is generally robust, it will still be used. No significant results were found for the repeated measures ANOVA when sampling for high in-group identification. Therefore across the two tests there were no significant differences. It is noteworthy that high identifiers were not affected by the audience manipulation. However, when examining the between-subjects effects, it can be shown that there is a significant main effect for status ($F(1,18)=5.563$, p= 0.030; $\eta^2=0.236$). Therefore the boost group performs significantly better than the threat group (see figure 2).

TABLE 4: Descriptive statistics of sample when sampling for high in-group identification

Descriptive Statistics

	STL	Audience	Mean	Std. Deviation	N
SCALE.RavenPretest	Threat	Ingroup	2.0000	1.00000	3
		Outgroup	2.0000	.00000	5
		Experimenter	2.6667	1.15470	3
		Total	2.1818	.75076	11
	Boost	Ingroup	4.2000	2.16795	5
		Outgroup	3.8000	2.28035	5
		Experimenter	2.0000	2.00000	3
		Total	3.5385	2.18386	13
	Total	Ingroup	3.3750	2.06588	8
		Outgroup	2.9000	1.79196	10
		Experimenter	2.3333	1.50555	6
		Total	2.9167	1.79169	24
SCALE.RavenPostTest	Threat	Ingroup	2.6667	1.15470	3
		Outgroup	1.0000	1.22474	5
		Experimenter	2.0000	1.00000	3
		Total	1.7273	1.27208	11
	Boost	Ingroup	4.4000	2.19089	5
		Outgroup	4.8000	3.70135	5
		Experimenter	3.3333	1.52753	3
		Total	4.3077	2.62630	13
	Total	Ingroup	3.7500	1.98206	8
		Outgroup	2.9000	3.28126	10
		Experimenter	2.6667	1.36626	6
		Total	3.1250	2.45503	24

Low in-group identification

In order to investigate within a low in-group identification sample if there was a difference in scores on the Raven's APM and whether or not this was influenced by status and audience the participant was visible to, , a repeated measures (ANOVA) was conducted. This was conducted the same as the above analysis the sample was first filtered for low in-group identification. The dependent measure was the Raven's pre-test/post-test. The independent measures were audience (in-group, out-group, and experimenter) and status (lift, threat). Box's M tests the null hypothesis that the observed covariance matrices of the dependent

variables are equal across groups $F(15, 1086.302)= 0.902$, $p=0.526$. The Box's M test is not significant, therefore we fail to reject the null hypothesis and conclude that the observed covariance matrices of the dependent variables are equal across groups. Next Mauchly's test of sphericity was used. The null hypothesis is that variances of the differences between conditions are equal. Because our dependent variable is only on two levels, sphericity is not an issue and can be assumed. The Levene's test for the Raven's pre and post test were both not significant therefore equality of variances can be assumed. Lastly the Shapiro-Wilk test of normality was used this is because within the STL condition (boost, threat) there are fewer than 50 cases. At pre-test and threat the results are as follows d.f.=16, W= 0.976, p=0.920, at pre-test and boost, d.f.=13, W=0.951, p=0.609, at post-test and threat d.f.=16, W=0.950, p=0.496, and at post-test and boost d.f.=13, W=0.498, p=0.569. None of these results are significant therefore normality can be assumed. There is a significant difference between pre test and post test Ravens ($F(1,23)= 8.931$, $p= 0.007$; $\eta^2=0.280$.) indicating a substantial audience effect. Within there are no other significant main or interaction effects when sampling for low in-group identification.

TABLE 5: Descriptive statistics of sample when sampling for low in-group identification

Descriptive Statistics

|	STL	Audience	Mean	Std. Deviation	N
SCALE.RavenPretest	Threat	Ingroup	3.6667	2.65832	6
		Outgroup	5.4000	1.51658	5
		Experimenter	4.6000	3.36155	5
		Total	4.5000	2.55604	16
	Boost	Ingroup	2.0000	1.15470	4
		Outgroup	2.3333	2.51661	3
		Experimenter	3.6667	1.86190	6
		Total	2.8462	1.86396	13
	Total	Ingroup	3.0000	2.26078	10
		Outgroup	4.2500	2.37547	8
		Experimenter	4.0909	2.54773	11
		Total	3.7586	2.38530	29
SCALE.RavenPostTest	Threat	Ingroup	4.3333	3.07679	6
		Outgroup	7.2000	.83666	5
		Experimenter	4.6000	3.36155	5
		Total	5.3125	2.84532	16
	Boost	Ingroup	3.0000	1.63299	4
		Outgroup	4.0000	3.00000	3
		Experimenter	4.5000	2.16795	6
		Total	3.9231	2.13937	13
	Total	Ingroup	3.8000	2.57337	10
		Outgroup	6.0000	2.39046	8
		Experimenter	4.5455	2.62159	11
		Total	4.6897	2.60636	29
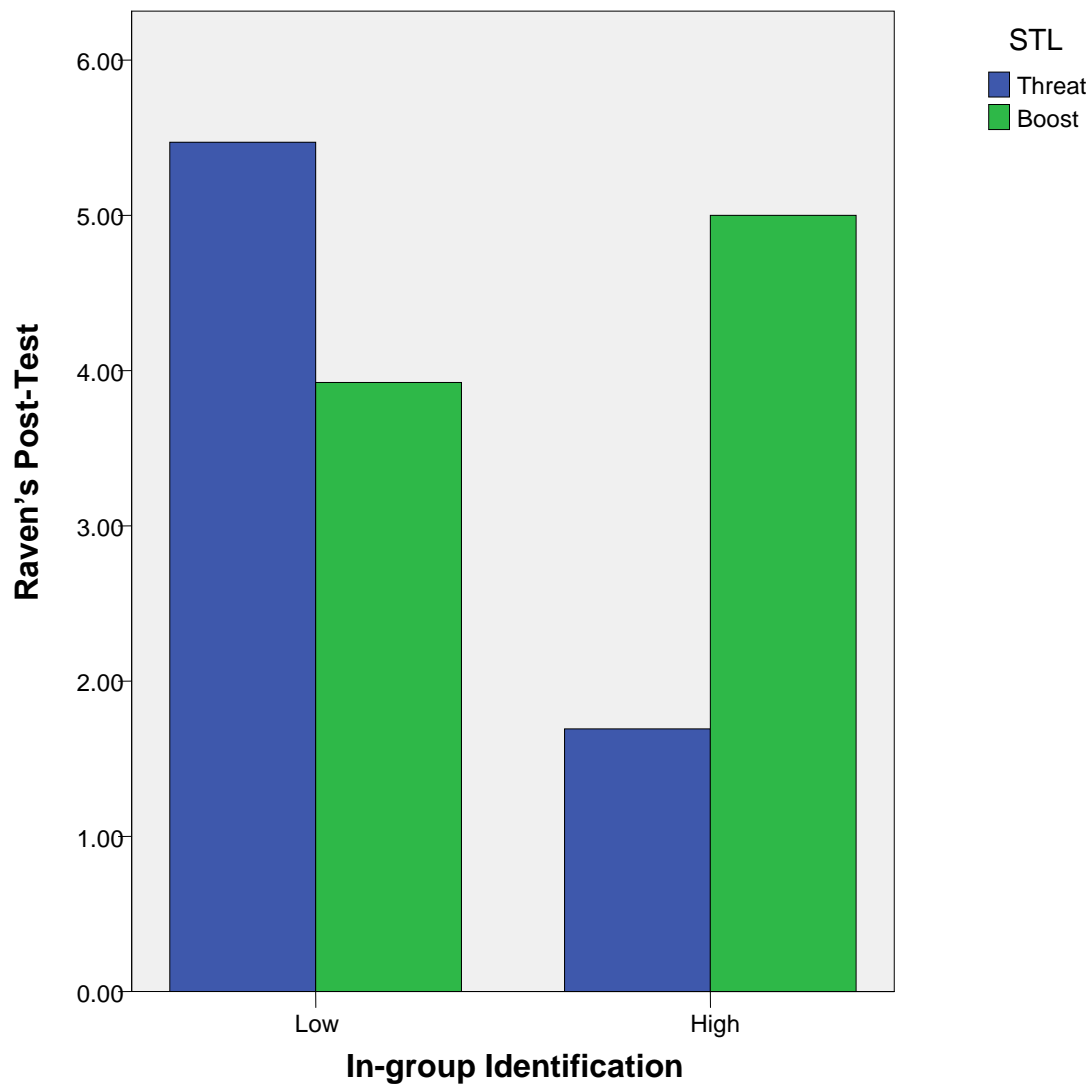
Figure 2: Bar graph showing Raven's post-test scores for threat and boost groups at low and high in-group identification.

Figure 2 shows that at low in-group identification the threat group perform the best, even better than the boost group at high in-group identification. The boost group at low in-group identification performs worse than the threat group as well as the boost group at high in-group identification. However the boost group at low in-group identification performs better than the threat group at high in-group identification.

Anxiety

In order to investigate whether differing self-reported levels of anxiety influenced the scores on the Raven's APM with regard to the status of the participants. A median split was created for the anxiety scale. A (2x2x3) Factorial ANOVA was conducted. The dependent variable is Raven's post-test, the independent variables are anxiety (low, high), status (boost, threat) and audience (in-group, out-group and experimenter). The Levene's test is not significant $F(11, 41) = 1.208$, $p = 0.313$. Therefore we fail to reject the null hypothesis of equality of variances and conclude that the assumption of homogeneity of variances is not violated. Lastly the Shapiro-Wilk test of normality was used this is because within the STL condition (boost, threat) there are fewer than 50 cases. At the Raven's post-test for both the in-group and out-group audience the results are significant d.f.=18, W=0.856, p=0.011; d.f=18, W=0.874, p=0.021, therefore normality cannot be assumed however they were not significant at the experimenter audience d.f.=17, W=0.906, p=0.086. Although two of the conditions were significant, the ANOVA is relatively robust test and therefore can still be used, although in this case the results should be treated cautiously. No significant main effects were found but there were significant interaction effect between status, audience and anxiety. $(F(2,41)= 3.362$, $p= 0.044$; $\eta^2=0.141)$.
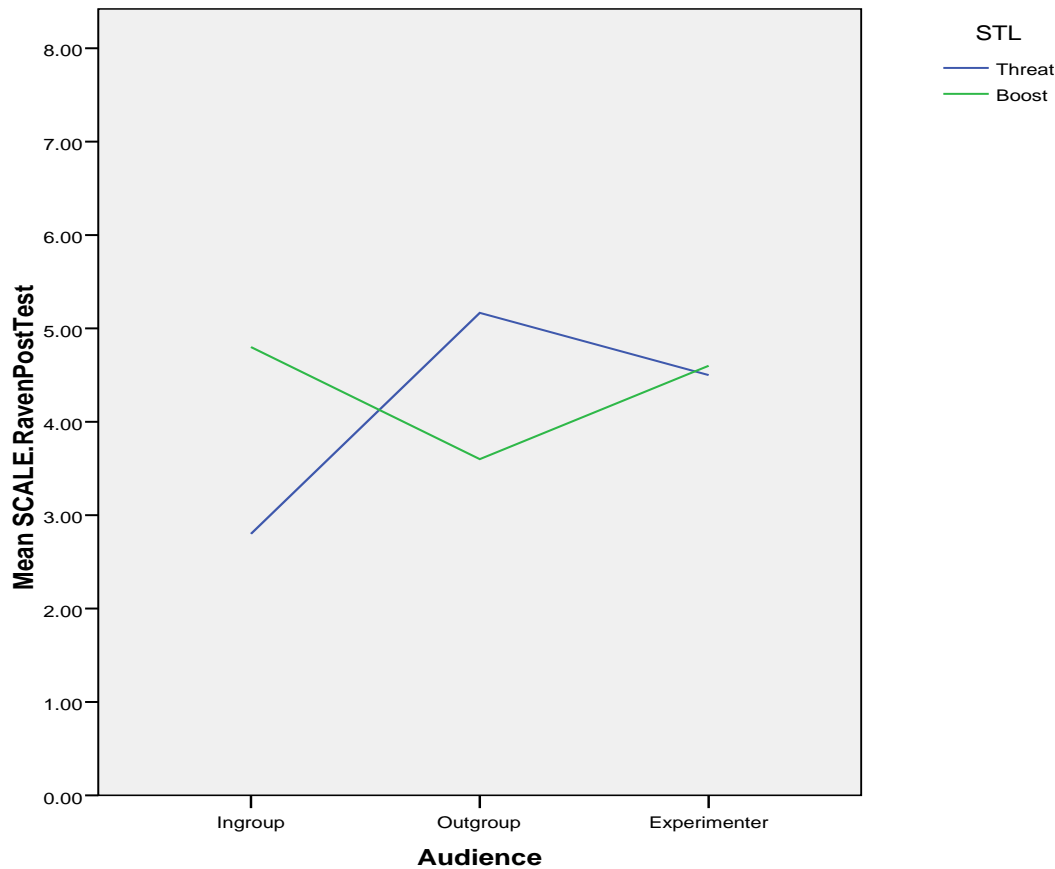
Figure 3: Graph showing Raven's post-test scores for threat and boost groups at low anxiety.

This graph shows that when low anxiety is reported by participants, then there is a reversal in the STL effect when participants believe they are visible to the out-group. That is, the threat group performs better than the boost group when visible to the out-group. When low anxiety is reported, those under threat performed best when visible to the out-group,(M=5.167, SE=1.107) and those under boost performed worst when visible to the out-group (M=3.600, SE=1.213). There is hardly any difference between the threat and boost groups when they are visible to the experimenter (M=4.5, SE=1.918; M=4.6, SE=1.213). Interestingly, those under threat group performed worst when visible to the in-group (M=2.8, SE=1.213), while those in the boost condition performed best when visible to the in-group (M=4.8, SE=1.213). It is only when the groups are only visible to the in-group that the traditional STL effect is observed.
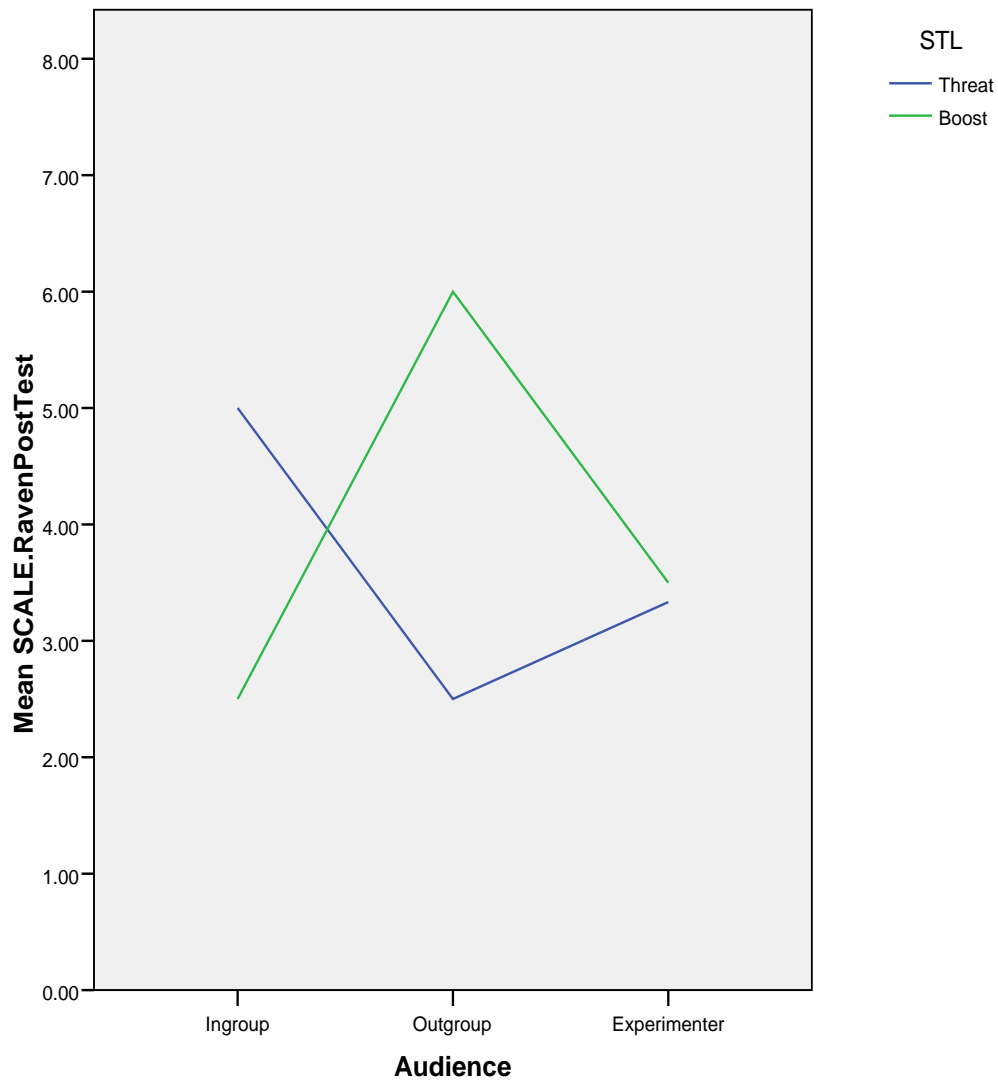
Figure 4: Graph showing Raven's post-test scores for threat and boost groups at high anxiety.

Figure four reflects APM post-test scores for those who reported high anxiety, and is almost the mirror image of the graph for those reporting low anxiety. This graph shows that when high anxiety is reported by participants, then there is the traditional STL effect only when they are visible to the out-group (boost) (M=6, SE=1.566), while those under threat perform worst when they are visible to the out-group (M=2.5, SE=1.356) and there is a reversal of the effect (in which those under threat

perform best) when visible to the in-group (M=5, SE=1.356). There is hardly any difference between the groups when they are visible to the experimenter (M=3.5, SE=1.356;M=3.333, SE=1.107) . In direct contrast to the pattern observed for those reporting low anxiety, the threat group performs best when visible to the in-group, while the boost group performs worst when visible to the in-group (M=2.5, SE=1.356).

Discussion

From the results it can be shown the status condition alone i.e. (boost, threat) has no significant effect on performance with regard to performance on the Raven's APM. Secondly it has been shown that regardless of the status condition to which the participant is assigned, their performance improves when they feel they are visible to an audience compared to when they are anonymous.

Lastly reversal effects were shown when adding in-group identification and anxiety into the model. When examining in-group identification, in particular the differences between high in-group identification and low in-group identification, it was demonstrated that participants that experienced high identification with their in-group experienced the traditional STL effects. However participants who experienced low identification with their in-group experienced the reverse effects of this, even though it is not significant, it was interesting to note these reversal effects where the threat group performed better than the boost group on the stereotyped task- the Raven's APM. This same effect was noted when examining self reported anxiety, the self-reported high anxiety sample experienced the traditional STL effects, whereas the low anxiety sample experienced the reversal effects, where once again the threat group outperformed the boost group. It is interesting to note that there were no significant effects for the audience which the participant was visible to in this comparison. The low in-group identification reversal finding is of particular interest, as traditional interventions such as blurring group boundaries to protect against the effects of STL may not help the low in-group identified participants and they may perform worse than if there was no intervention to begin with.

Status alone had no effect

The results of the study indicated that status alone had no effect on the performance of the Raven's APM. It was only when in-group identification was considered as part of the model that there were significant effects for status. This compares well with the results of Schmader (2002) who found that status alone had no effect, but when gender identification of the participants there were significant effects on a gender stereotyped maths task. This adds to the literature which suggests the stereotype threat is situated within a broader category of social identity threats. In this way more aspects of social identity should be considered when examining STL.

Scores improved when watched regardless of the STL condition

Secondly it has been shown that regardless of the status condition to which the participant is assigned, their performance improves when they feel they are visible to an audience compared to when they are anonymous. This means that the very fact that the participants feel they will be observed is more salient than the status condition. This can be explained by social facilitation theory whereby the presence of an audience (even passive) may improve performance (Zajonc, 1965; Bond & Titus, 1983). In the threat and boost groups when they feel they are being watched their results improve. However it should be noted that this is also partly because the reversal effects identified actually mask the STL effects in the combined model.

The participants scored significantly higher on the Raven's APM subtest when they were highly identifiable, compared to when they were anonymous members of their group. Therefore we would conclude that when a participant is highly identifiable they will perform better than when they are an anonymous member of a group (Forbes, 2007; Worchel *et al*., 1998). Ambady *et al.* suggest that, "if the salience of group identity is replaced by the salience of individual identity, the risk associated with negative stereotype activation might be attenuated and performance altered to reflect more accurately the unique capabilities of the individual rather than the stereotypes of the group" (Ambady *et al.,* 2004, p.402).

When the participants were anonymous members of their groups they performed significantly worse than when they were highly identifiable. Again this is consistent with the findings of Forbes (2007) and Worchel *et al*. (1998). In the study by Barreto and Ellemers (2000) "Low

identifiers do not pursue group goals when anonymous, accountability to the in-group may present a temporary motivation to do so" (Barreto & Ellemers, 2000, p.903). This refers to the fact that if the group member is accountable to the in-group they may have some motivation to pursue the group goals. These results co-incide with previous social identity and social loafing research (Doosje, Ellemers & Spears, 1995; Ellemers *et al.,* 1997; Jetten *et al*., 1997, cited in Barreto & Ellemers, 2000). This current finding suggests that when an individual is identifiable their productivity/work rate is improved regardless of the STL condition. In this way when they are identifiable it is suggested that they move away from their social identity, and focus on their individual identity. It has been argued that being viewed as an anonymous part of a group may lead to social loafing (Worchel *et al*., 1998). The results suggest that because of their improvement when identifiable, as individuals they are more inclined to be seen in a positive light and in this way would have more motivation or an investment in their individual identity to perform better on the Raven's APM than when they are alone in the testing room completing the Ravens' APM as a anonymous member of their social group (Forbes, 2007). "Being able to hide in the group reduces personal concerns about being evaluated, enabling the individual to reduce effort without facing censure" (Worchel *et al*., 1998, p 403 cited in Forbes, 2007).

Audience effects

It was interesting to note that there were no differences in audience effects i.e. whether being visible to a group of humanities students, science students or group of researchers and the participants performance on the Raven's APM. However merely the belief that you may be visible to one of these three groups did lead to a significant increase in performance on the Raven's APM when compared with when they were completing it alone in the testing room (anonymous). Therefore merely the belief that you are being watched is enough to increase performance. Finding that there were no difference in results could be due to the fact that the audience effects were coupled with the fact that the participant was personally identifiable as well as being identified as part of their social group. It could be argued that if they were only identifiable as a member of their group but visible to an audience it is possible that there may have been an influence on the audience effects. Perhaps the fact that they were personally identifiable to the audience meant that there was less pressure on them being evaluated as a group member and hence it did not matter which audience the participant was visible to.

Most of the studies examining audience effects towards the out-group and in-group, including those of anonymity and identifiability have been within the realm of group processes and intergroup relations with regard to social identity theory. This is the first study to the author's knowledge that takes the ideas from SIDE specifically anonymity, identifiability and audience effects and applies it to a stereotype threat context. When examining these aspects as highlighted by previous studies (e.g. Reicher & Levine, 1994; Reicher 1998; Lea, 2001; Klein, 2003; Klein, 2007). The results of this study did not match the findings of previous studies. According to the study by Lea (2001) greater visibility was linked to a greater expression of group norms. Therefore, it should follow that if you are visible then your results should more accurately reflect the group norms, more specifically your results should reflect whether your group has been negatively or positively stereotyped. From the Klein (2003) study and study three from Jetten, Hornsey, and Advares-Yorno (2006), one would expect that if you are visible to your in-group then you would display more in-group normative behaviour, i.e. if you are positively stereotyped your results should be better when viewed by your in-group than when view by your out-group.

 Along with these findings Klein *et al.* (2007) comments on the recent work on "respect" in SIT, from this it could be argued that when watched by the in-group it is important to perform according to stereotyped ways in order to earn the respect of the group, particularly if group identification is important. Klein *et al.* (2007) also suggests from the review that when you have been negatively stereotyped and are visible to the out-group it is also import to perform well. However, it should be noted that, the current study did not use group norms but used actual performance on the Raven's APM which is a IQ test, therefore the ability for the participants to perform in accordance with their group is dependent on their existing academic ability.

There are several scenarios which could possibly explain these patterns of results in STL research. It is more likely that when there is a belief about being watched by the in-group that the participants will perform in stereotypical ways in order to gain in-group approval. Secondly in terms of the out-group, it is more likely that participants would perform better when viewed by the out-group, as their social identity may be under more criticism, or that they have "more to prove" to the out-group regardless of the STL condition. However there

were no significant differences between the different audiences that the participants were exposed to and their results on the Raven's APM.

This could be for the following reasons. The first reason is that the audience manipulations, in terms of their distinctions (i.e. in-group, out-group, and experimenter) may have been unsuccessful. It is possible that experimentally the researcher was not able to create the effect that the participants truly felt that they were identifiable to distinct audiences. Secondly, it may just be that merely being identifiable to an audience is enough to lead to an increase in performance as shown, and that whom the participant is identifiable is unimportant as they are always trying to protect their identity whether it is their group or personal identity it does not matter. In this way it could be suggested that if they were identifiable to an out-group, then they would have to perform well in order to protect their in-group status and or superiority, while if they were identifiable to the in-group they would have to perform well in order to demonstrate to order in-group members that they too belong to their particular social group.

Audience effects when examining anxiety

There was a significant interaction effect between audience and anxiety when examining the results on the Raven's APM. However, it should be noted that these results are of the Raven's post-test only. At low anxiety there is the reversal in STL effects, this is where the threat group outperforms the boost group. When examining the audiences the participants are visible to the following is noted: threat performs best when visible to the out-group, (I'll show them attitude) and boost performs worst when visible to the out-group (no pressure). There is hardly any difference between the threat and boost groups when they are visible to the experimenter. Interestingly the threat group performs worst when visible to the in-group, while the boost group performs best when visible to the in-group. It is only when the groups are only visible to the in-group that the traditional STL effect is observed. This shows evidence that the groups will act in group normative behaviour, even though they have low anxiety, expectations of audience may induce expected behaviour. These results are similar to the findings of Lea (2001) and Klein *et al.,* (2003) which have been discussed above, that increased visibility will lead to group normative behaviour.

At high anxiety then there is the traditional STL effect only when the participants are visible to the out-group, while those under threat perform worst when they are visible to the out-group and there is a reversal of the effect (in which those under threat perform best) when visible to the in-group. This demonstrates under high anxiety threatened participants will perform best when visible to their in-group, perhaps suggesting that under high anxiety there is more at stake in terms of your group membership and needing your group to be evaluated positively. However, the boost group has no such pressure to perform and participants can therefore "hide in the group". When visible to the out-group and under high anxiety the participants act in group normative ways.

In-group Identification. The most exciting find from the study is the reversal effects of STL with regard to in-group identification as well as anxiety. Firstly we will examine the current study's results of STL and in-group identification with what has been reported in the literature. As indicated in the results section when the median-split was created for the in-group identification scale, there were interesting results for those participants who did not identify strongly with their in-group and therefore had low in-group identification compared to those participants who did identify strongly with their in-group and therefore had high in-group identification. When examining the results of the Ravens APM between the high in-group identification sample and the low in-group identification sample the results are as follows.

The stereotype threat group performed significantly better on the Ravens APM when the participants reported low in-group identification than when the participants reported high in-group identification. The stereotype boost group performed significantly better when the participants reported high in-group identification than when the participants reported low in-group identification. These increases in performance may be attributed to the level of in-group identification. When there is low in-group identification, the participants may not feel strongly attached to their social group and are therefore less affected by the stereotypes around the social groups. It this way low in-group identification may be protective against stereotype threat and encourage participants to perform better. However when there is high in-group identification, the participants may feel strongly attached to their social groups and therefore may perform in a stereotyped manner. Therefore in the stereotype threat group participants performed better than the stereotype boost group when the participants had low in-group identification but when the participants had high in-group identification the boost group performed significantly better than the threat group. The groups were only affected by the STL effect when there was high in-group identification. Our study found that high in-group identification led to the traditional STL effects while low in-group identification displayed a reversal in these effects.

A group of studies by Spears, Doosje and Ellemers (1997) examined the effect of threats to the status of the group with regard to self-stereotyping. Status of the group with relation to self-stereotyping was defined as how similar people felt towards other in-group members. In this way, self-stereotyping was more frequent for high identifiers than for low identifiers. Although the studies are not related to performance rather stereotyping the evidence supports the findings of the current study. The hypotheses that low identifiers would distance themselves from the group when under threat and high identifiers would stick to the group when under threat found some support in three of the four studies (Spears *et al.,* 1997). "Low identifiers in effectively distancing themselves from the group or its central tendency can be seen as relatively individualistic identity protection strategy. High identifiers on the other hand, is [sic] more collectivistic and loyal to the group" (Spears *et al*., 1997, p.550).

 This, it can be argued is what happened in the current study, by being highly identified you are more committed to the group and therefore will act with the normative behaviour or the way in which the group has been stereotyped that is demonstrate commitment to the group by performing in the stereotypical ways as the group has been stereotyped to.

Whereas when the group is low identified they are less committed to the group, feel more distant from the group and therefore are less compelled to perform in those stereotypical ways. Because there is less emphasis on the group the participants can be seen as protecting their own threatened identity in this way negatively stereotyped individuals performed better than the positively stereotyped individuals, this could be argued that the threatened individuals had something to prove to themselves or others through their performance while the boost group (positively stereotyped group) had nothing to prove and in this way their performance had less meaning to themselves and their identity as they are already seen in a positive light (Forbes, 2007).

The same reversal effects in in-group identification can be shown from Schmader, (2002). What was demonstrated was that women who were highly identified with their gender performed almost as well as the men when the stereotype was not made salient. However when the stereotype was made salient low identified males and females performed almost the same within the test. The highly identified women performed significantly worse than the highly identified men (Schmader, 2002). Hence "women showed poorer performance compared to men on a stereotype relevant task when their social identity was linked to their test performance, but only if they considered gender to be an important part of their self-definition" (Schmader, 2002, p.199).   In this way this study also supports the findings of the current research. When gender was important to the women, the women performed poorly and in doing so conformed to the stereotype whereas when it was not the participants performed almost the same. When gender was viewed as salient it was "on the line in the testing situation" (Schmader, 2002, p.199).

This demonstrates that being negatively stereotyped in a particular task is not sufficient to predict or alter one's performance. It is rather a combination of the threat and one's individual and social identity characteristics. Therefore your level of identification with your in-group would influence how you experienced the STL effects. If you were high identified then you would experience the traditional boost, threat scenario, if your group was positively stereotyped individuals would perform better than the negatively stereotyped individuals on the Raven's APM. If we examine the low in-group identification participants positively stereotyped group performed worse than the negatively stereotyped group.

Anxiety. Two studies Ford, Ferguson, Brooks and Hagadone (2004) examined the effect of humour in a stereotype threat situation. The authors were able to demonstrate in their second study that women under stereotype threat who had a coping sense of humour fared better on a maths test than those who did not. There was also no difference in performance on the maths test in the control group with regards to a coping sense of humour (Ford *et al.*, 2004). "Mediation analyses suggest that in the stereotype threat condition, state anxiety mediated the relationship between coping sense of humour and test performance. Women higher in coping sense of humour performed better because they felt less anxiety while taking the test" (Ford *et al.*, 2004, p.643).

These results support the findings of the current research that when anxiety is low the threat group performs better, while when there is high anxiety there is the traditional STL effect. This is similar to in-group identification findings. One explanation to this is that if the participants are very anxious they are therefore more concerned about their performance on the task as it has more relevance or meaning for them. As Steele *et al.* (2002) suggested the more self relevant the task the more the STL effects will be present. In this case this explains why the in-group identification and anxiety results mirror each other. It could also be suggested that high anxiety and high in-group identification can be viewed as markers of self-relevance for the task and therefore the stereotype effects will be more salient and stronger for those groups. For the groups where anxiety and in-group identification is low it could be argued that the STL is less self-relevant and in this way the effects are viewed as less salient for social identity, but may still be relevant to personal identity. Wout *et al.,* (2008) examined the source of identity threats in particular the difference between self and group threats. When there was a group threat presented to the women, the women that were highly identified with their gender underperformed compared with women who were less identified with their gender who were not affected by the group threat (Wout *et al.,* 2008).

Two other studies from Osbourne show similar effects Osborne (2001), wanted to examine whether or not anxiety played a role in explaining the test performance differences of participants who differed in terms of gender and race. Osborne (2001) used a sample from a high school data base and performed a number of analyses. What was found was that anxiety (self-reported measure) was significant in explaining the differences in performance on the test between whites and African Americans and whites and Latinos.

Osborne (2006), wanting to explain anxiety more holistically in terms of its relationship with STL sought to use some physiological measures when assessing anxiety. Male and female university students were sampled and placed into either a high or low stereotype threat conditions and given a maths test to complete. Osborne found that females who had been negatively stereotyped (i.e. high stereotype threat condition) exhibited more signs of anxiety from the physiological measures (skin conductance, surface skin temperature and diagnostic blood pressure) (Osborne, 2006). In this way the participants who were more threatened by the stereotype displayed more anxiety during the task.

O'Brien and Crandall (2003) examined the effects of arousal on stereotype threat. The authors predicted that participants under stereotype threat would perform poorly on difficult tasks and that their performance would improve on easier tasks when they are compared with participants who are not faced with the stereotype. Within the stereotype threat condition women performed better on the easy test compared to the control group. This could suggest that the women performed better as they were less anxious about their performance and therefore would not perform in the way the stereotype suggests. This supports the findings of the current research where low anxiety led to reversal effects in STL, that the threat group out-performed the boost group. When examining the difficult test the women performed worse on the test compared to the control group. Again a supporting finding for the current research where under high anxiety there is the traditional STL effect. Within the control group for the women there was hardly any difference in their test scores. There were no significant differences for men when comparing the stereotype threat men to the men in the control group. The difference in test scores for between men and women in the stereotype threat condition was more than double than the difference between men and women in the control group.

STL as strategic group membership investment

When reviewing the results of the participants in-group identification it could be shown that stereotype threat could be a strategic group membership investment. If in-group identification is high, this means that your social identity and investment in the group is high. Hence, you are more likely to act like a typical group member under high in-group identification. *"High identifiers on the other hand, is [sic] more collectivistic and loyal to the group"* (Spears *et al.,* 1997, p.550). If the group membership is viewed as being most important, then the group members will act in accordance with the groups social identity needs (Barreto & Ellemers, 2000). However, if your in-group identification is low, you would therefore have less investment in your group, and in this way you would be less likely to act as a typical group member especially when your identity (group or individual) is threatened. "Low identifiers in effectively distancing themselves from the group or its central tendency can be seen as relatively individualistic identity protection strategy (Spears *et al.,* 1997, p.550). Wout *et al.,* examined group versus individual threats and found that: when there was a group threat presented to the women, the women that were highly identified with their gender underperformed compared with women who were less identified with their gender as they were not affected by the group threat (Wout *et al.,* 2008). To summarise if your in-group identification is high, you are more likely to act in accordance with other group members and be respected by the group whereas if your in-group identification is low you have less of a commitment to be accepted by the group and would therefore act in your individual capacity. It is argued that "high identified group members are likely to be concerned with the welfare of the group regardless of the circumstances, whereas low identifiers are only persuaded to do so out of self-presentational concerns" (Barreto & Ellemers, 2000, p.891).

The first limitation of the study which is noted in the methodology section is that the researcher did not include the second status manipulation before the Raven's APM post test. This omission impacts on how much can be said or inferred about the STL effect within a repeated measures design. However, this does not impact substantially on the main findings. Including the second status manipulation would ensure that the STL effect was carried over throughout the repeated measures design. However, it can be assumed that the STL effect would carry over the pre and post-test conditions as there is no indication for the participants to neglect the initial information supplied in the first status manipulation.

A second limitation or shortcoming of the current study was in the creation of the two Raven's APM subtests. Originally the Raven's APM is structured from items 1-36. These items increase in difficulty and you are able to see patterns and learn from the previous items. When the Raven's were divided into two similar subtests of 12 items each which excluded the six easiest and hardest items they were not ordered in terms of easiest to hardest. They were in fact mistakenly ordered from hardest to easiest. This might explain the poor performance overall in the two Raven's APM subtests.

A third limitation of the study is that there was no use of a control group. In terms of research design including a control would have been the strongest design. It would also allow the researcher to be able to explain more around what exactly happens at stereotype threat and lift compared to a group that has not been introduced into the experimental condition. However it was thought that it was not possible to include a control group because this would require more participants and funding. The control group would have to complete the Raven's APM the same way as it was done by the experimental group (threat, lift). Also it would prove experimentally very hard as the participants would have to be told as much as the other participants within the threat and lift groups within inducing the stereotypes. Practically it was seen as too difficult to conceptualise and complete operationally in the given timeframe for the experiment. However it should be noted that although it was difficult to add to this study it would have been a desired inclusion.

A fourth limitation of the study is the small sample size. 53 participants across six conditions is low for an experimental study. It is suggested that a minimum of ten participants per condition, however it is recommended that twenty be used. This would increase the statistical power and decrease the chance of statistical errors.

A fifth limitation of the study is that the first Raven's APM was administered after the assignment of the participants to the threat and boost conditions. This oversight meant that the power advantages of the within-groups design were limited to the identifiability manipulation, while differences resulting from the STL manipulation could only be inferred from between-groups differences. This substantially increased the chance of type II error for the STL manipulation.

A further problem with most STL studies is that they are largely based on self-reported measures (Smith, 2004). It is argued that this may not be the best way to capture the information required to assess the effects of STL on performance. Non-self reported measures that have been used the findings have been mixed this may be due largely to problems within the methodology (Smith, 2004). Having said this, there has been a growing interest in social neuroscience and its application to STL (Derks, 2008; Schmader, Johns & Forbes, 2008). The criticisms are that self-reported measures do not capture the notion of STL fully, and that self-reported measures are in themselves limited for what one can say about a given construct. An example would be stress, is it better to have a self-reported measure of stress or measure the physiology of the person for signs of stress? Findings with regard to STL when using self-reported measures and non-self reported measures are mixed. In this way both ways of gathering data are not without there weaknesses and limitations. Self-reported measures were chosen because of the experimental context in which the participants were in they are inexpensive to use, they are non-invasive, easier to administer and less time consuming than non-self reported measures this is particularly important since the Raven's APM alone would take the participants half an hour to complete.

Conclusion

 The current study has demonstrated that being viewed as an individual or at least having the belief of being visible to others rather than being anonymous impacts on the performance of a task. In the current study when participants thought they were going to be watched they performed better on the Raven's APM than when they thought they were anonymous and their results were not linked to them. Secondly being watched or not was more important that the stereotype condition the participants were in. Whether they were in the threat or boost condition did not matter as all the participants results improved when they thought there results would be viewed by others. In this regard the way in which the participants view themselves or feel they are being viewed (evaluated) may well impact on their performance despite the stereotype condition. Perhaps there is then more scope for examining STL through the lens of social identity more carefully.

A further interesting finding from the results in terms of how social identity relates to STL is examining in-group identification. This showed that when in-group identification was more salient for participants STL effects were experienced in the performance on the Raven's APM. This result compares well with previous studies as mentioned previously by (Aronson *et al*., 1999, Schmader, 2002; O'Brien & Crandall, 2003). Therefore the current study once again highlights the importance of in-group identification when examining STL effects.

Examining Identifiability and how it relates to STL was the main aim of the study. Even though the were no clear differences in results according to which audience participants were visible to, the study did demonstrate that identifiability is an important construct when considering performance on a task. As mentioned above, regardless of the STL condition the participants performed significantly better when they thought that they were visible to a group.

Although there were limitations of the study, mostly due to the design and methodology which limited how much of the data could be examined, how it was examined and interpreted the findings from the current study are still important as it adds to the understanding of STL in terms of social identity theory. However I believe that more work can be done in this regard, as it was demonstrated that the belief of being watched or evaluated later by others was enough to increase the performance of the participants in the Raven's APM. Therefore

there must be something here, evaluation of other, managing one's reputation or identity in terms of others, or conforming to the expectations of others.

It was also interesting to note within the results that there were no significant findings when examining audience effects within the repeated measures design. However there seemed to be a distinct pattern in performance on the Raven's APM when the participants thought they were being watched by in-group members compared with out-group members. This is definitely an area which should be explored more in future research, and would also provide more insight and links between social identity theory and STL, particularly when examining the Social identity Model of Deindividuation effects (SIDE).

Other potential future research areas within identifiability and STL may include studies which more closely examine participants' group and personal identity. It would be interesting to examine whether the amount of identifiability the participant experiences relates to the identity which they adopt, in other words, does being seen as an anonymous part of a group mean that you would automatically operate from your group identity in terms of performance on certain tasks, and therefore if you are seen as an individual does that mean you would automatically operate from your personal identity in terms of performance on certain tasks?

STL is a performance modifier which is situational. This complex phenomenon has attracted much research in the last 14 years. Still not much can be said for what is exactly happening when someone is confronted with a stereotype and how this leads to either an increase or decrease in performance depending on the condition. Having said that it is now being understood more and more in terms of a form of social identity threat, therefore the more research within this area the better it is understood. It is hoped that this current study as modest as it is, may have added to the current research and understanding of STL in terms of social identity theory.

References

Alderton, D. L. & Larson, G. E. (1990). Dimensionality of raven's advanced progressive matrices items. *Educational and Psychological Measurement, 50*, 886-900.

Ambady, N., Paik, S. K., Steele, J., Owen-Smith, A. & Mitchell, J. P. (2004). Deflecting negative self-relevant stereotype activation: The effects of individuation. *Journal of Experimental Social Psychology, 40*, 401-408.

Aronson, J., Fried, C. B. & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*, 113 - 125.

Aronson, J., Lustina, M. J., Good, C., Keough, K. & Steele, C. M. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*, 29-46.

Arthur, W. & Day, D. V. (1994). Development of a short form for the raven advanced progressive matrices test. *Educational and Psychological Measurement, 54*, 394- 403.

Barreto, M. & Ellemers, N. (2000). You can't always do what you want: Social identity and self-presentational determinants of the choice to work for a Low-Status Group. *Personality and Social Psychology Bulletin, 26*, 891-906.

Bond, C. F. & Titus, L. J. (1983). Social facilitation: A meta-analysis of 241 studies. *Psychological Bulletin, 94*, 265-292.

Cortina, J. M. (1993). What is the coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98-104.

Derks, B. Inzlicht, M. &. Kang, S. (2008). The neuroscience of stigma and stereotype threat. *Group Processes Intergroup Relations, 11*, 163-181.

Finchilescu, G. (2002). Measurements. In C. Tredoux & K. Durrheim (Eds.) *Numbers, Hypotheses & Conclusions. : A course in statistics for the social sciences.*(pp201-229). Cape Town: UCT Press

Forbes, A. R. (1964). An item analysis of the advanced matrices. *British Journal of Educational Psychology, 34*, 223-236.

Forbes, J.D. (2007). *Stereotype-threat and the emotional sensitivity of psychology students*. University of Kwazulu-Natal (Unpublished Honours Thesis).

Ford, T. E., Ferguson, M. A., Brooks, J. L. & Hagadone, K. M. (2004). Coping sense of humor reduces effects of stereotype threat on women's math performance. *Personality and Social Psychology Bulletin, 30*, 643-653.

Haslam, S. A., Salvatore, J., Kessler, T. & Reicher, S. D. (2008). The social psychology of success. *Scientific American Mind A p r i l /M ay 2 0 0 8*, 24-32.

Jetten, J., Hornsey, M. J. & Adarves-Yorno, I. (2006). When group members admit to being conformist: The role of relative intragroup status in conformity self-reports. *Personality and Social Psychology Bulletin, 32*, 162-173.

Klein, O., Licata, L., Azzi, A. E. & Durala, I. (2003). "How European am I?'': Prejudice expression and the presentation of social identity. *Self and Identity, 2*, 251-264.

Klein, O., Spears, R. & Reicher, S. (2007). Social identity performance: Extending the strategic side of SIDE. *Personality and Social Psychology Review, 11*, 28-45.

Luhtanen, R. & Crocker, J. (1992). A collective self-esteem scale: Self-evaluation of one's social identity. *Personality and Social Psychology Bulletin, 18*, 302-318.

Lea, M., Spears, R. & de Groot, D. (2001). Knowing me, knowing you: Anonymity effects on social identity processes within groups. *Personality and Social Psychology Bulletin, 27*, 526-537.

Marx, D. M. & Stapel, D. A. (2006). Distinguishing stereotype threat from priming effects: On the role of the social self and threat-based concerns. *Journal of Personality and Social Psychology, 91*, 243-254.

McKown, C. & Weinstein, R. S. (2003). The development and consequences of stereotype consciousness in middle childhood. *Child Development, 74*, 498-515.

Oakes, P. J., Haslam, S. A. & Turner, J.C. (1994). *Stereotyping and Social reality*. Oxford: Blackwell Publishers.

O'Brien, L. T. & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin, 29*, 782-789.

Osborne, J. (2006). Gender, stereotype threat, and anxiety: Psychophysiological and cognitive evidence. *Electronic Journal of Research in Educational Psychology, 4*, 109-138.

Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement? *Contemporary Educational Psychology, 26*, 291 - 310.

Ouwerkerk, J. W., de Gilder D. & de Vries, N. K. (2000). When the going gets tough, the tough get going: Social identification and individual effort in intergroup competition. *Personality and Social Psychology Bulletin, 26*, 1550-1559.

Postmes, T. & Spears, R. (2002). Behavior online: Does anonymous computer communication reduce gender inequality? *Personality and Social Psychology Bulletin, 28* , 1073-1083.

Quayle, M. & Reicher, S. (2007), Social identity, identifiability and identification in participants experiences of stereotype threat. Paper presented at the PSYSSA Annual Congress, Durban.

Reicher, S. & Levine, M. (1994). On the consequences of deindividuation manipulations for the strategic communication of the self: Identifiability and the presentation of social identity. *European Journal of Social Psychology, 24*, 511-524.

Reicher, S., Levine, R. M. & Gordijn, E. (1998). More on deindividuation, power relations between groups and the expression of social identity: Three studies on the effects of visibility to the in-group. *British Journal of Social Psychology, 37*, 15-40.

Rosenthal, H. E. S. & Crisp, R. J. (2006). Reducing stereotype threat by blurring intergroup boundaries. *Personality and Social Psychology Bulletin, 32*, 501-511.

Rushton, J. P., Skuy, M., & Fridjhon, P. (2003). Performance on Raven's Advanced Progressive Matrices by African, East Indian, and White engineering students in South Africa. *Intelligence, 31*, 123–137.

Rushton, J. P., Skuy, M. & Bons, T. A. (2004).Construct validity of Raven's Advanced Progressive Matrices for African and Non-African engineering students in South Africa. *International Journal of Selection and Assessment, 12,* 220-229.

Salthouse, T. A. (1993). Influence of working memory on adult age differences in matrix reasoning. *British Journal of Psychology, 84*, 171-199.

Shapiro, J. R. & Neuberg, S. L. (2007). From stereotype threat to stereotype threats: Implications of a multi-threat framework for causes, moderators, mediators, consequences, and interventions. *Personality and Social Psychology Review, 11*, 107-130.

Smith, J.L., & White, P. (2001). Development of the domain identification measure: A tool for investigating stereotype threat effects. *Educational and Psychological Measurement, 61*, 1040-1057.

Spears, R., Doosje, B. & Ellemers, N. (1997). Self-stereotyping in the face of threats to group status and distinctiveness: The role of group identification. *Personality and Social Psychology Bulletin, 23*, 538-553.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38*, 194-201.

Schmader, T., Johns, M. & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115* 336-356.

Smith, J. L. (2004). Understanding the process of stereotype threat: A review of mediational variables and new performance goal directions. *Educational Psychology Review, 16*(3), 177-206.

Steele, C. M. (2003). Through the back door to theory. *Psychological Inquiry, 14*, 314 - 317.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*, 613-629.

Steele, C. M., Spencer, S. J. & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (ed.),*Advances in Experimental Social Psychology, Vol. 34* (pp. 379 - 440). Academic Press.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*, 99-103.

Tajfel, H. (1978). Interindividual Behaviour and Intergroup Behaviour. In H. Tajfel (ed.) *Differentiation Between Social Groups: Studies in the Social Psychology of Intergroup Relations* (pp. 27-98). London: Academic Press.

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, *33*, 1-39.

Unsworth, N. & Engle, R.W. (2005).Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence, 33*, 67–81.

Vigneau, F. & Bors, D. A. (2005) Items in context: Assessing the dimensionality of Raven's Advanced Progressive Matrices. *Educational and Psychological Measurement, 65*, 109-123.

Worchel, S., Rothgerber, H., Day, A., Hart, D. & Butemeyer, J. (1998). Social identity and individual productivity within groups. *The British Journal of Social Psychology, 37*, 389-413.

Walton, G. M. & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology, 39*, 456-467.

Wheeler, S. & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin, 127*, 797-826.

Wout, D., Danso H., Jackson, J. & Spencer, S. (2008). The many faces of stereotype threat: Group- and self-threat. *Journal of Experimental Social Psychology, 44*, 792-799.

Zajonc,R. B. (1965). Social facilitation. *Science, New Series*,*149*, 269-274.

Appendix A: Initial group membership tasks

LOGICAL PATTERN RECOGNITION TASK [IDEALLY SUITED TO SCIENCE STUDENTS' ABILITIES]: Find each number in the grid. Digits are adjacent and numbers run top-to bottom or left-to-right. EXAMPLE: The numbers 4492 and 2247 have been identified in the grid below:

4	8	**4**	1	8
7	6	**4**	0	6
2	0	**9**	9	2
1	**2**	**2**	**4**	**7**
9	0	8	6	5

FIND THE FOLLOWING NUMBERS IN THE GRID ALONGSIDE:

7	3	4	2	4	0	4	1	5	6	8	2
2	1	5	1	9	4	1	3	7	3	2	1
1	6	7	2	7	7	4	1	8	7	3	2
1	1	6	5	4	5	0	4	2	1	8	2
7	8	3	7	6	2	7	6	2	5	2	2
1	1	6	2	6	5	2	8	7	8	6	8
8	7	8	2	3	7	7	3	1	0	0	4
7	1	0	8	9	3	4	9	3	4	1	0
3	5	7	9	5	3	6	4	2	8	2	2
6	0	3	2	6	2	3	6	6	8	0	5
6	5	4	8	1	4	0	8	0	2	5	3
1	6	3	7	5	4	0	6	5	5	1	6

73424

773100

387141

45763

228212

72774

How many of the numbers did you find? _____

Research shows that the average **humanities** student finds only three of the six patterns within the time allocated whereas the average Science student finds all of them.

SYMBOLIC PATTERN RECOGNITION TASK [IDEALLY SUITED TO HUMANITIES STUDENTS' ABILITIES]: INSTRUCTIONS: Find the five objects in the picture below. When you find the picture circle it with the marker provided. You will have 90 seconds to complete this task.
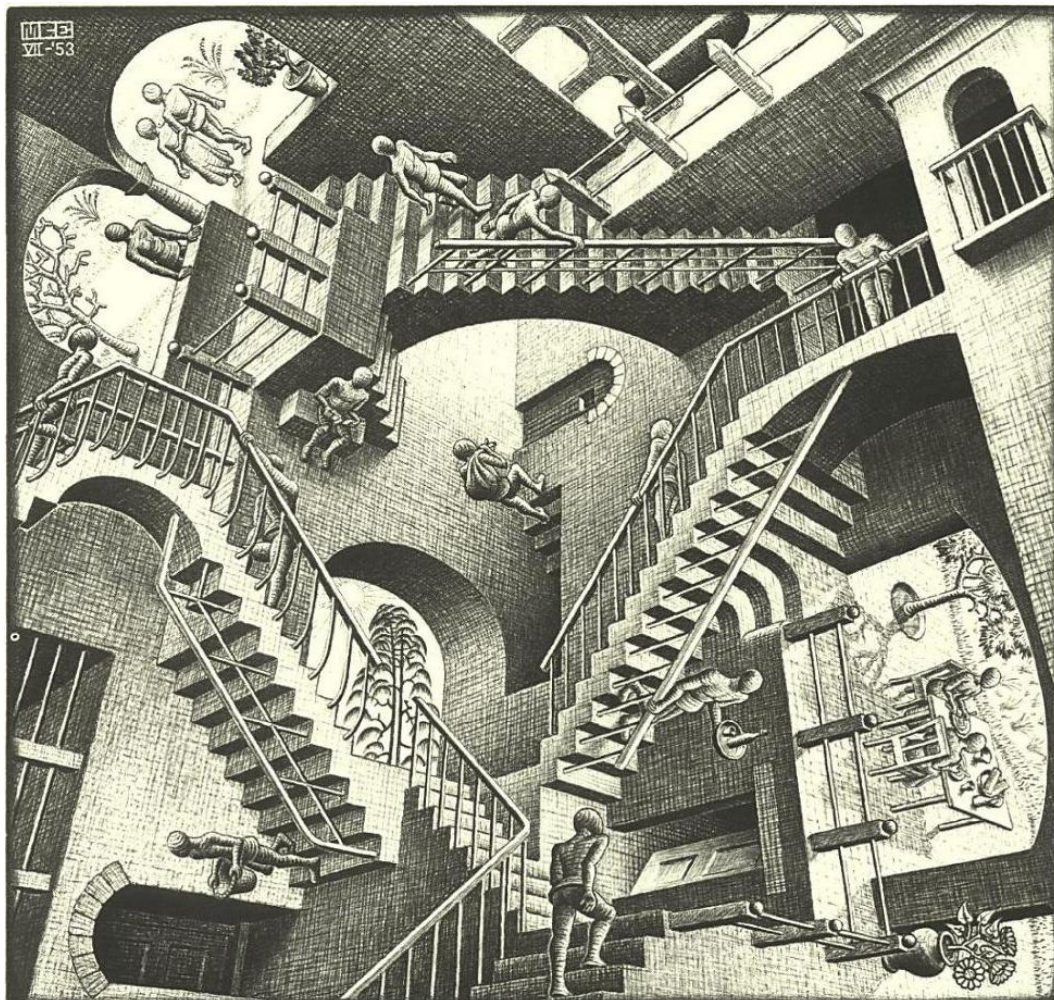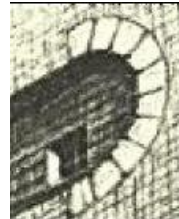
How many objects did you find?_____ Research shows that the average **Humanities** student finds all of the objects whereas the average Science student only finds two of the five objects in the time provided.

This form will provide you with information about the study – please read it carefully. Project title: Differences in academic ability between Humanities and Science students **Project aims:** This study aims to investigate differences in academic ability between Humanities and Science students. We are also interested in how Humanities students perceive Science students and vice versa.

Project investigator: This study is being undertaken by Jared Forbes a Research Psychology Masters student and supervised Mike Quayle from the School of Psychology, UKZN. For more information, or if you have any queries or complaints, contact Jared by email on 204501298@ukzn.ac.za or please call Mike on 033-2605016 or email quaylem@ukzn.ac.za **What is required of you:** If you choose to continue, you will be asked to complete several tests and questionnaires. Some tests relate to your academic ability and others relate to how you see yourself. Although some of the tests are difficult, they are not very stressful. You may discover information about yourself that you find stressful, but you will be given a chance to ask questions after the study and put your mind at ease about what you have discovered. **Participation will take about sixty minutes. Please note that you will complete two tests of academic ability, during the first one of these you will be completely anonymous but your performance on the other one will be video-taped. Once the study is completed the videotaped results will be discussed by a group of Humanities students, to get different opinions on participants' performance and strategies**. **How you might benefit:** Aside from the benefit of helping us to advance knowledge of psychology, you might also learn something about yourself and about psychological research. **Incentive:** You will earn R30 for participating (i.e. for about 60 minutes of your time).

Data: The results of the tests you complete will be analyzed for psychological research and the results may be presented at conferences and published in books and journals. The data will be stored indefinitely by the investigator and will be accessed by researchers working on the project. As already mentioned, your performance on the first task you will be completely anonymous and the remaining task you will be observed and discussed by a group of Humanities students **What if you don't want to participate?** If you would prefer not to participate that's absolutely fine. Participation is voluntary and you can withdraw at any time without giving a reason. **DECLARATION Note: By signing this form you are simply stating that you understand the nature of the research project and that you have agreed to participate without being forced or pressured by the researcher. You are not committing to anything that you cannot easily get out of. The form will demonstrate that the researcher has treated you fairly.**

.I……………………………………………………………………………(full names of participant) hereby confirm that I understand the contents of this document and the nature of the research project, and I consent to participating in the research project. I understand that I am at liberty to withdraw from the project at any time, should I so desire.

SIGNATURE OF PARTICIPANT   DATE

……

Debriefing

Information Sheet

Dear Participant,

You may be surprised to learn that not everything we told you about the study was completely accurate. However, because the results would have been affected if you had known exactly what we were researching, this was a necessary deception.

You were told that the study was examining differences in academic ability between students. This is not true. In reality the study was investigating stereotype threat. Stereotype threat is caused when a negative stereotype about an individual's group affects their performance of a stereotype-relevant task. An example of this would be women doing mathematics: if they believed that their professor held the stereotype that females have poorer mathematics ability: stereotype-threat theory predicts that their performance would decrease when they had to take a maths test in this circumstance.

The main aim of this study was to examine stereotype threat with respect to Science students, Humanities students and different ways of reasoning. The common stereotype is that Science students are better with numbers than Humanities students. Three types of booklet were handed out – one that presented the task as a positively and more suited for science students one and the other that presented it as a positive and more suited for Humanities students. The other booklet tried to present the task as neutrally as possible, i.e. not favouring a particular group. The data will be analyzed to examine whether activating the stereotype had any effect on the performance of the task. Also in the tasks you were told they would be examined by different people, such as a group of Science students, Humanities students or Researchers. This is not true, but this will allow us to investigate whether the extent to which you feel individually identifiable impacts on the extent to which you experience stereotype threat.

If there are any further questions please feel free to contact Jared Forbes by email 204501298@ukzn.ac.za or Mike Quayle by phone on 033-2605016 or email quaylem@ukzn.ac.za.

To be signed after completion:

The effect of identifiability to the in-group or out-group on stereotype threat on a test of the Raven's Advanced Progressive matrices

Having completed the study and debriefing I still consent to my data being stored and used for the study:

DATE   SIGNATURE OF PARTICIPANT

_____

Based on the available information, [BOOST GROUP] as a group have a good reputation with respect to the [EXPERIMENTAL TASK OR CONTEXT]

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree

*Based on the available information, [THREAT GROUP] as a group have a good reputation with respect to [EXPERIMENTAL TASK OR CONTEXT]

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree

Based on the available information, [EXPERIMENTAL TASK OR CONTEXT] is more suited to the skills of [BOOST GROUP] than [THREAT GROUP]

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree

*Based on the available information, more [THREAT GROUP] are likely to do well than [BOOST GROUP] in [EXPERIMENTAL TASK OR CONTEXT]

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree

Based on the available information, the worst performing participant will probably be an [THREAT GROUP MEMBER]

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree

How do you think the typical [BOOST GROUP MEMBER] would have experienced [the task] in terms of DIFFICULTY?

1	2	3	4
very hard	hard	Easy	very easy

How do you think the typical [BOOST GROUP MEMBER] would have experienced [the task] in terms of ENJOYMENT?

1	2	3	4
very unpleasant	Unpleasant	Enjoyable	very enjoyable

How do you think the typical [BOOST GROUP MEMBER] would have PERFORMED on [the task]?

1	2	3	4
very badly	badly	Well	very well

How do you think the typical [THREAT GROUP MEMBER] would have experienced [the task] in terms of DIFFICULTY?

1	2	3	4
very hard	hard	Easy	very easy

How do you think the typical [THREAT GROUP MEMBER] would have experienced [the task] in terms of ENJOYMENT?

1	2	3	4
very unpleasant	Unpleasant	Enjoyable	very enjoyable

How do you think the typical [THREAT GROUP MEMBER] would have PERFORMED on [the task]?

1	2	3	4
very badly	badly	well	very well

Please indicate your level of agreement by placing a cross (X) along the line:

How likely is it that the following will see this questionnaire?

A group of humanities student?
Very unlikely --- Very Likely

A group of science students ?
Very unlikely --- Very Likely

A group of Researchers?
Very unlikely --- Very Likely

How likely is it that the following will see this Video?
A group of humanities students?
Very unlikely --- Very Likely

A group of science students?
Very unlikely --- Very Likely

A group of Researchers?
Very unlikely --- Very Likely

Appendix H: Social Identity Inventory

In-group Identification scale

My group is an important part of who I am as a person

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree

I feel strong ties with [ingroup] as a group

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree

Being an [ingroup member] affects the way I am and how I think

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree

*Overall, my membership of [this group] has very little to do with how I feel about myself

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree

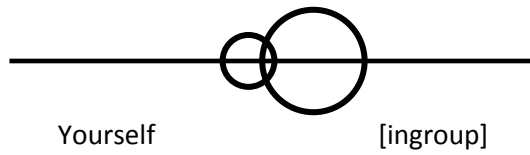**If someone says something bad about [my ingroup] it is like they are saying something bad about me**

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree

Choose the picture that best represents your own closeness to the [ingroup].

Yourself   [ingroup]

Yourself   [ingroup]

Yourself   [ingroup]

Yourself   [ingroup]

Yourself   [ingroup]

Yourself   [ingroup]

Yourself   [ingroup]
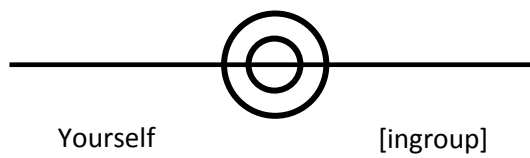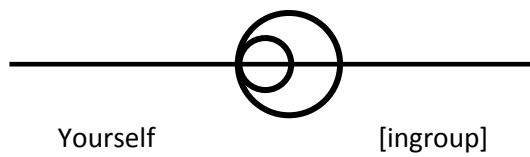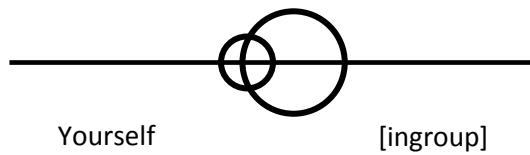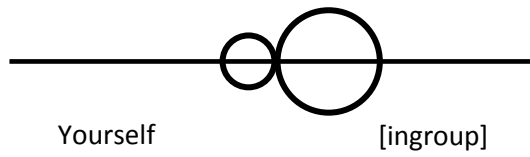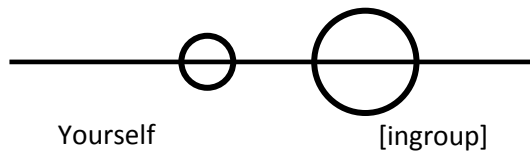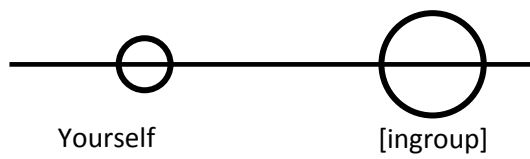
Identifiability scale

My results will reflect more on my group than on me as an individual

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree

I felt identifiable as an individual while I was doing the task

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree

I am being seen more as a group member than as an individual

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree

People who see these test results will be able to recognise me in other contexts

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree

I felt like I was personally in the spotlight while I was doing the task

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree

I felt like my group was in the spotlight while I was doing the task

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree

Intrinsic Motivation Inventory – Anxiety/Pressure/Tension Subscale

*I did not feel nervous at all while doing [the DV task]

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree

I felt very tense while doing [the DV task]

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree

*I was very relaxed in doing [the DV task]

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree

I felt pressured while doing [the DV task]

1	2	3	4	5	6	7
strongly disagree	disagree	somewhat disagree	neither disagree nor agree	somewhat agree	agree	strongly agree