

**Situated identity performance:**  
**Understanding stereotype threat as a social identity phenomenon**

Michael Quayle

University of KwaZulu-Natal

© Michael Quayle 2011. All rights reserved.

Research funded by the NRF under the Thuthuka/Researchers in Training programme and the UKZN research office under the competitive research grant programme.

### **Declaration**

Submitted in fulfilment / partial fulfilment of the requirements for the degree of  
Doctor of Philosophy (Human Sciences), in the Graduate Programme in Psychology,  
University of KwaZulu-Natal, Pietermaritzburg, South Africa.

I declare that this dissertation/thesis is my own unaided work. All citations, references and borrowed ideas have been duly acknowledged. It is being submitted for the degree of Doctor of Philosophy in the Faculty of Humanities, Development and Social Sciences, Pietermaritzburg, South Africa. None of the present work has been submitted previously for any degree or examination in any other University.

---

M F Quayle

---

Date

---

Prof. K Durrheim

---

Signature

### **Acknowledgements**

I wish to thank the two supervisors of this project: Prof. Steve Reicher while I was in St Andrews in 2005-2006 and Prof. Kevin Durrheim from 2006 until 2010. They have been instrumental in shaping this work.

The following students have worked on this project and have each played a role in data collection, data processing and have each used portions of the data for student submissions:

Jaqui Bowman

Megan du Toit

Jared Forbes

Karmini Naidoo

Tselisihang Motuba

Tamaryn Nicholson

Letitia Rambally

Tamlyn Seunanden

Luke Vurovecz

The notion of “resistance,” the method for calculating resistance metrics, the notion of stereotype threat and boost effects as outcomes of situated identity performance, the detailed analyses of each study and the cross-study analyses are novel to this PhD dissertation.



Many thanks to my wife, Kelly, for her constant support and encouragement.

Is there life after PhD? We're about to find out.



### Abstract

Stereotype threat or boost (STB) is a situational modifier of task performance that occurs when a group stereotype becomes relevant to the performance of a stereotype-relevant task. This dissertation aimed to re-imagine STB in light of social identity theory. Ten studies were undertaken that each manipulated status and either identifiability, conflict or permeability and explored the effects on the performance of the Ravens Advanced Progressive Matrices. Additional identity and socio-structural constructs were also measured and explored, including stability, legitimacy and ingroup identification.

The results showed that STB is not simply “activated” or “deactivated” when stereotypes become relevant to task performance. On the contrary, the specific features of identity, the contextual features of the social environment in which the identity performance takes place, and the performer’s strategic engagement with their identity resources and liabilities are important features of how STB impacts on performance, and how it is sometimes resisted and overturned by experimental subjects. Indeed, performance was generally not predictable on the basis of stereotype activation until resistance to the negative or positive status manipulations were also accounted for.

Although the STB literature is tightly focused on the case of negative stereotypes undermining performance, incongruent effects in which negative stereotypes enhance performance and positive stereotypes undermine it have also been reported. In the present studies incongruent STB effects were frequently observed. Underperformance in boost conditions was most consistently predicted by perceived intergroup conflict, while enhanced performance under threat was consistently predicted by perceived group boundary permeability. Additionally, underperformance in boost conditions was often a result of ‘slipstreaming’ rather than ‘choking under pressure,’ since participants were evidently counting on their generally secure identity in the experimental context to buffer poor performance on the experimental task. Improved performance in threat conditions was most likely when participants perceived themselves to be representatives of their group and when they believed that their improved performance would make a difference for their own reputation or the reputation of their group.

These findings challenge the common image of the passive subject in the STB literature and, instead, suggest that STB effects are an outcome of *situated identity performance*. This

model of STB effects understands task-performance in a specific performance context as an active and strategic expression of situated identity oriented not only to the social features of the performance context (as argued by most SIT theorists), but also to the their own reading of that context, their total identity liabilities and resources (including individual ability and alternative identities) and their strategic motivations in the context.



## Table of Contents

CHAPTER 1.	INTRODUCTION	1
CHAPTER 2.	STEREOTYPE THREAT	4
CHAPTER 3.	PROBLEMS WITH CONVENTIONAL STEREOTYPE THREAT THEORY AND THE PROMISE OF SOCIAL IDENTITY THEORY FOR ADDRESSING THEM	71
CHAPTER 4.	AIMS, METHOD AND DESIGN	106
CHAPTER 5.	RESULTS I: STATUS, PERFORMANCE AND RESISTANCE	147
CHAPTER 6.	RESULTS II: IDENTIFIABILITY	173
CHAPTER 7.	RESULTS III: CONFLICT	211
CHAPTER 8.	RESULTS IV: PERMEABILITY	238
CHAPTER 9.	RESULTS V: PREDICTING INCONGRUENT STEREOTYPE-PERFORMANCE EFFECTS	275
CHAPTER 10.	GENERAL DISCUSSION	311
CHAPTER 11.	CONCLUSION	345

**Table of study abbreviations**

Topic	Detail	N	Abbreviation
Social Identity Inventory Pilot (Pietermaritzburg, 2008)	Pilot of measures for the Social Identity Inventory	23	SII Pilot 1
Social Identity Inventory Pilot (Pietermaritzburg, 2009)	Pilot of measures for the Social Identity Inventory	36	SII Pilot 2
Identifiability (St Andrews, 2006)	Testing the effect of identifiability on STB and a pilot of the general method	54	Identifiability Study 1
Identifiability (Pietermaritzburg, 2007)	Replication of Leyens (eg. Leyens, Desert, Croizet, & Darcis, 2000)	218	Identifiability Study 2
Identifiability (Pietermaritzburg, 2008)	Testing the effect of identifiability on STB with a repeated measures design	51	Identifiability Study 3
Identifiability (Pietermaritzburg, 2010)	Testing the effect of identifiability on STB while manipulating individual reward	300	Identifiability Study 4
Conflict (Pietermaritzburg, 2008)	Testing the effects of conflict on STB with Science students	102	Conflict Study 1
Conflict (Pietermaritzburg, 2009)	Testing the effects of conflict on STB with Science students	122	Conflict Study 2
Conflict (Pietermaritzburg, 2010)	Testing the effects of conflict on STB with Humanities students	138	Conflict Study 3
Permeability (Pietermaritzburg, 2009)	Testing the effect of permeability on STB with Varsity College students	100	Permeability Study 1
Permeability (Durban & Pietermaritzburg, 2010)	Testing the effect of permeability on STB with Varsity College students	177	Permeability Study 2
Permeability (Pietermaritzburg, 2010)	Testing the effect of permeability on STB in a minimal groups paradigm	151	Permeability Study 3
Achievement motivation (Pietermaritzburg, 2009)	Testing the relationship between achievement motivation and STB; piloting the Shipley as a pretest measure of ability	93	Achievement Motivation Study 1

## Chapter 1. Introduction

Stereotype threat or boost (STB) is a shift in task performance that occurs when a negative or positive group stereotype becomes relevant to the performance of a stereotype-relevant task (Steele, 1997; Walton & Cohen, 2003). Stereotype threat is based on the finding that negative stereotypes undermine task performance (Steele, 1997, 2010; Steele, Spencer, & Aronson, 2002) and there is clear evidence that positive stereotypes can improve it (Walton & Cohen, 2003).

STB can affect any individual who is a member of a relevant group that is stereotyped with respect to a task or task-relevant domain. It has been empirically shown to apply to a wide variety of groups in varied settings. For example, it impacts on children as young as five years old (Ambady, Shih, Kim, & Pittinsky, 2001), women in math (eg. Spencer, Steele, & Quinn, 1999), men in psychology (eg. Leyens et al., 2000), whites in sports (eg. Stone, Lynch, Sjomeling, & Darley, 1999), people from low socioeconomic backgrounds in standardized tests (eg. Croizet & Claire, 1998), the memory performance of the elderly (eg. Hess, Auman, Colcombe, & Rahhal, 2003), white men in math compared to Asian men (eg. Aronson, Lustina, Good, Keough, & Steele, 1999), the academic performance of black students and many others (see Walton & Cohen, 2003 for a more comprehensive review).

When STB was first formulated, there was a great deal of hope that it could, at least partially, explain group-based differences in performance in many domains, for example, the race gap(s) in American academic performance generally and gender gaps in American math and science (Spencer et al., 1999; Steele & Aronson, 1995). This was a tantalizing prospect, as STB promised a socio-cognitive explanation of how social constructions such as stereotypes may cross from imagination to reality and result in genuine empirical performance differences between groups.

As empirical work in the field progressed, a large and increasing set of moderators and mediators were proposed and partially confirmed including effort, anxiety, self-handicapping, evaluation apprehension, confidence, stereotype endorsement, perceptions of the test, feelings about the self (Smith, 2004) domain identification (Steele, 1997), racial identity (Davis, Aronson, & Salinas, 2006), level of education, and stigma consciousness (Hess, Hinson, & Hodges, 2009a), amongst others. However, no mediators or moderators have been found to completely or *reliably* influence STB (Smith, 2004) and therefore the field is left with a large and confusing set of variables that have been found to influence STB in some studies but not others.

As the potential mediators and moderators of the effect piled up, some researchers began to challenge the importance and universality of the effect (eg. Sackett, Hardison, & Cullen, 2005) and to question the extent to which STB can explain or predict intergroup differences in real-life performance settings. There were also challenges to the validity, size and impact of the effect (eg. Wicherts, 2005). Although quite comprehensive meta-analyses had shown the effect to be relatively robust, and had included unpublished studies (Nguyen & Ryan, 2008; Walton & Cohen, 2003), there were some criticisms that the file-drawer effect (eg. Wax, 2008; Wicherts & de Haan, 2009) combined with increased type-I error due to violations of the assumptions of ANCOVA in much STB research (Wicherts, 2005) had resulted in gross over-estimates of the importance of the effect in explaining intergroup differences.

Even more strangely, there has been a small but growing body of studies showing that stereotype *lift* or *boost* may often *decrease* performance (eg. Cheryan & Bodenhausen, 2000; Cocchiara & Quick, 2004; Keller & Bless, 2008), a circumstance often attributed to ‘choking under pressure.’ Conversely, at least one recent study has even found that in certain circumstances stereotype *threat* can result in *increased* performance under certain

circumstances (Crisp, Bache, & Maitner, 2009). Indeed, a closer look at Nguyen & Ryan's (2008) meta-analysis shows that stereotype threat had a *positive* effect on performance in approximately an eighth (14 of 116) of the collated studies. Of these, only six were published and in these six the increase in performance under threat was explained as an anomaly (eg. Brown & Day, 2006) or was not visible in the original results due to transformations, such as dividing the number of problems answered correctly by the number of problems attempted (eg. Cohen & Garcia, 2005).

These mixed findings leave the field with the theoretically unattractive finding that both positive and negative stereotypes can increase or decrease task performance. The complex web of mediators and moderators that loosely predict when stereotype threat or boost might cause performance decrements or increments is desperately lacking in parsimony. Without some way of understanding the circumstances in which positive or negative stereotypes result in positive or negative shifts in performance the field of STB cannot fully describe the impact of group stereotypes in real-life performance settings.

In this dissertation I will argue that the empirical effect known as "stereotype threat" is an outcome of *situated identity performance* in which the subject's reading of the situation in terms of their own identity resources, plans and ambitions intersect with the constraints imposed on them in the social setting. I will argue that their responses in a stereotype-threat situation are not automatic and passive, but actively produced and strategic within the bounds of the situation and their own goals or ambitions in the STB situation and in broader social settings. This argument will be achieved by systematically applying social identity theory to STB and exploring how the predictions of STB are violated by particular groups in particular circumstances, and how these violations may relate to these participants' reading of the experimental context and the broader social setting.

## Chapter 2. Stereotype threat

Exploring the effects of social structures and variables on task performance has long been a theme of experimental social psychology. Indeed, long before social psychology was conceptualized as a discipline, Carpenter (1852) was already exploring the effect of hypnotic suggestions about identity on actual behavior. It was nearly fifty years later that Tripplett (1898) was undertaking his studies on the effect of social competition on the task of riding a bicycle, which are commonly presented in first-year social psychology textbooks as the first ever experiments in the field of social psychology (Cherry, 1995).

In the 1920's, Floyd Allport was exploring the impact of social facilitation on the performance of a free-association task (Allport, 1920) and audience effects on tests of analogies and word naming very similar to IQ tests (Allport, 1923), finding that audience effects improved results for the lowest performers but not the highest. This thread of research was revived and continued by Zajonc's (1965) review of social facilitation effects.

By then Gordon Allport (1954) was exploring the effect that features of social settings such as stereotypes and stigma may have on performance via internalization, extensively discussing their impact on standardized testing. However, Katz, Roberts & Robinson (1965) found that black students performed better in IQ tests when they believed that the tests were not assessing intellect, demonstrating that group-based performance differences are not necessarily the result of stable internalizations, but are sensitive to interpretations of both the meaning of the task and the purpose of the task performance.

By the 1980's and 1990's it had become increasingly evident that the physical co-presence of audience or actors was not necessary for social features to impact on task performance and behavior: simply priming people with social representations could substantially influence behavior and task performance through priming effects (see Wheeler & Petty, 2001 for a comprehensive review). For example, Berkowitz (1984) showed that

viewing representations of social behavior such as violence can result in similar behavior or behavioural intentions and Bargh (1996) showed that participants (regardless of age) primed with a stereotype of the elderly walked more slowly.

To summarize, by the early 1990's it was evident that: (1) features of the social environment can influence task performance; (2) categorization and stereotypes are important in this regard; (3) some combination of beliefs about minority groups (such as stigma) and beliefs about tasks and task-performance environments can depress the performance of minority group members; and (4) these shifts in performance are often contextually specific. These ideas were brought together by Claude Steele and his colleagues and labeled "stereotype threat."

### **2.1. Steele's role in the development and maintenance of the theoretical framework of stereotype threat**

In the definition and discussion of stereotype threat that follows, the accounts of Stereotype Threat developed in papers authored and co-authored by Steele will be privileged. This is because Steele was the senior partner with Spencer and then Aronson that led to the seminal 1995, 1997, 1999 and 2002 papers in which the theoretical framework of stereotype threat was laid out (Aronson et al., 1999; Spencer et al., 1999; Steele, 1997; Steele & Aronson, 1995; Steele et al., 2002).

Steele has also been the first author on the most highly cited stereotype threat papers. According to Google scholar, by May 2010, Steele & Aronson (1995) had received 1851 citations in academic and popular publications; and Steele (1997) had received 1799. The next closest stereotype threat publication was Spencer, Steele and Quinn (1999) with 750 citations which, in any case, was co-authored by Steele.

Of the 61,500 Google hits for "stereotype threat" recorded on the 11 May 2010, 24,000 mentioned Steele compared to only 12,700 mentioning Aronson and 8,310

mentioning Spencer. Steele has also (individually) received the *APA Award for Distinguished Senior Career Contributions to the Public Interest* and the *Award for Distinguished Scientific Contributions* largely for his contribution to Stereotype Threat ("Claude M. Steele: Award for distinguished senior career contributions to the public interest," 2003).

Additionally, apart from his role as first author in the review that cemented the notion of stereotype threat in the literature (Steele et al., 2002), he has now released a book on stereotype threat (Steele, 2010) making the notion of stereotype threat more accessible to the public and also extending its practical reach in ways that will be discussed (and critiqued) a little later. To both academics working in the field and to the general public, Claude M. Steele is regarded as the originator and custodian of “stereotype threat” and his work will be treated as theoretically central to the stereotype threat framework in this discussion.

## **2.2. Origins of the notion of stereotype threat**

As described in several biographical and autobiographical narratives of the origins of stereotype threat ("Claude M. Steele: Award for distinguished senior career contributions to the public interest," 2003; Steele, 1999, 2003a, 2010), Steele had a long-standing interest in the race and gender gaps in academic performance in which members of minority groups, particularly African Americans, have been shown to underperform in relation to other groups. Some have ascribed this ‘performance gap’ to hereditary race and gender differences (eg. Johnson & Bouchard, 2007; Rushton & Jensen, 2005a; Stafford, 1972). Others have argued that low performance is a result of long-term internalization of social representations and cumulative effects of stigma (eg. Allport, 1954).

However, a key moment according to Steele (2003a), was when he and his colleagues noticed that women were also underperforming in difficult math courses in patterns that closely matched the underperformance of minorities in academics more generally. Steele and



his colleagues began to be intrigued by the similarities in performance by these two minority groups across these domains (Steele, 2003a). Steele narrates that they noticed that women were performing comparably to their male colleagues in less advanced math courses and began to wonder whether some combination of features of the testing environment and minority identity might intersect to produce conditions in which minority underperformance was likely.

### **2.3. Identity predicaments and contingencies**

Steele and his colleagues began to consider the burden of holding a negatively stereotyped identity in performance environments, and how holding such an identity might impact on performance. This situation was initially referred to as the “predicament” of stereotype threat (Spencer et al., 1999; Steele & Aronson, 1995) and later relabeled an “identity contingency” (Purdie-Vaughns, Steele, Davies, Dittmann, & Crosby, 2008; Steele, 2010) because it represents an intersection of features that – together – influence performance<sup>1</sup>.

If identity and performance are contingent on an intersection of features of the context and identity resources (or liabilities) available to the individual in a setting, and if one of the components of the predicament can be shifted, then the effect on performance should disappear. If underperformance was reduced when identity was made irrelevant to task performance then, Steele and his colleagues argued, underperformance in certain conditions could be attributed to the identity predicament.

---

<sup>1</sup> Although the word "contingency" was used in Steele (1997), the phrase "social identity contingency" seems to be first used in stereotype threat literature in Steele (2003, p. 317), but was used by Levinson (1999) four years earlier discussing philosophy of education. She in turn cited Rorty (1989) who discussed the notion of identity contingencies (or at least contingent identities) before her. The next time it is picked up in the work of Steele and his colleagues seems to be in Purdie-Vaughns, Steele, Davies, Dittmann, & Crosby (2008), but it is a dominant feature of Steele's (2010) book *Whistling Vivaldi* and so is likely to become a popular way of defining the stereotype threat situation.

Steele and his colleagues began to experimentally explore whether underperformance of minority group members could really be related to contingent intersections of identity, stereotypes and domains. There are at least three ways to experimentally shift this identity predicament: the first is to ask participants to undertake two tasks – one in which their group is negatively stereotyped and another in which it is not and then to compare performance across tasks. Steele and his colleagues (Spencer et al., 1999; Steele, 2003a) tried this by asking men and women to complete easy and difficult math tests. They found that women underperformed on the difficult test but not the easy one, even though they were matched for skill with male counterparts. They argued that stereotypes about women's math abilities were only activated when the material was particularly difficult.

The second way of undermining the threatening identity predicament is to expose participants to a testing situation in which their group membership is either made relevant or made irrelevant. Steele and his colleagues tried this by asking black and white students to complete the same difficult verbal test either under conditions in which they believed it would be diagnostic of ability (making the stereotype of black intellectual inferiority relevant) or non-diagnostic of ability (Steele & Aronson, 1995). Once again, participants underperformed only when they were members of the negatively stereotyped group in conditions in which features of that identity were relevant to the perceived meaning of the test and testing situation. After controlling for SAT scores (understood as an estimate of individual ability) the scores of black participants were comparable to non-black participants in non-diagnostic conditions.

Note that others had already realized that believing a test to be diagnostic of ability could depress performance generally (see Baumeister & Showers, 1986), and Katz (eg. Katz, Roberts, & Robinson, 1965) had already established that black students particularly underperformed in these conditions. What Steele and his colleagues added was the crucial

point that the underperformance was *contingent* on both the perceived meaning of the test (as diagnostic or non-diagnostic) and the personal relevance of a group stereotype that made this representation of the test personally threatening or non-threatening.

The third way of shifting the identity predicament is to ask participants with multiple conflicting identities to undertake the same test from either a stigmatized or positive identity position. This was tested by comparing the math performance of male white students when they were either compared to Asian males (who are very positively stereotyped in this domain) or not specifically compared to anyone and therefore able to inhabit the generally positive group stereotype of white male students in that domain (Aronson et al., 1999). As predicted, participants generally performed poorly when they perceived themselves to be negatively stereotyped compared to a comparison group and relatively well when the negative comparison was not relevant. Interestingly, these results only applied to high performers, and more so for those who were highly identified with their group.

Taken together, these results provide strong support for the argument that intersections of identity, group stereotypes and understandings of the meanings and purposes of tasks can modify task performance, in other words, that poor performance in a task can be related to an identity predicament or social identity contingency.

These findings were radical for several reasons. First, this model of underperformance undermines pathologising accounts of minority *individuals* since “the experience of this threat is not seen to depend on a particular state or trait of the target such as believing in the stereotype or holding low expectations that might result from chronic exposure to the stereotype” (Steele et al., 2002, p. 390).

Second, this line of research undermines pathologising accounts of minority *groups* that locate reasons for poor performance in the general biological or cultural make-up of minority groups (eg. Rushton & Jensen, 2005a). This understanding of underperformance

therefore limits contagion of stigma from particular domains of performance to generalized views of groups because “it is the specific meaning of the stereotype that determines the situations, the people, and the activities to which the stereotype is relevant and, thus, determines on which behaviors, on which people, and in which settings [the predicament will be experienced]” (Steele et al., 2002, p. 390).

Third, finding a contingent reason for underperformance is liberating because threat “can be experienced by the members of any group about whom negative stereotypes exist— generation “X,” the elderly, white males, etc.” (Spencer et al., 1999, p. 6) or “the elderly, the young, Methodists, Blacks, Whites, athletes, artists” (Steele et al., 2002, p. 390). This universality has the effect of clearing some of the shame from belonging to a stigmatized group, since the negative experience is universally possible even if not evenly shared across groups in practice.

Fourth, an identity predicament or contingency operates on *metastereotypes*, or the individual’s perception of the stereotypes of the other, and does not require actual stereotypes to be present in a threatening context at all. Therefore, “identity contingencies can profoundly affect a person...without her encountering a single prejudiced person” (Steele, 2010, p. 212). This definition therefore allows the exploration of the *effects* of prejudice while neatly sidestepping some profound and intractable debates in social psychology about the nature of prejudice, how to measure it, and the resistance of prejudice to change (cf. Allport, 1954; Arkes & Tetlock, 2004; Banaji, Nosek, & Greenwald, 2004; Blumer, 1958; Devine, 1989; Fiedler, Messner, & Bluemke, 2006; Gaertner & Dovidio, 2005).

Finally, as argued by Aronson, Lustina, Good, Keough, Steele & Brown (1999, p. 44)), “this situationist view of minority underperformance is an encouraging one because it locates the problem not exclusively within the person, but within the social circumstances

confronting the person .. [and] underscores how changing those circumstances, even subtly, can have dramatically positive effects on performance" (p. 44)

It should be noted here that there is some dissent amongst Steele and his colleagues about these definitions. In earlier work, *identity predicaments* and *identity contingencies* were used as roughly interchangeable terms to refer to the same intersections of identity, context and task demands (Spencer et al., 1999; Steele, 1997; Steele & Aronson, 1995). However, Purdie-Vaughns, Steele, Davies, Ditlmann, & Crosby (2008) argue that stereotype threat is a subtype of a broader class of social contingencies, or that "being perceived through the lens of a negative stereotype is one of many different types of identity contingencies a person may face" (p. 627). However, two years later Steele (2010) reverts to using the phrase "identity contingency" to refer to the specific circumstances that produce stereotype threat.

In the same paper, Purdie-Vaughns, Steele, Davies, Ditlmann, & Crosby (2008) make several other important suggestions that are also overlooked in Steele's (2010) capstone book<sup>2</sup>: Firstly, that "contingencies are not always negative," even for negatively stereotyped identities, "for instance, if one's differentness creates positive visibility" (p. 617); and second that "threatening and positive contingencies may even be present simultaneously" (p. 617). Although the work of Steele and his colleagues never takes seriously the fact that stereotype threat may be a subtype of identity contingency, or the multiple, flexible and relational nature of identity contingencies, these are important themes that will be developed in this dissertation.

#### **2.4. Defining stereotype threat**

The concept of "stereotype threat" has now been fully accepted in the social psychology literature, to the point that authors are able to refer to it virtually without

---

<sup>2</sup> Although hints of these positions emerge in interviews *about* the book (eg. Gates & Steele, 2009)

introduction or definition (eg. Hoyt, 2005; Wicherts, Dolan, Carlson, & van der Maas, 2010). However, there are actually several quite different ways in which the concept has been defined and operationalized in the literature, even within the relatively closed circle of Steele and his colleagues.

#### **2.4.1. Performance-based definitions of stereotype threat**

The most complete definitions of stereotype threat combine group membership, stereotype valence, contingent features of a setting, the relevance of a stereotype to a task, and experiential features of the intersection of these features and consequent performance on a stereotype-relevant task. For example, Ambady, Paik, Steele, Owen-Smith, and Mitchell (2004) argue that "under stereotype threat, when an individual risks confirming a negative self-relevant stereotype, activation of the stereotype can inhibit performance on a subsequent, related task" (p. 401). Similarly, Deaux, Bikmen, Gilkes, Ventuneac, Payne, and Steele (2007) define stereotype threat as follows:

"When negative stereotypes about a group's abilities and potential are "in the air," they can undermine the performance of members of that group. When a stereotype is believed to be relevant to a domain of performance, it poses the threat that the person will be judged or treated in terms of the stereotype. The impact of that threat is reduced performance on domain-relevant tasks," (p. 386)

However, many definitions of stereotype threat in the literature omit one or more of these features.

#### **2.4.2. General, non-specific threat**

Some early work tended to define stereotype threat loosely and generally, as a "strictly situational threat of negative stereotypes" (Steele, 1997, p. 617), a "self-evaluative threat" (Spencer et al., 1999, p. 14; Steele & Aronson, 1995, p. 797) or "a threat in the air" (Steele, 1997, p. 614). This general definition is agnostic of cause or mediation and

specifically “does not depend on cuing an internalized anxiety or expectancy” (Steele, 1997, p. 617).

#### **2.4.3. Emotion or concern about being judged by, or reduced to, the negative stereotype**

Other definitions, generally used concurrently with the first, depend on understanding stereotype threat as an affective response evoked by task performance when a negative stereotype is relevant to that performance. Here stereotype threat is "a concern with the significance of one's performance in light of a devaluing stereotype" (p. 798) that results from a “fear [of] being reduced to that stereotype” (Steele, 1997, p. 614). It is a threat arising from concern that “others' judgments or their own actions will negatively stereotype them in the domain<sup>3</sup>” (Steele, 1997, p. 613), and the targets experience “apprehension” about facing “judgment based on societal stereotypes about one’s group” (Spencer et al., 1999, p. 6). It is a “concrete, real-time threat of being judged and treated poorly in settings where a negative stereotype about one's group applies” (Steele et al., 2002, p. 385).

By 2005, Davies, Spencer & Steele (2005) defined stereotype threat as “the risk of being judged by, or treated in terms of, those negative stereotypes [which] can evoke a disruptive state among stigmatized individuals,” creating a “situational predicament [that] can undermine stigmatized individuals’ performance and aspirations in any targeted domain” (p. 277). Notice that this definition begins to commit to a particular set of mechanisms by which negative stereotypes, and the sense of threat they may invoke, can undermine performance.

#### **2.4.4. Emotion or concern about potentially confirming the negative stereotype**

A third class of definitions focuses on the concerns or negative emotions evoked by the threat of *confirming* the stereotype by low performance (Aronson et al., 1999). However,

---

<sup>3</sup> A curiously circular definition.

there is ambiguity in these definitions about whether the threat or apprehension refers to potentially confirming that the stereotype about *your group* is true, or rather to confirming that the stereotype about the group is *true of yourself*. For example, Adams, Garcia, Purdie-Vaughns & Steele (2006) define stereotype threat as “the concern that one’s performance will confirm or be interpreted in light of cultural stereotypes about inferiority in a specific performance domain” (p. 603).

Most definitions used by Steele and his colleagues specify the latter, arguing that “stereotype threat refers to ... apprehension over confirming, or eliciting the judgment that the stereotype is *self-characteristic*” (Steele & Aronson, 1995, p. 810, emphasis added) or “confirming it as a self-characterization, both to one’s self and to others who know the stereotype” (Steele & Aronson, 1995, p. 808). This risk “increases the cost of failure because such failure threatens to confirm the alleged limitation in ability both in the eyes of others and perhaps in one’s own eyes as well” (Cohen, Steele, & Ross, 1999, pp. 1302-1303). This results in “extra pressure in situations where their behavior can confirm the negative reputation that their group lacks a valued ability ... [which] we call ... stereotype threat” (Aronson et al., 1999, p. 30).

However, a much later definition co-authored by Steele argues that “stereotype threat [is] a threat of confirming negative *ingroup* stereotypes” (Carr & Steele, 2009, p. 853, emphasis added) and, more specifically, “the threat of confirming a negative stereotype about one’s group” (Carr & Steele, 2009, p. 853).

#### **2.4.5. The problem with defining stereotype threat without reference to performance**

Strangely, the definitions above are primarily concerned with a loosely specified disruptive affective or cognitive response to being negatively stereotyped without reference to actual performance. This class of definitions has allowed some investigations of stereotype



threat to not even measure performance at all. For example, Carr and Steele (2009) measured perseverance, stereotype suppression, performance expectancy, domain identification, mood, motivation and concern with making mistakes, but not task performance. By defining stereotype threat primarily as a cognitive-emotional threat response they were able to make an argument that stereotype threat had been experimentally induced, even though they could not show that the threatened group had scored worse on the threatened task. Similarly Stone (2002) tested the effects of belonging to a negatively stereotyped group on self-handicapping, but participants were never required to actually complete the expected task itself<sup>4</sup>.

The problem with this type of definition is that stereotype threat is only an interesting and powerful explanatory concept if it is linked to performance deficits. There is nothing interesting or surprising in the notion that being the target of a negative stereotype makes you feel anxious, threatened or unhappy. Conversely, it would be relatively unsurprising if people responded to negative stereotypes with increased arousal that *increased* their task performance, for example by performing better to prove the naysayers wrong.

The model only becomes politically powerful when it is also argued that stereotypes (a) uni-directionally threaten performance and (b) are contingent on features of the context that are amenable to interventions. Indeed, Steele and his colleagues (Steele, 1997, 2010; Steele & Aronson, 1995; Steele et al., 2002) put a great deal of effort into demonstrating that negative stereotypes have a contingent and unidirectional impact on task performance in key performance areas, since it is this feature that makes “stereotype threat” theoretically, practically and politically viable.

---

<sup>4</sup> Interestingly, this study has frequently been cited *as if stereotype threat actually impacted on performance* (eg. Beilock et al., 2006; Schmader, Johns & Forbes, 2008) despite the fact that no performance task occurred.

At the same time, the performance-linked definition is limiting because it can only apply to a certain class of activities and social settings. Therefore, Steele and his colleagues increasingly seem to be turning towards definitions that hinge on the hypothesized affective-cognitive response to threat that is thought to cause underperformance. If task performance can be reframed as a possible subtype of activities that might be undertaken under conditions of identity threat then the reach of stereotype threat to be extended to more generalized social phenomena, such as the likelihood of white passengers sitting next to a lone black male in an airplane (Steele, 2010) or the personal distance that participants choose on threatening or non-threatening conversations with outgroup members (Goff, Steele, & Davies, 2008). Adams, Garcia, Purdie-Vaughns & Steele (2006) suggest that identity threat is “an extension of the more specific predicament of stereotype threat” (p. 603) in which it is not concerns about *task performance* that is in doubt, but the performance of social identity more generally in which “some aspect of the environment signals the danger that a person might be evaluated on the basis of a threatened social identity” (p. 603). In this way stereotype threat can be repositioned as a fundamental social-psychological process that contributes to a very broad range of social issues, including intergroup contact and microsegregation, geographical segregation, ceiling effects in employment, and so on.

The argument that will be laid out in this dissertation is in agreement with Steele and his colleagues that stereotype threat is a subtype of a class of social-psychological and experiential processes that has explanatory power in a very wide range of social phenomena. However, Steele and his colleagues will be challenged on two fronts: firstly, on their premature commitment to unidirectionality; and, secondly, on their commitment to a

particular set of emotional and affective responses as the primary *cause* of the empirical effect<sup>5</sup>.

#### **2.4.6. A working definition of stereotype threat**

In this study, stereotype threat will be defined as a situational modifier of task performance that occurs when a group stereotype becomes relevant to the performance of a stereotype-relevant task. It should be noted that neither the direction of the effect, nor mechanisms by which the effect is produced are specified in this definition, and the reasons for this will become evident as the review of literature is fleshed out.

#### **2.5. The mainstream model of stereotype threat**

In the earliest attempts to specify a model of stereotype threat, there was more focus on what stereotype threat is *not* than what it is. For example, Steele (1997) argued that it “not tied to the psychology of particular stigmatized groups” (p. 617), it does not require the physical presence of outgroup members, but can be experienced while alone through the mental representations of a threatened context, and it does not need “one [to] believe the stereotype nor even be worried that it is true of oneself” (p. 618). By 2002 Steele, Spencer and Aronson were still not in a position to pin down the model and argued that stereotype threat has “profound effects” on a wide range of behaviour, including standardized test performance, and “beyond that to the very nature of ... personal and social identity” and that these effects are “mediated through multiple processes that, from one experience of the threat to the next, will vary in their degree of involvement” (Steele et al., 2002, p. 435). However, by 2010, Steele was ready to commit to a set of causal mechanisms, and argued that stereotype threat results in a negative cycle in which awareness of a negative stereotype

---

<sup>5</sup> As Steele himself has pointed out, “Stereotype Threat Does Not Live by Steele and [colleagues] Alone” (2004) and the theoretical model of stereotype threat has developed over a number of years and has been contributed to by a large number of independent authors. However, there are important orthodoxies and assumptions in the field that are directly related to Steele and colleagues’ work that continue to shape the theorizing in the field. This justifies the focus on their work in this review.

triggers: (1) a dilution of concentration; (2) which results in anxiety and frustration; (3) which results in reduced performance. Steele argued that reduced ability to perform results in further dilution of concentration and the cycle therefore repeats and intensifies in a feedback loop.

### **2.5.1. Targets of stereotype threat**

A large number of studies have demonstrated that stereotype threat affects members of a very large number of groups in a wide variety of performance domains. For example, many studies have explored the paradigmatic case of African Americans and other racial or ethnic minorities in general academic performance and standardized tests (eg. Aronson, Fried, & Good, 2002; Cohen, Garcia, Apfel, & Master, 2006; Gonzales, Blanton, & Williams, 2002; Steele, 1997; Steele & Aronson, 1995). Other non-ethnic minorities, such as students from low socioeconomic backgrounds (Croizet, Desert, Dutrevis, & Leyens, 2001; Croizet & Claire, 1998) or American 'Southerners' (Clark, Eno, & Guadagno, 2010) have also been shown to be vulnerable to stereotype threat.

Groups that are usually dominant in a domain have been shown to experience stereotype threat when compared to groups with even higher contextual status, such as white males in math when compared to Asian males (Aronson et al., 1999). Men have also been shown to suffer performance decrements under stereotype threat in affective tasks or tasks requiring social and emotional intelligence (Cadinu, Maass, Lombardo, & Frigerio, 2006; Koenig & Eagly, 2005; Leyens et al., 2000) or childcare (Bosson, Haymovitz, & Pinel, 2004).

A large body of work has demonstrated that stereotype threat impacts on women in mathematics (Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003; Carr & Steele, 2009; Spencer et al., 1999), science and engineering (Bell, Spencer, Iserman, & Logel, 2003; Good, Woodzicka, & Wingfield, 2010; Murphy, Steele, & Gross, 2007), general work performance (Bergeron, Block, & Echtenkamp, 2006), entrepreneurial intentions (Gupta & Bhawe, 2007),

leadership aspirations (Davies et al., 2005), and negotiation skills (Kray, Reb, Galinsky, & Thompson, 2002). Women in one study were twice as likely to run over jaywalkers in a driving simulator when they were under stereotype threat (Yeung & von Hippel, 2008). Other studies with workplace and employment relevance have shown stereotype threat to impact on black managers' feedback-seeking and feedback-discounting (Roberson, Deitch, Brief, & Block, 2003) and on temporary employees' general performance when their job competence was called into question (von Hippel et al., 2005).

Older people have been shown to be sensitive to stereotype threat (Abrams et al., 2008) in memory performance (Chasteen, Bhattacharyya, Horhota, Tam, & Hasher, 2005; Hess et al., 2009a; Hess, Emery, & Queen, 2009b) and physical function (Hausdorff, Levy, & Wei, 1999).

Stereotype threat has been shown to influence performance in sports, such as golf and putting tasks (Beilock, Jellison, Rydell, McConnell, & Carr, 2006; Stone et al., 1999) and student athletes have been shown to experience stereotype threat when their identity as athletes is made relevant during academic tests (Dee, 2009; Yopyk & Prentice, 2005).

Stereotype threat has also been shown to impact on the cognitive performance of ecstasy and cannabis users on cognitive tests (Cole, Michailidou, Jerome, & Sumnall, 2006; Looby & Earleywine, 2010) and random survey participants on a test of political knowledge when participants were matched with an interviewer of a different race (Davis & Silver, 2003). People diagnosed with mental illness have also been shown to experience the effect (Henry, von Hippel, & Shapiro, 2010; Quinn, Kahng, & Crocker, 2004; Steele, 2010).

This massive body of research is generally supportive of Steele and colleagues' (Steele, 1997, 2010; Steele & Aronson, 1995; Steele et al., 2002) claim that stereotype threat is a general phenomenon that ultimately affects all people from all walks of life in one context or another. However, it is also evident that certain groups, such as women and

African Americans, carry a greater burden of negative stereotypes and are likely to experience stereotype threat in a wider variety of contexts, and in more important contexts with more serious consequences.

### **2.5.2. Triggers of stereotype threat**

Anything that can trigger categorization and thereby activate stereotypes can trigger stereotype threat. These range from blatant triggers, for example giving people scientific or media reports on intergroup differences in the domain (Hoyt & Blascovich, 2010), to subtle triggers, such as requiring participants to report their race (Alter, Aronson, Darley, Rodriguez, & Ruble, 2010), write a biographical narrative (Quinn et al., 2004) or view pictures of groups with low minority representation (Purdie-Vaughns et al., 2008) prior to testing. Some key research in the field has been more subtle still, and simply manipulated whether the test is presented as ‘diagnostic’ of a stereotype-relevant ability or as a puzzle or task (Steele & Aronson, 1995). This, it is argued, is enough to tap powerful stereotypes of racial or gender differences in specific settings and invoke stereotype threat (Steele et al., 2002).

### **2.5.3. Consequences of stereotype threat**

Several consequences of stereotype threat have been identified and will be briefly described here. Most will be discussed in more detail later as mediators and moderators of the theoretically key stereotype threat – performance link.

#### **2.5.3.1. *Reduced task performance***

The key defining effect of stereotype threat – the one that has made it famous – is the reduction of performance for targets of negative stereotypes in high-stakes settings. According to *reducingstereotypethreat.org* there have been at least 300 experiments empirically exploring the stereotype threat effect (Stroessner, Good, & Webster, 2010). Meta-analyses have broadly supported the argument that negative stereotypes reduce task

performance for targets of the negative stereotype in stereotype-relevant tasks in the way predicted by stereotype threat (Nguyen & Ryan, 2008; Walton & Cohen, 2003; Wei, 2009b). Therefore, there is compelling evidence that negative stereotypes can and do invoke underperformance for targets in stereotype-relevant tasks.

### *Challenges*

Although few disagree that the negative stereotype threat effects on performance are reliable and replicable in laboratory conditions, there have been some criticisms arguing that negative stereotypes do not impact on performance in real-life settings in the same way. For example, Cullen, Hardison and Sackett (2004) compared actual relationships between (presumably) independent predictors and performance on negatively stereotyped tasks with samples of 49,374 college students and 5,397 US army personnel – large samples by any standards. They concluded that the “study did not support the proposition that the phenomenon of stereotype threat is an appreciably important explanatory mechanism for the Black–White differences in cognitive ability test scores typically observed” in applied settings (p. 229). The study was repeated with a subset of the sample (N = 21,164) and some methodological improvements with similar results.

In a similar vein, Stricker and Bejar (2004) found that manipulating the difficulty of items presented in a computer-adaptive Graduate Record Examination (GRE) test had little impact on test performance or affective indicators of stereotype threat (N=343). Stricker and Ward (2004) examined patterns of performance on the Advanced Placement (AP) examinations amongst 2,993 high-school students and found that asking students to report their category membership before or after the test had no impact on students’ results. This is surprising, given that Steele and Aronson (1995) had found that asking participants to report their demographics prior to a test had invoked stereotype threat and reduced their performance. Stricker and Ward (2004) argued that “motivation to perform well was

probably heightened in the high stakes [real-life] settings” of their studies and they hint that this may result in reduced stereotype threat effects.

However, Danaher and Crandall (2008) reanalyzed Stricker and Ward’s (2004) data and came to a very different conclusion, arguing that “their conservative criterion for evidence led them to overlook significant stereotype threat effects with real practical implications” (p. 1639). Predictably, Stricker and Ward’s (2008) rejoinder supported their original conclusions, arguing that Danaher and Crandall’s conclusion was “unwarranted because of flawed estimates and extrapolations,” selective descriptions, and debatable and incorrect criticisms (p. 1656).

#### **2.5.3.2. *Stereotype boost and lift***

It must be noted that it has also become clear that positive stereotypes can improve performance for members of *positively* stereotyped groups, such as for Asian students in mathematics (Aronson et al., 1999). Early studies of the stereotype-performance link demonstrated that ‘stereotype susceptibility’ to positive stereotypes can boost performance (eg. Shih, Pittinsky, & Ambady, 1999), and an early meta-analysis convincingly demonstrated that being aware of a negatively stereotyped outgroup confers an advantage with approximately half the effect size of stereotype threat referred to as ‘stereotype lift’ (Walton & Cohen, 2003). There have even been studies where negative stereotypes have resulted in *increased* performance for members of the negatively stereotyped group (eg. Crisp et al., 2009), termed “stereotype reactance” by Kray and her colleagues (Kray, Thompson, & Galinsky, 2001).

However, these positive performance shifts are generally overlooked in the ‘official’ versions of stereotype threat promoted by Steele and colleagues. For example, Steele’s (2010) latest book summarizing advances in the first fifteen years of stereotype threat research does not discuss stereotype susceptibility, lift or boost at all despite citing Walton and Cohen’s



(Walton & Cohen, 2003) meta-analysis in which the lift effect was shown to be an intrinsic feature of environments in which some people are negatively stereotyped (indeed, it was titled *Stereotype Lift*). Nor do Steele and colleagues (Steele, 2010; Steele et al., 2002) discuss ironic stereotype activation effects, such as negative stereotypes improving performance (eg. Crisp et al., 2009; Kray et al., 2001) or positive stereotypes decreasing performance (eg. Cheryan & Bodenhausen, 2000). These ironic effects will be discussed in more detail after the mainstream model of stereotype threat has been described in full.

### ***2.5.3.3. Affective and physiological effects***

The impact of negative stereotypes on performance has traditionally been considered a “hot” process, mediated by affect rather than a “cold” automatic process (Wheeler & Petty, 2001). Anxiety has often, but not always, been shown to be associated with experiences of stereotype threat (Bosson et al., 2004; Chung, Ehrhart, Holcombe Ehrhart, Hattrup, & Solamon, 2010; Ford, Ferguson, Brooks, & Hagadone, 2004; Johns, Inzlicht, & Schmader, 2008; Marx & Stapel, 2006a; Ployhart, C., & McFarland, 2003; Sawyer & Hollis-Sawyer, 2005). Marx and Stapel (Marx & Stapel, 2006a) argue that targets of stereotype threat are likely to experience increased anxiety before the test, but increased frustration afterwards. Several studies have shown that targets of stereotype threat generally experience increased physiological arousal, including increased blood pressure, heart rate, skin conductance and temperature (Blascovich, Spencer, Quinn, & Steele, 2001; Hoyt & Blascovich, 2010; Murphy et al., 2007; Osborne, 2006; Osborne, 2007). Blascovich and colleagues argue that prolonged exposure to stereotype threat situations could partially account for the increased blood pressure observed in African American populations in the USA compared to white populations (Blascovich et al., 2001).

#### **2.5.3.4. *Disidentification and reduced aspirations***

A key longitudinal component of Steele and colleagues' model is that a common way of protecting one's self from the unpleasant experience of stereotype threat is to progressively disinvest from the domain in which the negative stereotype is active (Davies, Spencer, Quinn, & Gerhardtstein, 2002; Steele, 1997, 2010; Steele & Aronson, 1995; Steele et al., 2002). This is a way of protecting self-esteem, since if one is not personally invested in a domain then failure is not as personally threatening.

Although many studies have found domain identification to be an important factor in the stereotype-performance relationship (Smith, 2004), predictions about disidentification over time have not easily been tested since the most common designs in experimental social psychology are cross-sectional. However, there is some evidence from longitudinal and field studies that disidentification does occur in the ways predicted by the stereotype threat model (Chang, Eagan, Lin, & Hurtado; Massey & Fischer, 2005; Osborne & Walker, 2006).

#### **2.5.3.5. *Effort, interest and achievement motivation***

In Steele and Aronson's original sketch of stereotype threat (1995) they suggested that stereotype threat may result in reduced effort and achievement motivation. However, by 2002 they were forced to concede that there was no evidence that stereotype threat results in reduced effort. Later, it began to appear that stereotype threat may, ironically, result in too much effort (Nussbaum & Steele, 2007). 'Over-effort' became a central argument in Steele's (2010) later formulation of the stereotype threat model.

Others (eg. Walton & Cohen, 2007) have shown that stereotype threat reduces achievement motivation. Smith, Sansone and White (2007) showed that stereotype threat reduced task interest for those high in achievement motivation, but *increased* task interest for those lower in achievement motivation.

So far, then, the results are complex. What is clear is that being the target of a task-relevant stereotype changes one's relationship to the task, although it is not yet clear what form such changes may take.

#### **2.5.3.6. *Risk aversion/prevention focus***

Seibt and Förster (2004) suggested that being subject to positive or negative stereotypes results in shifts in regulatory focus. Specifically, that being the target of a negative stereotype induces a “prevention focus of vigilance” resulting in better analytical performance, while being the target of a positive stereotype results in a “a promotion focus state of eagerness” resulting in more creativity (p. 38). Others have confirmed these results, demonstrating that being the target of a negative stereotype may result in a prevention focus (Grimm, Markman, Maddox, & Baldwin, 2009; Keller & Bless, 2007, 2008), performance-avoidance goals (Brodish & Devine, 2009; Chalabaev, Sarrazin, Stone, & Cury, 2008; Smith, 2006) and inflexible perseverance rather than creativity (Carr & Steele, 2009). Indeed, several of these studies have found that negative stereotypes may *increase* performance for tasks that require a ‘negative’ prevention focus, straightforward analytic effort, or perseverance rather than creativity. These shifts in orientation are not limited to the local and immediate task-performance environment, but may have far-reaching impacts on long-term goals and aspirations in the context (von Hippel, Issa, Ma, & Stokes, 2010).

#### **2.5.3.7. *Identity management, stereotype suppression, and working-memory depletion***

Several studies have shown that being exposed to a negative self-relevant stereotype diverts cognitive resources to stereotype and identity management functions – cognitive resources that are required for the completion of the task. These identity management functions include stereotype-suppression (Logel, Iserman, Davies, Quinn, & Spencer, 2009; Schmader, Johns, & Forbes, 2008), denial of the importance of the stereotype (von Hippel et

al., 2005), self-defense strategies including vigilance, agitation and diffused attention (Förster, Jens, Higgins, & Werth, 2004), defensive projection (Govorun, Fuegen, & Payne, 2006), mistrust (Marx, Brown, & Steele, 1999), decreased belonging and desire to participate (Murphy et al., 2007), and increased emotional regulation (Johns et al., 2008). At the same time, targets of negative stereotypes are also likely to experience increased performance monitoring (Schmader et al., 2008) and vigilance for evidence that their behavior may be confirming the negative stereotype (Schmader, 2010). Many argue that the net effect of these additional processes is the depletion of working memory (Beilock, Rydell, & McConnell, 2007; Hess et al., 2009a; Rydell & Boucher, 2010; Rydell, McConnell, & Beilock, 2009; Schmader, 2010; Schmader, Forbes, Shen, & Berry Mendes, 2009; Schmader & Johns, 2003; Schmader et al., 2008; Wraga, Helt, Jacobs, & Sullivan, 2007).

#### **2.5.3.8. *Multiple effects***

It is evident from this discussion that at this point, fifteen years after the publication of Steele and Aronson's (1995) seminal study, the field has failed to identify a structured set of effects and experiences that consistently accompany the stereotype threat effect. Rather, it is increasingly agreed that the effects are multiple and contingent.

#### **2.5.4. Preconditions, limits, moderators and buffers of stereotype threat**

Research in stereotype threat has explored a very wide range of contextual and individual moderators of stereotype threat. Unfortunately, results have been mixed and none have been shown to be completely reliably or predictable. Spencer, Steele and Quinn (1999) laid out a fairly comprehensive set of parameters for the effect:

"It assumes that the test taker construes the test as a fairly valid assessment of ... ability, that they still care about this ability at least somewhat, and that the test be difficult. Stereotype threat effects should be less likely if the test is either too easy or too difficult (either in item content or time allotted) to be seen as validly

reflecting ability. Also, if the test taker has already disidentified with [the domain], in the sense of not caring about their performance, stereotype threat is not likely to drive their performance lower than their lack of motivation would. Thus, it is only when the test reflects on ability and is difficult and the test takers care about this ability that the stereotype becomes relevant and disturbing as a potential self-characterization" (p. 25)

The parameters laid out in this early work reflect fairly well the broad categories of preconditions, limits, moderators and buffers that would be explored in the following decade.

#### **2.5.4.1. *Features of the context***

One of the most exciting features of stereotype threat is that it proposed a model “rooted in intersubjectivity” (Steele & Davies, 2003, p. 313) and proposed, in theory at least, that the empirical effect was the result of a complex intersection of contextual, individual and social features. However, as the machinery of experimental social psychology methodically picked apart the problem, the focus increasingly shifted to individuals, and to individual differences in cognitive and affective responses that might predict and explain reduced performance under stereotype threat. Despite Steele’s early focus on stereotype threat as a *predicament* and his later attempts to refocus attention on stereotype threat as a *contingency* (Purdie-Vaughns et al., 2008; Steele, 2010), the contextual features of stereotype threat and – more importantly – the interfaces between context and individual performance – have not featured strongly in empirical stereotype threat research.

#### **2.5.4.2. *Features of the task***

Stereotype threat has been shown to affect performance on a wide variety of tasks and activities, such as math and language performance (Steele, 1997; Steele & Aronson, 1995), affective processing (Leyens et al., 2000), golf-putting (Beilock et al., 2006), driving (Yeung & von Hippel, 2008), and political knowledge (Davis & Silver, 2003; McGlone, Aronson, &

Kobrynowicz, 2006) amongst others. Even highly respected psychometric tests such as the Raven's Advanced Progressive Matrices (Brown & Day, 2006; Croizet et al., 2004; Mayer & Hanges, 2003; McKay, Doverspike, Bowen, & Martin, 2002) and the Implicit Association Test (Frantz, Cuddy, Burnett, Ray, & Hart, 2004) have been shown to be subject to stereotype threat effects. However, it is a central tenet of most stereotype threat models that some tasks are more likely to invoke STB than others.

#### *Difficulty of the task*

Steele (1995) argued that task difficulty is an important pre-requisite for stereotype threat to occur and this was confirmed by Spencer, Steele and Quinn (1999), who found that women underperformed under stereotype threat on difficult but not easy math tests. This has been confirmed empirically in many studies (see Smith, 2004, for reviews; Steele et al., 2002). However Stricker (1999; 2004) tested this hypothesis with items from the graduate Record Exam (GRE) and found that manipulating test difficulty had no effect on performance or indicators of stereotype threat.

Quinn & Spencer, 2001 argue the test does not need to be *objectively* difficult, but that it will only invoke stereotype threat if it is "at very edge of the [target's] abilities: If the test is easy and the test taker experiences little difficulty with it it is unlikely that stereotype threat will interfere with performance .... [but] if the test is so difficult that the test taker knows it is out of the bounds of his or her skills, stereotype threat is also unlikely to show up in performance differences" (p. 58).

Steele, Spencer and Aronson (2002) argue the impact of task difficulty on stereotype threat relies on both the objective difficulty of the items and the extent to which factors combine in the performance situation to produce frustration. They suggest that this frustration may impact on performance by giving weight to the negative stereotype and thereby amplifying stereotype threat. Additionally, they argue, the more difficult the task, the less

leeway there will be for the types of distractions and discomforts produced by stereotype threat. They were confident enough to argue that “the ingredient most critical to producing a stereotype threat effect on test performance is frustration on the test. If it is not a sufficient ingredient, it is very likely a necessary one” (p. 391).

Nguyen and Ryan’s (2008) meta-analysis confirmed this early hunch, showing that “stereotype-threatened minorities performed more poorly than did nonthreatened minorities when cognitive ability tests were highly difficult ... than when tests were moderately difficult” (p. 1324). However, an unanticipated consequence of Steele and colleagues’ early insistence that stereotype threat will primarily affect performance on difficult tasks meant that very few researchers have explored the effects of stereotype threat on *easy* tasks.

Although Spencer, Steele and Quinn (1999) simply found no difference in performance between threatened and nonthreatened women on easy math tests, O’Brien and Crandall (2003) found that women under stereotype threat performed *better* than men on an easy math test but exhibited conventional underperformance on a difficult test. Nguyen and Ryan’s (2008) meta-analysis provides some support for this finding, showing that “women underperformed when a math test was highly difficult ..., more so than when a math test was moderately difficult ... [but] ... when the test was easy, women tended to improve their test performance slightly” (p. 1324).

While these results generally support the early claims that stereotype threat is most likely to occur with difficult items, it seems strange not to treat women’s *overperformance* on easy items under threat as an interesting feature of the phenomenon. Instead, a protective boundary is drawn around ‘stereotype threat’ by arguing that it can only occur when items are difficult and, by corollary, categorizing improved performance for certain groups on certain items as a different kind of phenomenon. The almost universal use of difficult tasks in

stereotype threat research has therefore potentially resulted in a narrowing of scope that excludes potentially interesting and important features of the phenomenon.

#### *Perceived meaning of the task*

The earliest research in the field invoked stereotype threat simply by convincing participants that a test is “diagnostic” of ability (Croizet & Claire, 1998; Steele & Aronson, 1995). On the other hand, convincing participants that a test is something more innocuous, like a problem-solving task, can eliminate the underperformance associated with stereotype threat in certain situations (Steele & Davies, 2003). It is therefore clear that the same task can be experienced in different ways, and that some ways of framing a task are more likely to result in stereotype threat than others.

#### **2.5.4.3. *Features of the stereotype***

Despite some serious disagreements in the social psychological literature about what stereotypes are, where they reside, how they are produced and transmitted and what they are for (eg. Fiedler & Walther, 2004; Hilton & von Hippel, 1996; Macrae, Stangor, & Hewstone, 1996; Oakes, Haslam, & Turner, 1994), they are generally invoked uncritically in stereotype threat research. However, several features of stereotypes have been found to moderate stereotype threat effects.

#### *Stereotype self-relevance*

The key feature that differentiates stereotype threat from simple priming effects is that targets must perceive the stereotype to be both self-relevant and relevant to the performance of a task in order for performance to be affected (Marx et al., 1999). Priming effects, on the other hand, invoke behavior regardless of the relevance of the stereotype to the target, such as in the case of college students walking more slowly when primed with a stereotype of the aged (Bargh et al., 1996; Wheeler & Petty, 2001).



The importance of self-relevance of the stereotype for stereotype threat effects is axiomatic, but also empirically well established and relatively uncontroversial. For example, Deaux, Bikmen, Gilkes, Ventuneac, Joseph, Payne, and Steele (2007) found that first-generation black Afro-Caribbean immigrant students did not underperform when a test was presented as diagnostic, although African American students exhibited classic stereotype threat effects. Deaux and her colleagues argued that Afro-Caribbeans avoided stereotype threat because, although they perceived themselves as black, they subtyped themselves as fundamentally different to African Americans. They were therefore able to avoid perceiving negative stereotypes about blackness and academic performance as self-relevant, and therefore were able to avoid stereotype threat effects.

Schmader (2002a) found that women's levels of domain identification moderated the extent to which they experienced stereotype threat and the extent to which their task performance was negatively affected. When you look at the items used by Schmader to assess ingroup identification, such as "Being a woman/man is an important part of my self image," it is evident that they are likely to also be assessing the extent to which group stereotypes are generally experienced as self-relevant.

Marx and Stapel (2006b) manipulated whether participants visualized a stereotype target from first-person or third-person perspective. This, they argued, influences the extent to which participants not normally threatened by the stereotype would empathetically experience it as self-relevant. As predicted, the performance of stereotype targets suffered when primed by writing about a stereotype from both a first-person and third-person perspective, since the stereotype relevance was indelibly part of their social identity no matter what form the prime took. Those belonging to groups not normally included in the stereotype, however, were only affected by priming/stereotype threat effects when they wrote about the prime from the first person and thereby empathetically made the stereotype self-relevant.

*Relevance of the stereotype to the task*

As discussed above, the theoretically distinctive feature of stereotype threat is that it requires a stereotype to be activated, self-relevant and relevant to the task in order to produce performance decrements. Specifically, “when the performance is defined so that the negative stereotype is not relevant to it, the performance of the stereotyped group matches that of the nonstereotyped group” (Steele et al., 2002, p. 394). For example, Brown and Pinel (2003) showed that stereotype threat negatively impacted on performance when participants believed that the test exhibited gender differences in performance, but were eliminated when a test was presented as gender fair and therefore unrelated to gender stereotypes. Meta-analyses have shown the effect of task-relevance on the stereotype-performance link to be robust (Nguyen & Ryan, 2008).

*Stereotype prevalence and legitimacy*

Surprisingly, Steele, Spencer & Aronson (2002) argued that a well established social stereotype is *not* necessary to activate the stereotype threat effect. They gave the example that if you load your plate at a family buffet, only to realize that others have still not eaten, you might overhear others saying that your branch of the family is “so selfish.” Steele and his colleagues argued that in this situation you might spend the rest of the time being obviously generous to counter the “reputational threat” that the misunderstanding generated (p. 384). However, elsewhere they argued that the more powerful and negative the stereotype, the more likely it is that the effect will be experienced (eg. Steele, 1997) and the more ubiquitous the stereotype is in society, the more likely it is to result in long-term effects.

A great deal of stereotype threat research has focused on ubiquitous stereotypes deeply embedded in social life, such as those associated with race, gender and age. Many of these stereotypes, such as the expectation that blacks are intellectually inferior or that women have inferior math skills, are not only well known, but are backed by a great deal of scientific

and pseudo-scientific research giving them legitimacy. For example, psychologists are still arguing in scientific journals that black/white differences in IQ and male/female differences in visuo-spatial and math ability are rooted in genetic inferiority (eg. Johnson & Bouchard, 2007; Rushton & Jensen, 2005a, 2005b; Rushton, Skuy, & Fridjhon, 2003).

However, not all stereotypes are equally legitimate – and legitimacy is unstable, since stereotypes are socially produced and the outcome of constant negotiation and renegotiation in society (Reicher, Hopkins, & Condor, 1997). For example, ‘Irish jokes’ are still popular (a Google search returns 1.6 million hits), even though it is unlikely that anyone really believes them to be true. This was not always the case, and the trope of the ‘Irish joke’ has developed from scientific and popular Darwinian discourse in the 18<sup>th</sup> and 19<sup>th</sup> centuries arguing that the Irish were a distinct and deficient race (Boltwood, 2001). Therefore, although stereotypes about the Irish still exist and are relatively well known, they are no longer as legitimate as they used to be and are no longer supported by the weight of scientific and pseudoscientific dogma.

The impact of stereotype prevalence and legitimacy has not received a great deal of attention in stereotype threat research. However, Nguyen and Ryan’s (2008) meta-analysis broadly supported the notion that these might be important features, finding that effect sizes for performance decrements under stereotype threat differed across stereotypes. Race stereotypes (which carry more stigma than gender stereotypes) elicited a larger negative effect on performance. These are important findings, because they are reminders that stereotype threat effects are tied in to social structures that cannot be considered to be merely individual perceptions or cognitions.

The importance of legitimacy has been directly tested very seldom. Schmader, Johns, and Barquissau (Schmader, Johns, & Barquissau, 2004) found that "women who believe that status differences between the sexes are legitimate were more likely to endorse gender

stereotypes" and that stereotype endorsement was related to performance deficits under stereotype threat. Shih, Pittinsky, and Trahan (2006) found that "emphasizing the social construction of race buffers individuals from stereotype threat effects" (p. 1), demonstrating that the perceived legitimacy of the category is as important as the perceived legitimacy of the stereotype.

### *Stereotype activation*

Key to the stereotype threat phenomenon is the notion that a stereotype must be 'activated' in order to impact on performance. Although this lies at the heart of the theory, it is not clear that researchers know what stereotype activation is or how it happens. Early studies manipulated the 'diagnosticity' of experimental tasks and assumed that this also activated negative stereotypes (eg. Steele & Aronson, 1995). This claim was backed by the differences in word-completion tasks and other measure of implicit stereotype activation between those in diagnostic and non-diagnostic conditions. Other studies have manipulated the proportion of representation of minorities in the experimental context or in experimental materials to manipulate category activation and therefore activate stereotypes (Purdie-Vaughns et al., 2008; Sekaquaptewa & Thompson, 2003). Others have manipulated the "stereotypical content" of priming materials such as television adverts to activate stereotypes (eg. Davies et al., 2002).

One feature of stereotype activation that has received a great deal of attention is whether the primes are implicit or explicit. Implicit primes generally activate stereotypes without the conscious awareness of the target, such as when the diagnosticity of the test is manipulated (eg. Croizet & Claire, 1998; Steele & Aronson, 1995; Stone et al., 1999), the experimenter belongs to the positively stereotyped group (Stone & McWhinnie, 2008), there is an unbalanced demographic ratio in a setting (Murphy et al., 2007), or participants are asked to indicate their race on the front page of a test (Steele & Aronson, 1995).

Explicit primes, are much more blatant and, it is argued, generally make the stereotype relevant in such a way that the target will probably be consciously aware of it, such as when Yopyk and Prentice (2005) asked participants to write a narrative about a recent sporting competition in order to invoke their negatively stereotyped identities as student athletes. Other studies are more explicit still, both explicitly invoking the identity and spelling out the negative stereotype. For example, Leyens et al. (2000) told their participants that “it is a well-known fact that men are not as apt as women to deal with affect in general and specifically to process affective information as effectively” (p. 1193).

Nguyen and Ryan’s (2008) meta-analysis specifically explored the extent to which implicit and explicit methods of activating stereotypes moderated or mediated performance. For racial stereotypes they found that “subtle stereotype threat cues produced smaller stereotype threat effects” but that “moderately explicit threat-activating cues produced a greater mean effect size than blatant cues for minority test takers” (p. 1328). In fact, they argued that blatant negative stereotype primes might “ironically invoke behaviors that are inconsistent with the stereotype,” but this was not discussed in detail or integrated with the STB model. However, for gender stereotypes they found that “explicit threat-activation cues (both blatant and moderate) generally produced smaller mean effect sizes than subtle cues” (p. 1328). They concluded that “moderately explicit cues” are most likely to invoke performance decrements under stereotype threat. These cues are “direct enough to draw targets’ attention, ambiguous enough to cause targets to engage in detrimental off-task thinking..., but not too blatant to make some targets become motivated to “prove it wrong” (p. 1328).

Clearly there is a complex relationship between the nature of the category, the nature of the performance task and the way that stereotypes are invoked. Most intriguingly, here is another instance of behaviours that are “ironically ... inconsistent with the stereotype” (2008)

being excluded from the ambit of the phenomenon because they are inconsistent with the stereotype threat model, which assumes that exposure to self-relevant negative stereotypes should consistently result in reduced task performance.

#### **2.5.4.4. Identity**

Although research has largely focused on ‘stereotype activation’ in the tradition of priming research (eg. Jamieson & Harkins, 2007; Wheeler & Petty, 2001), Steele and his colleagues have consistently emphasized that they see stereotype threat as a predicament of *identity* in which a person’s identity in a performance context and social features of that context, such as negative stereotypes, combine to produce hurdles that negatively impact on performance in that context (Steele, 1997, 2010; Steele & Aronson, 1995; Steele et al., 2002). Although identity is not clearly defined or articulated in the stereotype threat literature, several features of identity have been explored in some detail.

##### *Domain identification*

Domain identification represents the extent to which an individual cares about the domain of performance in which stereotype threat occurs. Steele, Spencer and Aronson (2002) argue that:

The strength of stereotype threat should also depend on how much the person identifies with the domain of activity to which the stereotype applies. The term ‘identification’ refers here to the degree to which one’s self-regard, or some component of it, depends on the outcomes one experiences in the domain. The more one cares about a domain in this way, the more important one’s fate in the domain is likely to be and the more upset one is likely to be over the prospect of being negatively stereotyped in the domain (p. 390).

Steele (1997) had already argued that domain identification is a prerequisite for sustained success in “school and its subdomains” and many other important settings.

However, he argued, being negatively stereotyped in a domain makes it more difficult to identify with it and increases the psychic cost of maintaining the identification required to sustain high performance at high levels. If stereotype threat chronically reduces performance, then it is likely that targets will respond by disidentifying from the domain as a way of protecting their self-esteem (Steele, 1997, 2010; Steele & Aronson, 1995; Steele et al., 2002).

The general hypothesis that high domain identifiers are more susceptible to stereotype threat has been tested in “surprisingly ... few studies” (Nguyen & Ryan, 2008, p. 1316). Steele and his colleagues reported on four studies that looked at domain identification and concluded that “stereotype threat effects occur only, or at least most strongly, among those people who are identified with the intellectual domain in which they are being tested” (Steele et al., 2002, p. 395). Smith’s (2004) meta-analysis found some evidence for moderation of stereotype threat effects on performance, but not for mediation. (However, it should be noted that most of the studies reported on did not properly test for mediation). Walton and Cohen’s (2003) meta-analysis also found “tentative evidence” (p. 461) that domain identification magnifies stereotype threat effects.

Nguyen and Ryan’s (2008) more extensive meta analysis did not find conclusive evidence that domain identification moderates or mediates stereotype threat effects in race stereotypes. However, when looking at gender-based stereotype threat effects they found that highly math identified women experienced smaller stereotype threat effects ... than did moderately low math-identified women” (p. 1324), although these results were not highly reliable. Indeed, Keller (2007) found that low domain identifiers performed *better* when they were negatively stereotyped. Nevertheless, these results only provide partial support for Steele and colleagues’ predictions that high domain identification is a precondition for stereotype threat effects.

However, the interesting thing about domain identification is that it suggests a mechanism by which stereotype threat chronically affects performance and shifts minority students' orientations to domains (Steele, 1997, 2010; Steele & Aronson, 1995; Steele et al., 2002). The vast majority of research in stereotype threat has used laboratory based cross-sectional paradigms and so has not been able to test this hypothesis, but Massey and Fischer (2005) longitudinally surveyed over 4000 students at 28 institutions and "uncover[ed] a clear process of disidentification in response to minority stereotyping" (p. 19).

#### *Ingroup identification*

In their early work, Steele and colleagues considered the importance only of identification to the *domain* but not identification with the stereotyped ingroup (Aronson et al., 1999; Cohen et al., 1999; Marx et al., 1999; Spencer et al., 1999; Steele, 1997; Steele & Aronson, 1995), and it is never well articulated how *domain* identification and *group* identification overlap. Although it was relatively peripheral to their argument, Steele, Spencer and Aronson (2002) did begin to consider the notion that ingroup identification may be an important factor in stereotype threat effects.

Specifically, they argued that "in general, the more one is identified with the group about whom the negative stereotype exists, or the more one expects to be perceived as a member of that group, the more stereotype threat one should feel in situations where the stereotype applies" (p. 391), but they also argued that the extent to which ingroup identification buffers or amplifies the stereotype threat effect depends on the nature of the group identity and the cultural resources that the group has developed to deflect the negative consequence of prejudice and stigmatization. They argued that African Americans, due to their long history of stigmatization, have group defenses available that may be accessed by increased group identification.



Schmader (2002) found that women with higher levels of ingroup identification were more susceptible to stereotype threat and others have shown that targets of stereotype threat tend to distance themselves from the threatened identity (eg. Crisp et al., 2009). For example, in Steele and Aronson's (1995) earliest studies they showed that black students under threat distanced themselves from stereotypically African American interests and pursuits, such as rap, jazz, basketball and boxing.

### *Multiple identities*

Although stereotype threat research has often focused on inescapable group memberships written on the body, such as race and gender, Steele and his colleagues argued that the impact of stereotype threat should be "reduced ... by the perception that one has personal features, or counterstereotypic capacities that will deflect the application of the stereotype to oneself" (Steele et al., 2002 p. 391). They gave the example that an identity as a clinical psychologist may buffer a man against the stereotype of male emotional insensitivity and thereby reduce his sensitivity to stereotype threat in that domain.

This argument has received empirical support. Shih, Pittinsky and Ambady (1999) explored stereotype threat amongst female Asian math students and found that they performed worse in a math task than a control group when their identity as women was made relevant, but better when their math-positive identity as Asian-Americans was made relevant. Similarly, Rydell, McConnell, and Beilock (2009) showed that female college students were less susceptible to stereotype threat when their alternative identities as college students were made relevant than when only their identity as women was primed.

Shih, Bonam, Sanchez, and Peck (2007) compared monoracial and biracial targets of stereotype threat, and found that "monoracial participants showed significant performance changes in reaction to race salience [but] multiracial individuals did not" (p. 125), which they

attribute to the additional identity resources available to biracial targets that buffer them from stereotype threat.

There is therefore some empirical support for Steele and colleagues position that stereotype threat is a predicament of identity, or “identity contingency” in which a target of a negative stereotype has to bring whatever identity resources they have at their disposal to bear on the performance of a task in a particular context. However, the vast majority of stereotype threat research focuses on single experimenter-defined identities and strips research participants of other identities that might be relevant to them in the experimental context (cf. Danziger, 1994).

#### *2.5.4.5. Features of the individual task-taker*

Although stereotype threat was envisioned as a contextual and socially contingent phenomenon (Steele, 1997; Steele & Aronson, 1995; Steele et al., 2002), the majority of empirical studies in the field have focused on individual difference variables rather than social and structural variables at macro-individual levels.

#### *Stereotype knowledge*

McKown and Weinstein (2003) showed that knowledge of the types of cultural stereotypes relevant to stereotype threat develops between the ages of 6 and 10, and that children from stigmatized groups become aware of broadly held negative social stereotypes at earlier ages than non-stigmatized children. Additionally, they were able to show that “children from stigmatized groups who were aware of broadly held stereotypes withdraw more effort and perform worse ... [than] children who are not aware of broadly held stereotypes” (p. 511). To some extent it is the expectation that targets must have some knowledge of the negative stereotype that theoretically differentiates stereotype threat from self-fulfilling prophecy theories like the ‘Pygmalion effect’ (cf. Gyll, Madon, Prieto, & Scherr, 2010; Rosenthal, 1994; Rosenthal & Jacobson, 1968).

### *Stereotype endorsement*

As mentioned previously, Steele and colleagues have firmly argued that neither endorsement nor internalization (cf. Allport, 1954) is necessary for stereotype threat to disrupt performance (Steele, 1997, 2010; Steele & Aronson, 1995; Steele et al., 2002). Although Leyens and his colleagues (Leyens et al., 2000) found that stereotype endorsement was unrelated to performance decrements under stereotype threat, other studies have shown that stereotype threat effects were more pronounced for those targets who personally endorsed the stereotype (Chatard, Selimbegovic, Konan, & Mugny, 2008; Marx & Goff, 2005; Schmader et al., 2004). Keifer and Sekaquaptewa (2007) showed that even implicit stereotype endorsement impacted on women's math performance, and particularly so under stereotype threat conditions. Interestingly, Blanton, Christie and Dye (2002) found that stereotype endorsement *mediated* how participants evaluated their performance, suggesting the counter-intuitive possibility that agreeing with a negative stereotype about your *group* can actually make you feel better about your own performance because it changes the frame of reference. In other words, if you perform moderately overall, but are still near the top of the negatively stereotyped group then you can still feel like a top performer.

### *Stigma consciousness*

Stigma consciousness refers to chronic self-consciousness of one's stigmatized status (Pinel, 1999) and therefore goes beyond mere awareness that a negative stereotype exists and might be held by others in the testing environment. Steele, Spencer and Aronson (2002) mention that there is suggestive, but not definitive, evidence that stigma consciousness amplifies the stereotype threat effect.

This was confirmed by Brown and Pinel (2003), who found that participants high in stigma consciousness performed worse than non-stigma-conscious persons under stereotype threat conditions but not in control conditions. When actual grades were compared, stigma-

conscious students tended to underperform as predicted by stereotype-threat, but non-stigma-conscious minority members did not (Brown & Lee, 2005). However, Hess, Hinson & Hodges (2009a) found that increased stigma consciousness was associated with *increased* performance amongst older adults when negative age stereotypes were invoked.

Mendoza-Denton, Shaw-Taylor, Chen and Chang (2009) found that women high in stigma consciousness performed worse only when it was ambiguous whether a male interviewer was prejudiced. They performed equally well when he was explicitly chauvinistic or egalitarian. Therefore, the findings suggest that stigma consciousness might either increase or decrease performance under stereotype threat depending on other contextual and individual factors.

#### *Status concerns*

Josephs, Newman, Brown and Beer (2003) argue that stereotype threat effects are a subtype of a more general process of status maintenance or enhancement. Using testosterone as an indicator of status-enhancement orientation, they found that participants highest in testosterone were most sensitive to stereotype threat. They conclude that status concerns amplify stereotype threat effects, and the authors have found similar results in later studies (Josephs, Sellers, Newman, & Mehta, 2006; Newman, Sellers, & Josephs, 2005).

#### *Internal locus of control*

Cadinu, Maass, Lombardo, and Frigerio (2006) found that although their performance was better in general, people reporting an internal locus of control “showed a sharper decrease in the stereotype threat condition compared to individuals with external locus of control beliefs” (p. 183). Gupta and Bhawe (2007) found that the performance of women with “more proactive personality” were more significantly disrupted by stereotype threat than women with less proactive personality. However, others have not found locus of control to

significantly influence stereotype threat effects (eg. McKay, Doverspike, Bowen-Hilton, & McKay, 2003).

#### *Coping sense of humour*

Ford, Ferguson, Brooks and Hagadone (2004) were interested in the effect of coping sense of humour, or "the use of humor as a strategy for coping with stressful experiences" (p. 645), on stereotype threat effects. They found that targets with high coping sense of humour did not underperform in stereotype threat conditions while those with low coping sense of humour did. Since humour functions to subvert or maintain the legitimacy of social orders (eg. Billig, 2001a; Griffiths, 1998), such as the status hierarchies underlying stereotype threat, it is possible that those high in coping sense of humour were less vulnerable to stereotype threat because of their ability to undermine the legitimacy of the stereotype and the testing situation.

#### *Self-monitoring*

'Self-monitoring' is "the desire and ability to control one's self-expressions in order to cultivate a desired public image" (Inzlicht, Aronson, Good, & McKay, 2006a, p. 325). Inzlicht and his colleagues found that participants high in self-monitoring "may be resilient to threatening environments because they react to negative stereotypes with increased (and not decreased) performance" (p. 323).

#### *Achievement motivation*

'Achievement motivation' refers to generalized and long term motivation to perform well. The study of achievement motivation in education has a long history and by the 1960's it was already a well established concept (eg. Rosen & D'Andrade, 1959). It is closely related to 'domain identification,' which—being defined as interest and motivation to succeed in a specific domain—can be considered a more specific subtype of achievement motivation (Keller & Bless, 2007). Smith (2004) argued that a combination of high achievement

motivation and negative stereotypes in a domain leads to avoidance goals, lower interest in the task and ultimately lower performance. Similarly, Smith and White (2007) found that participants higher in achievement motivation were more likely to respond to stereotype threat with performance avoidance goals and that these goals influenced task performance via task absorption.

Although Chalabaev, Sarrazin, Stone, and Cury (2008) did not measure chronic achievement motivation, they also found that participants were likely to respond to stereotype threat with performance-avoidance goals. However, they were not able to demonstrate that these task avoidance goals impacted negatively on performance.

Keller (2007) argued that "overmotivation is particularly likely to emerge in individuals with a high chronic level of achievement motivation (e.g. high domain identifiers)" (p. 328) and that stereotype threat effects on performance are largely due to overmotivation. His results generally confirmed these hypotheses.

### *Education*

Andreoletti and Lachman (2004) explored the effects of negative aging stereotypes on memory performance. Their results suggested that participants "with more education are more resilient when faced with negative age stereotypes about memory ... [but] those with less education show greater susceptibility to the detrimental effects of age stereotypes" (p. 129).

It is possible that education acts as a buffer for negative stereotypes about memory because highly educated participants are more likely to doubt the legitimacy of the stereotype, although this is conjecture. Alternatively, being educated might provide an alternative identity with much more positive stereotypes in the performance domain of memory and thereby buffer targets from threat.

*Affects the vanguard*

Steele and his colleagues have always insisted that stereotype threat is most likely to negatively impact on the ‘vanguard,’ in other words, to impede the performance of the most gifted and motivated minority members (Steele, 1997, 2010; Steele et al., 2002). This, Steele (1997) suggests, is because the vanguard are likely to be most domain identified and therefore most susceptible to threat effects. They are also more likely to be performing at the edge of their capabilities in real-life contexts (Steele, 1997; Walton & Spencer, 2009), and therefore more likely to suffer the negative impact of stereotype threat on performance.

There is fairly sound empirical support for the predictions that stereotype threat influences targets with higher domain identification and more profoundly on difficult tasks. However, direct empirical evidence for the more general contention that stereotype threat effects are most pronounced for the ‘vanguard’ is scarce.

Indeed, there is some emerging evidence contradicting this prediction. Crisp, Bache and Maitner (2009) argued that women who have chosen a traditional masculine environment and successfully survived to a fairly high level would have developed strategies and resources to cope with negative stereotypes leveled against them. As predicted, their female engineering student participants performed *better* on difficult math tests under stereotype threat conditions.

*Features specifically not necessary to invoke stereotype threat*

Steele and his colleagues have been very clear on two conditions that are specifically not necessary for stereotype threat to be experienced, although they might amplify the effects. First, generalized or historical stigma is not necessary: stereotype threat “can arise or become disruptive as a result of situational pressures alone” (Aronson et al., 1999, p. 29). Second, it is not necessary for targets of stereotype threat to believe in the stereotype or worry that it could

be true, although they must be aware that the stereotype applies to them and is relevant in the situation (Steele, 1997).

### **2.5.5. Hypothesised cause(s) of stereotype threat**

One of the enduring problems with stereotype threat is that it is a fascinating empirical effect still looking for a cause. And, despite being articulated as a social and contextual effect on group performance, the culture of experimental social psychology has demanded individualistic explanations embedded in the affective and cognitive processes of individuals. In Steele and Aronson's (1995) first sketch of stereotype threat they suggested a number of potential causes, including distraction, narrowed attention, anxiety, self-consciousness, withdrawal of effort, overeffort, attributional ambiguity, self-doubt activation and stereotype avoidance. They suggested that "several such mechanisms may be involved simultaneously, or different mechanisms may be involved under different conditions" (p. 809), but tentatively said that their "best assessment is that stereotype threat caused an inefficiency of processing much like that caused by other evaluative pressures" (p. 809). Later Steele (1997) argued that "the emotional reaction [that stereotype threat] causes could *directly* interfere with performance" (p. 614, emphasis added).

#### **2.5.5.1. Anxiety**

The emotional state most commonly suspected of causing performance decrements under stereotype threat is anxiety, but empirical results have been very mixed. Most studies used self-report measures of anxiety, and some found heightened anxiety under stereotype threat but no consequent mediational effect on underperformance (eg. Hess et al.; Spencer et al., 1999). While some studies have found little effect of anxiety at all (eg. Good et al., 2010; Keller & Dauenheimer, 2003; Leyens et al., 2000; Stangor, Carr, & Kiang, 1998), others have found evidence of partial mediation, or at least interesting partial relationships (eg. Sawyer & Hollis-Sawyer, 2005). For example, Chung, Ehrhart, Holcombe Ehrhart, Hattrup, and



Solomon (2010) found a weak relationship between state anxiety, stereotype threat and task performance in a model of real-life performance on job promotion exams. Ford, Ferguson, Brooks, and Hagadone (2004) found that state anxiety mediated test performance only through “coping sense of humor.” Other studies have found that anxiety is implicated in stereotype threat effects, but only in some conditions (Osborne, 2001; Ployhart et al., 2003; Stone et al., 1999). Smith (2004) reported that across 17 studies only two found “small but reliable evidence for anxiety as a partial mediator” while the other 15 did not find evidence that anxiety mediates stereotype threat effects.

Many of the studies above largely used self-report measures of anxiety. Others have used physiological measures and found that physical markers of anxiety such as blood pressure, skin conductance, and skin temperature were elevated for participants under stereotype threat (Blascovich et al., 2001; Bosson et al., 2004; Osborne, 2006; Osborne, 2007). However, in these studies self-report measures of anxiety were generally unrelated to physiological measures. Since anxiety is only one of many affective states associated with increases in these physiological markers, the results could mean that participants under threat were suppressing their anxiety and were unaware of it (Johns et al., 2008), or that they were accurately reporting their levels of anxiety but the physiological markers were associated with different affective states, such as frustration or anger.

Marx and Stapel (Marx & Stapel, 2006a) concluded that targets of stereotype threat feel anxiety before and during a test but frustration once they’re finished. However, many studies that have found no mediation by anxiety have placed the anxiety measure immediately prior to the stereotype threatened task and still had inconclusive results (eg. Schmader, 2002a).

### 2.5.5.2. *Arousal*

Several studies have explored arousal as a potential mediator or cause of stereotype threat and, again, the results have been mixed. Some studies (eg. O'Brien & Crandall, 2003) have found patterns of arousal and performance that suggests that arousal directly impacts on performance and that underperformance under stereotype threat is a result of over-arousal (in other words, being on the wrong end of the inverted U of the arousal-performance curve).

Others have made arguments that arousal impacts on performance indirectly. Ben-Zeev, Fein and Inzlicht (2005) explored the relationship between arousal, attributions about arousal and performance and found some evidence of a mediating relationship. Schmader and her colleagues (Schmader et al., 2008) have made a similar argument that arousal (stress) impacts on performance under stereotype threat conditions in conjunction with cognitive processes such as self-monitoring or vigilance and suppression of negative stereotypes and attributions. A later paper by Schmader and colleagues (Schmader et al., 2008) treated anxiety and arousal as essentially the same construct (“anxiety arousal”) and conclude that “anxious arousal and negative thinking do not predict cognitive impairments in isolation. Rather, affective and cognitive variables combine in a metacognitive way such that performance is impaired in high-stakes situations where one’s initial affective experience is interpreted in light of concurrent negative thoughts” (p. 595).

### 2.5.5.3. *Other emotions or affective states*

Despite mixed empirical results, anxiety and arousal are still receiving most attention in stereotype threat research. However, many other emotions and affective states might be associated with the effect. For example, Keller and Dauenheimer (2003) explored the mediating effects of anxiety and the dimensions of “agitation–quiescence” and “dejection–cheerfulness,” concluding that dejection was the primary mediator of the stereotype threat effect. Smith’s (2004) meta-analysis reported that investigators have explored the role of

evaluation apprehension, self esteem and depression (with no evidence of moderation or mediation) and performance confidence (with mixed results). Out of 18 studies that included a self-confidence measure, only three found “some evidence for partial mediation” (p. 190).

#### **2.5.5.4. *Performance expectancies***

One of the earliest causal suggestions was that stereotype threat may lower performance expectations for the targets of negative stereotypes, thereby reducing their likely task performance (Steele et al., 2002). Stangor, Carr and Kiang (1998) were the first to empirically test this and found that invoking threat did indeed reduce participants’ performance expectations, but they did not assess actual performance differences under stereotype threat and could not comment on whether performance expectancies actually influenced performance. Some studies have replicated this finding that performance expectancies were sensitive to stereotype threat, but could not be shown to impact task performance (Kellow & Jones; Kray et al., 2001). Others found that stereotype threat had no effect on performance expectancies (eg. Spencer et al., 1999). However, there are some studies that have shown that stereotype threat both impacts on performance expectancies and at least partially mediate task performance for some groups (Cadinu et al., 2003; Rosenthal, Crisp, & Suen, 2007; Stone et al., 1999). The evidence that performance expectancies generally mediate performance shifts under stereotype threat is therefore neither reliable nor strong.

#### **2.5.5.5. *Beliefs about the test***

In contrast to Allport (1954) who argued that stereotypes affect performance via internalized self-expectations Marx, Brown and Steele (1999) argued that "stereotypes do not appear to affect self-expectations; instead they appear to foster a climate of mistrust that results in depressed performance" (p. 491). They argue that targets of stereotype threat cannot

be sure that their results will be fairly judged and that this uncertainty results in uncertain performances.

This argument has been a feature of Steele and colleagues' work since their first studies in stereotype threat manipulated 'test diagnosticity' as a way of invoking stereotype threat (eg. Steele & Aronson, 1995). Their logic was that groups such as African Americans and women have such a long history of stigma and prejudice that simply believing that a generally negatively stereotyped ability was being tested with a diagnostic test would be enough to activate the negative stereotype and threaten performance. However, despite its importance as a trigger of stereotype threat, Smith's (2004) review concluded that "test perceptions [do] not emerge as a complete or partial mediator of the stereotype-threat–performance relationship" (p. 191).

It should be noted that beliefs about a *test* are closely linked to beliefs about whatever constructs the test is thought to measure. For example, beliefs about the 'diagnosticity' of an IQ test are related to one's beliefs about the legitimacy of IQ as a construct and the stability or malleability of the construct within individuals over time.

#### **2.5.5.6. Effort**

Steele and Aronson (1995) originally suggested that the negative impact of stereotype threat on performance may be related to withdrawal of effort or over-effort. They measured self-reported effort, but found no relationship between effort and performance. Smith (2004) reviewed 18 studies that used both self-report and behavioural measures of effort and concluded that "effort was not shown to completely or partially mediate the stereotype-threat-performance relationship" (p. 184). Other studies have also failed to find significant moderating or mediating relationships for effort for stereotype threat related performance (Brown & Pinel, 2003; Gonzales et al., 2002; Koenig & Eagly, 2005; O'Brien & Crandall, 2003).

However, there is some evidence that reduced effort may be related to stereotype threat effects. McKown & Weinstein (2003) did find that children in diagnostic stereotype threat conditions reported withdrawing more effort and this group also performed significantly worse in a threatened task. Massey and Fischer (2005) explored stereotype threat in real-world educational settings and found that students reporting stereotype threat withdrew effort and reported less identification with the performance domain over time.

More studies have shown *increased* effort under stereotype threat. Aronson, Lustina, Good, Keough, and Steele (1999) found that targets of threat reported significantly more effort but were unable to demonstrate that the increased self-reported effort was directly related to performance decrements under threat. Croizet, Despres, Gauzins, Huguet, Leyens, and Meot (2004) found some physiological evidence that effort increased under threat, but were also unable to directly relate this to task performance differences. Beilock, Jellison, Rydell, McConnell, and Carr (2006) showed that stereotype threat conditions increased effort in a golf-putting task requiring fluid practiced movement and thereby reduced performance by directing “too much attention ... to processes that usually run more automatically” (p. 1059). Jamieson and Harkins (2007) demonstrated that stereotype threat increased motivation and argued that increased “mere effort” under threat resulted in inflexible brute-force approaches to problems that resulted in reduced performance when flexibility in approach was required and increased performance when “mere effort” was appropriate.

The relationship between stereotypes, effort and performance, therefore, is unclear. Sometimes stereotype threat results in *decreased* effort and reduced performance (Massey & Fischer, 2005; McKown & Weinstein, 2003); and other times in *increased* effort and decreased performance (Beilock et al., 2006; Jamieson & Harkins, 2007). Therefore despite the central position of effort in early outlines of stereotype threat (eg. Steele & Aronson,

1995), it is far from clear how effort is related to performance under stereotype threat, if at all, or what variables may moderate this relationship.

#### 2.5.5.7. *Self-handicapping*

'Self-handicapping' is defined as a tendency "to search for external explanations for ... weak performance," such as reporting that one did not get enough sleep the night before the test, as a way of shifting attribution for failure to external sources. It was considered as an important feature by Steele and Aronson (1995), who found that stereotype threat had significant effects on three of their four self-handicapping measures. Others have replicated these results, finding that self-handicapping increases under threat conditions (Keller, 2002; Schimel, Arndt, Banko, & Cook, 2004). Stone (2002) measured actual practice prior to an expected performance task (although actual performance was not measured) and found that those under threat practiced significantly less than those in other conditions. However, some studies (eg. Keller & Dauenheimer, 2003) have found no evidence for increased self-handicapping under threat, even in real-life stereotype threat situations.

There is an important distinction to be highlighted here: most studies exploring self-handicapping under stereotype threat have treated it as a set of strategies for rhetorically deflecting the consequences of poor performance from innate qualities of self or inescapable group categories to transitory situational factors. In most cases, the behaviours that were reported on to assess 'self-handicapping' were out of the scope of the study, and so only the *report* could change under stereotype threat. However, Stone (2002) explored self-handicapping by measuring differences in practice during the experiment itself, thereby capturing self-handicapping as an actual behavioural strategy that could both impact on performance and provide a rhetorical account for that poor performance.

Despite the relatively reliable shifts in self-handicapping under stereotype threat, and Smith's (2004) conclusion in her review that "to date, behavioral self-handicapping has not

been adequately tested as a potential stereotype-threat–performance mediator" (p. 185), interest in self-handicapping in stereotype threat seems to have largely declined. This is unfortunate, since it is one strand of empirical work in stereotype threat that (implicitly) acknowledged that targets may actively be seeking ways to positively differentiate themselves in the threat situation rather than passively reacting as subjects.

#### **2.5.5.8. *Regulatory focus and strategy***

A fairly convincing finding is that participants under threat may approach tasks differently and apply different strategies (Hess et al., 2003). These strategy shifts may be related to shifts in ‘regulatory focus’ in response to positive or negative stereotypes. For example, Smith and White’s (2002) results showed that participants under ‘stereotype nullified conditions’ (ie. explicitly not under stereotype threat) were able to focus more intently on task performance and consequently performed better. Inzlicht, McKay, and Aronson (2006b) found that “individuals belonging to stigmatized groups are less able to regulate their own behavior when they become conscious of their stigmatizing status or enter threatening environments” (p. 262). This, they argued, results in diminished concentration, diminished ability to tolerate discomfort and therefore to diminished task performance.

Seibt and Förster (2004) argued that positive stereotypes activate a “promotion focus state of eagerness” whereas negative stereotypes activate a “prevention focus state of vigilance” (p. 38). They argued that these different regulatory foci result in different strategies of engagement with tasks when exposed to positive compared to negative stereotypes. Specifically, targets exposed to positive stereotypes tended to perform faster, more creatively and less accurately than those in the negatively stereotyped group who exhibited better analytical performance. These patterns were supported by Grimm, Markman, Maddox and Baldwin (2009) who showed that stereotype threat could be negated by changing the reward structure of the task and inducing more flexible performance (cf.

Jamieson & Harkins, 2007; Jamieson & Harkins, 2009). The link between stereotype threat and inflexibility/perseverance has also been supported by Carr and Steele (Carr & Steele, 2009) and is increasingly gaining empirical support.

#### **2.5.5.9. *Vigilance and suppression***

Schmader, Johns, and Forbes (2008) offered a slightly different notion of ‘vigilance,’ arguing that stereotype threat invokes “a tendency to actively monitor performance ...and... suppress negative thoughts and emotions in the service of self-regulation” (p. 336). They argue that this vigilance and suppression “disrupts performance on sensorimotor tasks directly” and “consume[s] executive resources needed to perform well on cognitive and social tasks” (p. 336). This class of potential mediators had been suggested previously (eg. Quinn & Spencer, 2001) and received early empirical support (Steele et al., 2002). However, Schmader and her colleagues (2008) argue that physiological stress (such as anxiety and arousal), vigilant monitoring processes and stereotype suppression processes act on performance distally via ‘working memory efficiency.’ On the other hand, Inzlicht, Aronson, Good & McKay (2006a) found that participants high in self-monitoring (and, presumably, vigilance) performed *better* under stereotype threat.

#### **2.5.5.10. *Working memory***

Schmader and Johns (2003) were among the first to argue that stereotype threat impacts on task performance by impacting on ‘working memory,’ defined as “the ability to focus one’s attention on a given task while keeping task irrelevant thoughts at bay” (Engle, 2001 as cited in Schmader & Johns, 2003, p. 440). Specifically, they found that priming negative stereotypes did seem to reduce working memory capacity, and that this reduction mediated stereotype threat effects. This model of stereotype threat disrupting task performance via disruptions to working memory has some empirical support (Beilock et al., 2007; Rydell et al., 2009).



However, a key criticism of this model is that it does not readily account for stereotype threat effects in domains that do not rely heavily on working memory for success, such as golf putting (eg. Stone et al., 1999). Beilock and her colleagues (Beilock et al., 2006) supported this critique by demonstrating that golf putting was indeed sensitive to stereotype threat effects (cf. Stone, 2002; Stone, Perry, & Darley, 1997), but also showed that negative stereotypes can disrupt performance “by inducing too much attention to execution rather than too little” (p. 1069).

Other studies have explicitly tested the working memory hypothesis and failed to find support (eg. Hess et al., 2009a; Jamieson & Harkins, 2007; Jamieson & Harkins, 2009), including a neuroimaging study that found that “mental load induced by negative stereotype messages is not generated in working memory, but rather in regions associated with social and emotional processing” (Wraga et al., 2007, p. 18).

#### ***2.5.5.11. Failure of single-variable models***

The empirical evidence seems clear that it is unlikely that the effect of stereotype threat on performance is directly or independently caused by any of the contenders examined in the literature so far. Smith’s (2004) review ended with the conclusions that “how stereotype threat does its damage to performance remains unclear” and that “no single mediator should be prematurely dismissed” (p. 192). Although research has continued unabated since then, it seems that we are not much closer to understanding the causes, mediators or moderators of stereotype threat effects.

At the same time, it is also clear that the variables already examined in the literature (such as anxiety, arousal, performance expectancies, effort, self-handicapping, regulatory fit, mental vigilance, and working memory) are implicated in stereotype threat effects at least some of the time. It is possible that the inconsistent patterns observed to date could be integrated with a multiple-mechanism/multiple-effect approach (eg. Wheeler & DeMarree,

2009). However, all of the candidate models explored to date have focused exclusively on individual-level variables. It is possible that as-yet unexplored moderators at different levels of analysis may synthesize these conflicting results. The integrated model approach will be discussed first.

#### ***2.5.5.12. Integrated models of stereotype threat***

Current studies in stereotype threat are increasingly examining the effects of multiple mediators and moderators working in tandem. For example Schmader, Forbes, Shen & Mendes (2009) explored how affect, metacognitive interpretations of affect, and compromised mental efficiency may combine to result in reduced performance under stereotype threat. Some integrated models of stereotype threat have been proposed, and these show some promise.

##### *The integrated process model of stereotype threat*

Schmader and Johns (2008) proposed an “integrated process model” that argues that the effect of negative stereotypes on performance is distally mediated by three distinct but interrelated processes: (1) affective and arousal responses to the threatening situation; (2) active behavior and thought monitoring processes, including regulatory focus responses; and (3) stereotype and thought suppression processes. It is argued that any or all of these will be impacted by stereotype threat in different performance contexts and that any sub-processes can negatively impact on working memory (defined as optimal task focus) and thereby degrade performance.

This model is a substantial improvement over single-variable explanations and supporting evidence seems to be relatively strong. However, it treats stereotype threat primarily as a temporary state and does not consider the impact of negative stereotypes and stigma over long periods of time. It also treats the individual as a tightly bounded system, and does not consider the social features of stereotype threat beyond being inputs into a

bio-cognitive system. Jamieson and Harkins (2007; 2009) have produced evidence that working memory is not necessarily implicated.

*Steele's (2010) model*

Steele's (2010) model fleshes out Steele and colleagues' earlier work (especially Steele et al., 2002). It is less precise than Schmader and Johns' (2008) model, but is probably more applicable to more contexts. He argues that Stereotype threat, over time, increases targets' vigilance for potential prejudice in the social environment. This increased watchfulness and rumination can itself divert attention and mental resources away from the task and negatively impact on results and "general functioning" (p. 126). This makes targets more anxious and aroused, which feeds back into the cycle and promotes heightened vigilance and decreased task-focused attention. "A full-scale vicious cycle ensues, with great cost to performance and general functioning" (p. 126). Note that, although Steele (2010) does implicate working memory amongst the mental capacities that might be diverted away from the task, he does not commit himself to a single set of mental resources as Schmader and Johns (2008) do.

Steele (2010) argues that "if people are under threats from stereotypes or other identity contingencies for long periods, they may pay a tax. The persistent extra pressure may undermine their sense of well-being and happiness as well as contribute to health problems caused by prolonged exposure to the physiological effects of the threat" (p. 127). He therefore remains committed to the notion that people chronically exposed to stereotype are likely to 'disidentify' with the domain as a means of self-protection (cf. Steele et al., 2002). He concludes:

My mission ... is to broaden our understanding of human functioning, to get us to keep in mind that, especially in identity-integrated situations, people are not only coping with the manifest tasks of the situation, but are also busy appraising threat

and protecting themselves from the risk of being negatively judged and treated. Perhaps the chief discovery of our research is that this protective side of the human character can be aroused by the mere prospect of being negatively stereotyped, and that, once aroused, it steps in and takes over the capacities of the person—to such an extent that little capacity is left over for the work at hand" (Steele, 2010, p. 213 - 214).

However, it is evident that Steele's model does not account for the broad range of empirical effects observed in stereotype threat research, as will be discussed in more detail below. It seems that the model of stereotype threat has been kept intentionally simple and general because it is the centrepiece of his political engagement to secure educational reform in the US (cf. Steele, 2010). As such, it is important to his political project that his model retains its simplicity and charm in order to appeal to a large lay audience of teachers.

*Shapiro and Neuberg's (2007) multi-threat model*

Perhaps the most innovative attempt to account for the complex, inconsistent, and often contradictory findings in the stereotype threat literature is Shapiro and Neuberg's (2007) multi-threat framework. They pose two serious questions to the field: first, are threats to the self the same as threats to the group? And second, are threats from the ingroup the same as threats from the outgroup or even the same as threats from the self? If, as they suggest, these different sources and targets of threat are phenomenologically different, then they combine to form six distinct combinations of threat source and target.

Shapiro and Neuberg (2007) argue that the stereotype threat literature has tended to treat "stereotype threat" as a single phenomenon, when in fact the stereotype threat literature covers the full spectrum of these combinations of source and target. This, they suggest, could account for some of the apparent contradictions in the literature. They conclude that stereotype threat research and intervention should pay careful attention to the nature of

intergroup relations invoked in the context, and of the precise nature and target(s) of the threat invoked. However, possibly due to its complexity, this model has not been well cited. At the time of writing<sup>6</sup> Google scholar reported that Shapiro and Neuberg (2007) had received 27 citations, compared to 49 citations for Schmader, Johns and Forbe's (2008) more individualistic 'integrated process model' that was published the following year.

### **2.5.6. Interventions and remedies**

One of the key features of the stereotype threat model that has fuelled its public and academic popularity is that it offers a reason for the underperformance of minorities that is not rooted in intractable features of biology or society. Instead it is a situationist theory that demonstrates that a fairly fragile contingencies of identity and situation contributes substantially to underperformance, and that many features of the contingency are amenable to intervention. As Steele and Aronson first argued:

"This approach is Lewinian; it is also hopeful. Compared to viewing the problem of Black underachievement as rooted in something about the group or its societal conditions, this analysis uncovers a social psychological predicament of race, rife in the standardized testing situation, that is amenable to change" (Steele & Aronson, 1995, p. 810)

Despite the difficulties in actually demonstrating that stereotype threat substantially influences performance in real-life domains, it has at least been shown to be sensitive to a large variety of interventions.

#### **2.5.6.1. Individual interventions**

Although the stereotype threat model has a strong focus on contextual and social factors, a large number of studies have explored individualistic interventions. These generally

---

<sup>6</sup> June 2010.

intervene at the level of individual appraisals for internal experiences such as arousal and affect, cognitions, values and beliefs.

### *Reappraisal*

Steele, Spencer and Aronson (2002) argued “to reduce this threat, individuals have to do something that disarms the appraisal hypothesis that they are under threat or that, if they are, it will significantly affect them” (p. 433). This idea has been taken up in the literature and has been shown to have some merit. For example, targets of threat have been shown to increase their performance when they are given a plausible reason for negative experiences such as arousal and anxiety (Johns et al., 2008; Schmader et al., 2008).

Reappraisal may take the form of teaching targets about the experience of stereotype threat and its consequences (Johns, Schmader, & Martens, 2005) or teaching students that their difficulties in school are not abnormal and are a normal part of performing in the setting (Aronson, Cohen, & McColskey, 2009a, p. ii; Good, Aronson, & Inzlicht, 2003). Ben-Zeev, Fein and Inzlicht (2005) told participants that their arousal and negative affect would be caused by a (bogus) ‘subliminal noise generator’ and observed that participants given this non-self-threatening *reason* for their experiences under stereotype threat performed equivalently to controls and the positively stereotyped group.

### *Beliefs about the malleability of the stereotyped skill or ability*

The types of skills and characteristics tested in typical stereotype threat situations range in modifiability from golf putting (Beilock & McConnell, 2004; Stone et al., 1999) to standardized IQ test performance (Brown & Day, 2006; Mayer & Hanges, 2003) and stereotype threat has been shown to impact skills and domains across this spectrum. However, it would be possible to believe either that golf putting is highly learnable or that people have a fixed aptitude or potential that can be achieved, but not exceeded, through

practice. The same range of malleability beliefs is possible with IQ, which can be regarded as a genetic and heritable capacity or as a set of skills that can be learned.

Several studies have demonstrated that targets of negative stereotypes are more likely to underperform when they believe the stereotyped skill or ability to be innate or entity-like rather than malleable or incremental (eg. Aronson et al., 2002; Good et al., 2003; Steele et al., 2002; Thoman, White, Yamawaki, & Koishi, 2008). Steele and Aronson (2002) argue that believing the stereotyped ability to be learnable, malleable or incremental “makes [targets’] stereotype about their limited ability less relevant to their performance [which] in turn, may have the further effect of enabling them to have greater trust in the setting” (p. 394).

Therefore, interventions that emphasize the malleability of the stereotyped skill or ability may be successful. However, there are limits to the extent that any intervention can change deep-seated social beliefs about the malleability of the abilities of interest in the stereotype threat literature, such as math ability and intelligence.

#### *Affirmation and self-affirmation*

Several studies have supported the use of affirmation or self-affirmation to counteract the negative effects of stereotype threat. Croizet, Desert, Dutrevis, and Leyens (2001) “led participants to indicate [on a questionnaire] that they were participating because they were good students (i.e., they liked helping graduate students and were curious).” (p. 304-305). Martens, Johns, Greenberg, and Schimel (2006) asked participants “to rank order a list of 11 ‘characteristics and values’ in order of personal importance.” These included “sense of humor, creativity, physical attractiveness, social skills, and relations with friends/family.” They were then asked to “write about why their most valued characteristic ... was personally important and describe a time when it had been particularly important to them” (p. 238). Participants in Cohen, Garcia, Apfel, and Master’s (2006) study were “presented [with] a list of values (such as relationships with friends or family or being good at art) [and] ... asked to

indicate their most important value" (p. 1308). In each of these studies, the affirmation exercise buffered participants' performance against stereotype threat.

However, it is possible that these affirmation procedures activated alternate identities that helped participants to avoid stereotype threat effects, or helped participants to individuate themselves as atypical group members. These strategies will be discussed in more detail below.

#### **2.5.6.2. Contextual interventions**

One of the clearest findings in the stereotype threat literature is that members of negatively stereotyped groups are more likely to underperform when they believe that the task or the task context is 'diagnostic' or evaluative. A large number of studies have shown that convincing stereotype targets that the task is non-diagnostic can negate stereotype threat effects (eg. Bell et al., 2003; Carr & Steele, 2009; Croizet & Claire, 1998; Elizaga & Markman, 2008; Gonzales et al., 2002; Kellow & Jones, 2005; Lawrence, Marks, & Jackson; Martens et al., 2006; Marx & Stapel, 2006c; McKown & Strambler, 2009; McKown & Weinstein, 2003; Ployhart et al., 2003; Steele & Aronson, 1995; Stone et al., 1999). For example, negatively stereotyped participants no longer underperformed on the Raven's Advanced Progressive Matrices when it was presented as a 'set of puzzles', rather than as an IQ test (Brown & Day, 2006). However, even when a test is explicitly portrayed as diagnostic of ability, if test-takers can be convinced that it is fair to members of minority groups then they may be able to avoid stereotype threat effects (Good, Aronson, & Harder, 2008).

However, there is a limit to the extent to which important performances such as the SAT, or final examinations can be presented as non-diagnostic in real life. To some extent this contextual feature is so deeply embedded in the social structures of performance that it is unlikely to be amenable to intervention in key performance areas.



The way that minority groups are treated, represented and portrayed in the setting can also be shifted. Steele, Spencer and Aronson (Steele et al., 2002) argue that to minimize stereotype threat, members of negatively stereotyped groups must have evidence that the authorities in a setting treat people with fairness, impartiality and respect" (p. 431). This may include ensuring that there is adequate representation of minority groups in positions of power which has been shown to alleviate stereotype threat effects (Marx et al., 1999; Marx & Goff, 2005).

Purdie-Vaughns, Steele, Davies, Ditlmann, and Crosby (2008) showed that perceptions of diversity did indeed influence black participants' sense of threat (although task performance was not measured). Ironically, assurances of a "colorblind" policy actually increased threat when this expressed value was contradicted by the low representation of minorities in photographs of the organization. However, black participants were more inclined to express trust when materials specifically stated that diversity was valued in the setting, even when minority representation was low.

The fact that participants in Purdie-Vaughns and colleagues' (2008) study felt most threatened when an explicit "colorblind" policy was contradicted by evidence of prejudicial hiring policies on the ground hints that *ambiguity* about prejudice and discrimination in an organization is particularly threatening. Mendoza-Denton, Shaw-Taylor, Chen, and Chang (2009) found that "the task performance of women who are chronically concerned about gender discrimination suffered when a male interviewer's gender attitudes were ambiguous, relative to when his attitudes were either explicitly chauvinistic or explicitly egalitarian" (p. 275). These studies suggest that well-intentioned efforts to reduce overt prejudice in organizations may, ironically, increase the burden of threat for members of minority groups because it will be rare for prejudice to be eliminated, but very common for it to become more subtle and ambiguous.

### **2.5.6.3. *Relational interventions***

Steele, Aronson and Spencer (2002) argued that “the relationships a person has, and the ways in which people relate to each-other in a setting can affect the sense of identity safety a potentially identity-threatened person experiences in the setting” (p. 425). Supporting this argument, they reported that “among the strongest predictors of college GPA for African American students is the number of white friends they have” (Graham, Baker, & Wapner, 1984, cited in Steele et al., 2002). Their explanation was that having friends from other groups refutes other cues of prejudice in the setting, but it is also possible that it produces shared ingroup identity or a sense that the intergroup boundaries are permeable, which will be discussed in more detail below.

They also argue that in academic settings the non-judgemental quality of teaching and feedback is vitally important (Steele et al., 2002). Cohen, Steele and Ross (1999) argued that those under stereotype threat are particularly vigilant for false feedback, and were most likely to respond positively when both high-standards and the mentors’ faith in the student to meet those standards were emphasized

### **2.5.6.4. *Social identity interventions***

Many interventions for stereotype threat focus specifically on aspects of social identity.

#### *Avoiding stereotype activation*

The simplest of these is simply to avoid activating the negatively stereotyped identity in the testing situation and thereby avoid invoking stereotype threat effects. The problem with this solution is that most environments in which stereotype threat is likely to occur are experienced as hostile by default, so just hoping that the stereotype will not be relevant is futile (Steele et al., 2002).

However, there is evidence that even in hostile and highly stereotyped environments, subtle changes to the accessibility of stereotypes can result in important performance shifts. For example, a field study of Advanced Placement tests showed that women performed better when asked to report their gender after the test rather than before (Danaher & Crandall, 2008; Stricker & Ward, 2004).

*Refute negative stereotypes*

Some studies have had success with the strategy of specifically refuting negative stereotypes. However, two meta-analyses have suggested that explicitly refuting negative stereotypes may, ironically, produce even stronger stereotype threat effects (Nguyen & Ryan, 2008; Walton & Cohen, 2003).

*Emphasize positive stereotypes*

Others argue that exposing potential targets of stereotype to positive stereotypes about their group can buffer them from the negative effects of stereotype threat, such as in Steele and colleagues' 'WISE' educational intervention (Steele, 1992; Wolfe & Spencer, 1996). For example, Hausdorff, Levy and Wei (1999) found that elderly people exposed to positive age stereotypes exhibited less stereotypically elderly behavior. Rydell, McConnell and Beilock (2009) found that making positive stereotypes relevant, even alongside negative stereotypes, eliminated underperformance. However, Ho and Sidanius (2010) found that black students high in 'public regard' (and therefore more aware of positive black stereotypes) and high in private regard were more susceptible to stereotype threat than students low in these beliefs. This suggests that positive stereotypes are helpful mainly when people do not feel the pressure to live up to them.

*Provide successful role-models and exemplars*

Providing role-models, exemplars and evidence that other ingroup members have succeeded in a context despite negative stereotypes has been shown to reduce the negative

effects of stereotype threat (eg. Blanton, Crocker, & Miller, 2000; Marx & Roman, 2002). Steele, Spencer and Aronson (2002) argued that "The mere presence of another person in a setting who is succeeding there, and who shares the social identity that is negatively stereotyped there, should reduce the threat that other people with the same identity feel in the setting" (p. 428).

Some studies have shown that simply knowing that another ingroup member has done well on the test in question can reduce stereotype threat related underperformance (eg. Blanton et al., 2000). McIntyre and colleagues (McIntyre et al., 2005; McIntyre, Paulson, & Lord, 2003) have demonstrated that the effect of having exemplars of successful ingroup members is incremental: the more exemplars there are, the less negative stereotypes are likely to affect performance.

The ideal exemplars, it seems, are those who are skilled in the specific performance domain, as opposed to those who are admirable in unrelated contexts (Marx, Stapel, & Muller, 2005). Additionally, they should successfully blend (rather than disavow) negatively stereotyped ingroup characteristics with those features seen as essential for performance in the domain. Putting it differently, the most effective role models are those who "have managed to preserve an ingroup identity that challenges or even defies [the dominant] culture [in the domain]" (Pronin, Steele, & Ross, 2004, p. 167).

For African Americans, the most exciting role model to emerge in recent years is Barack Obama, and two studies have directly explored his effect, as an exemplary exemplar, on stereotype threat for African Americans. The first, a field study, found that "when Obama's stereotype-defying accomplishments garnered national attention – just after his convention speech, and election to the presidency – they had a profound beneficial effect on Black-Americans' exam performance, such that the negative effects of stereotype threat were dramatically reduced" (Marx, Ko, & Friedman, 2009, p. 953). However Aronson, Jannone,

McGlone, and Johnson-Campbell (2009b) found that thinking of Obama before a test had no impact on performance under stereotype threat in a randomized experiment.

Although the evidence seems relatively clear that positive role models can reduce the negative effects of stereotype threat, it is not entirely clear how this comes about. Is a positive role model a challenge to the legitimacy of negative stereotype? Or does a successful role model demonstrate that it is possible to penetrate into the domain of the high status group? These issues will be discussed in more detail below.

#### *Alternative identity activation*

Several studies have reduced stereotype threat effects by activating positive identities instead of, or alongside, the negatively stereotyped identities. Most famously, Aronson, Lustina, Good, Keough, and Steele (1999) found that white men who were primarily aware of their identity as non-Asian underperformed compared to white men without this negative identity comparison. Similarly, Shih, Pittinsky & Ambady (1999) found that Asian women who were primarily aware of their identity as Asians performed better than women primarily aware of their identity as women.

Gresky, Eyck, Lord, and McIntyre (2005), found that invoking multiple identities successfully reduced stereotype threat effects for highly domain-identified targets regardless of their valence, and the more complex the better. Others have specifically tried to manipulate targets' sense of shared identity in the performance context and have found this to be a successful stereotype threat reduction strategy (Kray et al., 2001; Walton & Cohen, 2007).

Activating positive stereotypes about college students in general has also been shown to buffer minority students against stereotype threat effects (McGlone & Aronson, 2006; Rydell & Boucher, 2010; Rydell et al., 2009). However, as well as invoking a more positive identity, these latter interventions also invoke a superordinate category, thereby invoking a shared social identity in the domain.

*Blur intergroup boundaries*

Rosenthal and colleagues (Rosenthal & Crisp, 2006; Rosenthal et al., 2007) specifically manipulated intergroup boundaries, and found that targets of stereotype threat performed significantly better when they were told to think of the way that group boundaries overlapped in the context. Shih, Bonam, Sanchez and Peck (2007) achieved similar results by manipulating the extent to which participants believed that race categories were socially constructed. These interventions demonstrate that stereotype threat effects are highly sensitive to specific features of intergroup relations (such as categorization beliefs) in the performance context.

*Individuate*

As well as activating alternate identities, stereotype threat can also be alleviated by helping targets to individuate themselves, presumably making their group identity less accessible and relevant to the performance context. This has been done by making targets individually identifiable (Ambady et al., 2004) or convincing them that individual rather than group differences are being investigated (Schmader et al., 2008). Of course, there's a fine line here, as individuation may hint that the test is diagnostic, which we know to trigger stereotype threat.

Although neither Ambady and colleagues nor Schmader and colleagues made this link, there is a promising literature showing that identifiability has powerful effects on social identity that will be explored in more detail below.

**2.5.6.5. Summary of interventions**

Although it is encouraging that such a wide variety of interventions have been shown to reduce stereotype threat effects, it is also sobering there is no 'silver bullet' that eliminates stereotype threat. Perhaps Shapiro and Neuberg (2007) are right that there are several distinct types of threat that will each require interventions at different levels. On the other hand, it is

possible that the apparent lack of parsimony in the variety of interventions that have been shown to be effective demonstrates that the theoretical frameworks for understanding stereotype threat are not working particularly well.

### **2.5.7. A brief summary of mainstream understandings of stereotype threat**

The stereotype threat literature is large, expanding and only marginally coherent given the large variety of mediators, moderators and interventions that have been identified as important, but not reliably so. As with many promising areas in social psychology, the engine of experimental reductionism has produced a fragmented reflection of the phenomenon that leaves us relatively certain that there is some substance to it, but unsure of what that substance may be. Despite his full participation in the process, Steele (eg. 2003) has heatedly resisted attempts to pin the phenomenon down to a specific and fully predictable set of antecedents, mediators, moderators and consequences, arguing that “stereotype threat is not an absolute thing that either generalizes or not .... it is a predicament that varies in strength depending on the specific elements of the testing situation” (p. 321).

Despite the lack of coherence in experimental empirical results, three axiomatic principles of stereotype threat (Steele, 1997, 2010; Steele & Aronson, 1995; Steele et al., 2002) are generally unchallenged: 1) many important tasks in important settings carry negative stereotypes for particular groups; 2) coping with contingencies of task, negative stereotype and identity results in an emotional and cognitive burden for members of the negatively stereotyped groups that the members of positively stereotyped groups do not need to confront; and 3) this identity contingency results in immediate reduced task performance for the members of negatively stereotyped groups that may result in long-term shifts in their relationship with the performance domain.

STB is a theoretically and empirically convincing situationist explanation for the underperformance of negatively stereotyped groups in specific performance contexts, such as

women in math or black university students more generally. It has made an important contribution to critiquing accounts of underperformance that ‘blame the victim’ in various ways, for example, their biology (eg. Halpern & Tan, 2001; Rushton & Jensen, 2005a), culture (see Suzuki & Aronson, 2005), or motivation (see Hyde & Kling, 2001). As a situationist account it demonstrates that features of the context – entirely out of the control of the target – can undermine performance, and that these situational features are amenable to intervention.

However, the genius of STB is that it shows that the performance of members of any group can be undermined in any context in which their group is negatively stereotyped. For example, the STB literature has demonstrated that even the math performance of white males – a group with generally high status – can be negatively affected if they believe that they are being negatively compared to a group with even higher status in the domain (Aronson et al., 1999).

The STB account of underperformance therefore universalizes and democratizes the experiences that result in underperformance and emphasize that any group could be subject to the effect in any domain, depending on the situational features and social beliefs perceived to be active within the performance context. However, while undermining victim-blaming accounts, STB is also a model that casts targets of negative stereotypes as helpless victims of circumstance. Steele argues not only that victims cannot directly avoid or combat stereotype threat effects, but that attempts to do so result in ‘over-effort’ that further undermines their performance (2010). Therefore, ironically, STB critiques models of minority underperformance that ‘blame the victim,’ but produces a model of subjectivity in which the target of negative stereotype is a helpless victim of the situation.



### **Chapter 3. Problems with conventional stereotype threat theory and the promise of Social Identity Theory for addressing them**

Stereotype threat has now become a mainstream explanation for the underperformance of women and minorities and is discussed in most best-practice manuals related to the education of women and minority students in the US (eg. Aronson et al., 2009a; Halpern et al., 2007; Perna & Thomas, 2006). However, despite its clear usefulness and widespread uptake, the literature has revealed several problems with its conceptualization and empirical support.

#### **3.1. Criticisms that stereotype threat is mainly a laboratory phenomenon**

Early field studies failed to identify stereotype threat effects on performance in real-life performance settings, leading some to suggest that stereotype threat is largely a laboratory phenomenon (eg. Cullen et al., 2004; Cullen, Waters, & Sackett, 2006; Stricker et al., 1999; Stricker & Ward, 2004). It was argued that real-life high-stakes settings produce additional arousal and motivation that allow participants to overcome the effects of stereotype threat. Several studies have now explored stereotype threat effects when performance is related to monetary rewards and confirmed that the stereotype threat effect is less predictable when real rewards are directly related to performance (McFarland, Lev, & Ziegert, 2003; Nguyen, O'Neal, & Ryan, 2003; Ployhart et al., 2003).

However, there is now a substantial body of evidence that has demonstrated stereotype threat effects with non-college samples in realistic environments (Huguet, Brunot, & Monteil, 2000; Huguet & Régner, 2007, 2009; Keller, 2002, 2007; Keller & Dauenheimer, 2003; Kellow & Jones, 2005). Field studies and field experiments have also demonstrated stereotype threat effects on actual school and college assessments, or the positive impact of threat-reducing interventions on actual school performance (Cohen et al., 2006; Essack, 2008;

Good et al., 2003; Good et al., 2008; Keller & Dauenheimer, 2003; Walton & Cohen, 2007) and in work environments (Roberson et al., 2003).

### **3.2. Inconsistent and incongruent stereotype-performance effects**

Although published evidence for stereotype threat effects on performance seems relatively strong, it is also true that many studies have *not* found significant stereotype threat effects on performance. This has led to criticisms that stereotype threat is an illusion resulting from the ‘filing drawer’ phenomenon in which studies finding significant results are more likely to be published (Wax, 2008). Wicherts and de Haan (2009) argued at a recent conference that their “meta-analysis of 55 published and unpublished studies of this effect shows clear signs of publication bias,” that “the effect varies widely across studies, and is generally small” and “does not generalize to non-adapted standardized tests, high-stakes settings, and less academically gifted test-takers.” They concluded that “stereotype threat cannot explain the difference in mean cognitive test performance between African Americans and European Americans” (p. 1). Although this conclusion is an unconventional reading of the literature, they are right – at least – that stereotypes do not *always* have negative effects on performance.

#### **3.2.1. Stereotype lift, boost and susceptibility: positive stereotypes may result in increased performance**

Although it is barely mentioned in ‘official’ versions of the stereotype threat model (Steele, 1997, 2010; Steele & Aronson, 1995; Steele et al., 2002), many studies and meta-analyses have demonstrated that being made aware of positive self-relevant stereotypes often boosts performance (Ambady et al., 2001; Armenta, 2010; Good et al., 2010; Josephs et al., 2003; Levy, 1996; Osborne, 2007; Shih, Ambady, Richeson, Fujita, & Gray, 2002; Shih et al., 1999; Wraga, Duncan, Jacobs, Helt, & Church, 2006; Wraga et al., 2007). This positive

effect of positive stereotypes is generally referred to as “stereotype boost,” but sometimes as “stereotype susceptibility” (eg. Shih et al., 1999).

Even better news for members of non-target groups is that simply knowing that an outgroup is negatively stereotyped in a domain – in other words, having an accessible downward comparison – is often enough to ‘lift’ task performance (Chalabaev et al., 2008; Chatard et al., 2008; Marx & Stapel, 2006c; Walton & Cohen, 2003). For example, Adams, Garcia, Purdie-Vaughns and Steele (2006) found that men who believed that an experimenter was sexist (and prejudiced against women) performed better than men who had no reason to have this belief.

### **3.2.2. ‘Choking under pressure’: positive stereotypes may result in reduced performance**

However, it has also been observed that being made aware of positive group stereotypes can *reduce* performance in much the same way as conventional stereotype threat (Keller & Bless, 2007, 2008; Kray, Locke, & Haselhuhn, 2007; Morrison, Wheeler, & Smeesters, 2007; Smith & Johnson, 2006). Suggestions for the cause of these ironic effects of positive stereotypes include: anxiety or ‘choking under pressure’ (Cheryan & Bodenhausen, 2000; Smith & Johnson, 2006); shifts in regulatory focus (Keller & Bless, 2007, 2008); or reduced self-confidence as a result of exposure to positive category stereotypes (cf. Ellemers & Barreto, 2006; Fryberg, Markus, Oyserman, & Stone, 2008), perhaps as a result of apprehension about matching up to high ingroup standards. Some (eg. Shih et al., 2002) have argued that performance is most likely to increase when positive stereotypes are presented implicitly and to decrease when stereotypes are activated explicitly. Although it does not seem to have been suggested in the literature, it is also possible that when identity is secure there may no real reason to perform well; in other words, positive group stereotypes may increase ‘social loafing’ type effects.

It is surprising that these ironic effects of positive stereotypes have not been incorporated into mainstream models of stereotype threat (eg. Steele, 2010), since ironic performance decrements under stereotype *boost* could easily mask congruent performance decrements under stereotype *threat*. Instead, stereotype threat, stereotype boost and stereotype lift are generally treated as isolated and independent phenomena, despite the fact that they are all likely to be in effect in any given real-life performance context.

### **3.2.3. Negative stereotypes sometimes result in an incongruent *increase* in performance**

While the evidence is strong that negative stereotypes often do result in reduced performance, there have also been many examples in the literature of negative stereotypes resulting in *increased* performance – reversing the predicted stereotype threat effect. As mentioned in the introduction, Nguyen and Ryan’s (2008) meta-analysis “showed zero or *positive* effects” for “one fourth of studies” (p. 1326). From Table 3 in Nguyen and Ryan (2008) it seems that there were positive effect sizes for stereotype threat on performance in 12% of the surveyed studies. The 90% confidence interval for the effect size for stereotype threat on performance in all studies was between  $-.85$  and  $.29$ , which they interpret as meaning that “true moderators” exist (p. 1323), but which also demonstrates that in some circumstances negative stereotypes result in increased performance.

Kray, Thompson, and Galinsky (2001) found that when the stereotype that males have superior bargaining skills was activated implicitly, women underperformed in a bargaining task as predicted by stereotype threat, but that when the stereotype was activated ‘blatantly’ it resulted in women outperforming men. They called this ironic effect *stereotype reactance* and this ironic effect has now been thoroughly replicated in bargaining and leadership tasks (Curhan & Overbeck, 2008; Hoyt & Blascovich, 2007, 2010; Kray et al., 2007; Kray, Reb, Galinsky, & Thompson, 2004). However, these authors have shied away from generalizing to

other contexts and have argued that reactance might be more likely to occur when rebelling against the stereotype is both possible and practical. For example, Kray et al. (2001) were “led to conclude that reactance effects might be limited to certain contexts” that were different to “noninteractive, academic context[s] in which the necessary trait or behavior is not ... easily obtained (e.g., analytic math ability)” (p. 955).

As Kray et al. (2001) predicted, other studies have found reversals of the stereotype threat effect in performance tasks requiring interactive flexibility and creativity, such as empathy (Evans, 2009) and library skills (Fritzsche, DeRouin, & Salas, 2009). However, stereotype threat reversals have now also been observed in memory tasks (Hess, Hinson, & Statham, 2004)<sup>7</sup>, psychology knowledge exams similar to actual course assessments (Williams, 2006), general academic tasks (Deaux et al., 2007; Harrison et al., 2009; McFarland et al., 2003) and even in difficult math tasks (Ambady et al., 2004; Crisp et al., 2009; Forbes, 2009; Grimm et al., 2009; Inzlicht et al., 2006a; O'Brien & Crandall, 2003; Oswald & Harvey, 2000), even in realistic classroom settings (Keller, 2007) and non-experimental field studies (Wei, 2009a).

#### **3.2.4. Recovering all four quadrants of stereotype-performance effects**

There is therefore emerging evidence on the fringes of the STB literature that positive and negative stereotypes can both result in increased or decreased performance in different circumstances, resulting in four quadrants of stereotype-performance effects (illustrated in Figure 1). Two quadrants represent performance shifts in stereotype-congruent directions, namely negative stereotypes resulting in reduced performance and positive stereotypes resulting in increased performance, and these have been relatively well represented as

---

<sup>7</sup> Although Hess, Hinson & Stratham (2004) actually report that older adults were unaffected by negative stereotypes when they were aware they were being stereotyped, a closer look at the data shows that in the ‘aware’ condition the negatively stereotyped group performed better than the positively stereotyped group and comparably to the positively stereotyped group in the ‘unaware’ condition, suggesting a true reversal effect.

stereotype “threat,” “boost” or “lift” in the STB literature. The ironic or incongruent effects of stereotype-performance effects have been less thoroughly covered, although there are enough clues in the literature to demonstrate their importance.

It seems clear that the conflicting results from the stereotype threat literature can only be understood once we acknowledge that stereotype-performance effects can be either congruent or incongruent in different circumstances. However, the key question is under what conditions are incongruent effects likely to occur?

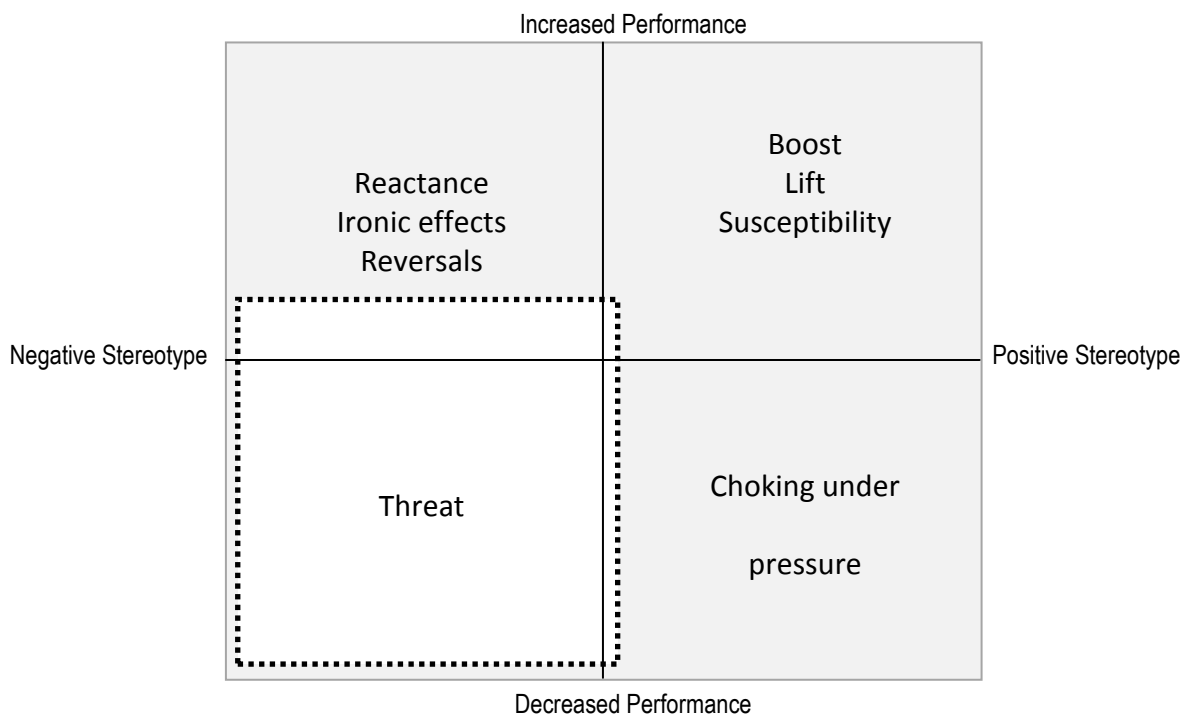


Figure 1: The four quadrants of stereotype-performance effects

#### 3.2.4.1. When threat results in increased intention to do well

Steele and others have argued that even when stereotype threat results in attempts to try harder and to disprove the stereotype, such ‘overtrying’ will generally contribute to reduced performance under stereotype threat (eg. Keller & Bless, 2008; Steele, 2010).

However, despite Steele’s (2010) insistence that overtrying will always reduce performance,

there is also good evidence that increased intention to do well under stereotype threat can translate into increased performance in some circumstances (Crisp et al., 2009; Fritzsche et al., 2009; Jamieson & Harkins, 2007; Jamieson & Harkins, 2009).

#### **3.2.4.2. *When there is an explicit, unambiguous activation of the stereotype***

Many instances of ironic reversals have occurred when stereotypes have been explicitly and unambiguously – rather than subtly – activated (Hess et al., 2004; Hoyt & Blascovich, 2010; Kray et al., 2004; Kray et al., 2001). Although not related to performance on a stereotyped task, Barreto, Ellemers, Scholten and Smith (2010) found that participants were more likely to resist inappropriate categorization or stereotypes when they were explicitly instantiated.

#### **3.2.4.3. *When there is ‘headroom’ for improvement***

Steele and colleagues (2002) applied the boundary condition of task difficulty to stereotype threat after it was noticed that targets of negative stereotypes commonly perform better on easy tasks (eg. Ben-Zeev et al., 2005; Dutrevis & Croizet, 2004; O'Brien & Crandall, 2003). Their argument that negative performance effects under threat are more likely on difficult tasks has the converse that positive performance effects under threat are more likely and possible on easy tasks. Kray, Thompson and Galinsky (Kray et al., 2001) concur, arguing that positive performance effects under threat are only likely when reacting against the stereotype is both possible and practical.

However, other studies have shown that stereotype threat reversals are possible on very difficult tasks (Ambady et al., 2004; Crisp et al., 2009; Deaux et al., 2007; Forbes, 2009; Grimm et al., 2009; Harrison et al., 2009; Inzlicht et al., 2006a; Keller, 2007; McFarland et al., 2003; O'Brien & Crandall, 2003; Oswald & Harvey, 2000; Williams, 2006) and stereotype threat is possible on moderate (Lawrence et al., 2010) and easy tasks (Yeung & von Hippel, 2008).

Although it is clear that Steele and colleagues' (Steele, 1997, 2010; Steele et al., 2002) boundary conditions that stereotype threat is likely to affect 'the vanguard' of top achievers mainly on very difficult tasks is likely to increase the likelihood of observing congruent effects on performance, it is now clear that even under these circumstances negative stereotypes can sometimes increase performance.

#### **3.2.4.4. *When there is 'fit' between regulatory focus and stereotype valence***

Several authors have argued that being confronted with a negative stereotype induces a defensive regulatory focus in which targets are oriented towards avoiding failure. This type of regulatory focus can, for example, increase effort but reduce creativity and mental flexibility. If this is the case then enhanced performance under stereotype threat is most likely when the requirements of the task match the defensive orientation to the task produced by the threat. This reasoning has been supported in several studies (Grimm et al., 2009; Jamieson & Harkins, 2007; Jamieson & Harkins, 2009; Seibt & Förster, 2004; Todd Maddox & Markman, 2010). For example, Jamieson and Harkins (2009) found that in GRE problems stereotype threat improved performance on tasks requiring brute-force mental effort, but decreased performance on tasks requiring more nuanced and creative logic. These results suggest that stereotype-performance effects are at least partly related to participants' strategic orientations to task performance.

#### **3.2.4.5. *When the social meaning of the context supports resistance***

It is now well known that convincing participants that a negatively stereotyped ability is malleable or trainable can reduce threat, but it has also been demonstrated that malleability beliefs also increase the probability of incongruence improvement in performance in response to negative stereotypes, which Kray (2007) calls 'reactance.' Wei's (2009a) working analysis of field data from the US National Assessment of Educational Progress found stronger evidence for ironic reactance than for congruent threat effects, with performance being



sensitive to the exact phrasing of the primes. His (2009b) meta-analysis and laboratory studies provide converging evidence that reactance is likely when primes are “also self-affirming (i.e., allowing students to state an opinion, make an argument against the stereotype, or re-affirm positive values about themselves)” (p. i), in other words, when participants are able to actively contribute to meaning-making in the context.

**3.2.4.6. *When the stereotype target is high in impression management or self-monitoring***

Curhan and Overbeck (Curhan & Overbeck, 2008) found that participants high in impression management responded to negative stereotypes with counter-stereotypic behaviour. Similarly, Inzlicht, Aronson, Good, and McKay (2006a) found that participants high in self-monitoring were more likely to fight back than those who respond passively to the stereotypes imposed on them. These constructs are related in that participants high in either construct are likely to actively process experimental materials against their own experiences and be aware of how their performance positions them in the performance context.

**3.2.4.7. *Identifiability or individuation***

Ambady, Paik, Steele, Owen-Smith and Mitchell (2004) predicted that 'individuated' women would be less affected by stereotype threat. However, in their study participants who were made individually identifiable actually outperformed all other groups.

**3.2.4.8. *Domain identification***

Keller (2007) and Williams (2006) have both found that participants with *low* domain identification performed better under stereotype threat. Keller argues that stereotype threat increases motivation and performance for group members with low domain identification, but produces over-arousal and increased mental load for high domain identifiers, resulting in reduced performance. However, McFarland, Lev, and Ziegert (2003) found that those who

here *highly* identified with the domain performed better under threat. Therefore, although it seems that domain identification might contribute to incongruent stereotype-performance effects, it is far from clear in the literature how, why or when this would occur.

**3.2.4.9. *When the stereotype target has situational power, self-efficacy or group resources for resistance***

Kray Reb, Galinsky and Thompson (2004) found that stereotype reactance only occurred when the targets of the negative stereotype were “not ... crippled by tangible power deficiencies” (p. 409). In other words, “blatantly telling women that they lack the attributes necessary to prevail at the distributive bargaining table produces stereotype threat in the absence of power and stereotype reactance when sufficient power is possessed” (p. 409). Similarly, Hoyt and Blascovich (2007) found that positive reactance responses were more likely for people with high self-efficacy. Crisp, Baiche & Maitner (2009) specifically chose a group of high-achieving women engineering students with the hunch that over the years they would have developed individual and cultural resources for performing in that domain and, as predicted, found powerful reactance effects in that environment. Forbes (2009) found that stereotype threat increased math motivation for women trained to have positive implicit math attitudes, increased working memory and increased math performance. Therefore, there is strong emerging evidence that access to situational power, personal identity resources (such as attitudes) and group identity resources (such as alternative categorizations or stereotypes) are likely to be important enabling factors for reactance.

**3.2.4.10. *Multiple identities, ambiguous categorization or permeability***

Deaux, Bikmen, Gilkes, Ventuneac, Joseph, Payne, & Steele (2007) found that first-generation black Caribbean immigrants to the US responded to perceived test diagnosticity (which is generally considered to invoke negative stereotypes about blackness) by performing better, while African American and second-generation black Caribbean

immigrants performed worse. Deaux et al. argued that this reversal was due to the first generation immigrants having access to a positive subtyped identity (black Caribbean) that buffered them from the negative stereotypes attached to the superordinate black identity.

Crisp, Baiche and Maitner (2009) compared the math performance of female psychology and engineering students under stereotype threat, and found that the performance of engineering students ironically improved under threat. They suggested that “female engineers may have recategorized or subgrouped themselves as a way to remain engaged with the task but to avoid the threatening comparison” (p. 11). Female psychologists, on the other hand, were double-minorities in that context, as both of their relevant identities (as psychology students and women) were negatively stereotyped in the domain and they succumbed to negative stereotypes as predicted by stereotype threat.

#### ***3.2.4.11. Hostile environment or conflict***

Deeply ironically, given the focus in modern life on gender and race fairness, a hostile environment can promote stereotype reactance (Oswald & Harvey, 2000). Oswald and Harvey’s study, attempts to smooth over prejudice with a “stereotype removal message” (p. 353) undermined resistance and reactance and resulted in congruent stereotype threat effects on performance, but participants exposed to the hostile gender message without apology reacted against the stereotype and performed better under threat.

### **3.3. Steele’s treatment of incongruent stereotype-performance effects**

Despite early signs of stereotype boost and lift in the literature (eg. Levy, 1996; Shih et al., 2002; Shih et al., 1999) and indications that negative stereotypes can sometimes increase performance (Kray et al., 2001; Oswald & Harvey, 2000) these effects were not discussed in Steele, Spencer and Aronson’s (2002) review. Indeed, of these five studies that all clearly show either boost/lift or stereotype threat reversals or both, only Shih, Pittinsky and Ambady (1999) were cited, and even then the positive effects of stereotypes on

performance were not discussed. Although Steele and his colleagues mentioned that targets of stereotype threat may engage in “counterstereotypic behavior” to disprove the stereotype, they seem to have been referring to behavior unrelated to the key performance task oriented to establishing credentials and distancing targets of negative stereotypes from the negative group – like a black man whistling Vivaldi to appear less threatening. They do not acknowledge or take seriously the notion that negative stereotypes may – at least sometimes – invoke improved performance.

Now, even after several years of mounting evidence that stereotype threat reversals are real – including two studies for which he was listed as a co-author (Adams et al., 2006; Deaux et al., 2007) – Steele (2010) was almost entirely silent on the topic in his latest book on stereotype threat. Given Steele’s massive influence on the field, this is likely to have a powerful influence on stereotype threat research and intervention in the next few years.

The one reversal he does mention is that observed in O'Brien and Crandall’s (2003) study of women’s math performance in which they noticed that women performed better under stereotype threat on easy math problems and worse on difficult ones. Instead of acknowledging that threat sometimes improves performance, Steele (2010) argued instead that task difficulty is a pre-requisite for genuine stereotype threat effects. In other words, he produced a boundary condition that splits the positive and negative outcomes of negative stereotypes; essentially treating positive and negative performance shifts under negative stereotypes as two separate phenomena.

Other authors, by naming positive effects of positive stereotypes ‘stereotype boost’ (Shih et al., 2002) and ‘stereotype lift’ (Walton & Cohen, 2003); and positive effects of negative stereotypes ‘stereotype reactance,’ (eg. Kray et al., 2001) or ‘stereotype susceptibility’ (eg. Ambady et al., 2001) have contributed to producing an artificially segmented and unnecessarily fragmented literature. Although some of these terms have the

advantage of linking in to broader literatures, they nevertheless promote the conceptualization of positive and negative reactions to negative stereotypes as qualitatively distinct phenomena resulting from completely separate processes although there is little evidence to support this assumption (see Marx, 2009 for a partial discussion).

### **3.4. What needs to change to understand the full range of stereotype-performance effects?**

The current trajectory of stereotype threat research is very constricted: focusing only on the politically interesting quadrant of stereotype threat has required an increasingly long list of boundary conditions and has produced the unwarranted impression that negative stereotypes inevitably and deterministically undermine task performance. A careful reading of the literature has shown that this is simply not true.

#### **3.4.1. Recovering the active subject**

The situationist stance of stereotype threat theory has produced a passive model of the subject in which targets of stereotype threat are believed to respond fatalistically and unstrategically to instructions, stereotypes and tasks. However, it seems clear that – at least in some cases – participants choose to actively reject the stereotypes applied to them in the experimental context, either by failing to live up to a positive stereotype or working hard to overthrow a negative stereotype. This observation forces us to pay more attention to understanding how experimental and non-experimental stereotype threat situations are *experienced* by participants, including consideration of their extra-experimental identities, their 'reading' of the experimental setting, their motivations for participation, their understandings of the politics of success or failure in that specific context, and the negatively stereotyped experimental task itself.

### **3.4.2. Understanding the social conditions under which submission or resistance to stereotypes is likely**

We already know that negative stereotypes sometimes improve and sometimes hinder performance, and we know the same applies to positive stereotypes. However, this knowledge is virtually useless, because the focus in the field on a single quadrant of effects has left us with little understanding of the social conditions under which submission or resistance to stereotypes becomes more likely.

The reversals observed in the literature so far suggest three broad classes of features that we should be concerned with: (1) individual identity resources and habits-of-being, such as self-efficacy (Hoyt, 2005), history of resistance against a particular stereotype (Crisp et al., 2009); self-monitoring (Inzlicht et al., 2006a), impression management (Curhan & Overbeck, 2008), or domain identification (Keller & Bless, 2007; Williams, 2006); (2) features of the specific task-performance environment, such as the extent to which the performance setting is constructed as hostile (Oswald & Harvey, 2000), the extent to which stereotypes and discrimination are blatant or implicit (Hess et al., 2004; Hoyt & Blascovich, 2010; Kray et al., 2004), the extent to which the setting provides ‘headroom’ for increases in performance (Kray et al., 2001; O'Brien & Crandall, 2003), or the extent to which participants are individually identifiable (Ambady et al., 2004); and (3) the socio-structural features of the intergroup context that allow strategies such as recategorization to avoid negative stereotypes altogether (Deaux et al., 2007), or that modify targets’ relationships with categories and stereotypes, such when decreased group boundary permeability increases ingroup identification (Ellemers, 1993; Ellemers, van Knippenberg, de Vries, & Wilke, 1988) or increased intergroup conflict increases ingroup identification and competitiveness (Haslam & Turner, 1992; Tajfel & Turner, 1979a).

### **3.4.3. Understanding the means by which resistance can take place**

There are indications that taking an active view of subjectivity in the stereotype-threat context may explain some of the anomalies that have previously been reported in the literature. For example, if we accept that stereotype targets are active subjects and that task performance is one of the many means by which identity can be produced in a context, we also have to accept that there are many ways in which participants can succumb to or resist categorizations and stereotypes in that context. Thus ‘disidentification from the domain’ (Steele, 1997) can be seen as a means of resistance, because it devalues an axis of comparison which is prejudicial to one’s social group. As such, it can be seen as a strategy of social creativity (Tajfel & Turner, 1979a), even though it results in undesirable performance outcomes from the perspective of educators.

Alternatively, when it seems that a group has a legitimate claim to the terrain in which they are negatively stereotyped, it becomes more likely and possible that individual group members will be able to perform better to stake their claim on the domain. For example, when negatively stereotyped minorities are told that intelligence is malleable (undermining the legitimacy of their exclusion) they are able to perform better (Good et al., 2003; Rydell, Hugenberg, Ray, & Mackie, 2007).

### **3.5. The need for a meta-theoretical identity framework for understanding STB effects**

The review of STB literature has already demonstrated that identity has been an important consideration in research to date. For example, at the very heart of STB theory is the role of negative task-related group status (ie. stereotypes). Several key features of identity have been explored as potential mediators or moderators of the effect, such as domain or ingroup identification (Armenta, 2010; Clark et al., 2010; Schmader, 2002a; Steele, 1997), stereotype endorsement (Chatard et al., 2008; Marx & Goff, 2005; Schmader et al., 2004;

Steele, 1997, 2010; Steele & Aronson, 1995; Steele et al., 2002), stigma consciousness (Brown & Lee, 2005; Brown & Pinel, 2003; Hess et al., 2009a; Mendoza-Denton et al., 2009) and multiple identities (Rydell et al., 2009; Shih et al., 2007; Shih et al., 1999).

However, a key problem with these investigations is that STB lacks a coherent meta-theory of identity and intergroup relations and so the way that identity has been represented has generally been fragmented and atheoretical. Additionally, with no metatheory to conceptualize group-level processes, STB research has become progressively more individualistic, focusing increasingly on internal cognitive mechanisms at the expense of the social and contextual features that were central to the development of the model. Social Identity Theory may provide a useful framework for conceptualizing identity in STB situations.

### **3.6. Social identity theory**

Social Identity Theory (SIT) was conceived as a framework for understanding human behavior in settings where status and power are unevenly distributed. Tajfel and Turner were critically interested in understanding the different social conditions that would sustain an inequitable *status quo* or would allow individuals and groups to escape social or material deprivation (Tajfel, 1981a; Tajfel & Turner, 1979b).

Although STB is primarily interested in task-performance, that task performance is by definition located in inequitable status hierarchies (as represented by positive and negative stereotypes). The types of task that are generally of interest to STB theorists are gatekeepers to the advancement of individuals in modern meritocracies, such as SAT's or university exams. These activities are crucial for individual and group advancement or, more cynically, the maintenance of the inequitable *status quo*. Therefore, the focus of interest in STB research is nested in the interests of social identity theory and SIT provides a useful meta-theoretical framework for understanding identity in STB contexts.



SIT is based on five central axioms: (1) Social life is structured by *groups* – mental representations of belonging that are defined both by individual cognition and social consensus. (2) Human interaction takes place on a continuum where it based mainly on personal characteristics at the one side, and group characteristics at the other. Behaviour at either extreme is rare, and most social interaction occurs at a point on the continuum where personal and group characteristics are both contextually relevant. (3) Individuals derive social identity from the group memberships relevant in any particular context. The social value of a group membership is derived from its relative position compared to other relevant groups. People are motivated to achieve positive self and social identity, and are therefore motivated to maximise positive group memberships. (4) This may be done either by leaving negative groups and joining positive ones (a strategy of individual social mobility), or by working to change either the perceptions of the group or the material status of the group (a strategy of social change). (5) The choice between enhancing social identity through social mobility or social change depends on three factors relating to the relevant group structures: their permeability, stability, and legitimacy. The presence of all of these factors is likely to encourage individual responses (i.e. social mobility) and their absence is likely to promote group action (i.e. social change) (Tajfel, 1978b, 1978c, 1981a, 1982; Tajfel & Turner, 1979b).

### **3.6.1. Groups**

The word ‘group’ here is used to refer to a mental representation of collectivity. Although it could also apply to a gathering of people who are physically proximal, this is not necessary for a group to exist. It is more common that it refers to a number of people who are physically scattered but are categorisable as group members by virtue of characteristics relevant to some aspect of a consensual representation. Tajfel (1978c) offers the very simple definition that a group is a collection of people who consider themselves to be a group, or are

forced to consider themselves a group by others. Therefore the existence of ‘a group’ in society relies on social consensus (from within or without the group):

"The consensus about 'who is who' will be in many ways shared by the group socially categorized in certain ways and the surrounding groups by which and from which it is perceived as distinct. A stronger statement can be made: the consensus may often originate from other groups and determine in turn the creation of various kinds of internal membership criteria within the ingroup."

(Tajfel, 1978c, p. 31)

Although any individual may have many group memberships, some will be more important for structuring social life than others. Each group membership may be relevant in different social contexts and may apply different constraints to social interaction.

For an individual group member, group membership has at least one of the following three components: cognitive, evaluative or emotional. The cognitive component is simply “the knowledge that one belongs to a group” (Tajfel, 1978c, p. 28) and is related to processes of categorization discussed below. The evaluative component is the “sense that the notion of the group and/or one’s membership of it may have a positive or a negative connotation” (ibid, p. 29) and is discussed in the section on group membership and social identity, below. The emotional component refers to the affective dimensions of the combination of cognitive and evaluative components in the sense that one’s group memberships “may be accompanied by emotions (such as love or hatred, like or dislike) directed toward one's own group and towards others which stand in certain relations to it” (ibid, p. 29).

### **3.6.2. Social Categorization**

The most basic statement that can be made about categorization is that it is a process of grouping by which the differences on relevant dimensions within a group are minimized and differences between groups are maximized (Oakes et al., 1994). However, it is not simply

a matter of arranging fixed elements in a perceptual field because the allocation of items to categories changes the way that the elements are perceived, as demonstrated in Tajfel and Wilkes's (1963) studies of line-length perception. "Social categorization" therefore also refers to the effects that group-based perception has on an individual's reactions to their own and to other groups. Consequently social categorization refers both to the cognitive processes by which people locate themselves and others in social environments and to processes by which social behaviour is produced:

"Social categorizations are conceived here as cognitive tools that segment, classify and order the social environment, and thus enable the individual to undertake many forms of social action. But they do not merely systematize the social world; they also provide a system of orientation for *self*-reference; they create and define the individual's place in society" (Tajfel & Turner, 1979a, p 40).

As a result of these active and socially locating features, "social categorization almost invariably involves value judgements" (Tajfel, 1978d) and biases for or against individuals on the basis of their perceived group memberships. This aspect will be discussed further below.

### **3.6.3. Relativity**

Firstly, SIT argues that all groups are relative to other groups: neither the boundaries of a group nor the meanings attached to group membership can make any sense unless there is some other group to compare it to and meaningful dimensions for comparison. At the very least the relative group definition will compare members in the group to those left out of it. "A group becomes a group in the sense of being perceived as having common characteristics or a common fate mainly because other groups are present in the environment" (Tajfel, 1978d, pp. 66-67). More importantly, the *meaning* of group membership is relative to the

other groups in the social environment. In any particular comparison “the characteristics of one's group as a whole (such as its status, its richness or poverty, its skin colour or its ability to reach its aims) achieve most of their significance in relation to perceived differences from other groups and the value connotation of these differences” (Tajfel, 1978d, p. 66).

#### **3.6.4. Relative deprivation and status differences**

Since categorization is inherently comparative it usually results in a state of relative deprivation for some groups and relative ascendancy for others and social processes of categorization are often directed towards this end (Tajfel, 1981b). Large relative differences between groups in status relevant dimensions are likely to reify the boundaries between groups, to increase the perceived dissimilarity between them, and to mask the relative nature of group definitions.

#### **3.6.5. Ingroup and outgroup comparisons**

However, inequitable group relationships are not only the *basis* for categorization but also its *effects*, because of two simple and apparently universal effects: ingroup bias and intragroup homogeneity.

##### **3.6.5.1. Ingroup bias**

Ingroup bias is “the tendency to favour the ingroup over the outgroup in evaluations and behaviour” and has been shown to be a “remarkably omnipresent” feature of social categorization and intergroup relations (Tajfel & Turner, 1979a, p. 38). Ingroup bias has been shown to operate almost unvaryingly, even when categorization is based on apparently meaningless dimensions, such as in minimal group studies (Tajfel & Turner, 1979a). Of course, over time, ingroup bias is likely to entrench perceptions of inter-group differences.

##### **3.6.5.2. Intragroup homogeneity**

The categorization and stereotyping process, by definition, requires that members of a stereotyped group are seen as similar in some way (Brown, 2000). Indeed, according to the

metacontrast principle, groups will only be perceived as groups when the differences between them are perceived as greater than the differences within them in a specific comparative context (Oakes et al., 1994). In order for this perception to be achieved, similarities between group members must be perceptually amplified and differences minimized (Tajfel, 1981b). Thus intragroup homogeneity refers to the increased likelihood that members of an ingroup or outgroup will be perceived as homogenous members of a class rather than as differentiated individuals and is both an outcome and an objective of categorization and stereotyping.

The more important particular group memberships are to individuals or in particular social contexts, the more likely it is that productions and contestations of homogeneity will become important in the intergroup setting (Brown, 2000; Tajfel, 1981b). For example, members of low status groups may work to differentiate themselves from their ingroup, to demonstrate that they are *not* prototypical of their low-status group (Doosje, Ellemers, & Spears, 1995; Pettit & Lount, 2010) while high status group members may work to demonstrate that low status group members are, indeed, all the same (Billig, 2001b; Eiser, 1996) in order to defend the favourable status hierarchy.

### **3.6.6. Stereotypes and their functions**

Categorization naturally depends on generalizations, and it is generally assumed that members of groups share characteristics relevant to the group definition. These generalizations about groups can be called 'stereotypes'. While Tajfel (1981a) argues that "they derive in large measure from, or are an instance of, the general [individual] cognitive process of categorizing" (p. 145) and that they may be held by individuals, they become stereotypes (rather than attitudes or opinions) only when shared by large numbers of people. Therefore groups and group processes are intrinsic to the social identity approach to stereotypes and stereotyping. As a result, individualistic cognitive programmes of research are limited in the extent to which they can understand stereotypes because they ignore their

group basis. Full understanding of stereotypes, argues Tajfel, will require an understanding of their group nature *and* of individual socio-cognitive processing (Tajfel, 1981a).

Tajfel argued that the most important aspect of stereotypes was not their role as individual representations, but the social functions that stereotypical representations perform in group terms. While he readily acknowledged the importance of understanding the functions of stereotyping in individual cognitive processing, he also argued that stereotypes have important functions in terms of ideological justification, developing positive group differentiation and linking individual cognition with social groups. Each of these will be discussed in turn.

#### **3.6.6.1. *Cognitive functions of stereotypes***

Since stereotypes are cognitive representations of group features, the types of information that stereotypes represent are naturally related to processes of categorization and are implicated in inter-group relations. Consequently, Tajfel argued, stereotypes fulfil important cognitive functions that cannot be addressed from an individualistic cognitive perspective; as well as serving as internal summaries of complex social phenomena they may serve as *explanations* for complex (large scale) social events and as *reasons* for action (Tajfel, 1981a). In other words stereotypes serve not only to *describe* complex stimuli, but also to *make meaning* of them.

#### **3.6.6.2. *Linking individual cognition to group-level processes***

Since stereotypes are, by definition, shared by large numbers of people, they are mental representations that are continuously consensually validated. There is no reason to believe that stereotypes should *feel* like stereotypes to the people that hold them, although some clearly do. On the contrary, since they are consensual mental representations of other groups they most often represent group relations in individual cognition in ways that feel like

natural facts (Tajfel, 1981a) and therefore naturalize patterns of categorization and group relations.

### ***3.6.6.3. Stereotypes, values and ideological justification***

As a result of their explanatory and naturalising properties, Tajfel (1981a) argues that stereotypes are important in determining and justifying social values. Because they are mental representation of groups (both one's own and others) that have the appearance of natural fact they validate both ideology and action. Tajfel is not simply arguing that self-aware individuals may reflexively use known stereotypes to justify ideology and actions that they know to be wrong, but that the consensual nature of stereotypes generally renders them invisible to their holders and therefore makes ideologies and actions seem right, natural and inevitable.

The implication is that stereotypes have a powerful role in reifying and preserving value systems and ideologies. To the extent that they consensually naturalize group relations they preserve and defend the status quo, and the more they can be constructed as stable, constant and clear-cut, the more effective they are in ordering individuals' social environments (Tajfel, 1981a).

However, they can also be important factors in social change. If stereotypes justify ideology, then challenging stereotypes may also destabilize ideology and entrenched patterns of intergroup relations. If "... a social categorization into groups is endowed with a strong value differential ... [then] encounters with negative or disconfirming instances would not just require a change in the interpretation of the attribute assumed to be characteristic of a social category. Much more importantly, the acceptance of such disconfirming instances threatens or endangers the value system on which is based the differentiation between the groups" (Tajfel, 1981a, p. 152).

#### **3.6.6.4. *Stereotypes and positive group differentiation***

Stereotypes are, by definition, representations that encode group differences as a result of categorization processes. Tajfel (1981a) argues that they are most likely to originate and be widely diffused either when they create a differentiation between groups that has strong positive benefits for one group, or when group identity is threatened or eroded.

#### **3.6.6.5. *Stereotype maintenance and change***

Tajfel argues that social consensus acts as positive feedback for stereotypes: the more consensual they are, the more natural and inevitable they appear, and the more resistant they will be to modification. Because of this pre-existing consensus, “less information than is the case for physical categories will be needed to confirm these judgements, and considerably more to disconfirm them in the face of what appears to fit in with what is generally accepted as a social ‘reality’” (Tajfel, 1981a, p. 146). Additionally, stereotypes guide action according to group categorizations and thereby result (materially) in the very differences that they posit. In other words, there is a feedback process between representation, action, social reality and perception.

#### **3.6.7. Group membership and social identity**

Social identity refers to “that *part* of an individual's self-concept which derives from his [sic] knowledge of his [sic] membership of a social group (or groups) together with the value and emotional significance attached to that membership” (Tajfel, 1978d, p. 63, emphasis original).

“... the assumption is made that, however rich and complex may be an individual's view of himself or herself in relation to the surrounding world, social and physical, some aspects of that view are contributed by the membership of certain groups or categories. Some of these memberships are more salient than



others; and some may vary in salience in time and as a function of a variety of social situations" (Tajfel, 1978d, p. 63).

Since the identity value of any particular group membership is derived from the processes of social categorization, discussed above, it is a thoroughly relative and comparative process. However, the process is not unbounded. Tajfel and Turner (1979a) argue that there are three important qualifications. First, it is not enough to be assigned to a group by others – group membership must be internalized. Second, the social situation must allow relevant intergroup comparisons. Third, the comparison group must be relevant to the dimension(s) of comparison.

Tajfel argues that people are not passive recipients of social identity, but that “at least in our kinds of societies, an individual strives to achieve a satisfactory concept or image of himself [sic]" (Tajfel, 1978d). One of the ways in which a positive self-concept can be attained and preserved is to manipulate various aspects of social identity to maximize its contribution to a positive identity. It is this need to positively differentiate in order to maximize self-concept that drives the creation and maintenance of social groups (Tajfel & Turner, 1979a).

Pulling this all together, a general model of social identity can be outlined as follows:

- (1) People are motivated to maximize their self-concept (or self-esteem).
- (2) They have different category memberships as resources that are each associated with different cultural capital that could either enhance or threaten their positive esteem in different contexts.
- (3) The value of an individual’s own group memberships are determined through social comparisons with other groups, where “positively discrepant comparisons between ingroup and outgroup produce high prestige; negatively discrepant comparisons between in-group and out-group result in low prestige” (Tajfel & Turner, 1979a, p. 40).

People will maximize positive identity by setting up favourable comparisons between their ingroup and relevant outgroups. This can involve strategically choosing favourable comparison groups or engaging in social creativity and setting up dimensions for comparison in which the ingroup is superior. If the social identity arising from this process of strategic comparison is unsatisfactory then individuals with the ability to do so are likely to either i) opt for individual social mobility and join other more favourable groups if possible, or ii), work towards social change by trying improve the status of the group(s) to which they already belong (Tajfel & Turner, 1979a).

The key feature of this framework for understanding individual behavior in intergroup settings is that "status is not considered here as a scarce resource or commodity, such as power or wealth; it is the outcome of intergroup comparison" (Tajfel & Turner, 1979a, p. 43). Therefore group and individual status is actively produced, defended and contested by ordinary people in ordinary settings and categorization and stereotyping are key resources for doing so.

### **3.6.8. The interpersonal-intergroup continuum and self-categorization**

Self categorization occurs when individuals perceive themselves to belong to a certain social category. It results in a qualitative transformation of identity and experience in a given context, "transforming self-conception and assimilating all aspects of ones [sic] attitudes, feelings, and behaviors to the ingroup prototype; it changes what people think, feel, and do ... and renders perceptions, attitudes, feelings, and behaviors stereotypical and group normative" (Hogg, 2001, p. 187).

SIT argues that all human interaction occurs at some point on a continuum between interpersonal and intergroup (Tajfel, 1978c; Tajfel & Turner, 1979a). In other words, in some situations people interact purely as individuals and their various group memberships have no relevance to, or impact on, their interaction. In other cases people act purely on the basis of

their group memberships with no reference to individual characteristics, for example, when soldiers engage in combat. In practice either extreme is rare (Tajfel & Turner, 1979a), but concerted group-based efforts towards large-scale social change are unlikely when behaviour is located at the interpersonal extreme of the continuum.

There are several factors that may contribute to locating behaviour on the continuum: First, some group memberships may be pervasive elements of an individual's social identity in a given intergroup context (Tajfel, 1978c). Second, there are situations (such as war) that are so powerful that interaction is forced to the intergroup extreme no matter how weak the individual's subjective sense of group membership. These conditions are related, because:

...social situations which will force the individuals involved to act in terms of their group membership will also enhance for them some group identifications which had previously not been very significant to them, or perhaps even create or bring to life group memberships which were previously only dormant or potential. As a consequence ... many of these people will be likely to perceive in common an increasingly greater number and variety of social situations as relevant in some ways to their group membership (Tajfel, 1978c, p. 39).

One such condition is intergroup conflict: the more intense the intergroup conflict, the more likely that interaction will be on the intergroup dimension (Tajfel & Turner, 1979a).

Behaviour at the intergroup extreme requires shared ingroup affiliations and interpretations of intergroup relations (Tajfel, 1978c) and will be generally independent of (1) individual characteristics of ordinary members of the ingroup or outgroup, (2) personal relationships that exist between ingroup and outgroup members in other contexts, and (3) the motivational states of individuals before, during or after the encounter(s). In addition, the nearer behaviour is to the intergroup extreme, more the group members will experience outgroup homogeneity (that is, they will perceive members of the outgroup as

undifferentiated instances of a class) and the more uniform will be behaviour towards the outgroup (Tajfel, 1978c).

### **3.6.9. The conditions for social change**

The intergroup processes discussed up to this point have been understood as thoroughly relative and plastic and have been discussed as if individuals have a great deal of flexibility in terms of choosing both their group memberships and the contexts in which they are compared to others. The world, however, is not such an idyllic place, and social identity theory is intended to address social conditions of inequity and injustice; conditions in which individuals often have little choice about the groups they belong to, the social value of their group memberships, or the groups to which they are compared. In these conditions, where a particular group membership is tightly bound to one's place in society, the terms 'inferior' and 'superior' are used as "rough (and by no means optimal) shorthand terms referring to the psychological correlates of a number of interacting dimensions of social differentiations, such as discrepancies between groups in social status, in power, in domination etc." (Tajfel, 1978a).

Social identity theory sketches out circumstances of intergroup relations under which social change is most likely or unlikely. These circumstances are determined by three dimensions related to qualities of the consensual representations of the relevant groups in society, namely: legitimacy, stability and permeability.

#### **3.6.9.1. Legitimacy**

Legitimacy refers simply to the extent to which existing intergroup relations seem just or right (Tajfel & Turner, 1979a). Intergroup relations that are perceived as entirely legitimate are often experienced as entirely natural and, often, inevitable. An example of a completely legitimate set of intergroup relations in contemporary society is the adult/child divide which, although a fairly modern invention (Aries, 1962), now has the illusion of

complete inevitability and naturalness. Complete legitimacy therefore results in an intergroup relationship that is blended seamlessly into everyday life. In conditions of legitimacy people may be aware that they are relating in a particular way to a member of a particular group but will have no reason to question either the form of the relationship or its results. Further, when intergroup relations are consensually experienced as legitimate they are continuously reinforced by their repetition. It is only when the legitimacy of a particular intergroup relationship is threatened that it stands out as something to talk about, to justify or to challenge.

Therefore if the legitimacy of a particular intergroup relationship is to be challenged it must first become evident that there is some basis for comparison (i.e. some similarity) between the two groups. In other words, the group definitions must be challenged to the extent that it becomes reasonable for members of the groups involved to begin to perceive the differences between them as illegitimate or unfair. This is a very important element of SIT because it identifies a vital starting point for social change; if inequitable intergroup relationships are to be changed they must first be perceived as illegitimate, and if they are to be perceived as illegitimate there must first be some basis for comparison between the two groups (Tajfel, 1978d).

### **3.6.9.2. *Stability***

Stability refers to the extent to which the intergroup status hierarchy is experienced as changeable or permanent. In some cases intergroup relations appear so permanent that it is impossible to imagine the social world being configured in any other way. For example, prior to the twentieth century it was unimaginable that women should be allowed to vote. Prior to the late eighteenth century and the French revolution it was unimaginable that ordinary men should be allowed to vote in the affairs of state either.

Before social change can occur, it must first be possible to imagine social conditions with different configurations of intergroup relations; in other words, there must be 'cognitive alternatives' (Tajfel & Turner, 1979a). When there are no cognitive alternatives, and it seems as if current intergroup relations are unassailable or immutable, the intergroup relations are described as 'secure' (Tajfel, 1978a; Tajfel & Turner, 1979a). In secure conditions there is generally complete and unwavering consensus about the legitimacy and stability of intergroup relations (Tajfel, 1978a).

However, it is unlikely that relations characterized by extreme inequity between groups can also be completely and effortlessly secure. Tajfel argued that it is unlikely that there are many situations with completely static, unchanging social relations, and even when they appear to be so, it is likely that this appearance is the result of constant and careful maintenance:

"A completely secure social identity for a group consensually 'superior' is nearly an empirical impossibility. The kind of psychological distinctiveness that would insure [sic] their unchallenged superiority must not only be gained; it must be preserved, and it can only be preserved if social conditions of distinctiveness are carefully perpetuated, together with the signs and symbols of distinctive status without which the attitudes of complete consensus about 'superior' distinctiveness are in the danger of disintegrating. In this sense, therefore, even in the most rigid caste system (be it racial or any other), the social distinctions which may appear very stable are related to a continuously dynamic psychological situation in which a superior group can never stop working at the preservation of its distinctiveness" (Tajfel, 1978a, p. 87-88).

On the other hand, in conditions where intergroup relations are insecure many cognitive alternatives exist, and it becomes easy to imagine a state of affairs where superior

groups fall from grace, inferior gain status, or where specific group memberships are no longer relevant. Tajfel & Turner (1979a) give the example that it is easy to see the current fortunes of a football team changing in the distant future, or even within a particular game. An environment where large-scale social change is likely is one where intergroup relations can be experienced as insecure, and where cognitive alternatives to the status quo can be easily imagined.

### **3.6.9.3. *Permeability***

Even where intergroup relations are experienced as illegitimate and secure, social change is unlikely if individual members of the subjugated group feel that it is possible to move from the inferior to the superior group and thereby attain substantial benefits in status or material conditions associated with membership of the superior group. The extent to which it is possible for an outgroup member to exit the inferior group or enter the superior group through individual effort is referred to as ‘permeability’. In conditions where intergroup structures are inequitable the degree of permeability determines the extent to which individual group members can change their own fortunes by shifting their individual group allegiance. This has a powerful influence in locating behaviour on the interpersonal-intergroup continuum because “if adequate conditions for the preservation of positive social identity are not offered by a group the individual will leave it – psychologically, objectively or both” (Tajfel, 1978d, p. 67). For example, in a racist society a black person may nevertheless increase their personal status by gaining a university degree and becoming a professional. Tajfel and Turner (1979a) refer to this individual strategy as ‘social mobility.’

In other conditions where group boundaries are impermeable, individual social mobility is impossible and the only means for members of the subjugated group to improve their social identity is to effect large-scale social change by redefining the symbolic and material relationship of superiority and inferiority between the two groups:

“In situations which are characterized by the structure of belief of 'social change' [i.e limited permeability], the issues become more complicated. In some conditions ... a social group can fulfil its functions of protecting the social identity of its members only if it manages to keep its positively valued distinctiveness from other groups. In other conditions this distinctiveness must be created, acquired and perhaps even fought for through various forms of relevant social action. Yet, in other conditions, some or most individuals will place their bets, implicitly or explicitly, on certain processes of 'objective' social change which, they hope, will lead finally to a structure of genuine social mobility; this might imply the distant goal of the dissolution of a group which at present is mainly defined through its negative attributes when it is compared to other groups” (Tajfel, 1978d, p. 67).

The permeability of group structures impacts on the variability or uniformity of ingroup attitudes and behaviours in that the less permeable the social structure, and the more relations are characterised as intergroup, the more likely it is that members of the outgroup will be treated as homogenous instances of a class. In other words, decreased permeability is likely to result in increased perceived outgroup homogeneity (Tajfel & Turner, 1979a).

#### **3.6.9.4. Interactions**

The dimensions of legitimacy, stability and permeability exist in a closely-coupled relationship such that changes to one will often alter the others. For example, “perceived illegitimacy is likely to determine, sooner or later, attempts to change the situation; and perceived instability (which can be translated as the development in a group of the awareness of cognitive alternatives to the existing situation) is likely to be associated, sooner or later, with the decrease in that group of the perceived legitimacy of the situation” (Tajfel, 1978c, p. 52).



### 3.6.10. The social identity model of change

If the membership of a particular social group makes a positive contribution to aspects of social identity then people will tend to maintain their membership if already a member, or attempt to join it if not. If the group makes a negative contribution to social identity then individuals will try to leave it by disidentification, *unless* leaving it is impossible for objective reasons, or leaving “conflicts with important values which are themselves a part of ... acceptable self image.” If changing group membership is an impossible or otherwise unavailable option, then it is possible to address the negative intergroup relationship by engaging in social creativity. This could involve either (a) redefining the constraints of the comparison, for example, by comparison with a different group or by finding new (favourable) dimensions for comparison, or (b) redefining the consensual group definitions such that “unwelcome features ... are either justified or made acceptable through a reinterpretation” (Tajfel, 1978d, p. 64). Alternatively the negative intergroup comparison can be addressed by social action that addresses the material and ideological inequalities between groups. Of course, reinterpreting the meaning of group membership and material social action are often related in that each may be implicated in the other (Tajfel, 1978d, p. 64).

To put it another way, "providing that individual mobility is unavailable or undesirable, consensual inferiority will be rejected most rapidly when the situation is perceived as both unstable and illegitimate" (Tajfel & Turner, 1979a, p. 45). At the same time, “when the dominant group or sections of it perceive their superiority as legitimate they will probably react in an intensely discriminatory fashion to any attempt by the subordinate group to change the intergroup situation” (Tajfel & Turner, 1979a, p. 46).

In short, group-based strategies of influence and change are most likely to be pursued in conditions where legitimacy, stability and permeability are all low or absent. Conversely,

when legitimacy, stability and permeability are all high then individual responses are most likely.

### **3.7. Responding to social identity threat**

Although it has not yet been acknowledged in the stereotype threat literature, there is some emerging evidence that targets of identity threat do experience and orient their task performance in active and strategic ways as predicted by social identity and self-categorization theoretical frameworks. For example, in a recent study participants in a high status group had higher blood pressure when performing in a context where the intergroup status hierarchy was unstable, presumably because of the threat to their superior group identity (Scheepers, Ellemers, & Sintemaartensdijk, 2009). Similarly, Petite and Lount (2010) have found that members of high status groups under identity threat perform particularly well when they know their performance will be evaluated against that of a low-status outgroup members in order, they argue, to “secure [their] social identity and self-categorization” (p. 18). Others have found that the task performance group members with low ingroup identification (primarily concerned with personal identity) benefitted from *self*-affirmation under social identity threat while high identifiers benefitted from *group*-affirmation (Derks, Scheepers, Laar, & Ellemers, 2010).

These studies provide evidence that task performance under stereotype threat, lift or boost conditions should be considered in light of how that performance intersects with individual and group identity concerns. Although there has not yet been a great deal of replication, these studies provide a strong challenge to individual cognitive understandings of the STB phenomenon.

### **3.8. Applying social identity theory to stereotype threat**

Most of the key features of social identity theory are already present in the STB literature, but are often implicit and unexamined. For example, the most successful stereotype

threat interventions include convincing targets that the negatively stereotyped attribute is malleable and learnable rather than essential and unchangeable (Burnette, Pollack, & Hoyt, 2010; Good et al., 2003) or making targets aware of successful minority role models who have penetrated the high status group or domain (eg. Marx et al., 2009; Marx & Roman, 2002). The common feature of these interventions is that both manipulate the perceived permeability of the high status group or privileged domain.

Some successful interventions have asked participants to critique the distinctness of group boundaries (Rosenthal & Crisp, 2006; Rosenthal et al., 2007) or to consider the categories as socially constructed (Shih et al., 2007). On reflection, these could be considered manipulations of the legitimacy of category boundaries and of the groups themselves.

Other studies have reduced or eliminated stereotype threat by making participants aware of their other social and individual identities (Aronson et al., 1999; Gresky et al., 2005; Shih et al., 1999) or encouraging participants to recategorize themselves with the high status outgroup in shared or superordinate membership categories (Kray et al., 2001; Walton & Cohen, 2007) (McGlone & Aronson, 2006; Rydell & Boucher, 2010; Rydell et al., 2009). These interventions capitalize on the key tenet of social identity theory that all individuals have multiple individual and social identities that can be called upon in a given context, and that identity can be optimized by creating the right conditions for people to strategically access their most viable identity resources.

The point is that the STB literature is well-seasoned with concepts that can be interpreted in social identity terms. However, so far these concepts have been employed in isolation and have not been considered in light of the social identity theory meta-theoretical framework. The aim of the present programme of research was to re-imagine situated task performance in social identity terms and to systematically explore the usefulness of the SIT framework in modeling and explaining stereotype-performance effects.

## **Chapter 4. Aims, method and design**

So far three important points have been made. Firstly, the stereotype threat literature has spawned a large number of mediators and moderators contending for inclusion in a universal predictive model of stereotype threat. However, none have been found to consistently influence stereotype threat across groups, tasks and contexts and no suggested causal models have received consistent empirical support. Secondly, a little digging has uncovered many cases of incongruent stereotype-performance effects in the literature, in which positive stereotypes have produced negative shifts in performance and negative stereotypes have improved targets' performance. Although these reversals are excluded from mainstream descriptions of stereotype threat (eg. Steele, 2010), they are common enough in empirical studies to suggest that resistance to stereotype threat effects is possible. There are even some hints suggesting the conditions under which resistance is most likely. Thirdly, the conditions for resistance observed in the literature so far are reasonably congruent with social identity theory, which provides a well developed framework for theorizing the sociostructural conditions in which resistance becomes both possible and likely.

### **4.1. Aims**

The broad aim of this study therefore is to re-imagine stereotype threat and boost (STB) as a social identity phenomenon by exploring the impact of key sociostructural variables on task performance under stereotype threat and boost conditions. However, there are too many variables of interest in social identity theory to explore in a single dissertation, including identifiability, permeability, conflict, stability, legitimacy, and cognitive alternatives. Although it is highly likely that any or all of these could impact on STB effects, the present study focuses on the key sociostructural variables that, according to Tajfel and Turner (1979b) are most likely to impact on the immediate production of self-identity and the positioning of the self on the interpersonal–intergroup continuum in lived experiences.

The two variables with the most primal influence on the production of self-identity in a given context, according to Tajfel and Turner (1979b), are intergroup conflict and intergroup permeability with legitimacy, stability and cognitive alternatives being more distal. Given that later studies found that identifiability – the extent to which a person is identifiable as an individual or able to act as an anonymous member of a group – has a profound effect on the way that self-identity is positioned on the interpersonal–intergroup continuum in any given context (Klein, Spears, & Reicher, 2007; Reicher, Levine, & Gordijn, 1998; Reicher, Spears, & Postmes, 1995), it was included in this programme of research as an important feature of how self identity may be produced from social and individual identity resources in STB situations. Therefore, although there are other SIT variables that may very well impact on SIT that deserve attention in future studies, this dissertation focuses primarily on the variables of intergroup permeability, intergroup conflict and individual identifiability because of their clear impact on the local and immediate production of self-identity.

The specific aims of this project were as follows:

1. To investigate whether STB effects are replicable in the South African tertiary education context by invoking comparisons between different educational institutions and different groups within institutions such as faculties or disciplines.
2. To experimentally explore the impact of perceived intergroup permeability on STB in this context.
3. To experimentally explore the impact of perceived intergroup conflict on STB in this context.
4. To experimentally explore the impact of personal identifiability on STB in this context.

5. To explore the importance of other key individual and sociostructural social identity variables in descriptive and correlational models of STB effects, including: indicators of group-position such as ingroup identification and ingroup and outgroup closeness; structural features of the intergroup context, such as legitimacy, stability, group differentiation; and indicators of individual's experiences, aims and abilities within the specific intergroup setting, such as effort, achievement motivation, social desirability, and category salience.
6. To attempt to use these results to develop a preliminary model re-imagining STB as a social identity phenomenon.

#### **4.2. Method and general design**

Although it seemed likely that most social identity constructs discussed in the literature review might impact on task performance in intergroup contexts, the present programme of research focused on the three that seemed most likely to (a) locate the performance context on the interpersonal-intergroup continuum and (b) impact on the meaning of the task performance in terms of intergroup status. Although identifiability, conflict and permeability were chosen as the most likely candidates, it is very likely that perceived legitimacy and stability will also be important. However, it was not possible to tackle all of these constructs in one dissertation so difficult choices had to be made.

The programme of research consisted of a series of factorial experiments that each invoked STB effects by manipulating perceived group status along with one of the key SIT variables resulting in 2 x N factorial designs. Although this procedure allowed for systematic comparison of the effects of different SIT variables on STB, it was limited to the extent that it did not allow a unified understanding of how SIT variables influence STB in combination.

Although each study included scales measuring SIT variables that were not manipulated, this research strategy still resulted in a piecemeal and relatively unintegrated understanding of the complex and holistic interaction of SIT variables in the STB experience. Nevertheless, the method does allow a preliminary sense of the applicability of SIT to STB phenomena and gives a rough idea of how each variable might impact on STB experiences.

#### **4.2.1. Conventional STB designs**

The most common paradigm for stereotype threat research has drawn on groups with long histories of stigmatization and strong and unambiguous stereotypes in the performance domain, such as African-American students in a University setting; or Women in mathematics. The participants are then asked to participate in a performance situation in which a stereotype is either activated or not activated (implicitly or explicitly). Since neither the groups nor the stereotypes can be randomly assigned to participants, participants cannot be randomly assigned to conditions and such designs are therefore quasi-experimental. While these designs presumably have larger effect sizes, as they draw on powerful stereotypes in realistic conditions, they lack full experimental control and therefore cannot fully explore the empirical relationships between variables or logically exclude confounds.

#### **4.2.2. A paradigm for a (more) controlled method**

A less common method occasionally used in the stereotype threat literature has a great deal of potential as a blueprint for true experimental designs in stereotype threat research. Although we focus a great deal of attention in social psychology to the 'big' groups with powerful and relatively fixed stereotypes, such as race and gender, it is also true that many groups are targets of multiple stereotypes, even within the same performance domain. Indeed, in many social interactions the content of stereotypes and their relevance to specific situations are hotly contested and are the focus of a great deal of social action and interaction (Reicher, Hopkins & Condor, 1997). Therefore, for most groups, it should be possible to identify a

performance domain in which contradictory stereotypes may hold true. For example, Stone, Lynch, Sjomeling & Darley (1999) recruited black and white participants and informed them either that the study was investigating “sports intelligence” (favouring white participants) or “natural athletic ability” (favouring black participants). As a result, (almost) true random assignment to experimental conditions was possible even though group membership was fixed<sup>8</sup>.

There are several advantages to this technique: firstly, the ability to expose each group to both threat and boost conditions increases power; secondly, the improved random assignment counters many threats to internal validity; and thirdly, the method extends the reach of stereotype threat research by invoking less extreme forms of stigma than conventional STB studies.

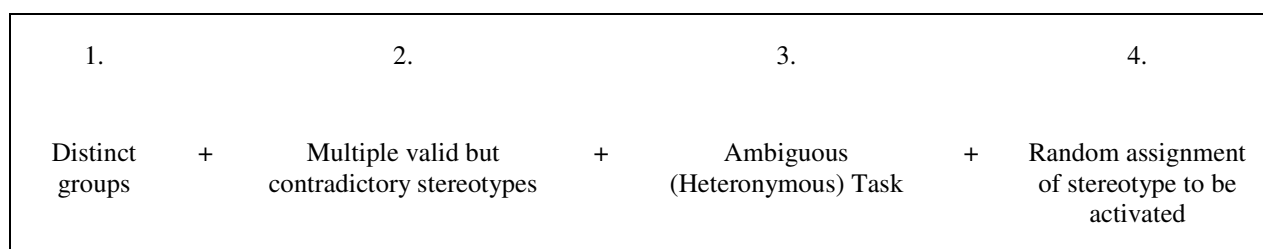
#### **4.2.3. General extension of paradigm**

This method can be expressed as a general form that can easily be applied to a wide variety of social groups and performance contexts (see Figure 2). Specifically, distinct groups need to be produced or identified that have multiple valid but contradictory (ie. positive and negative) stereotypes. An ambiguous task must then be found that can be understood as matching either the positive or negative group stereotype depending on how it is described. Participants can then be randomly assigned to having either the positive stereotype activated, resulting in stereotype boost, or the negative stereotype, resulting in stereotype threat. Each of these components of the paradigm will be discussed in detail below.

---

<sup>8</sup> I say “almost” true random assignment, because although participants are randomly assigned to threat or lift conditions, it is still a leap of faith to believe that “white threat” and “black threat” or “white boost” and “black boost” are equivalent psychological experiences, as discussed below.





*Figure 2: General form of the design*

#### **4.2.3.1. *Distinct Groups***

Firstly, at least two distinct groups must be identified. These groups should, as far as possible, be exhaustive and mutually exclusive within the intended sample, although the possibility of exploring the effects of crossed-categorization is a potential strength of the design.

#### **4.2.3.2. *Multiple valid but contradictory stereotypes***

The groups selected in the first step should additionally be subject to multiple stereotypes with opposite valence. In other words, there should be both positive and negative stereotypes that could plausibly apply to the groups in the performance setting. Since the first psychological studies of stereotypes it has been known that it is common for groups to be associated with both positive and negative stereotypes (Billig, 1996; Katz & Braly, 1933), and it is those that are uni-dimensionally stereotyped that are exceptional. For example, South Africans asked to complete a negotiation task could be negatively stereotyped as violent and thuggish (with apartheid police as an exemplar) or positively stereotyped as reconciliatory and skilled negotiators (with Nelson Mandela as an exemplar). Personal Assistants (PA's) could be seen either as highly organized and efficient, or as unintelligent and servile. Equally, an academic or professor could be stereotyped as brilliantly intelligent or as forgetful and ridiculous. These shifts in stereotype valence may be produced by reframing the experimental task, as Cadinu, Maass, Lombardo & Frigerio (2006) did when they described a task as

testing “social intelligence” (producing stereotype boost for women and threat for men) or “logical intelligence” (producing stereotype threat for women and boost for men). Similarly Stone et al. (1999) presented a golf-related putting task as testing “sports intelligence” (producing stereotype threat for black participants and boost for white participants) or “natural athletic ability” (producing stereotype threat for white participants and boost for black participants).

#### **4.2.3.3. *Ambiguous (heteronymous) task***

Heteronyms are words identical in form, but different in meaning (such as “row” meaning either to argue or paddle). Similarly, most experimental tasks can plausibly be understood in a great variety of ways although this is something that we consensually choose to overlook when we interpret our results. Indeed, even with a construct as deeply entrenched in psychological research as “intelligence,” there is little agreement about what the performance of an intelligence test really represents (Flynn, 2007; McClelland, 1973). This is a feature of the experimental method that can easily be exploited in stereotype threat research, since a single task can be used to invoke different stereotypes depending on how it is framed. However, participants are active and thoughtful, and we must be sure that both ‘frames’ are plausible understandings of the dependent measure task.

#### **4.2.3.4. *Random assignment of stereotype to be activated***

The fact that different stereotypes can be accessed by reframing a single task allows the random assignment of participants to boost and threat conditions even though groups are fixed. However, this is not true random assignment, since there may be underlying group differences in task ability (independent of task framing) that cannot be dealt with by random assignment. More subtly, group identity is more than simply the experience of a stereotype – and group identity cannot be randomly assigned. Therefore we must be careful of assuming that the experience of threat for one group is equivalent to the experience of threat for the

other, as different group identities offer different resources to participants in their experiences of the experimental setting. For example, as previously discussed, Stone et al. (1999) induced stereotype threat conditions by telling white participants that they were completing a test of “natural athletic ability” and black participants that it was a test of “sports intelligence.” Apart from the experience of identity threat, there is little to suggest that these experiences are phenomenologically similar. This may be dealt with by limiting designs to a single group who are randomly assigned to boost or threat conditions (ie. half of the full design), by using repeated measures designs or matched control groups, or by treating the group differences in identity resources and experience as an advantage.

#### **4.2.4. Specific designs used in the present studies**

This general paradigm was used to develop three specific designs: one comparing faculties (humanities/science; or humanities/commerce); one comparing institutions (the University of KwaZulu-Natal, a large public university and Varsity College, a small private university) and one comparing minimal groups (using the fictitious categories of active estimators and reactive estimators). All three instantiations of the general paradigm employed the Raven’s Advanced Progressive Matrices as an ambiguous dependent variable and took a similar form, despite minor differences in the categories and stereotypes employed. The explanation below will focus on how the general paradigm was adapted to the comparison of science and humanities students, which was the most commonly used comparison in the programme of research.

##### ***4.2.4.1. Applying the general paradigm to the specific comparison between Humanities and Science students***

Science and Arts/Humanities students were informed that performance on the dependent measure required either symbolic pattern recognition (favouring Arts/Humanities students) or logical pattern recognition (favouring science students). These specific categories

and stereotypes have been successfully invoked in previous studies (Blanz & Aufderheide, 1999; Ellemers & Barreto, 2006; Morton, Haslam, Postmes, & Ryan, 2006; Spears, Doosje, & Ellemers, 1997). The dependent measure was the Ravens Advanced Progressive Matrices which, although a well-regarded test of fluid intelligence or eductive ability, has previously been shown to be sensitive to stereotype threat and boost effects (Brown & Day, 2006; Croizet et al., 2004; Desert, Preaux, & Jund, 2009; Mayer & Hanges, 2003; McKay et al., 2003; McKay et al., 2002). As a graphical pattern-matching task (see Figure 3), it seems reasonable that it could be considered a test of “symbolic pattern recognition” or “logical pattern recognition” (cf. Cadinu et al., 2006).

#### 4.2.4.2. *Manipulation*

Participants were given an adaptation of the following description: “Humanities students are skilled at finding intuitive patterns in symbolic figures whereas Science students are skilled at finding logical patterns in datasets,” with the ingroup always described in the first position. Participants then read the following sentence, edited so that phrases in brackets matched their status assignment: “Raven’s Advanced Progressive Matrices requires [symbolic/logical] pattern recognition and is therefore more suited to the skills of [Humanities/Science] students whereas [Science/Humanities] students are substantially disadvantaged in the task.” These manipulations result in the design summarised in Table 1, below.

	DV presented as biased towards SYMBOLIC pattern recognition	DV presented as biased towards LOGICAL pattern recognition
Science	Threat	Boost
Arts/Humanities	Boost	Threat

Table 1: The general design tailored to categories of humanities and science students

#### **4.2.5. Advantages and disadvantages of this design paradigm**

Most STB studies invoke powerful and inescapable categories such as race and gender. Although these categories are linked to powerful stereotypes and therefore result in designs that maximize effect sizes and power, they also result in: (1) quasi-experimental designs, since neither the fixed categories nor the powerful stereotypes are manipulable; and (2) designs that are one-sided in that one of the groups will always be exposed to threat and the other will always be exposed to boost. In contrast, although the design paradigm used in the current studies is likely to be less powerful, it does allow (almost) true random assignment and also ensures that each group experiences both threat and boost conditions. This allows the analysis to be more sensitive to specific features of identity as they apply to the experience of STB. Experimental models and designs that focus only on extreme forms of stigma, and those that expose groups only to single conditions, cannot explore the full range of identity processes under threat or boost conditions.

#### **4.2.6. Ethics**

##### ***4.2.6.1. Stressful or upsetting procedures***

The design of this study required half of the student participants to be assigned to a low status group that was exposed to a negative group stereotype related to their faculty, discipline or educational institution. However, given that stereotype threat is by nature an unpleasant effect that requires status differences between groups, there is no way to experimentally explore it without at least some participants experiencing at least a mildly unpleasant condition. The potential harms were limited by selecting relatively innocuous categories and stereotypes that were unlikely to spill-over outside of the experimental context. While this was ethically preferable to designs that invoke inescapable and

overwhelming social stereotypes such as race and gender (eg. Steele, 1997), it did result in some participants experiencing a mildly unpleasant situation.

Additionally, in some studies participants were exposed to manipulations invoking perceived conflict, reduced permeability and increased identifiability that may have been experienced as stressful or unpleasant. Again, these were relatively innocuous, unlikely to cause harm to participants, and there was no way to study the key constructs without invoking these manipulations

#### **4.2.6.2. Deception**

The use of deception is one of the most intractable ethical issues in psychological research. Some argue that deception violates the principles of informed consent and autonomy and should not be employed at all, but empirical research with actual research participants has also shown that “subjects who have participated in deception experiments versus nondeception experiments enjoyed the experience more, received more educational benefit from it, and did not mind being deceived or having their privacy invaded” (Christensen, 1988, p. 664). Despite the controversy that it generates, it is generally agreed that its use is only justified if (1) there is no viable alternative for researching the phenomenon; (2) the deception is unlikely to cause harm to the participants; and (3) the research question is important enough to justify it. (Athanasoulis & Wilson, 2009; Christensen, 1988; Herrera, 2001; Hertwig & Ortmann, 2008). In the present programme of research, deception was required on several levels.

#### *Concealing study aims to avoid demand characteristics*

To avoid introducing confounding demand characteristics, participants were not informed of the true purposes of the study until the debriefing stage. Instead, they received a description that accurately informed them of the nature of the tasks they would be required to perform, but received a ‘smokescreen’ description of the scientific aims of the study.

*Manipulating and invoking ingroup stereotypes*

The second form of deception involved invoking negative or positive stereotypes about some aspect of participants' student identity (for example, their status as humanities or science students). With the exception of the minimal groups manipulation, the positive or negative stereotypes invoked relied on well-known tropes that participants would have already encountered in their everyday lives, and so did not expose participants to strange or unusual negative conditions.

However – unlike most stereotype threat studies that activate overwhelming categories such as race – the groups activated in the study were generally relatively innocuous, and unlikely to be highly stigmatized outside of the experimental context. Therefore, it was hoped, the negative stereotype manipulations for the half of the participants in the low status group would be less likely to 'overflow' and influence them outside of the experimental context. Although one study invoked the stereotype that men are less empathetic than women, this gender stereotype is neither particularly harmful nor strong, and was unlikely to cause lasting harm to participants.

*Manipulating identifiability, permeability, or conflict*

All participants were exposed to a manipulation attempting to convince them that the boundaries between the ingroup and outgroup were more or less permeable, that there was more or less conflict between their ingroup and outgroup, or that they were identifiable or anonymous in the experimental context. Of these, the conflict manipulations was most potentially problematic.

The conflict condition simulated an interaction between the participants' ingroup and outgroup. In the high conflict condition the 'feedback' received from the virtual outgroup was hostile in the sense that it was uncooperative and virtual outgroup members assigned the

minimum possible number of points to actual participants. The impression that the outgroup were hostile was corrected in the debriefing process.

#### *Justification for deceptions*

These stressors and deceptions were necessary to the design of the study and justified by the importance of understanding stereotype threat for our understanding of the underperformance of minority group members in all important social settings, including education and the workplace. No alternative designs capable of exploring these constructs could be identified. To minimize harmful consequences, participants were informed during the informed consent procedure that they “may be exposed to some mildly unpleasant information” and were given precise descriptions of the tasks that they would be required to perform to allow them to make an informed choice about participation. It was expected that discovering the true nature and purpose of the research during debriefing was very unlikely to cause harm or shame.

#### *Correcting deceptions through debriefing*

Participants were all fully debriefed after the experiment with a written debriefing document in which they were told the true nature of the study and why deception had to be used. They were all given the opportunity to discuss their experience and ask questions. Importantly, the debriefing procedure explained the concept of stereotype threat and suggested strategies and resisting the effect in participants’ everyday lives. Such knowledge has been shown to provide some protection from the stereotype threat effect (Johns et al., 2005) and so can be considered a benefit of participation.

#### **4.2.6.3. Anonymity/Identifiability**

Although anonymity has become an important ethical value in research, it is a very uncommon feature of social life. Therefore the strict adherence to this requirement decreases the realism of social psychology experiments and, in some cases, may obscure important



features of the phenomena under scrutiny. Since (a) stereotype threat is a very serious social phenomenon that affects the performance of minority group members in a wide variety of social settings and (b) identifiability was expected to be an important element of the stereotype threat experience, violations of the anonymity principle were justified in order to further understand the mechanisms by which stereotype threat operates.

Participants in identifiable conditions were told that their results would not be anonymous and, in some cases, that images or video footage of their performance would be recorded. Although their names may remain on paper-and-pencil questionnaires (stored in a locked office in the UKZN School of Psychology), all data was stripped of identifying information during the data-entry process and all video footage and photographs were immediately deleted if recorded at all.

#### **4.2.6.4. *Informed consent***

Despite the necessary deception discussed above, participants were warned during the informed consent process that some aspects of the experiment might be stressful and were given precise details about the nature of the tasks they would be required to complete, although they received a deceptive description of the social context and scientific purpose of the tasks. Later studies used a dual-stage informed consent procedure which gave participants the choice to withdraw their data if they were unhappy with the deception they were exposed to after debriefing. None did so.

#### **4.2.6.5. *Incentives:***

Participants in all studies were offered between R20 and approximately R40 (£4) and, in some cases, the opportunity to be entered into a lucky draw for prizes to the value of approximately R500 to R2000 as incentives for participation. The exception was for participants in Identifiability Study 4, who believed that they would receive a reward between R16 and R30 dependent on their performance. Since participation sometimes took over an

hour and participants were drawn from relatively low risk populations these incentives were unlikely to overwhelm participants' autonomy.

#### **4.2.6.6. *Benefits***

The concept of stereotype threat was explained to participants by the researcher during debriefing and participants were given strategies for identifying and resisting the effect in their everyday lives. Such knowledge has been shown to provide some protection from the stereotype threat effect (Johns et al., 2005).

#### **4.2.6.7. *Ethical clearance***

Ethical clearance for each study was applied for from the relevant research ethics committee (REC) prior to undertaking recruitment or data collection<sup>9</sup>.

### **4.3. Measures**

Since task-performance as conceived in STB has not been a focus of SIT research, and social identity has generally been considered only in terms of ingroup identification in STB research, there was little basis for making strong hypotheses about which SIT variables would be most important in modeling the task performance/status relationship. Although difficult decisions had to be made about which variables to manipulate, there were theoretical grounds for believing that a large number of additional SIT constructs might inform the performance/status relationship. Therefore, considering the exploratory nature of this programme of research, each study assessed a battery of SIT constructs in addition to the dependent measure, in the hope that these measures would be useful in developing SIT models of STB.

---

<sup>9</sup> Ethical clearance certificate references: University of St Andrews School of Psychology REC: PS1058; University of KwaZulu-Natal Faculty of Humanities, Development and Social Sciences REC: HSS/1238/010D; HSS/0134/08D; HSS/0136/08H; HSS/0145/08H; HSS/0142/08H; HSS/0131/08D; HSS/0132/08D.

This programme of research evolved over a period of five years, as did the measures included in each study. Over time the measures were gradually improved in terms of scope and relevance, validity, and reliability. While this was appropriate in terms of improving the design of each progressive set of studies, it makes the description of measures complicated. Additionally, since participants' time and goodwill was a limited resource, it was not possible to include scales for all of the relevant constructs from STB and SIT literature that could possibly have contributed to theory building. Therefore, each year scales were added or removed from the inventory of measures on the basis of their relevance in preliminary analyses as the research progressed. In the clear light of retrospect it is evident that some measures should have been included from the start, and others should never have been included at all. Nevertheless, each of the measures included in the analysis will now be discussed in turn.

Unless otherwise stated, each item was presented in Likert-scale format on a seven-point response scale from "strongly disagree" to "strongly agree." This format seems to provide a good balance between sensitivity and overwhelming participants with options; and odd-numbered scales with neutral points generally have psychometric advantages over even-numbered scales (Cox, 1980; Dawes, 2008; Preston & Colman, 2000). Note that in the following description of measures, items designed to be reverse-scored have been noted with an asterisk and phrases in square brackets were to be modified to match the specific features of each study.

Reliability of measures was estimated with Cronbach's alpha, which estimates the internal consistency of a scale and can be interpreted as a signal-to-noise ratio (DeVellis, 2003). Following the guidelines suggested by de Vellis (2003) for reliability thresholds for scales to be used for research purposes only, in this study  $\alpha < .6$  was considered unacceptable except for very tentative exploratory analysis;  $.60 \leq \alpha < .65$  undesirable;  $.65 \leq \alpha < .70$

minimally acceptable;  $.70 \leq \alpha < .80$  respectable;  $.80 \leq \alpha < .90$  very good and  $\alpha > .9$  an indication that the scale could be shortened.

However, Cronbach's alpha is sensitive to scale length and many scales in this programme of research were very short (three or four items), so the reported reliabilities of these short scales generally indicate more respectable reliability than the guidelines above suggest. In the very rare cases that adequate levels of reliability could only be achieved with two item 'scales,' reliability has been calculated by SPSS with a Spearman-Brown correction. For simplicity, cross-study reliability analyses will be presented alongside the description of each measure to avoid repetition in later chapters.

During the measure development and piloting phase all measures were assessed for unidimensionality and, to a lesser extent, independence using factor analysis. A detailed description of the measure development procedure is available on request (Quayle, 2009).

#### **4.3.1. The dependent measure: the Raven's Advanced Progressive Matrices (APM)**

The Raven's Advanced Progressive Matrices (APM) is a well-regarded graphical test of fluid intelligence or 'eductive ability' (Raven, 1989) that is relatively culture-fair compared to other intelligence tests (see Figure 3 for an example of an item) (Brouwers, de Vijver, & Hemert, 2009). However, in this programme of studies its qualities as a test of intelligence are irrelevant. More important is that it has previously been shown to be sensitive to stereotype threat and boost effects (Brown & Day, 2006; Croizet et al., 2004; Mayer & Hanges, 2003; McKay et al., 2003; McKay et al., 2002) and, as a graphical pattern-matching task, it seems ambiguous enough that it could be considered a test of "symbolic pattern recognition" or "logical pattern recognition" (cf. Cadinu et al., 2006). In the final analysis, STB was identified as a group-based difference in APM performance in response to the experimental manipulations.

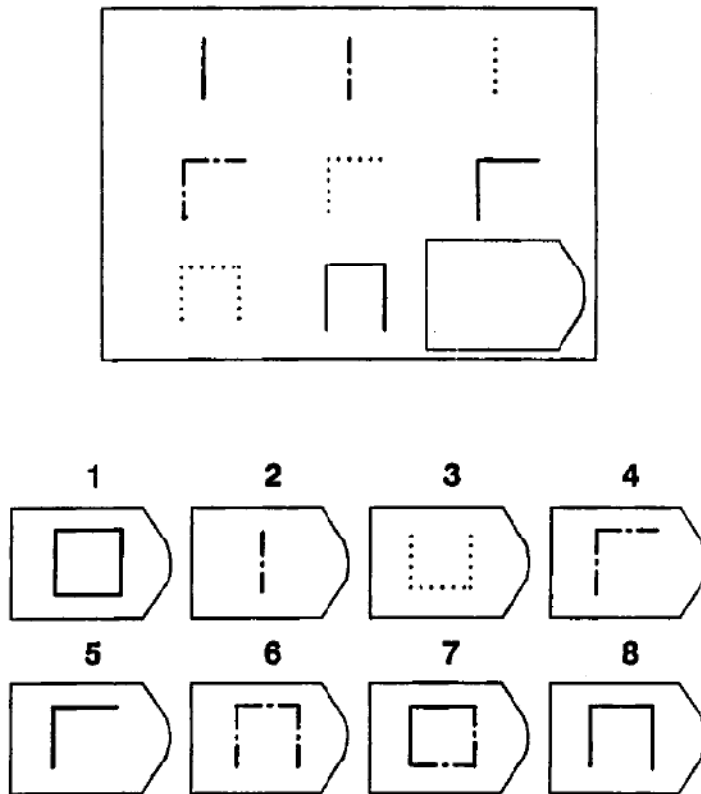


Figure 3: Example of APM item (from Raven, 2000)

The full scale APM consists of 36 items (Set II), not including the 12 practice items (Set I), and generally takes about 45 minute to administer, which is about the amount of time that participants could reasonably be expected to take for the complete study, including manipulations and other measures. For the sake of brevity, the first study (Identifiability Study 1) naively therefore used only Set I, which is usually used only for practice. Although the full range of problem-types is covered in Set I the items are less difficult than Set II (Rushton et al., 2003), resulting in a ceiling effect for the university sample used in the first study.

Therefore the following studies used validated short forms of the APM. Initially the Bors and Stokes short form was used (1998), but in some samples this proved to be too

difficult, introducing a floor effect, and most studies used the items recommended by Bors and Stokes (1998) along with the two easiest items from the Arthur and Day short form (1999) to provide a more suitable range for the target participants.

The exception to this is Identifiability Study 3, which required two equivalent and parallel versions. These were developed by following the procedure outlined by Bors and Stokes (1998) as closely as possible and matching items in each scale to develop two short-forms that should have been very closely matched in terms of difficulty (on the basis of an analysis of all available published item scores on the APM). Actual items used in the APM short forms used in each of the studies are displayed in Table 2.

In all cases, items were scored as incorrect (0) or correct (1) and averaged to arrive at a final score between 0 and 1. These raw scores were analyzed directly without further processing due to the fact that (a) the samples in all cases were culturally and linguistically heterogeneous, (b) in South Africa, at least, no norms are available for the current profile of University students, and (c) all studies with the exception of Identifiability Study 1 sampled comparable participants. For any cross-study comparisons z-scores were used. The reliability for the short form of the APM was consistently acceptable across the studies in this program of research, as displayed in Table 3.

Identifiability Study 1	Conflict Study 1	Identifiability Study 3 Form A	Identifiability Study 3 Form B	All following studies
1.1	1.1	1.1	1.1	1.1
1.2	1.2	1.2	1.2	1.2
1.3	2.3	2.1	2.6	2.4
1.4	2.10	2.2	2.8	2.8
1.5	2.12	2.7	2.9	2.10
1.6	2.15	2.11	2.14	2.12
1.7	2.16	2.12	2.16	2.15
1.8	2.18	2.13	2.17	2.16
1.9	2.21	2.15	2.18	2.18
1.10	2.22	2.19	2.21	2.21
1.11	2.28	2.23	2.22	2.22
1.12	2.30	2.25	2.24	2.28
	2.31	2.27	2.26	2.30
	2.34	2.28	2.29	2.31
				2.34

*Table 2: APM items used in the dependent measure for each study, where the prefix refers to Set I or II and the suffix to the item number within the set.*

	Pilot 1	Pilot 2	C1	ID3 A/B	C2	P1	AM1	P2	P3	C3	ID4
M	-	.549	.43	.288/.304	.404	.390	.391	.384	.38	.332	.419
SD	-	.271	.217	.207/.179	.181	.200	.174	.203	.20	.179	.209
$\alpha$	-	.846	.764	.685/.690	.640	.720	.625	.735	.78	.656	.731
N	23	36	102	53	122	100	93	177	151	138	300

*Note.* C1/C2/C3 = Conflict Studies 1-3; ID3/ID4 = Identifiability studies 3-4; P1/P2/P3 = Permeability studies 1-3; AM1 = Achievement Motivation Study 1

*Table 3: Reliability for the short form APM across studies*

#### **4.3.2. Baseline measure: the Shipley Institute of Living Scale**

Previous research demonstrates that STB is often weak and difficult to detect (Nguyen & Ryan, 2008). Many studies have resorted to including covariates such as SAT scores to increase statistical power enough to detect the effect (eg. Steele, 1997), although this has led to serious methodological criticisms (Sackett et al., 2005; Wicherts, 2005). However, in South Africa no equivalent of SAT scores exist. Therefore, instead of relying on self-report measures of estimates that are incomparable anyway, such as high-school results, it was

decided to include a baseline measure of the same construct assessed by the APM prior to the manipulations.

The Shipley Institute of Living Scale (Abstraction Subtest) is twenty item test of general abstraction ability developed in 1940 (Pollack, 1963) and unchanged since then (included in Appendix A). A typical item gives the prompt “AB BC CD D\_” for which the answer is “E.” Answering this item requires the test-taker to spot the pattern that each pair is in sequential alphabetical order and starts with the second letter of the previous pair. Since the SILS requires pattern recognition in sequences of letters (and, surprisingly, most items require knowledge of alphabetical order rather than English language skills), it is ideal as a baseline measure. Although the measure is highly comparable to the APM, the format is different enough to avoid practice effects.

Senior (2001) reported a median correlation of .72 with the APM based on data from twenty studies. Bors and Stokes (1998) used the abstraction subtest of the Shipley to validate their short form of the APM and reported a correlation of .73, which is highly comparable to Senior’s results. To test its suitability in the local context, the SILS was piloted on a sample of 36 postgraduate psychology students alongside the 13-item short form of the APM used in most of the studies (combining the Bors and Stokes, 1998 version with the easiest items from the Arthur & Day, 1994 short-form). In the pilot the SILS achieved excellent reliability ( $\alpha = .852$ ) and correlated well with APM scores ( $r = .628$ ).

Reliability for the SILS was high across all studies in which it was used (see Table 4). Although the correlation between the SILS and APM was generally much lower than that reported in the literature, it must be remembered that the relationship was intentionally disrupted by the experimental manipulations. Only the second pilot was not subject to the status manipulation, and the correlation with APM scores was much more comparable to Senior’s (2001) review.



	Pilot 1	Pilot 2	C1	ID3 A/B	C2	P1	AM1	P2	P3	C3	ID4
M	-	.764	-	-	.660	.635	.653	.610	.650	.644	.675
SD	-	.172	-	-	.174	.167	.171	.202	.199	.191	.177
$\alpha$	-	.852	-	-	.773	.780	.780	.849	.847	.840	.810
$r(\text{APM})$	-	.628**	-	-	.441**	.575**	.273**	.679**	.553**	.408**	.434**
N	23	36	102	53	122	100	93	177	151	138	300

*Note.* C1/C2/C3 = Conflict Studies 1-3; ID3/ID4 = Identifiability studies 3-4; P1/P2/P3 = Permeability studies 1-3; AM1 = Achievement Motivation Study 1

*Table 4: Reliability and scale statistics for SILS across studies*

### 4.3.3. Stereotype threat and boost manipulation check

The aim of this measure was to assess the extent to which participants had read and understood the content of the status manipulation – in other words, the extent to which they were aware of – and willing to repeat – the positive or negative stereotype that was being applied to their group. Initially five items were developed but the scale was extended to eight items after mixed performance in the first two studies. The items in the full scale were: (1) Based on the description above, [BOOST GROUP] as a group have a good reputation with respect to [the task]; (2) \*Based on the description above, [THREAT GROUP] as a group have a good reputation with respect to [the task]; (3) Based on the description above, [the task] is more suited to the skills of [BOOST GROUP] than [THREAT GROUP]; (4) \*Based on the description above, [the task] is more suited to the skills of [THREAT GROUP] than [BOOST GROUP]; (5) Based on the description above, [BOOST GROUP MEMBERS] are likely to do better than [THREAT GROUP MEMBERS] on [the task]; (6) \*Based on the description above, more [THREAT GROUP MEMBERS] are likely to badly compared to [BOOST GROUP MEMBERS] on [the task]; (7) Based on the description above, the best performing participant will probably be a [BOOST GROUP MEMBER]; (8) \*Based on the description above, the worst performing participant will probably be a [THREAT GROUP

MEMBER]. The labels in parentheses were tailored to the specific manipulations experienced by each participant in each study.

There are three ways of combining these items. Firstly, a scale of stereotype comprehension can be formed by reverse-coding items two, four, six and eight and then averaging with items one, three, five and seven. With this scaling system, high scores represent high comprehension of the stereotype description and vice versa (pilot  $\alpha = .956$ ;  $M = 5.383$ ;  $SD = 1.311$ ). Secondly, a scale of comprehension of own-group performance expectations (ie. boost or threat) could be calculated by: (a) for those assigned to the threat condition, reverse-coding items one, three, five and seven and averaging with items two, four, six and eight; and (b) for those assigned to the boost condition, reverse-coding items two, four, six and eight and averaging with items one, three, five and seven. If the instructions are properly comprehended, this scale should be high for those assigned to the boost condition and low for those assigned to threat. This allows the scores of high and low status participants to be directly compared as a manipulation check or, as will be discussed below, for the calculation of resistance scores. Reliability for this scale was generally good across studies, although in most studies either item 6 or item 8 or both had to be dropped to achieve good reliability, as displayed in Table 4.

	Pilot 1	Pilot 2	C1	ID3	C2	P1	AM1	P2	P3	C3	ID4
M	4.757	4.341	4.144	3.883	3.850	4.271	3.862	4.302	4.195	3.920	4.199
SD	.148	1.889	1.982	1.516	1.881	1.430	1.821	1.576	1.584	1.752	1.149
$\alpha$	.696*	.979	.816*	.481* <sup>^</sup>	.608	.629 <sup>§</sup>	.953	.702 <sup>§§</sup>	.910 <sup>§§</sup>	.887 <sup>§§</sup>	.900 <sup>§§</sup>
N	23	36	102	53	122	100	93	177	151	138	300

*Note.* C1/C2/C3 = Conflict Studies 1-3; ID3/ID4 = Identifiability studies 3-4; P1/P2/P3 = Permeability studies 1-3; AM1 = Achievement Motivation Study 1

\* Initial five-item scale; <sup>^</sup>excluding item 1; <sup>§</sup>Excluding item 8; <sup>§§</sup>Excluding Items 8 & 6;

*Table 5: Reliability and scale statistics for the STB manipulation check across studies (coded as own-group performance expectations)*

#### 4.3.4. Ingroup identification

The ingroup identification measure examined the extent to which participants identified and aligned themselves with their ingroup. Although many ingroup identification scales have been developed, none seemed completely suitable given that the samples for the programme of research were expected to include many second-language English speakers and the simplicity of items was therefore vital. Over the course of the two pilot studies and the first two studies, the eight most appropriate items that could be located were sourced as follows: “My group is an important part of who I am as a person” (adapted from Crisp & Beck, 2005; Ellemers, Kortekaas, & Ouwerkerk, 1999); “I feel strong ties with [ingroup] as a group” (Doosje et al., 1995); “Being an [ingroup member] affects the way I am and how I think” (adapted from Costarelli, 2007); “It feels bad when people say bad things about [my ingroup]” (adapted from Verkuyten & Nekuee, 1999); “I am NOT proud to be a [ingroup member]” (novel item) and “I prefer not to see myself as an [ingroup member]” (novel item).

Reliability was generally adequate (see Table 6) except for the conflict studies, in which it was theoretically expected that the intergroup conflict manipulation would disrupt ingroup identification. This was indeed the case. For example, the reliability analysis was repeated for Conflict Study 3 on the four subsets defined by crosstabulating conflict and

status. Reliability for the low-conflict+low-status condition was  $-.533$  but for the low-conflict+high-status subset it was  $.708$  (dropping items three and four).

Cronbach's alpha has been shown to be a very poor indicator of reliability when its assumptions are violated (eg. Shevlin, Miles, Davies, & Walker, 2000; Vehkalahti, Puntanen, & Tarkkonen, 2006; Vehkalahti, Puntanen, & Tarkkonen, 2009; Voss, Stem, & Fotopoulos, 2000) and, in this case, the error variance was clearly not randomly distributed across conditions because it was related to the conflict manipulation. Although it is an unconventional position, Voss, Stem and Fotopoulos (2000) argue that blind adherence to Nunally's (1978) reliability thresholds in these conditions is inadvisable and that actual reliability for given scale can be substantially higher than alpha coefficients suggests. The problem is that in these conditions there is no way of telling whether the scale is reliable or not. Therefore, since reliability for this scale was generally good, the variable was very tentatively used for exploratory analysis in the conflict studies despite the poor performance of Cronbach's Alpha.

	Pilot 1*	Pilot 2*	C1*	ID3*	C2	P1	AM1	P2	P3	C3	ID4
M	4.348	4.582	4.878	4.595	5.72	4.982	4.100	5.059	4.615	4.663	5.428
SD	1.574	.939	1.637	1.647	.774	1.069	.676	1.023	1.175	.840	.976
$\alpha$	.853	.804	.747	.627	.513 <sup>^</sup>	.723 <sup>§</sup>	.711	.681	.723	.425 <sup>~</sup>	.681
N	23	36	102	53	122	100	93	177	151	138	300

*Note.* C1/C2/C3 = Conflict Studies 1-3; ID3/ID4 = Identifiability studies 3-4; P1/P2/P3 = Permeability studies 1-3; AM1 = Achievement Motivation Study 1

\* Provisional scales under development; <sup>^</sup> Items 3, 4 & 5 only; <sup>§</sup> Dropping items 2; <sup>~</sup> dropping items 3 & 4;

*Table 6: Reliability and scale statistics for ingroup identification across studies (coded as own-group performance expectations)*

#### 4.3.5. Ingroup closeness, outgroup closeness and intergroup closeness

Ingroup closeness, outgroup closeness and intergroup closeness were assessed using Schubert and Otten's (2002) graphical group closeness measures (displayed in Figure 4).

Although these are single item measures, they have been thoroughly validated (see Schubert & Otten, 2002) and – because they do not rely on verbal descriptions – are likely to be more reliable in multilingual samples and for cross-study (cross-context) comparisons than verbal scales.

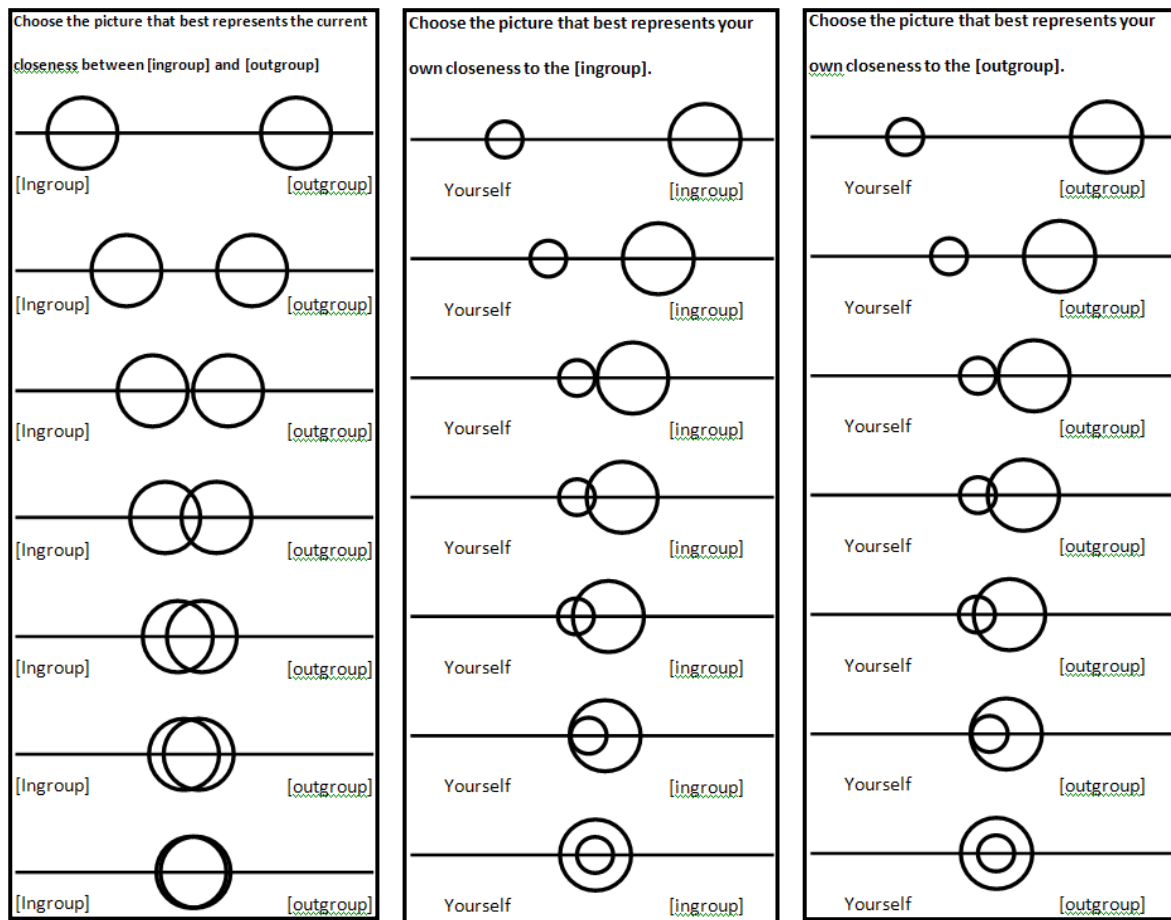


Figure 4: Schubert and Otten's (2002) graphical measures of intergroup, ingroup, and outgroup closeness

#### 4.3.6. Group Differentiation

The group differentiation scale assessed the extent to which participants believed that the ingroup was different to, and distinct from, the outgroup. In the initial three item scale, three items were taken from Jackson's (2002) 13 item Self-Categorization scale ( $\alpha = .84$ ), namely: "[Ingroup members] are different to [outgroup members]"; "[Ingroup members] have a number of things in common with each other"; and "People in [the outgroup] are a lot alike in many respects."

After the scale returned mixed reliability in the first pilot, Identifiability Study 3, and Conflict Study 1, additional items were developed and piloted (Pilot 2) and the best four

items were selected for the following studies, namely: “There are important differences between the [ingroup] and [outgroup],” “[Outgroup] members are different from [ingroup] members,” “In general [ingroup members and outgroup members] are very similar to each other,” and “[ingroup members and outgroup members] have a lot in common.” However, this scale was not included in all of the remaining studies as it did not appear to be adding much value in preliminary analyses.

	Pilot 1*	Pilot 2 <sup>#</sup>	C1	ID3	C2 <sup>§</sup>	P1 <sup>§</sup>	AM1	P2	P3	C3	ID4
M	4.935	3.995	5.361	5.099	4.526	3.407	-	-	-	-	-
S D	1.354	.834	1.348	1.473	.712	.968	-	-	-	-	-
$\alpha$	.716	.820	.204	.726	.662	.650	-	-	-	-	-
N	23	36	102	53	122	100	93	177	151	138	300

*Note.* C1/C2/C3 = Conflict Studies 1-3; ID3/ID4 = Identifiability studies 3-4; P1/P2/P3 = Permeability studies 1-3; AM1 = Achievement Motivation Study 1

\* Initial 3 item scale; <sup>#</sup>Extended 8 item scale; <sup>§</sup>Improved 4 item scale;

*Table 7: Reliability for group differentiation scale across studies*

#### 4.3.7. Identifiability

Identifiability refers to the extent to which a person believes they are acting entirely anonymously, as representatives of an identifiable group, or as individually recognizable agents. Identifiability has been shown to have important social-identity-related effects that may be relevant to SIT, such as deindividuation (cf. Klein et al., 2007; Lea, Spears, & de Groot, 2001; Reicher et al., 1998) and, indeed, has been shown to impact on STB effects in the guise of “individuation” (Ambady et al., 2004).

Since no measures could be found in the literature, three items were developed to assess group-level identifiability (the extent to which participants felt that they were acting as identifiable representatives of the group), namely: “I felt like my group [ie. ingroup] was in the spotlight while I was doing the task”; and “I felt as if I was being tested on behalf of my

group”; “I felt like my group would be judged on the basis of my performance.” For studies in which identifiability was the topic of manipulation, an additional item was added to increase reliability: “People who saw this questionnaire would know that I am a [ingroup member].” The reliability for this scale across studies is reported in Table 8 .

	Pilot 1	Pilot 2	C1	ID3	C2	P1	AM1	P2	P3	C3	ID4
M	-	3.500	4.330	4.165	4.789	4.580	-	-	4.673	5.090	5.208
SD	-	1.240	1.342	1.143	1.203	1.272	-	-	1.465	1.121	1.435
$\alpha$	-	.815	.586	.647 <sup>+</sup>	.697	.707	-	-	.780	.644	.775
N	23	36	102	53	122	100	93	177	151	138	300

*Note.* C1/C2/C3 = Conflict Studies 1-3; ID3/ID4 = Identifiability studies 3-4; P1/P2/P3 = Permeability studies 1-3; AM1 = Achievement Motivation Study 1

+ three-item scale

*Table 8: Reliability for group identifiability scale across studies*

Three items were developed to assess individual identifiability in most studies, namely: “I felt that I could be identified 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”; and “People who saw this questionnaire would be able to trace me as an individual.” A further item was included as a reliability buffer for studies manipulating identifiability, namely: “I felt completely anonymous while completing [the task].\*.” The reliability for the individual identifiability scale across studies is reported in Table 9.



	Pilot 1	Pilot 2 <sup>#</sup>	C1	ID3	C2	P1	AM1	P2	P3	C3	ID4
M	-	3.694	4.343	4.458	4.043	4.123	-	3.754	3.960	3.627	4.361
SD	-	1.094	1.200	1.249	1.092	1.054	-	1.245	1.307	1.333	1.319
$\alpha$	-	.760	.599	.646	.606	.642 <sup>^</sup>	-	.690	.661	.758	.675
N	23	36	102	53	122	100	93	177	151	138	300

*Note.* C1/C2/C3 = Conflict Studies 1-3; ID3/ID4 = Identifiability studies 3-4; P1/P2/P3 = Permeability studies 1-3; AM1 = Achievement Motivation Study 1

#five-item scale; ^dropping items 3 & 4;

*Table 9: Reliability for individual identifiability scale across studies*

Reliability for both the group and individual identifiability scales was generally acceptable, and improved over the course of the programme of research as the items were fine-tuned. Note that the items reported above are the final form of these scales achieved by the end of the programme of research, but differences introduced across studies were minor.

#### **4.3.8. Legitimacy**

Legitimacy refers to the extent to which the status difference between groups is perceived to be legitimate, just or valid (Ellemers, Wilke, & Van Knippenberg, 1993; Mummendey, Klink, Mielke, Wenzel, & Blanz, 1999). The social identity literature was searched for validated items that were simple and easily understandable, and the following items were selected: “the difference between [ingroup] and [outgroup] is justified and right” (adapted from Costarelli, 2007; Weber, Mummendey, & Waldzus, 2002); “the difference between [ingroup] and [outgroup] makes sense” (“is reasonable” in Weber et al., 2002); “the difference between [ingroup] and [outgroup] is unfair<sup>\*</sup>” (adapted from Hornsey, Spears, Cremers, & Hogg, 2003; Jost & Burgess, 2000); and “when people think the [high status group] is better than the [low status group] they are not seeing things as they really are<sup>\*</sup>” (original/novel item). Note that these items went through minor revisions over the course of

the programme of research and are reported here in their final form. Reliability was acceptable across studies, as displayed in Table 10.

	Pilot 1 <sup>#</sup>	Pilot 2 <sup>#</sup>	C1 <sup>#</sup>	ID3 <sup>#</sup>	C2	P1	AM1	P2	P3	C3	ID4
M	3.864	3.217	4.74	4.241	4.469	4.082	-	-	4.043	2.806	3.922
SD	1.365	.986	1.386	1.512	.720	1.125	-	-	1.155	1.224	1.092
$\alpha$	.824	.861	.749	.697	.627	.724	-	-	.753	.608 <sup>+</sup>	.607
N	23	36	102	53	122	100	93	177	151	138	300

*Note.* C1/C2/C3 = Conflict Studies 1-3; ID3/ID4 = Identifiability studies 3-4; P1/P2/P3 = Permeability studies 1-3; AM1 = Achievement Motivation Study 1

<sup>#</sup>eight-item scale; <sup>+</sup>excluding item 3.

*Table 10: Reliability for the legitimacy scale across studies*

#### **4.3.9. Stability**

Stability refers to perceptions that the status hierarchy between groups is firmly entrenched and resistant to transformation, or fragile and likely to change (Bettencourt & Hume, 2001; Ellemers, Van Knippenberg, & Wilke, 1990; Mummendey et al., 1999; Scheepers, 2009; Scheepers et al., 2009).

Items were adapted from Mummendey, Klink, Mielke, Wenzel & Blanz (1999) who asked participants about the stability of “the relationship” between the groups. However, since stability is specifically oriented towards intergroup status, the word “relationship” seems quite non-specific. Therefore three of these items were adapted to ask about “the difference” between groups rather than “relationship” to make them more specifically related to the intergroup status hierarchy. The initial items were: “I think the difference between [ingroup] and [outgroup] will remain stable for the few next years”; “the current gap between [ingroup] and [outgroup] is just temporary<sup>\*</sup>”; and “the current difference between [ingroup] and [outgroup] will not change easily.” After inconsistent performance in the pilots and the first two studies, three novel items were added, namely: “even if they try their best, [low

status group as a group] will not overtake [high status group] in terms of status and privileges”; “It’s unlikely that [high status group] will lose their good reputation”; and “No matter what they do, [low status group] will never have as much status as [high status group].” Reliability was generally acceptable to good across studies, as displayed in Table 11.

	Pilot 1	Pilot 2	C1	ID3	C2	P1	AM1	P2	P3	C3	ID4
M	4.91	4.132	4.993	4.805	4.003	4.522	-	-	3.666	3.700	4.001
SD	1.365	.743	1.480	1.490	.872	1.030	-	-	1.081	1.094	1.126
$\alpha$	.671	.619	.629	.473	.761	.717	-	-	.738	.737	.728
N	23	36	102	53	122	100	93	177	151	138	300

*Note.* C1/C2/C3 = Conflict Studies 1-3; ID3/ID4 = Identifiability studies 3-4; P1/P2/P3 = Permeability studies 1-3; AM1 = Achievement Motivation Study 1

*Table 11: Reliability for the stability scale across studies*

#### **4.3.10. Permeability**

Permeability refers to the extent to which members of one group can transfer membership to the other group and become fully accepted members with all the benefits or liabilities of group membership (Bettencourt & Hume, 2001; Ellemers et al., 1988; Ellemers et al., 1990; Mummendey et al., 1999). Although it is quite possible for permeability perceptions to differ qualitatively by direction (for example, a science student may believe that science students could successfully transfer to humanities degrees but not vice versa), it is the extent to which the participant believes that they personally could transfer (or be relegated) to the outgroup that is most important in the context of STB and this series of studies. Therefore the permeability scale focuses on ingroup to outgroup permeability.

Since a thorough review of the literature did not identify any useable scales of permeability, an item pool was developed from which four items were selected on the basis of pilot study results, namely: “It would be difficult for an [ingroup member] to adjust to being an [outgroup member]”; “An [ingroup member] would feel anxious about becoming an

[outgroup member]”; “An [ingroup member] would feel confident about moving to [the outgroup]”; and “[Ingroup members] would fit in well with [the outgroup].” For studies manipulating permeability (in which a reliable measure was critical) item two was dropped and a further three items were included, namely: “an [ingroup member] can easily become an [outgroup member]”; “An [ingroup member] would be successful as an [outgroup member]”; and “it is difficult to move from [ingroup] to [outgroup].” Although the scale had (just) adequate reliability for all but Conflict Study 2, it is clear that the extended version was much more successful than the shorter four-item version.

	Pilot 1	Pilot 2	C1	ID3	C2	P1	AM1	P2	P3	C3	ID4
M	4.424	4.414	.481	3.730	4.311	4.319	-	4.36	4.865	2.297	3.869
SD	1.557	.685	1.826	1.600	1.057	1.234	-	1.15	.983	1.207	1.449
$\alpha$	.700	.865 <sup>+</sup>	.723	.707	.549 <sup>#</sup>	.794 <sup>+</sup>	-	.795 <sup>+</sup>	.746 <sup>+</sup>	.655	.599 <sup>^</sup>
N	23	36	102	53	122	100	93	177	151	138	300

*Note.* C1/C2/C3 = Conflict Studies 1-3; ID3/ID4 = Identifiability studies 3-4; P1/P2/P3 = Permeability studies 1-3; AM1 = Achievement Motivation Study 1

+extended scale; #Excluding item 2; ^excluding items 1 &2, leaving just a two-item “scale”(alpha reported with the Spearman-Brown correction)

*Table 12: Reliability for the permeability scale across studies*

#### 4.3.11. Conflict

Conflict is the extent to which intergroup relations are perceived to be hostile and conflictual or cooperative and peaceful (Jackson, 2002; Tajfel, 1978a; Tajfel & Turner, 1979a). For the generic scale, three items were selected and adapted from Tropp’s (2003) “Emotional states in intergroup contexts” scale which, in turn, had been adapted from Zuckerman and Lubin (1965, as cited in Tropp, 2003), Spielberger, Gorsuch, and Lushene (1970, as cited in Tropp, 2003), and Stephan and Stephan (1985, 1992, as cited in Tropp, 2003). These items assess affective aspects of conflict and rivalry, namely: “I feel IRRITATED when I think about interacting with [the outgroup]”; “I feel UPSET when I

think about interacting with [the outgroup]”; “I feel ANGRY when I think about interacting with [the outgroup].” To include cognitive perceptions of conflict, one novel item was included, namely: “There is cooperation between the [ingroup] and [outgroup].<sup>\*,\*</sup>”

Although an extended scale was prepared for studies in which conflict was to be manipulated, the additional items actually detracted from reliability and were not retained. Reliability was good to excellent across the studies, although the performance of item 4 was unpredictable, as displayed in Table 13.

	Pilot 1	Pilot 2	C1	ID3	C2	P1	AM1	P2	P3	C3	ID4
M	2.894	-	2.781	3.103	1.936	2.591	-	2.098	2.541	1.969	2.284
SD	1.353	-	1.469	1.476	.943	.914	-	1.205	1.165	1.254	1.055
$\alpha$	.893	-	.735	.886	.769 <sup>+</sup>	.918 <sup>+</sup>	-	.730 <sup>+</sup>	.784 <sup>+</sup>	.770	.742
N	23	36	102	53	122	100	93	177	151	138	300

*Note.* C1/C2/C3 = Conflict Studies 1-3; ID3/ID4 = Identifiability studies 3-4; P1/P2/P3 = Permeability studies 1-3; AM1 = Achievement Motivation Study 1

+dropping item 4

*Table 13: Reliability for the conflict scale across studies*

#### **4.3.12. Effort**

The effort/importance subscale assessed the perceived importance that participants attached to performing well in the task and the amount of effort they put into completing the task. Three items were adapted from the effort/importance subscale of Deci and Ryan’s (2005) intrinsic motivation inventory (IMI). These were: “I put a lot of effort into the Raven’s Advanced Progressive Matrices task”; “I didn’t try very hard to do well at the Raven’s Advanced Progressive Matrices task<sup>\*</sup>”; and “while doing the Raven’s Advanced Progressive Matrices task, when I came to difficult problem I did my best to work it out.” After inconsistent performance in the first pilot and first two studies a fourth (novel) item was

added, namely: “while doing the Raven’s Advanced Progressive Matrices task, when I came to a difficult problem I did not mind guessing.\*”

Reliability for this scale is reported in Table 14 and was generally adequate to good across the studies. However, the scale did tend towards negative skewness which was not ideal.

	Pilot 1	Pilot 2	C1	ID3	C2	P1	AM1	P2	P3	C3	ID4
M	4.884	5.026	5.205	4.747	5.143	5.052	5.508	5.585	5.100	3.251	5.444
SD	1.260	1.000	1.415	1.451	1.211	1.136	.971	1.045	1.246	1.965	1.040
$\alpha$	.756	.822	.419	.381	.726	.705	.631	.697 <sup>+</sup>	.770	.716	.643
N	23	36	102	53	122	100	93	177	151	138	300

*Note.* C1/C2/C3 = Conflict Studies 1-3; ID3/ID4 = Identifiability studies 3-4; P1/P2/P3 = Permeability studies 1-3; AM1 = Achievement Motivation Study 1

+dropping item 4

*Table 14: Reliability for the effort scale across studies*

#### **4.3.13. Category salience**

This scale assessed the extent to which participants’ category membership was salient while they were completing the dependent measure. All four items were novel, namely:

“While I was doing the Ravens Advanced Progressive Matrices I was not aware that I am an [ingroup member]”; “While I was doing Ravens Advanced Progressive Matrices I felt very much like [an ingroup member]”; and “While I was doing Ravens Advanced Progressive Matrices I thought of myself as an [ingroup member].” Reliability for this scale is reported in

Table 15 and was good to excellent across studies.

	Pilot 1	Pilot 2	C1	ID3	C2	P1	AM1	P2	P3	C3	ID4
M	.4101	-	5.082	4.483	5.133	5.203	-	-	3.918	2.716	4.588
SD	1.915	-	1.820	1.792	1.562	1.406	-	-	1.371	1.511	1.673
$\alpha$	.766	-	.738	.802	.822	.738	-	-	.827 <sup>+</sup>	.729	.611
N	23	36	102	53	122	100	93	177	151	138	300

*Note.* C1/C2/C3 = Conflict Studies 1-3; ID3/ID4 = Identifiability studies 3-4; P1/P2/P3 = Permeability studies 1-3; AM1 = Achievement Motivation Study 1

+dropping item 1, alpha for the two-item “scale” calculated using the Spearman-Brown correction

*Table 15: Reliability for the category salience scale across studies*

#### **4.3.14. Achievement motivation**

Following the 2008 conflict and identifiability studies, it was realized that achievement motivation may be a viable proxy for the SAT scores used by many STB studies as a covariate that ‘reveals’ the STB effect. Fifteen items were piloted on 36 postgraduate psychology students and, on the basis of pilot results, the following four items were selected: “it is important for me to excel in most things I do” (Lochbaum, Stevenson, & Hilario, 2009); “I worry when I feel like I might not succeed at a task” (novel item); “being average does not bother me<sup>\*</sup>” (novel item); and “hard work is something I prefer to avoid.<sup>\*</sup>” (Cassidy & Lynn, 1989). Following poor performance in studies, two additional items were added: “I worry that I may not do as well as I possibly can” (Lochbaum et al., 2009); and “I like to be admired for my achievements” (Cassidy & Lynn, 1989).

Reliability for this scale was generally very poor across studies and, in most cases, could not be improved by dropping items (Table 16). The construct of achievement motivation has therefore been invoked as little as possible in the analysis.

	Pilot 1	Pilot 2	C1	ID3	C2	P1	AM1	P2	P3	C3	ID4
M	-	-	-	-	5.885	5.728	4.785	5.713	5.559	5.007	5.792
SD	-	-	-	-	.864	.765	.712	.822	.956	.730	.802
$\alpha$	-	-	-	-	.528*	.284*	.516*	.336*	.630	.407	.541
N	23	36	102	53	122	100	93	177	151	138	300

*Note.* C1/C2/C3 = Conflict Studies 1-3; ID3/ID4 = Identifiability studies 3-4; P1/P2/P3 = Permeability studies 1-3; AM1 = Achievement Motivation Study 1

\*Four-item scale

*Table 16: Reliability for the achievement motivation scale across studies*

#### 4.3.15. Social desirability

Social desirability refers to the extent to which participants are motivated to answer questions and act in ways that produce or maintain a positive impression in the experimental context. Social desirability is often measured with the Marlowe-Crowne social desirability scale (Crowne & Marlowe, 1960), which is a 33 item scale that asks participants to self-report on behaviours that are relatively normal, but socially unacceptable to varying degrees. The idea is that the average person will have engaged in most of the behaviours, but participants high in social desirability tendencies will be less willing to admit it.

One of the many problems with the original Marlowe-Crowne scale is that 33 items places a substantial additional response burden on participants. Therefore the shortest validated form of the Marlowe-Crowne scale that could be located was used (Hays, Hayashi, & Stewart, 1989) with some amendments to vocabulary to make it more relevant to the South African participants. Participants were given the following instructions: “Listed below are a few statements about your relationships with others. How much is *each* statement TRUE or FALSE for you?” The following items were listed: “I am always polite, even to people who are unpleasant”; “there have been occasions when I took advantage of someone\*”; “I sometimes try to get even with people rather than forgive and forget\*”; “I sometimes feel



resentful when I don't get my way\*"; "No matter who I'm talking to, I'm always a good listener."

Reliability for this scale was generally poor, but Hays and colleagues (Hays et al., 1989) argue that binary coding of extreme versus moderate scores will nevertheless reliably distinguish between participants with particularly high and low levels of the construct.

	Pilot 1	Pilot 2	C1	ID3	C2	P1	AM1	P2	P3	C3	ID4
M	-		-	-	3.785	4.602	4.214	4.584	4.464	2.882	3.786
SD	-		-	-	.655	1.104	1.139	1.000	.760	.792	.847
$\alpha$	-		-	-	.545	.588	.660	.585	.374	.477	.569
N	23	36	102	53	122	100	93	177	151	138	300

*Note.* C1/C2/C3 = Conflict Studies 1-3; ID3/ID4 = Identifiability studies 3-4; P1/P2/P3 = Permeability studies 1-3; AM1 = Achievement Motivation Study 1

*Table 17: Reliability for the social desirability scale across studies*

#### **4.3.16. Symbolic and realistic threat**

In earlier studies in the programme of research measures of symbolic and realistic threat were included. Intergroup threat can be understood as the extent to which two groups are perceived to be competing for the same material or symbolic resources. In conditions of high perceived threat, ingroup members will believe that outgroup members' competition for resources or competing values threatens the continued existence or viability of the ingroup (Stephan et al., 2002; Stephan & Stephan, 1996).

Three items measuring realistic threat were developed by the author, namely; "The [outgroup] is gaining resources at the expense of [ingroup]"; "It will take me longer to find a job because of [outgroup members] in the job market"; "If [outgroup] gets too successful then [ingroup] will really struggle to succeed."

Two items assessing symbolic threat were adapted from Stephan, Boniecki, Ybarra, Bettencourt, Ervin, Jackson, McNatt, and Renfro (2002), namely: "[Outgroup members] don't

understand the way that [ingroup members] view the world”; and “[outgroup members] don’t realize the true importance of [a core ingroup activity].” A further two items were adapted from Durrheim, Dixon, Tredoux, Eaton, Quayle and Clack (2009) and Dixon, Durrheim, Tredoux, Tropp, Clack, and Eaton (2010), namely: “most [outgroup members] will never understand what [ingroup members] are like”; and “the values that are important to [ingroup members] are under threat, because of the influence of [outgroup] values.” Scale statistics for the symbolic and realistic threat scales are displayed in Table 18 and Table 19.

	Pilot 1	Pilot 2	C1	ID3	C2	P1	AM1	P2	P3	C3	ID4
M	4.148	3.879	4.461	4.56	4.556	4.102	-	-	-	-	-
SD	1.636	5.404	1.581	1.661	4.729	4.353	-	-	-	-	-
$\alpha$	.776	.859	.766	.618	.670	.741	-	-	-	-	-
N	23	36	102	53	122	100	93	177	151	138	300

*Note.* C1/C2/C3 = Conflict Studies 1-3; ID3/ID4 = Identifiability studies 3-4; P1/P2/P3 = Permeability studies 1-3; AM1 = Achievement Motivation Study 1

*Table 18: Reliability for the symbolic threat scale across studies*

	Pilot 1	Pilot 2	C1	ID3	C2	P1	AM1	P2	P3	C3	ID4
M	-	3.086	-	-	5.554	5.980	-	-	-	-	-
SD	-	3.936	-	-	2.575	2.526	-	-	-	-	-
$\alpha$	-	.823	-	-	.525*	.555*	-	-	-	-	-
N	23	36	102	53	122	100	93	177	151	138	300

*Note.* C1/C2/C3 = Conflict Studies 1-3; ID3/ID4 = Identifiability studies 3-4; P1/P2/P3 = Permeability studies 1-3; AM1 = Achievement Motivation Study 1

\*Dropping item 1

*Table 19: Reliability for the realistic threat scale across studies*

#### 4.4. Strategy and methods for analysis

All designs manipulated status and one other social identity variable to arrive at a 2 x K factorial design. Analyses started by testing the most basic possible model assessing the

influence of the two manipulated factors on APM performance. The analyses then proceeded to explore the influence of multiple continuous (or at least ordinal) predictors in the model. Since multiple regression and ANOVA are both special cases of the general linear model and are generally equivalent (Miles & Shevlin, 2001; Thompson, 1986), and regression was more appropriate for the more complex models (Miles & Shevlin, 2001), the simple factorial designs were also analyzed using regression techniques to maintain consistency across analyses.

Centering of continuous independent variables is highly recommended (Kraemer & Blasey, 2004), so analyses were routinely conducted with centred data. However, when the model fit and coefficients were similar when the analysis was repeated with non-centred variables, these results were preferred to aid interpretation.

Model fit for all regression analyses was assessed using the scatterplots, q-q and leverage plots, but unless there were clear problems with the model these are not reported due to space constraints. Variables were transformed where necessary to correct skewness and kurtosis.

#### **4.4.1. Descriptives**

The procedure followed when analyzing the data for each study was first to calculate scales, clean the data and check that all scales were reliable. Exploratory data analysis was then performed to check the scales for ceiling and floor effects and skewness, and correlations were calculated to check for potential problems with multicollinearity, using Garson's (2010) rule-of thumb that one correlation higher than .90, or several higher than .70 indicates potential problems with multicollinearity. Again, due to severe space constraints, these diagnostics were not reported unless problematic.

#### **4.4.2. Experimental hypothesis testing**

Second, as described above, the primary experimental hypotheses were tested using multiple regression with interaction terms entered as product effects (Aguinis, 2004; Miles & Shevlin, 2001; West, Aiken, & Krull, 1996). These product terms were automatically calculated with the `lm()` function in R (R Development Core Team, 2010).

## **Chapter 5. Results I: Status, performance and resistance**

The first aim of this programme of research was to invoke the classic stereotype threat effect. All studies manipulated the perceived status of the participants' group membership and assessed the impact on between-groups differences in the performance task, which was generally a short form of the Raven's Advanced Progressive Matrices (APM). According to the stereotype threat model this manipulation should have invoked stereotype threat and boost (STB) and should have resulted in significant differences in task performance. The exceptions to this design template were study Identifiability Study 2 which used an emotional word recognition task instead of the APM, and Identifiability Study 3 which used a repeated measures design.

### **5.1. The effects of status on APM performance**

Eleven studies were completed in total and, as displayed in Table 20, none revealed significant differences in performance by status. Indeed, only the first study (Identifiability Study 1) came close to finding a significant effect,  $p = .104$ .

Abbreviation	N	APM Performance by Status			
		<i>t</i>	<i>F</i>	df.	<i>p</i>
Identifiability Study 1	54	1.653	-	52	.104
Identifiability Study 2 <sup>+</sup>	218	-	.936	1,214	.334
Identifiability Study 3 <sup>++</sup>	53	-.361	-	51	.720
Identifiability Study 4	300	-.233	-	277	.816
Conflict Study 1	102	.716	-	100	.476
Conflict Study 2	122	.876	-	70	.384
Conflict Study 3	138	1.038	-	126	.301
Permeability Study 1	100	.346	-	98	.730
Permeability Study 2	177	.546	-	165	.586
Permeability Study 3	151	-.122	-	149	.903
Achievement Motivation Study 1	93	-.868	-	91	.388

<sup>+</sup>DV was emotional sensitivity; <sup>++</sup>Used a repeated measures design

*Table 20: Results for performance by status for all eleven studies*

Given that these studies avoided strong stereotypes such as race and gender for ethical and methodological reasons it is likely that all of these studies had low power. However, the study with the largest sample size,  $N = 300$ , had a particularly high estimate that the observed difference in performance by status could have occurred by chance alone,  $p = .816$ .

The majority of these studies used virtually identical dependent measures (with very minor variations in the actual items used in the short-form of the APM in the earlier studies) and similar manipulations of status (with minor variations tailoring the general script to specific categories). The exceptions are Identifiability Study 1, which used Set I of the APM as a dependent measure and was so easy that it suffered an extreme ceiling effect, Identifiability Study 2 which did not use the APM as a dependent measure at all, and Identifiability Study 3 which was a repeated measures design and too different from the other studies to consider directly comparable. If the social identity manipulations are overlooked,

the data of eight studies are highly compatible and can be combined to form a large single sample,  $N = 1092$ <sup>10</sup>.

When the analysis of APM performance by status assignment was repeated on the combined sample the results were similar,  $t = .518$ ,  $df. = 1090$ ,  $p = .647$ . Therefore the evidence is clear that status had no significant direct effect on performance on the Ravens Advanced Progressive Matrices in these studies.

## 5.2. Manipulation checks

There are at least two possible explanations for these resoundingly non-significant findings. The first is that the manipulations may not have reliably and/or effectively invoked stereotype threat. The second possibility is that in these contexts stereotype threat and boost conditions were successfully invoked but – for whatever reason – did not result in significant shifts in performance on the APM. The first two studies lacked an effective manipulation check, but the balance of the studies included a manipulation comprehension scale of between five and eight items assessing participants' comprehension of the status manipulation. All questions began with the phrase “based on the description above” followed by a question about whether the boost group or threat group were more likely to do well on the stereotyped task and were answered on a 7-point Likert scale from Strongly Disagree to Strongly Agree. For example: “Based on the description above, *humanities students* as a group have a good reputation with respect to Raven's Advanced Progressive Matrices” (where the category in italics was modified to be congruent with the participant's group membership and manipulation; see section 4.3.3 for a description of measure development). This scale generally had acceptable to good reliability (see Table 22) and allowed some

---

<sup>10</sup> A meta-analysis would be a needless trade-off here, as the measures, manipulations and sample characteristics were virtually identical across the studies. Additionally, a critical key to this analysis is the significance of an interaction effect which is difficult to model with meta-analytic techniques.

analysis of whether the status manipulations were understood and recalled accurately. The placement of the manipulation check was immediately after the status manipulation and immediately before the dependent measure (the APM) in each study, so it had the additional advantage of cementing the negative or positive stereotype by requiring the participant to recall, process and actively report it immediately prior to undertaking the stereotyped task.

Item
1. Based on the description above, [BOOST GROUP] <sup>#</sup> as a group have a good reputation with respect to the Raven's Advanced Progressive Matrices
2. Based on the description above, [THREAT GROUP] <sup>#</sup> as a group have a good reputation with respect to the Raven's Advanced Progressive Matrices
3. Based on the description above, the Raven's Advanced Progressive Matrices is more suited to the skills of [BOOST GROUP] than [THREAT GROUP]
4. Based on the description above, the Raven's Advanced Progressive Matrices is more suited to the skills of [THREAT GROUP] than [BOOST GROUP]
5. Based on the description above, [BOOST GROUP MEMBERS] are likely to do better than [THREAT GROUP MEMBERS] on the Raven's Advanced Progressive Matrices <sup>+</sup>
6. Based on the description above, more [THREAT GROUP MEMBERS] are likely to badly compared to [BOOST GROUP MEMBERS] on the Raven's Advanced Progressive Matrices <sup>+</sup>
7. Based on the description above, the best performing participant will probably be a [BOOST GROUP MEMBER] <sup>+</sup>
8. Based on the description above, the worst performing participant will probably be a [THREAT GROUP MEMBER]

*Note.*<sup>+</sup> Items did not appear in Identifiability Study 3 or Conflict Study 1; <sup>#</sup> Category labels were modified according to participants' status assignment.

*Table 21: Updated items for the STB Manipulation Check, answered on a 7-point Likert scale from Strongly Disagree to Strongly Agree*

As described in section 4.3.3, these items were designed so that they could be combined to form a scale assessing the participant's knowledge of how members of their own group would be expected to behave (expected ingroup performance) or an estimate of manipulation comprehension in which high scores represent high comprehension of the stereotype description and vice versa, as discussed in 4.3.

If the manipulation was successful then we would expect participants to report *similar* (high) levels of *comprehension* regardless of which status condition they were allocated to,



but we would expect significant *differences* between status conditions in expected ingroup performance. These manipulation checks are reported in Table 22, below.

Study	N	Items	Cronbach's Alpha	Expected ingroup performance			Manipulation Comprehension		
				<i>t</i>	df.	<i>p</i>	<i>t</i>	df.	<i>p</i>
Measure pilot 1	23	6	.978	-	-	-	-	-	-
Measure pilot 2	36	8	.956	-	-	-	-	-	-
Identifiability Study 1	54	-	-	-	-	-	-	-	-
Identifiability Study 2	218	-	-	-	-	-	-	-	-
Identifiability Study 3	53	4*	.541	-	-	-	-1.939	51	.058
Identifiability Study 4	300	8	.806	14.607	277	<.001	-3.616	277	<.001 <sup>+</sup>
Conflict Study 1	102	6	.816	-1.506	100	.135	-9.169	100	<.001
Conflict Study 2	122	8	.723	-10.763	70	.000	1.875	70	.645
Conflict Study 3	138	8	.892	2.8	121	.006	.342	121	.733
Permeability Study 1	100	8	.769	-9.959	98	<.001	4.604	98	<.001
Permeability Study 2	177	8	.834	-3.907	164	<.001 <sup>#</sup>	-3.8	164	<.005 <sup>##</sup>
Permeability Study 3	151	8	.684	-17.804	149	<.001	-2.027	149	.044
Achievement Motivation Study 1	93	8	.887	-11.603	91	<.001	.220	91	.826

*Note.* \*Dropping item 1 from a 5 item scale. Remaining items not suitable for calculating expected ingroup performance

<sup>+</sup>Levene's test for violation of the assumption of homogeneity of variance was significant,  $F = 5.006, p = .026$ .

<sup>#</sup> Levene's test for violation of the assumption of homogeneity of variance was significant,  $F = 15.220, p < .005$ .

<sup>##</sup> Levene's test for violation of the assumption of homogeneity of variance was significant,  $F = 20.138, p < .005$ .

*Table 22: Reliability of status manipulation comprehension check across studies and significance of difference in manipulation check by status for all studies containing both the manipulation and the measures*

As expected, there were highly significant differences between low and high status conditions in expected ingroup performance for all but one study (Conflict Study 1) which did not fall far off the mark ( $p = .135$ ). This confirms that participants in the two conditions generally understood and were willing to report that their groups were expected to perform

differently on the APM task. However – surprisingly – in six of the nine studies there were also significant effects on manipulation comprehension by status. Since the text of the manipulations were identical for high and low status conditions (except for the swapping of relevant category labels), it is not immediately obvious why participants in low and high status conditions reported significantly different levels of comprehension of the stereotyped expectations of their group's task performance.

These results, then, suggest that the manipulations were successful in communicating expectations of stereotypical performance differences between high and low status groups. Therefore, we cannot conclude that the failure to detect differences in APM performance by status is simply due to poor manipulations. Although it is possible that the participants comprehended the manipulations but did not believe the stereotypes, the STB literature is clear that belief in a stereotype is not necessary for performance shifts to occur (Huguet & Régner, 2009; Marx et al., 1999; Steele et al., 2002). Therefore, there is mounting evidence here that detectable performance shifts did not occur despite the clear invocation of performance-related stereotypes, which indicates a failure of the simple STB model in these specific circumstances.

### **5.3. Shortcomings with conventional manipulation checks**

Manipulation checks are *de rigueur* in experimental social psychology. In general they are used in two ways: to filter out participants who fail to understand the manipulations or apply themselves to the requirements of the experimental tasks (eg. Smith, 2006) or as validity checks (eg. Sawyer, Lynch, & Brinberg, 1995).

The late Bruce Faulds used to argue that studies of memory and learning using lab rats were fatally confounded by the (un)natural selection pressures on the breeding pool caused by twitchy undergraduate students dropping, flinging or squashing the unfortunate creatures when bitten. The rats that survived to breed again, therefore, were those that quickly

learnt not to bite their reluctant masters. If this were true, it would undoubtedly influence the outcome of learning trials since the most, stubborn, subversive and rebellious ‘participants’ (and their descendants) would have been removed from the sampling frame.

This anecdote has a direct analogue in the common use of the manipulation check as a means of excluding participants from analysis if they “fail” the manipulation check (eg. Steele, 2003b). While this is very common practice it makes an unacknowledged assumption: that the qualities or processes related to “failure” on the manipulation check are independent of the constructs and models being investigated. However, in practice the manipulation check is often closely related to the construct of interest and so this assumption is not always as reasonable as it appears.

This methodological practice of dichotomizing response on the manipulation check and simply excluding participants who behave in unexpected ways will necessarily select out participants of particular kinds and remove particular genres of social behavior from the analytic frame. This is exactly the point of the practice: to remove inattentive or non-compliant participants from the analysis that would introduce statistical noise.

Of course, unexpected results on the manipulation check could result from many factors: fatigue, boredom, failure to understand, low investment, too much knowledge of experimental methods, disagreement, and many other possibilities that will differ according to the specific features of the study, the sample and the experimental context. Therefore, *not* distinguishing between participants on the basis of a manipulation check would also be foolish.

However, scores on a manipulation check may be understandable as psychological or behavioural orientations to the experimental context and manipulations. It is therefore possible to include them in statistical models and thereby differentiate between the performance of “good” and “bad” subjects, but to do so in much more nuanced and useful

ways than the traditional practice of simply dropping those who “fail the manipulation check.” Including the manipulation check in the statistical model, if it is theoretically meaningful to do so, will distinguish between the performance of participants who scored differently on the manipulation check<sup>11</sup>

#### 5.4. Resistance

The challenge here is making sense of two pieces of information: the condition to which the participant was assigned, and their answer to the manipulation check, which assesses the extent to which they were willing or able to repeat the negative or positive stereotypes to which they were exposed in the manipulation. Critically, all eight questions in the manipulation check started with the stem “based on the description above” and the questions are unambiguously assessing basic comprehension of the instructions – a task that the student samples should be well-drilled in.

When the manipulation clearly states that humanities students are likely to excel in the Raven’s Advanced Progressive Matrices, it should be a relatively simple task to immediately report which group is expected to perform best “according to the description above.” Since the manipulation is quite extreme, items such as “Based on the description above, the best performing participant will probably be a [boost group member]” should be answered on the extremes (that is, at the poles of the Likert response scale) to accurately reflect the descriptions in the manipulation. Therefore, a participant who carefully reads the manipulation and accurately reports it *should* have an extreme score on the expected ingroup performance scale.

Therefore, if participants were neutrally and passively reporting their comprehension, any variation from the extreme would be a result of participants not reading the instructions

---

<sup>11</sup> But, of course, at the risk of including genuine noise in the analysis.

carefully, forgetting them, or getting confused with the wording of the eight items in the manipulation check. However the generally high reliability of the scale is acceptable, suggesting that this noise is within acceptable limits. Also, true random allocation to conditions should distribute these errors randomly across status conditions, dispatching them to the residual term in the statistical model.

But as I have discussed previously, participants are not passive or neutral reporters of their internal states. As agents they are not just trying to passively understand instructions, but are actively engaging with them as lived experience. This kind of agency has long been acknowledged in psychological research, for example, as “social desirability bias” in which participants inflate or deflate their responses to present themselves in socially desirable ways. In social desirability literature at least, this “bias” is usually understood either as “self-deceptive enhancement” in which participants are unaware that their answers are not “true” reflections of their internal states or as “impression management” in which they have some awareness that they are bending the truth a little (eg. Crowne & Marlowe, 1960; Paulhus, Harms, Bruce, & Lysy, 2003). Of course, this means that the notion that participants answer questionnaires in ways that are sensitive to their social position in the research environment and the way that their answers may be read by others has been acknowledged in experimental psychology for over fifty years, although the full import of this knowledge on experimental results outside of the narrow constraints of social desirability bias is rarely considered.

However, using social desirability as an example runs the risk of trivializing this insight because social desirability is generally treated as means of removing or correcting the responses of unsuitable (dishonest or incapable) participants rather than as an important part of the psychological experience of the phenomenon of interest. It also implies that the participants who, consciously or unconsciously, are concerned with their social position in the research setting are defective in some way, perhaps narcissistic, defensive or delusional.

On the contrary, there are good arguments that resisting or challenging instructions can be a powerful means of resistance for the powerless and that, in social psychology, this resistance should be examined rather than discarded. For example, Lemaine and his colleagues asked groups of boys to compete by building a hut in the forest over the course of a few days (described in Lemaine, 1974). However, in each study one group of boys “chosen at random ... did not receive an important element for the execution of the task” (p. 35). Therefore, although the prize would be awarded to the group with the best hut, one group was doomed to inferiority because of the social conditions of the research environment in which they found themselves. A common strategy employed by the low status group was to “introduce another criterion of judgement than that or those which were implicitly contained in the instructions” (p. 35), such as focusing their efforts on making an attractive garden for the hut, and then to convince the other participants and the judges (the researchers) that it was allowable as a basis for comparison. In these studies, the subjugated boys’ resistance to strictly adhering to the instructions and their ability to reinterpret and dispute them was a critical part of the group’s social-psychological response to the demands of the experimental context.

However, even *passively* failing to understand or “properly” carry out instructions can be a means of resistance for subjugated groups. For example, servants may “[fight] back on the job by working at their own pace, misunderstanding instructions they [do] not wish to obey ... and occasionally by spitting in Madame’s soup” (Klein, 1999 p. 194). Similarly, students who perform poorly at school may do so as an act of rebellion against the system (Willis, 1981).

Therefore, it is likely that asking participants eight times to actively repeat the negative or positive stereotype attached to their group is a qualitatively different experience for participants assigned to low and high status groups. It is tempting to label these varied

responses as “bias” akin to social desirability bias. However, on the contrary I am making the argument that the social positioning of the participant in the research environment is a critical element of the stereotype threat experience, and the active, meaningful, and strategic elements of participants’ responses to the manipulation check are critical to understanding STB. This is particularly important since stereotype threat research should be – by definition – critically interested in the embedded identity of participants in the performance context. In this view, failure to understand simple instructions may be a means of figuratively spitting in the researcher’s soup; and refusal to accurately echo the description of one’s inferior status may be a means of resisting one’s subjugation.

Of course, there are bound to be deviations from accurate comprehension due to inattention, poor encoding, forgetting and so on, but these regular sources of variation are likely to be evenly distributed across status conditions. These conventional explanations cannot explain the significant difference in *comprehension* between participants assigned to threat and boost conditions in most studies in this programme of research. Instead, systematic between-groups differences in comprehension between participants assigned to threat and boost conditions are understood here as resistance to the experimental manipulations .

#### **5.4.1. Calculating a resistance metric**

Knowing that participants in low and high status conditions, on average, perform differently on a comprehension manipulation is not in itself theoretically or empirically powerful. However, after grappling with this problem for some time it became clear that the items in the status manipulation check scale can be rearranged to represent resistance much more directly.

Assuming that the manipulation is clear and unambiguous and that participants are reasonably competent, the expected ingroup performance scale *should* be low for participants assigned to the low-status (threat) condition. The extent that their estimate of expected



ingroup performance is *higher* than the lowest extreme may represent their resistance to the manipulation. Conversely, participants in the high-status (boost) condition *should* have high levels of expected ingroup performance and the extent to which their estimate of own-group performance is *lower* than the upper extreme represents resistance. Therefore resistance to the status manipulation may be calculated as follows:

For low status participants:                      status resistance = 0 + expected ingroup performance

For high status participants:                    status resistance = 8 - expected ingroup performance

(Where expected ingroup performance is measured on a seven-point scale)

One interesting feature of this measure of resistance is that it measures positive and negative resistance on the same scale. In other words, the experience of defending against a negative stereotype and that of being reluctant to accept a positive stereotype can be assigned the same value. However, these experiences are likely to be psychologically different and if they have any relationship to performance at all, it is likely that the effects might be different. For example, someone resisting a positive stereotype may not feel good enough to live up to it, but someone resisting a negative stereotype may feel that the stereotype is underestimating their true potential. Therefore, if the resistance variable has any significant effect on performance at all it is likely to do so as an interaction between manipulated status and measured resistance rather than as a direct effect.

#### ***5.4.1.1. Exploring the validity of the resistance metric***

While it makes sense that the extent to which a participant contradicts the stereotypes presented to them represents some kind of resistance to the manipulations, common-sense is a poor validity check. To explore the validity of the resistance metric, correlations were examined between resistance and other social identity and motivational measures available across studies.

As I have already argued, it is likely that resistance will be a psychologically different experience or state for those assigned to low and high status conditions, so correlations between resistance and any social identity variable that appeared in more than a third of the studies were run on subsets of low and high status conditions separately and then also on all the data combined, as displayed in Table 23.

In different circumstances a factor analysis or regression model may have been more appropriate, but the fact that different measures were used in different studies means that a single model including all of the available social identity variables results in a sample size of 0 due to casewise deletion. The combined correlations were only interpreted when the subset results were also significant and in the same direction, since when subscale correlations were in opposite directions it indicated that the combined  $r$  value was obscuring important interactions.

Table 23: Correlations between resistance and social identity variables

	FILTERED FOR STATUS			Number of datasets
	LOW (THREAT)	HIGH (BOOST)	COMBINED	
Group identifiability	-.187**	.027	-.096**	6
Intergroup closeness	-.176**	-.039	-.097*	4
Achievement motivation	.258**	-.113	.104*	5
Group differentiation	.301**	.111	.207**	3
Stability	.165**	-.105*	-.011	6
Ingroup identification	.017	-.100*	-.066*	8
Permeability	.035	-.205**	-.052	7
Effort	-.047	-.311**	-.167**	7
Social desirability	-.055	-.153**	-.104**	7
Legitimacy	-.040	-.200**	-.129**	6
Category salience	-.045	-.274**	-.159**	5
Ingroup closeness	.095	.065	.093**	4

*Note.* Combined correlations only considered interpretable if the correlations for low and high status subsets are both in the same direction and both significant.

\* Significant at the 0.05 level; \*\* Significant at the 0.01 level

The most obvious trend in the pattern of correlations is that, as expected, the patterns of correlation are clearly different for threat and boost conditions, confirming that resistance is a qualitatively different experience for participants assigned to low and high status conditions.

Specifically, for participants in the low status condition, high resistance was associated with higher reported group differentiation and decreased perceived closeness between the ingroup and outgroup (intergroup closeness). They were also more likely to report feeling like representatives of their group in the testing context (group identifiability) and that the group status hierarchy was relatively stable. Low status participants with high

resistance were also more likely to have higher levels of individual achievement motivation than those exhibiting low resistance. Although this pattern is difficult to interpret precisely, it seems that participants who resisted agreed that the groups were different and that these differences were stable, but were attempting to differentiate themselves from the negatively stereotyped group (suggested by the negative correlation between resistance and group identifiability for this subset). In other words, by resisting the status manipulation, participants in the low status group seem to be taking the position that “I agree that my group is bad, but I am better than my group.”

Consistent with the notion of resistance as a willingness to challenge the experimental manipulations, high resistance participants in the high status condition were likely to score lower on the social desirability scale than those exhibiting low resistance. They were also more likely to challenge the notion that their group was advantaged in the task (legitimacy) and more likely to report concern that the group as a whole may not be able to maintain its superiority (stability). However, they also seemed to be less likely to consider themselves prototypical members of the high status group. Specifically, those with higher resistance were less likely to report a sense of belonging to the high status group (ingroup identification); and more concerned that their membership of the group was not secure (permeability).

The remaining relationships are more difficult to interpret. Although high-status participants exhibiting resistance to their positive categorization were much more likely to report lower effort, it is impossible to know whether they were really disengaged from the task or whether this is an example of self-handicapping, that is, giving reasons to explain possible poor performance that deflects possible attributions for poor performance away from important, internal, stable traits or abilities such as intelligence or ability (eg. Berglas & Jones, 1978; Keller, 2002; Stone, 2002). The self-handicapping explanation seems to be supported by the strong tendency for participants who resisted the high-status manipulation to

report substantially less awareness of their categorization while they were completing the performance task (category salience), since reporting less awareness that they were categorized as members of the high performing group would reduce their culpability for (potentially) failing to live up to the expectations related to their group membership.

Given the pattern of results above, it is puzzling that there is a positive overall correlation between resistance and ingroup closeness. Although it is weak and does not achieve significance for either subset in isolation, it does suggest that participants with high resistance were marginally more invested in the fortunes of their group. Perhaps this explains why they were more likely to engage in active identity work in relation to the experimental manipulations in comparison to their low resistance counterparts.

Despite some puzzling results, the pattern of results is consistent with the conclusion that the resistance metric is indeed measuring some kind of defiance to the experimental manipulations.

#### **5.4.2. Exploring the statistical validity of the resistance metric**

The procedure of combining reverse-coded scores for participants assigned to one condition and raw scores for participants in the other is methodologically equivalent to flipping the direction of item wording for participants in different conditions to ensure that items encode constructs in the same direction across groups. However, since both were included in the final model and they are clearly not entirely independent, it is potentially possible that the interaction effects observed between the manipulated and manipulation-resistance variables were artifacts of the calculation procedure. This was tested using a bootstrapping procedure.

First a new dataset of 100 cases was generated with equal numbers of participants in the cells of a 2x2 design simulating the design of most studies in the programme of research. For the sake of clarity, this description will assume that the first simulated factor represents

manipulated status and the second represents manipulated identifiability. This dataset was then populated with random scores generated from a normal distribution for simulated APM performance, perceived identifiability, and item scores for the eight manipulation comprehension items. The mean and standard deviation for each randomly generated variable or item was randomly generated from a random distribution with the mean and standard deviation equivalent to that of that scale or item in the real data across all studies.

Once random item scores had been generated, the simulated scales were computed from items, and the resistance scores were calculated from the scale-scores and manipulations using the same scripts that were used to complete these calculations on the real data. Any potential artifacts introduced by the data-processing and calculation procedures would therefore have been present in the simulated data.

Second, a bootstrapping procedure was followed in which (1) a random sample of 100 cases<sup>12</sup> was drawn with replacement and (2) a fully saturated linear regression model was fitted to each bootstrapped sample with “status”, “status-resistance”, “conflict” and “conflict-resistance” on “APM performance.” A bootstrap with 10,000 replications was then performed with the `boot()` procedure in R (R Development Core Team, 2010) and the composite  $F$  for the bootstrapped regression models was recorded. To ensure that the results were not skewed by random properties of the original small randomly generated sample, this entire procedure was repeated five times on five separate randomly generated datasets. The bootstrapped 95% confidence intervals for the  $F$  statistic and the associated  $p$  values for the upper limit  $F$  are recorded in Table 24<sup>13</sup>.

---

<sup>12</sup> Small bootstrap samples were used to maximize the potential of the procedure to identify bias or artifacts.

<sup>13</sup> Since the model  $F$  value for a regression model cannot be less than 0 the lower limit can simply be ignored.

In this procedure, the highest bootstrapped 95% confidence interval for the model  $F$  value was 1.303, which (at 15 and 83 degrees of freedom) is associated with a  $p$  value of 0.219. However, by definition, the  $p$  value associated with the 95% confidence interval for  $F$  on random data (that is, under the null hypothesis) *should* be .05. Since the  $p$  values derived in this confirmation procedure were consistently higher (and therefore further from spurious significance) than expected, we can conclude that if any artifacts are being introduced by the novel coding of the resistance variables they are not likely to be increasing the chance of type I error for the saturated resistance model.

	Upper 95% confidence interval for $F_{15, 83}$	Upper limit $p$
Replication 1	0.362	0.985
Replication 2	1.026	0.437
Replication 3	0.573	0.888
Replication 4	1.211	0.280
Replication 5	1.303	0.219
Average	0.895	0.572

*Table 24: 95% confidence intervals for  $F$  and associated  $p$  values for bootstrapped simulation of model including resistance variables with random data*

### 5.4.3. Exploring a model including status and resistance

When resistance was included with status in a regression model using all the data combined, along with the status\*resistance interaction, a significant model emerged,  $F_{3, 1082} = 2.896$ ,  $p = .034$ ,  $r^2 = .008$ . The main effect for status was marginally significant,  $t = 1.769$ ,  $p = .093$ ,  $b = .055$ , and the status\*resistance interaction,  $t = -2.244$ ,  $p = .025$ ,  $b = -.021$ , was significant as hypothesized. Resistance was not directly significant at all,  $t = 1.679$ ,  $p = .950$ , which is consistent with the argument that the psychological experience of resistance is different for participants in boost and threat conditions and is therefore likely

to have different effects on performance. Critically, the interaction between status and resistance was more clearly significant than the direct effect of status.

**5.4.3.1. *Exploring the effect of status and status\*resistance on performance in more detail***

Since the status by resistance interaction was significant, the effect of status (ie. the effect of STB) could only be interpreted in relation to resistance. The status\*status-resistance interaction plot (Figure 5) revealed that participants with low resistance were likely to exhibit small traditional stereotype threat effects, with those in the boost condition performing better than those in the threat condition. For participants exhibiting high resistance there was a much larger difference in APM performance between boost and threat conditions, but this difference was in the opposite effect to that predicted by STB, with participants in the boost group performing much *lower* than participants with similar levels of resistance in the threat group.



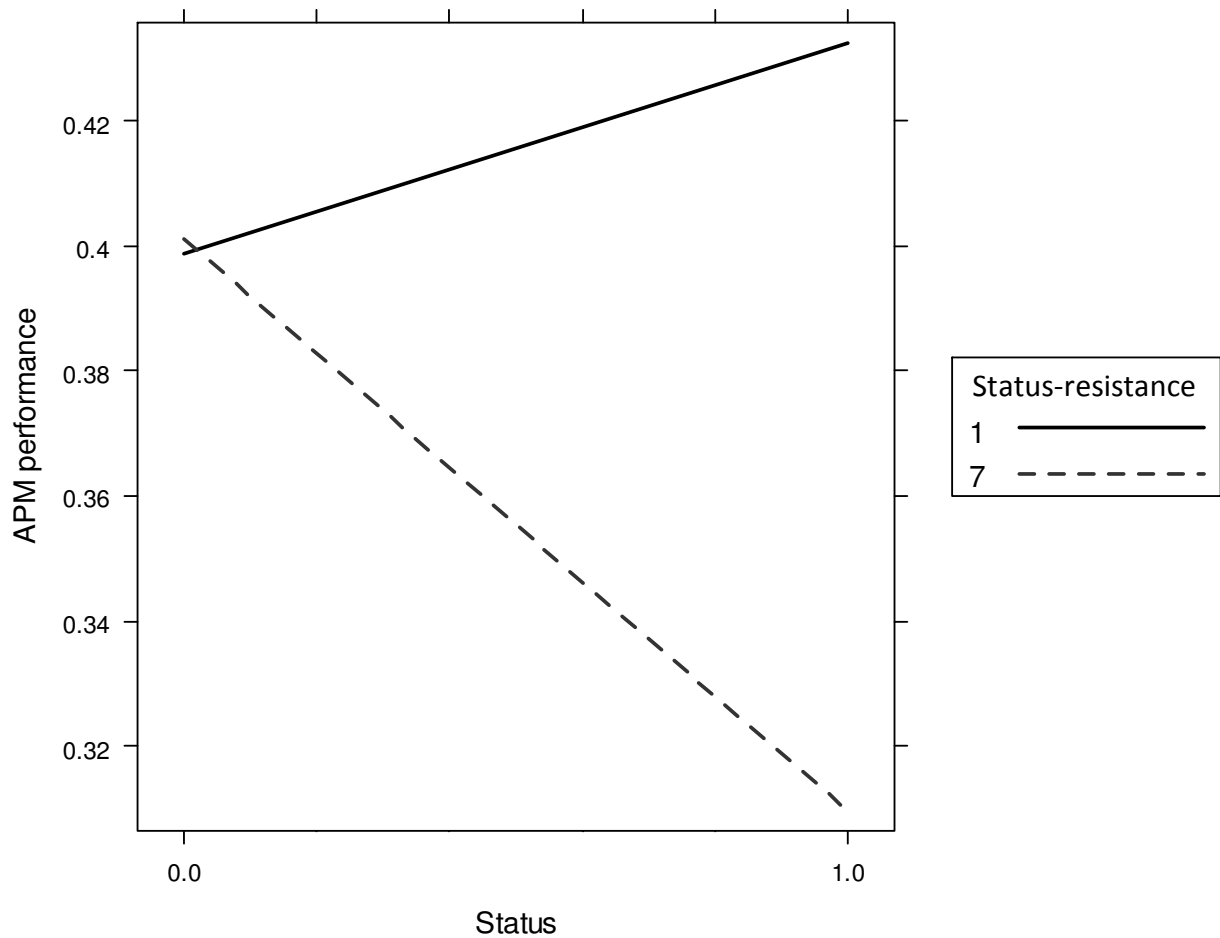


Figure 5: Interaction plot for status\*resistance on APM performance for the combined data

#### 5.4.3.2. Exploring the effect of sex on the status-resistance relationship

Although there are many advantages to using the APM as a dependent measure in the design developed for this programme of studies, one clear disadvantage is that it has been shown to consistently favour males (Abad, Colom, Rebollo, & Escorial, 2003; Colom, Escorial, & Rebollo, 2004; Lynn & Irwing, 2004). Additionally, whether due to social, environmental or genetic factors, many studies have shown differences between males and females on dimensions that have relevance to the present study, such as in personality traits such as agreeableness and conscientiousness (closely related to social desirability), and

values such as power, achievement and self-direction (eg. Gerber, 2009; Schmitt, Realo, Voracek, & Allik, 2008; Schwartz & Rubel, 2005). For these reasons it was not unlikely that participants' responses to the experimental manipulations might differ by sex.

Therefore a second regression analysis was run with status, resistance, sex and all two-way interactions on APM performance. The model itself was highly significant,  $F_{6,1071} = 4.741, p < .001$ . The main effect for sex was highly significant,  $t = 3.418, p = 0.001, b = .129$ , congruent with the results cited above suggesting that men generally outperform women on the APM by a small margin. The main effect for status achieved full significance in this model,  $t = 2.505, p = .012, b = .087$ , as did the status\*resistance interaction,  $t = -2.598, p = .009, b = -.024$ . However, the resistance\*sex interaction was significant,  $t = -2.084, p = .037, b = -.019$ , and the status\*sex interaction closely approached significance,  $t = -1.892, p = .059, b = -.047$ , which suggests that the effects of status and resistance were different for female and male participants.

Although it would be preferable to explore these effects by decomposing the three-way interaction between status, resistance and sex, this term was not significant when it was added to the model. Therefore, to explore these patterns further, the model regressing status + resistance + status\*resistance on APM was repeated separately for males and females.

Although the overall model for males was significant,  $F_{3,497} = 1.453, p = .011$ , none of the effects had significant coefficients although the interaction term was marginally significant,  $t = -1.673, p = .095, b = -.025$ . On the other hand, the overall model for females did *not* achieve significance,  $F_{3,573} = 1.453, p = .227$  despite the terms for status,  $t = 2.016, p = .044, b = .086$ , and the status\*resistance interaction,  $t = -1.989, p = .047, b = -.024$ , both achieving or closely approaching significance. The interaction for males and females has been plotted from these segregated models in Figure 6.

These plots reveal that male participants performed better than female participants overall. Additionally, for male participants there was little systematic difference between performance in low and high status conditions when resistance was low. The only cell contributing to the marginally significant interaction for males was for participants exhibiting high resistance to the boost condition who performed worse than participants in the other three cells. Certainly, for men, the lowest performing group were those in the boost condition with high resistance.

Females, in contrast, exhibited performance much more congruent with the theoretical framework of resistance to STB developed above, such that when resistance was high the STB effect was completely reversed. Specifically, women exhibiting high resistance did *better* under threat and *worse* under boost while women with low resistance performed as predicted by STB.

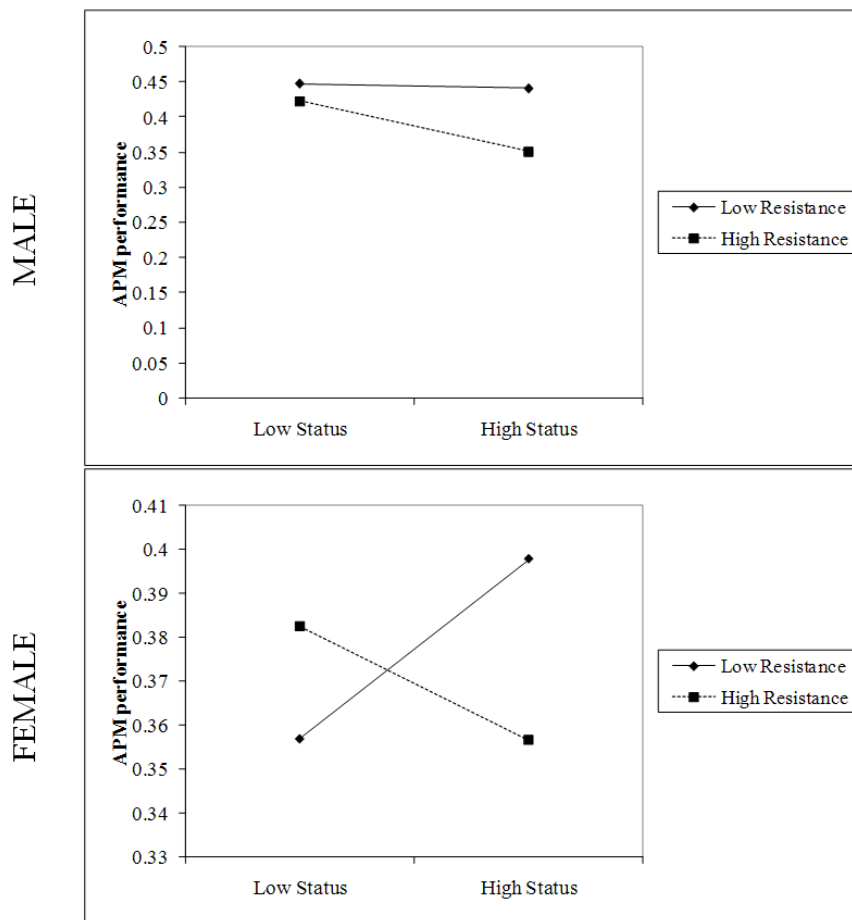


Figure 6: Status\*resistance interaction on APM performance plotted separately for females and males

### 5.5. Discussion of resistance

There are clearly problems with this analysis, particularly in the post-hoc nature of the development of the metric and the use of less than optimal statistical procedures (such as correlations instead of multiple regression) to accommodate inconsistencies between studies in the dataset. However, the resistance measure has high face validity and meaningful convergent and divergent validity with the available identity and motivational measures. The regression model for status and resistance on APM performance is both consistent with the way that “resistance” has been conceptualized and with a social identity understanding of stereotype threat.

Despite its problems, this analysis is generally supportive of the notion that resistance may be an important feature of STB effects that deserves further investigation. This relationship has been masked in previous studies by the practice of using measures of ‘comprehension’ as manipulation checks and not including them in models exploring stereotype-performance effects. This practice can result in researchers simply dropping those who “fail” a manipulation check and thereby losing the opportunity for understanding how negative stereotypes influence the performance of those who reject them.

While the metric that has been derived here is post-hoc and probably open to substantial improvement, the notion of resistance to stereotype threat is theoretically important and opens up new possibilities for exploring how, and under what conditions, negative stereotypes result in reduced performance and positive stereotypes result in performance boosts. More excitingly, this theoretical innovation may offer a means to explore the conditions in which practical resistance to STB is possible.

Although the effects were not strong, it must also be acknowledged that the minor differences in manipulation across studies and – especially – the different combinations of SIT variables that were manipulated alongside status (that were ignored in this analysis) must introduce a great deal of statistical noise. It is possible that a cleaner design would have resulted in stronger effects. Of course, it is also possible – and even likely – that there are mediators and moderators of the STB-resistance-performance relationship in addition to sex that have not been considered in this overall analysis. In fact, the social identity model makes a strong case that individuals will use whatever identity resources are available to them in the performance context to make sense of their particular experience. Therefore, it is very likely that other identities (such as race, educational background, affluence, or any other potential identity strength or weakness in the context) may enter into participants’ experience of the STB context which is, of course, congruent with Steele’s (2010) notion of stereotype threat as

an identity *contingency*. However, these results suggest that the layering of identities in the context and the agency of the target make the experience more complex than Steele suggests.

The analysis now turns to exploring the impact of the social identity variables of identifiability, permeability and conflict on STB effects. Each of these variables was explored with at least three studies, and a chapter will be devoted to exploring the impact of each variable.

## **Chapter 6. Results II: Identifiability**

Four studies explored the effects of identifiability on STB. The first two were exploratory and were undertaken during the development of the general experimental paradigm and so did not use the full set of manipulations or measures. Therefore only the third (to some extent) and fourth identifiability studies allowed the calculation of the resistance metric. However, the results are interesting and deserve some attention because they do shed some light on the effect of identifiability and anonymity on STB effects.

### **6.1. Identifiability Study 1**

#### **6.1.1. Research question**

This study was the first in the programme of research and explored whether identifiability (the extent to which participants felt individually identifiable or anonymous in the experimental situation) impacted on their performance of the Raven's Advanced Progressive Matrices (APM) under stereotype threat or boost (STB) conditions. It was predicted that the main effect for threat would be significant, as would the interaction between threat and identifiability. It was not possible to calculate resistance to the status manipulation or to include it in the analysis because the status manipulation check had not yet been developed.

#### **6.1.2. Sample**

Students registered for Arts degrees at the University of St Andrews were sampled. Some students registered for Commerce degrees participated, but were dropped from the final sample, resulting in a final sample size of 54. Of these: 12 reported their sex as male and 42 as female; 40 were first-language English speakers and 14 reported a different mother-tongue. All participants received a cash incentive of £4.

### 6.1.3. Design and procedure

The study employed a 2 (stereotype threat or boost) by 2 (identifiable or anonymous) factorial design. Set 1 of the Raven's Advanced Progressive Matrices (APM) was used as a dependent measure. Although the APM is generally used as an intelligence test, in this context between-group differences in performance would indicate STB effects.

#### 6.1.3.1. *Identifiability manipulations:*

Participants in the high identifiability condition were asked to write their name on every page of the materials and were told that their performance would be video-recorded with their face visible in the footage. Participants in the low identifiability condition were asked to write a code only known to themselves on each page of the materials and were told that their performance would be video-recorded with only their hands visible in the footage.

#### 6.1.3.2. *STB manipulation:*

All participants were told that the APM is a measure of 'g,' which predicts post-degree performance in the workplace. Participants in the threat condition were told that the study was exploring "real world market-ability." The banner logo on each page displayed the equation " $e = mc^2$ ." They were informed that 'g,' is associated with logical thinking, accuracy and the ability to find "clear, simple solutions to complex problems." Reasons were then given why science degrees produce graduates higher in 'g' than Arts degrees and, therefore, why science graduates are more likely to experience post-degree real-world success. Students in the stereotype boost condition were given materials with a banner logo on each page displaying a pile of books and the caption "post-degree adapt-ability." They were told that people high in 'g' are flexible thinkers who are able to arrive at "best-case" solutions and are more concerned with "the big-picture than precision." Again, reasons were then given why arts degrees produce graduates higher in 'g' than science degrees and, therefore, why arts graduates are more likely to experience post-degree real-world success.



#### 6.1.4. Key findings

Neither the predicted main effect for status,  $F_{3, 50} = 2.725, p = .105$ , nor the status\*identifiability interaction effect,  $F_{3, 50} = .890, p = .350$ , were significant. However, accounting for gender and language in a 2x2x2x2 ANOVA with STB (threat or boost), identifiability (anonymous or identifiable), sex (male or female) and first language (English or foreign) as between-subjects factors resulted in a highly significant model,  $F_{15,38} = 2.585, p = .009, \eta_p^2 = .505$  with adjusted  $r^2 = .310$ .

Obviously with a sample size of 54, a 2x2x2x2 model risks overfitting the data. Since removing sex or first language resulted in substantial reductions in model fit ( $\Delta r^2 = .186$  and  $.359$  respectively) it was explored whether these variables could be combined.

A clear theoretical similarity between these categories is that – in the context of the University of St Andrews – both categories represent minority status of different forms.

Therefore the conditions can be recoded as follows:

1. Non-minority: Male and English
2. Single-minority: Male and non-English; Female and English
3. Double-minority: = Female and non-English

When this was entered into a 2x2x3 ANOVA with STB (threat or boost), identifiability (anonymous or identifiable) and minority status (non-, single- or triple-minority) as between-subjects variables the model was highly significant,  $F_{11,42} = 3.540, p = .001, \eta_p^2 = .481, \text{ adjusted } r^2 = .345$ , a drop in  $r^2$  of only 0.024 from the model including sex and language as separate factors. Although this is still more factors than is comfortable for a sample of this size, it is a slight improvement.

The main effects for minority-status,  $F_{2, 51} = 3.656, p = .018, \eta_p^2 = .174$ , and identifiability,  $F_{1, 52} = 5.422, p = .025, \eta_p^2 = .114$  were both significant, but the main effect for STB was not,  $F_{1, 52} = 1.113, p = .298$ . The influence of identifiability was to generally

increase APM by a small margin (mean difference between high and low identifiability conditions = .812,  $p = .025$ , 95%CI for mean difference:  $.108 \leq d \leq 1.517$ ).

The interaction between minority status and identifiability was marginally significant,  $F_{11,42} = 2.707$ ,  $p = .078$ ,  $\eta_p^2 = .114$ , as was the interaction between identifiability and threat,  $F_{11,42} = 3.635$ ,  $p = .063$ ,  $\eta_p^2 = .080$ . However, the three-way interaction between minority status, identifiability and threat was highly significant,  $F_{11,42} = 7.211$ ,  $p = .001$ ,  $\eta_p^2 = .294$ , meaning that the simple effects and lower-order interactions should be interpreted with care, if at all. The interaction plots for this three-way interaction are displayed in Figure 7. Since including confidence intervals would have made the interaction plot unreadable, these have been displayed separately in Figure 8.

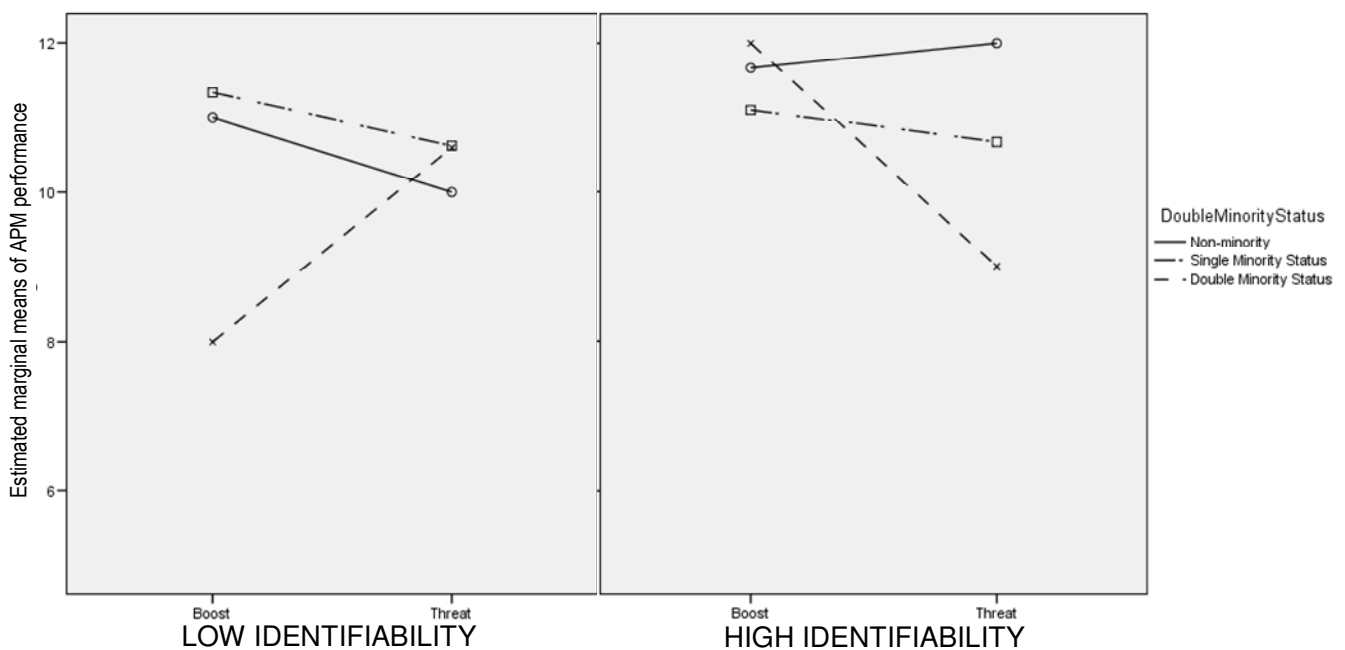


Figure 7: Interaction plots for status\*identifiability\*minority status interaction

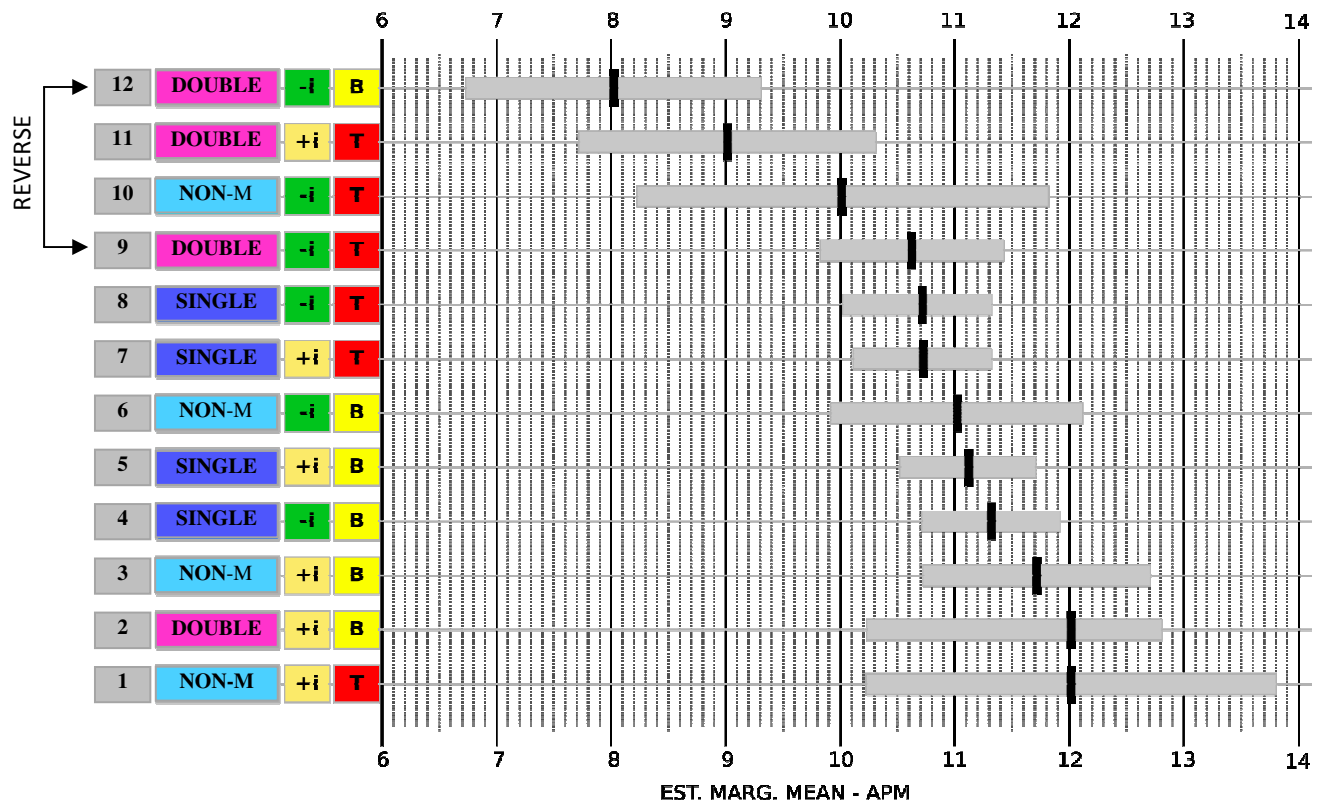


Figure 8: Estimated marginal means and 95% confidence intervals for three-way interaction between minority status, identifiability and STB (key: single, double or triple = minority status; -i = low identifiability; +i = high identifiability; B = stereotype boost)

The most obvious finding is that – contrary to STB predictions – the lowest performing group were in the boost condition (12) and the highest performers were under threat (1). However, the lowest performers were double minority participants (that is, female second-language speakers) under boost and the highest performers were male first-language speakers under threat. It is possible that sex differences in APM performance were responsible for this pattern. But this explanation is undermined by the fact that the equivalent group of low identifiability double minority participants under threat (9) performed substantially and significantly better than their counterparts under boost (12) and the tied highest performing group (2) were double minority members (in the high identifiability boost condition).

It is apparent, looking at the interaction plots (Figure 7), that increased identifiability was generally associated with increased performance. However, the most dramatic shift from low to high identifiability conditions was the reversal in the effect of the status manipulation for the double-minority participants. Specifically, double-minority participants in the high identifiability condition performed as predicted by STB, with those under boost substantially outperforming those under threat (and those under boost performing at the ceiling of the measure), but the double-minority participants in the low identifiability condition exhibited a reversal of the STB effect, with those under threat substantially out-performing those under boost. It is very unlikely that this reversal was driven by “choking under pressure” (eg. Cheryan & Bodenhausen, 2000) since the only difference between these groups was identifiability. Since increased identifiability presumably increases pressure, and the high-identifiability group were the top performers, it is more likely that the drop in performance was driven by reduced investment than increased pressure.

#### **6.1.5. Discussion of Identifiability Study 1**

When language and sex were coded as double minority status, the highest performing group was actually in a threat condition (non-minority, high identifiability). The lowest performing group was in the boost condition (double minority, low identifiability). These results suggest that the status manipulation did impact on performance, but not in a deterministic way, and only in conjunction with other key identity variables, such as sex and language.

The effect of identifiability was to subtly increase performance in most conditions. However, for double minority-participants, identifiability was associated with a complete reversal of the STB effect, such that double-minority participants in the low identifiability *boost* condition were the lowest performers overall, but low identifiability *threat* participants performed comparably with participants in the other threat conditions. While double minority

threat participants in the *boost* condition under low identifiability conditions were the *lowest* performers, they were also the (tied) *highest* performers under high identifiability conditions.

Traditionally in psychology, group-based differences in performance are attributed to some genetic or cultural trait common to group members, such as attributing females' lower APM performance to stable visuo-spatial deficits relative to males. In this case such a simple trait-based explanation is untenable, since the most important grouping variable is language, and the APM is generally considered to be relatively language and culture fair (Brouwers et al., 2009). Therefore, we need to look more carefully at the experience of the experiment for double minority students compared to single minority students to understand these patterns of difference.

Fortuitously, a study on acculturation was being completed at St Andrews at the same time as the stereotype threat experiment and several surveys of foreign students in the UK were completed at roughly the same time, so we have a fairly good snapshot of the experiences of the cohort of foreign students at the university at the time of the study. Firstly, in the UK as a whole, there were estimated to be over 300,000 full-time international students attending British Universities of whom 68% were non-EU citizens paying foreign student fees that were, on average, three times higher than local fees (HESA, 2007 in Ramos, 2010). A survey of foreign students in the UK at the time of the study found that non-EU students reported difficulties getting work, much lower proportions had paid jobs than British students (15% compared to 44%) and more of the foreign students with jobs reported working for basic essentials rather than to supplement disposable income. Nearly half (46%) of international students reported having "some financial difficulties" (UNITE, 2006). Therefore, it seems that foreign students in the UK at the time had both higher expenses and fewer ways of supplementing their income than local students.

At St Andrews in particular, foreign students reported problems with acculturation and familiarity with experiences of discrimination. Experiences of discrimination were negatively correlated with self-esteem and foreign students' sense of relative deprivation<sup>14</sup> and hostility in relation to local students tended to increase over the course of their degree as their willingness to participate in the host community decreased (Ramos, 2010).

It is clear then that the experience of foreign and local students at the university of St Andrews was substantially different. Most importantly, the meaning of the £4 cash incentive to local and foreign students is likely to have resulted in the selection of participants with quite different motivations for participation in the two groups: local students, who – according to contemporary research – probably had lower expenses, better access to employment and enough disposable income and were more likely to have participated for altruistic reasons and interest; and foreign students with triple the student fees to pay and less access to employment. As the recruitment poster stated, £4 at the time could buy 25 cans of Tesco Value baked beans and was therefore not an insubstantial sum of money for a foreign student. Given that multiple studies would be running in the St Andrews School of Psychology at any one time, and all offered similar incentives, it is not unlikely that research participation was a viable income stream for many foreign students<sup>15</sup>.

The reversal in performance for double minority students in the boost condition between low and high status conditions is certainly consistent with this explanation: it is possible that when they were anonymous and positively stereotyped they had nothing to prove; they were simply not motivated to invest the effort required to perform well on the APM, and were able to 'slipstream' and maintain an acceptable sense of identity by relying

---

<sup>14</sup> Labelled "outgroup privilege" by Ramos (2010).

<sup>15</sup> And this resonates with my own experience as well, since my wife and I were living on a Commonwealth scholarship of £675 a month at a time when someone on the minimum wage would have earned £5.05 an hour, or £808 a month and local PhD students were receiving a stipend of £1000. We also participated in as many psychology experiments as possible to supplement our income.

on their positive group identity and the fact that they would not be individually accountable for their individual performance. However, when under threat, it seems that the equivalent group (double-minority; low identifiability) were challenged either to defend their group or demonstrate that they were personally not exemplars of the negative stereotype attached to their group and therefore invested more effort and performed much better. Similarly, when identifiable as individuals double-minority participants were exemplary performers under boost, as predicted by STB.

Although this explanation fits the patterns observed in the analysis, it is not possible to verify since effort was not included as a variable in this study, although measures of effort and the extent to which participants felt that they were acting as individuals or representatives of their group were added in consequent studies to further explore this possibility.

Given the small sample size and the post-hoc nature of these interpretations, these results can only open questions for debate rather than make any substantial theoretical claims. However, the finding that layerings of identity can combine with identity-relevant features of the performance context (such as identifiability) to transform the impact of positive or negative stereotypes in a performance context is theoretically important. It forces us to consider that status (in the form of activated negative or positive stereotypes) is not the only constraint on participants' social position in the experimental context; and that some combinations of constraints are more likely to result in passive disengagement while others are likely to galvanize active resistance.

## **6.2. Identifiability Study 2**

### **6.2.1. Research question**

The second study undertaken in the program of research was also exploratory, and undertaken while an overall design strategy was being decided on. It replicated a study by Leyens, Desert, Croizet, & Darcis (2000) that explored the impact of stereotype threat or

boost (STB) for male and female psychology students in affective processing tasks, but this time including an identifiability manipulation as an independent variable. While the design is not comparable to the other studies in this programme of research, the results contribute to the understanding of the impact of identifiability on STB and so will be reported here.

### **6.2.2. Design and sample**

The study employed a 2x2x2 factorial design with sex (male, female), stereotype activation (activated or not activated) and identifiability (anonymous or individually identifiable) as between-subjects factors. When the stereotype that women are superior to men in affective processing tasks was activated, males were exposed to stereotype threat conditions and females to stereotype boost conditions. Participants in the condition in which neither the category nor the stereotype were activated served as the control.

#### **6.2.2.1. Sample**

Two-hundred and eighteen undergraduate psychology students participated in the study. This was a purposive sample, as their identity as male and female psychology students in conjunction with the stereotype that women are better than men at processing affective information was used as the basis for invoking stereotype threat. Consistent with the demographics of psychology majors, 165 participants reported their sex as female and 53 as male. Of these, 103 participants were assigned to the stereotype-activated condition (threat for males or boost for females) and 115 to the control condition.

#### **6.2.2.1. Dependent measure: emotional sensitivity**

The dependent measure was a count of errors of omission and commission in identifying positive and highly affective words from a list that also included words with positive valence that were not affective. For example, participants were asked “Is *cherish* emotional?” and had to answer ‘yes’ or ‘no’. The primary dependent measure was the difference between the number of high emotion words correctly identified as emotional and



the number of low emotion words incorrectly identified as emotional, resulting in a scale in which higher numbers indicate better affective judgement which is naturally corrected for guessing.

A nine item domain identification measure adapted from Smith and White (2001) was also included in the materials ( $\alpha = .840$ ;  $M = 4.971$ ;  $SD = .822$ ). Items included “psychology is one of my best subjects” and “How likely would you be to take a job in a psychology related field?” and were answered on 7-point response scales.

#### **6.2.2.2. *Stereotype threat and boost manipulation***

Participants in the control condition were given materials that explained that the study was about “cognitive processing of verbal information.” Participants in the STB condition were told that the study would test their levels of empathy. To increase the personal relevance and importance of the task they were also (falsely) told that it was a pilot for a selection procedure for psychology postgraduate students. In the stereotype activation condition the materials specifically explained that “male students ... face a tougher task as female students are more empathetic and are better suited to recognizing emotion.” This description invoked stereotype threat for men and boost for women. In order to further emphasize the stereotype, materials for participants in the stereotype-activated condition were lightly scented with women’s perfume (cf. Holland, Hendriks, & Aarts, 2005). Unfortunately no manipulation check was included to test the efficacy of the status manipulation.

#### **6.2.2.3. *Identifiability manipulation***

Students in the high identifiability condition were asked to write down their name and student number on the answer sheet and to sign it. Students in the low identifiability condition were asked to make up and write down a secret code that only they would be able to recognize.

*Identifiability manipulation check*

Participants were asked to report whether they would be anonymous or identifiable as individuals in the study. No participants miscategorized themselves, and therefore a chi-square on assigned and reported identifiability is highly significant,  $\chi^2 = 160$ ,  $df. = 1$ ,  $p < .001$ . However, 58 participants did not answer the question – a puzzlingly high non-response rate. Although there was no overall significant association between non-response and status assignment,  $\chi^2 = 3.779$ ,  $df. = 2$ ,  $p = .151$ , adjusted residuals hint that with a larger sample size it might emerge that participants under threat were more likely to fail to report their identifiability (standardized adjusted residual = -1.9, which approaches significance), which is a trend akin to resistance to the status manipulation discussed in the previous chapter.

**6.2.3. Key findings**

A saturated multiple regression on the number of errors of commission and omission by sex, stereotype activation and identifiability was not significant,  $F_{3,214} = .8338$ ,  $p = .477$ ,  $AIC = 158.184$ . None of the main or interaction effects approached significance. Although, the significance of the model improved when the identifiability manipulation was added to the model, the adjusted model fit did not<sup>16</sup>,  $F_{7,210} = 1.095$ ,  $p = .3676$ ,  $AIC = 160.903$ ,  $\Delta AIC = -2.719$ . Nevertheless, although the overall model was not significant, the sex\*stereotype-activation approached significance,  $t = -1.781$ ,  $p = .076$ ,  $b = -.266$ , (indicating an almost significant STB effect) and the sex\*stereotype-activation\*identifiability interaction was marginally significant,  $t = 1.686$ ,  $p = .093$ ,  $b = .374$ . The model improved substantially when domain identification was included,  $F_{15, 189} = 1.704$ ,  $p = .053$ ,  $AIC = 144.670$ ,  $\Delta AIC = 16.232$ , with the four way sex\*stereotype-activation\*identifiability\*domain-identification

---

<sup>16</sup> Lower AIC represents better model fit corrected for the number of predictors.

interaction achieving marginal significance,  $t = 1.703$ ,  $p = .090$ ,  $b = .079$ . Also significant was the sex\*identifiability\*domain-identification interaction,  $t = -2.251$ ,  $p = .026$ ,  $b = -.087$ , and the sex\*identifiability interaction,  $t = 2.367$ ,  $p = .019$ ,  $b = 3.377$ .

However, despite the fact that this model is the best fit for the data, the four-way interaction cannot be easily interpreted. Refitting the model without the four-way interaction marginally decreased the significance of the model,  $F_{14,190} = 1.602$ ,  $p = .082$ , but improved the model fit very slightly,  $AIC = 144.67$ ,  $\Delta AIC = 1.12$ . Since the theoretical interest here was the impact of identifiability and domain identification on threat and boost, the saturated model with stereotype activation, identifiability and domain identification (centered) on affective word recognition was fitted for males and females separately to simplify interpretation despite the increased risk of familywise error.

This model was not at all significant for males,  $F_{7,43} = .781$ ,  $p = .607$ , and only the stereotype-activation\*identifiability interaction approached significance,  $t = 1.893$ ,  $p = .065$  (see Figure 9). Although the model as a whole was not significant, the marginally significant interaction represents the stereotype threat effect in conjunction with identifiability and, given its theoretical importance (being the two manipulations in the study), has been plotted in Figure 9. Interpreting this interaction was justified because of the significant four-way interactions in the significant model including both males and females (cf. Bedeian & Mossholder, 1994).

Since the low stereotype-activation condition was the control condition and the high stereotype-activation was the threat condition in which the stereotype that men are poor processors of affective information was activated, the expected pattern according to the stereotype threat framework would be for affective word recognition to be significantly lower in the threat condition. This is exactly what is observed for participants in the low identifiability condition. However, in high identifiability conditions in which participants

were individually identifiable, participants scored substantially *better* in the threat condition than the control condition. In other words, for males activating the stereotype that men are less empathetic than women had opposite effects depending on whether they were identifiable or anonymous.

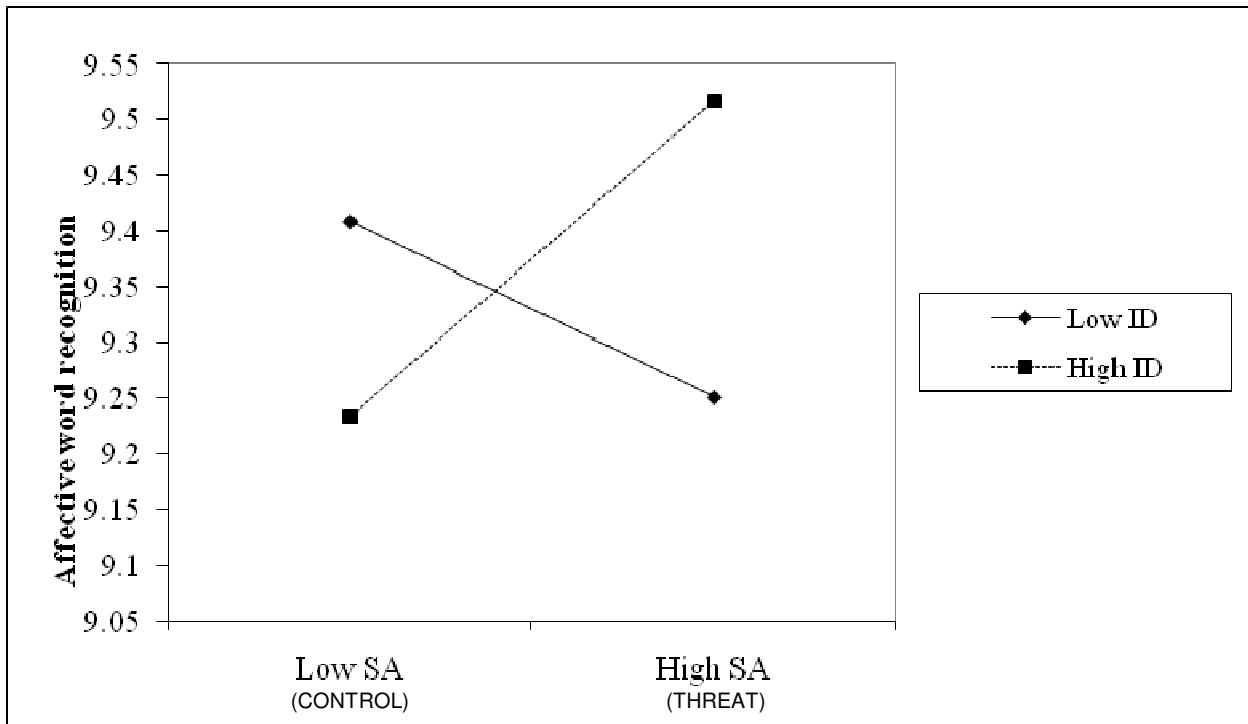


Figure 9: Interaction plot for stereotype activation (SA) and Identifiability (ID) on affective word recognition scores for males only

For females the model was highly significant,  $F_{7,146} = 2.873$ ,  $p = .008$ ,  $r^2 = .121$ , with significant main effects for stereotype-activation (ie. stereotype boost),  $t = 2.428$ ,  $p = .016$ ,  $b = .178$ , identifiability,  $t = 2.003$ ,  $p = .047$ ,  $b = .138$ , and domain-identification,  $t = 2.170$ ,  $p = .032$ ,  $b = .019$ . The two-way stereotype-activation\*identifiability interaction was marginally significant,  $t = -1.780$ ,  $p = .077$ ,  $b = -.197$ , while the identifiability\*domain-identification interaction was highly significant,  $t = -3.303$ ,  $p = .001$ ,  $b = -.041$ .

Since the stereotype-activation\*identifiability interaction was the most theoretically important and the only effect that could be compared to the male subset, it has been graphed

in Figure 10. For the female subset, the high stereotype-activation was the *boost* condition in which the stereotype that women men are skilled processors of affective information was activated. The predicted pattern of performance for females, according to the STB framework, is that they should have performed better under stereotype activation than in the control condition. Indeed, this is exactly the pattern observed in the low identifiability condition. However, the STB effect disappears entirely in the high identifiability condition due to participants in the control condition performing much higher than their low-identifiability counterparts – and, indeed, at the same level as those participants in the boost condition.

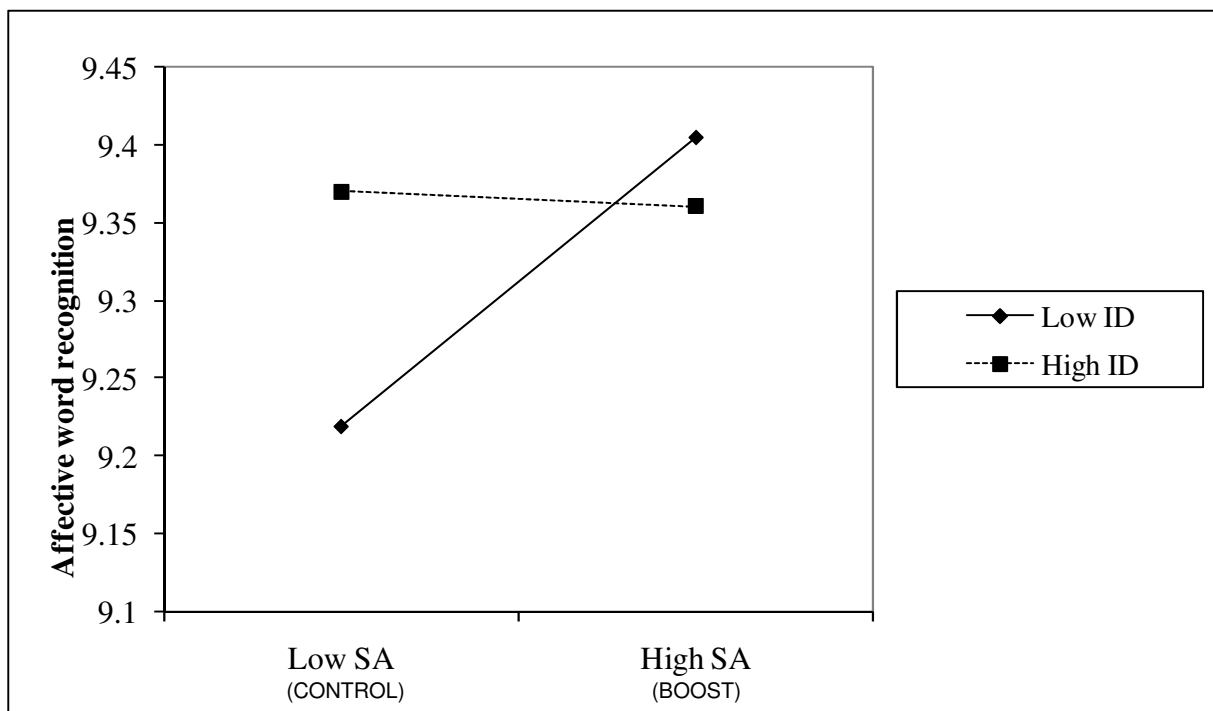


Figure 10: Interaction plot for stereotype activation (SA) and Identifiability (ID) on affective word recognition scores for females only

The three way stereotype-activation\*identifiability\*domain-identification interaction was highly significant,  $t = 2.560$ ,  $p = .012.$ ,  $b = .053$ . This three-way interaction is plotted in

Figure 11. The graph reveals that participants in all conditions performed better in the boost condition than the control condition, with the exception of identifiable participants with low domain identification who exhibited a reversal of the expected boost effect and scored substantially worse under boost than in the control condition. The top performers were those with low identifiability and low domain-identification in the control condition, while the worst performers were the equivalent group (control condition; low domain-identification) when they were not individually identifiable.

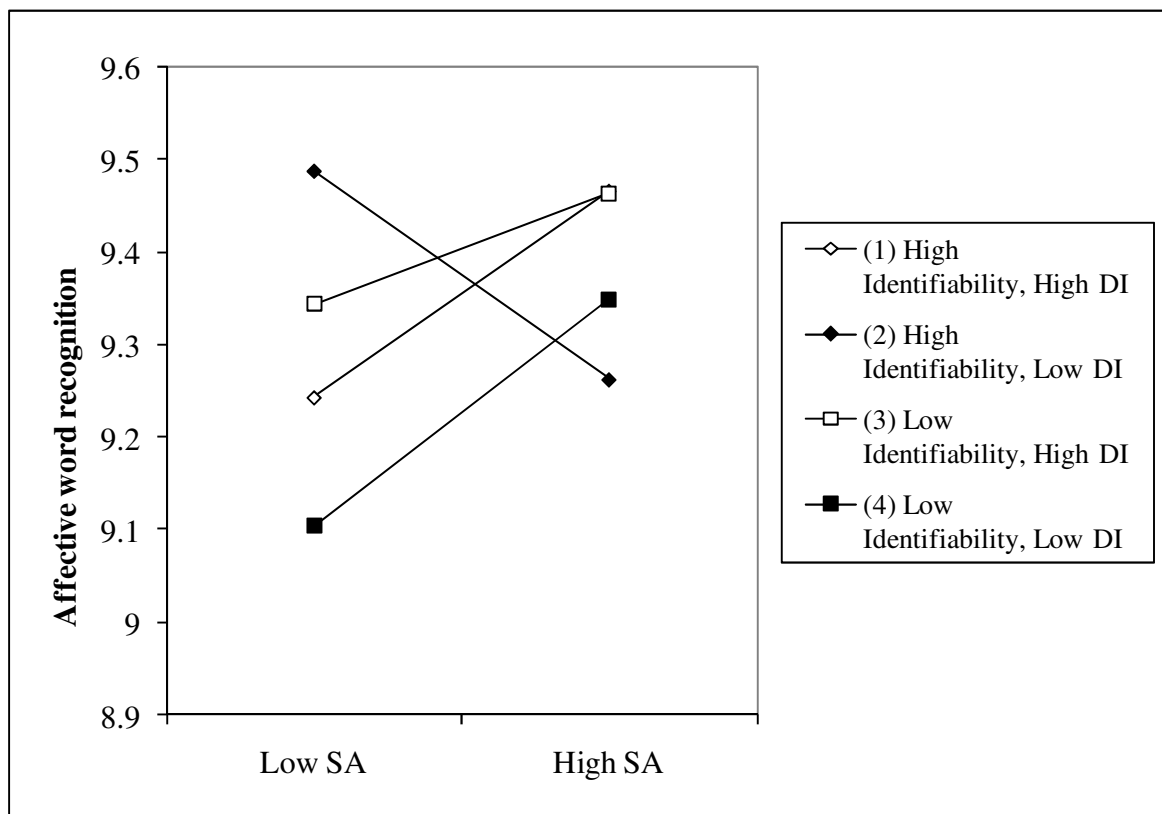


Figure 11: Interaction plot for stereotype activation (SA) \* Identifiability \* Domain Identification (DI) on affective word recognition scores for females only

#### 6.2.4. Discussion of Identifiability Study 2

These results confirm the most general findings of the first identifiability study (Identifiability Study 1), namely that identifiability modifies participants' experiences of the performance context and may moderate the effect of stereotypes on performance. However,

the specific patterns by which identifiability and threat influenced performance were quite different: in the first study STB had conventional effects when identifiability was high, but reversed effects for double-minority participants when identifiability was *low*. This reversal was largely driven by the underperformance of double-minority participants in the boost condition, possibly due to low investment. Note that in the first study, double minority status was independent from threat and boost in the sense that double minority members could be assigned to either positively or negatively stereotyped conditions.

In the second study, sex was integral to the STB manipulation, since females could only be assigned to the control condition or the boost condition and males could only be assigned to the control condition or the threat condition. In this second study, STB had conventional effects in *low* identifiability conditions, but reversed effects (for males) or no effect (for females) when identifiability was *high*. For males the reversal effect seemed to be driven largely by men performing particularly well when both identifiable and under threat, and women performing well when identifiable regardless of their threat assignment. When the influence of domain identification was included, it emerged that for women the STB reversal was particularly evident for participants with low domain-identification in the high identifiability condition.

Together, the results of these studies suggest that identifiability can disrupt the predicted STB effects in at least two ways: first, when identifiability is low (particularly when investment is also low, as indicated in these studies by double minority status and domain identification), the experience of being positively stereotyped may give someone enough identity security in the performance context that they have nothing to prove and can therefore produce a less-than-optimal performance with little risk to their self-concept in the performance environment (as in the first study). Second, when identifiability is high, members of the negatively stereotyped group with high investment in the domain may be

motivated to perform particularly well (as in the second study), either in defense of their negatively stereotyped group or to demonstrate that they are not prototypical exemplars of the negative group stereotype.

### **6.3. Identifiability Study 3**

#### **6.3.1. Research question**

Although the first two studies demonstrated that identifiability has some effect on the stereotype-performance relationship, it is difficult to know how to understand this experience of identifiability in social identifiability terms. Were participants concerned about their ‘moral career’ (Goffman, 1961) in the department or institution in which the research took place, their own position in the ingroup, or their own and their group’s position with respect to the outgroup?

The third study in the programme of research was similar in concept to the first: humanities students were randomly assigned to threat or boost conditions by informing them that the Raven’s APM requires either “logical patterns in datasets” (a task at which they have low status in relation to Science students) or “intuitive patterns in symbolic figures” (in which they should have high comparative status). This time, instead of manipulating simple identifiability, the *type* of audience to which participants believed they would be visible was manipulated, with some believing that they would be identifiable to the ingroup, some to the outgroup and others to a group of experimenters.

Additionally, by this time a comprehensive battery of social identity measures had been developed and piloted, including: the status comprehension/manipulation check (and therefore the means to assess resistance to the status manipulation); graphical measures of ingroup, outgroup and intergroup closeness (Schubert & Otten, 2002); measures of individual identifiability and the extent to which individuals felt they were representatives of their group in the testing context (group identifiability); ingroup and outgroup differentiation; category



salience during task performance; and the key sociostructural variables legitimacy, stability, permeability, conflict. Reliability analyses for these variables appear in section 4.3.

### **6.3.2. Sample**

Sixty students registered at the University of KwaZulu-Natal participated in the study. Of these, the results of seven had to be dropped as they did not fit the requirements of the study (they were not humanities students) leaving 53 participants in the sample. This is not as small as it sounds, as a quasi-repeated measures design was used to increase the power of the study.

Of these participants, 27 were assigned to the threat condition and 26 to boost. The audience condition was also balanced: 18 participants believed their performance would be observed by ingroup members; 18 by outgroup members; and 17 by a group of researchers. In terms of demographics: 40 participants reported their sex as female and 13 as male; 50 participants reported their race as black African, 2 as white and 1 as 'other.'

### **6.3.3. Design and procedure**

The dependent measures were two equivalent short forms of the Raven's Advanced Progressive Matrices (APM) developed for this study using the methods described by Bors and Stokes (1998) and Arthur and Day (1994). Although every effort was taken to ensure that the means and variances were equivalent for the two short forms based on all published item analyses for the APM that could be found in the literature (see Forbes, 2007), with this particular sample there was a significant albeit small difference in performance on the two short forms,  $t = 2.385$ ,  $df. = 52$ ,  $p = .021$ , mean difference = .05 (one third of a standard deviation). However, the two short forms were randomly counterbalanced as pretest and posttest measures to avoid order effects. Therefore the small difference in difficulty between short forms was randomly distributed across conditions and therefore contributed to noise but should not have confounded the design.

Perceived group status (invoking threat or boost) was a between-groups independent variable that was manipulated *prior* to the APM pretest and therefore both the pretest and post-test were subject to STB. The audience manipulation was a within-subjects manipulation that was invoked following the pretest and should therefore be evident in post-test results only. The pretest therefore represents an anonymous baseline condition and the post-test represents an identifiable condition and both should be subject to STB. The post-test would additionally be influenced by the nature of the audience to which participants believed they would be identifiable (ingroup, outgroup or experimenters). Since it was essential for participants to complete the dependent measure immediately after the manipulations to maximize effect size and statistical power, the battery of identity measures could only be completed after the APM post-test. This means that the analysis must consider the extent to which these identity measures may also reflect participants' efforts to position themselves in the experimental context with respect to their experience of the manipulations and their own APM performance. Although there are problems with this design, the data yields some useful insights that contribute to our understanding of STB.

The data were first analyzed using a linear mixed model in SPSS. Linear mixed modeling does not assume independence of observations and is designed to deal with multi-level and nested effects and is therefore appropriate for this design<sup>17</sup>. Subsequently, after identifying a significant model for the full repeated measures design, a multiple regression model was fitted by simply ignoring the pretest results. Although this results in a loss of statistical power, the post-test was subject to all manipulations and this analysis allowed

---

<sup>17</sup> Difference scores or conventional repeated-measures models were not appropriate because one variable was manipulated within the repeated measures design and one outside. Conventional GLM models are not designed to accommodate this.

detailed exploration of the relationships between audience, stereotype activation and social identity variables.

#### **6.3.4. Linear mixed model analysis**

A linear mixed model analysis revealed that the status manipulation had no direct impact on APM performance,  $F(1,51) = .007, p = .936$ . However, there was a significant interaction between status and identifiability,  $F(2,51) = 3.647, p = .033$ , indicating that the status manipulation had different effects on performance at different levels of identifiability. However, the model improved when ingroup identification was included as a covariate in the model (AIC dropped slightly from 448.708 to 445.991), with the main effect for status (ie. STB) achieving significance,  $F(1,49.324) = 6.175, p = .016$  and the interaction between status and ingroup identification emerging as a highly significant term,  $F(2, 49) = 5.616, p = .006$ .

When estimated marginal means were plotted for this interaction (see Figure 12, below), it emerged that there was little difference between anonymous and identifiable conditions for those participants in the threat condition, but a substantial difference in the effect of stereotype boost between anonymous and identifiable conditions. When participants were anonymous, performance of participants in the boost condition was substantially *lower* than participants in the threat condition, which is a reversal of the predictions of stereotype threat. The highest performance was achieved in the high-identifiability boost condition. Identifiability generally seemed to increase performance across the board, and anonymity resulted in a reversal of the STB effect.

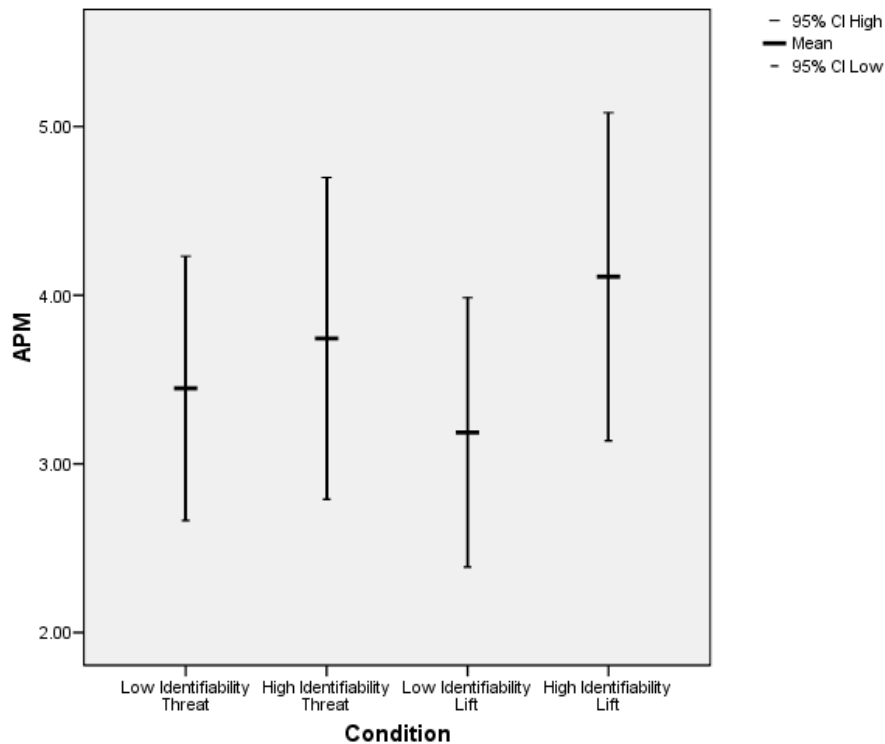


Figure 12: Estimated marginal means for the significant status\*identifiability interaction

When participants were identifiable they performed better under boost and worse under threat, as predicted by STB literature. However, when completing the materials anonymously they performed substantially worse in the boost condition. The lowest performers across the board were those in the boost condition while completing the materials anonymously and the best were those in the high identifiability boost condition.

### 6.3.5. Visibility to specific audiences

In this study the identifiability condition was divided into three subconditions: participants were informed that their video-recorded task performance would be scrutinized by either a group of ingroup-members, outgroup members or experimenters. The next stage in the analysis therefore considered the impact of visibility to specific audiences over and above the effects of identifiability discussed above.

However, the small sample size severely limited the analytic procedure. Ideally specific audience visibility should have been dummy-coded in two variables to represent the three conditions and then entered into one model along with status and resistance status. The problem with this procedure is that the saturated model would have tested 12 effects with just 39 denominator degrees of freedom.

Therefore, despite the increased risk of type-I error, dummy-coded ingroup-visibility, outgroup-visibility and experimenter-visibility were tested in separate models, each testing seven effects with 43 denominator degrees of freedom. Although these sample sizes were tiny, they (just) exceeded the most lenient recommended minimum sample size for moderated multiple regression of five participants per effect (Tabachnick & Fidell, 1989).

The conditions invoking ingroup- or experimenter-visibility had no significant effect on APM performance, although the model for experimenter-visibility came close to marginal significance and deserves attention in future studies. The model regressing status, status-resistance and outgroup-visibility on APM performance (with a log transformation to correct skewness and dropping two outliers) was just significant,  $F_{7, 43} = 2.232$ ,  $p = .050$ ,  $b = .267$ . The model was driven by the significant 3-way status\*status-resistance\*outgroup-visibility interaction,  $t = 2.947$ ,  $p = .005$ ,  $b = .359$  (plotted in Figure 13).

The interaction plot shows that stereotype-performance effects were only observed when participants believed that their performance would be scrutinized by outgroup members. The comparison group (labeled “not visible to outgroup”) consisted of participants who believed that their performance would be observed either by members of the ingroup or experimenters. The lack of stereotype-performance effects in these conditions therefore either meant that only outgroup-visibility invoked stereotype-performance effects at all, or that the stereotype-performance effect was in opposite directions for participants in ingroup- and experimenter-visibility conditions and cancelled out when it was combined.

However, models regressing status, status-resistance and status\*status-resistance separately on these two subsets did not yield results even remotely approaching significance, suggesting that the lack of stereotype-performance effects in the “not visible to outgroup” condition really did indicate that participants were most likely to exhibit stereotype-performance effects when they believed their performance would be visible to outgroup members rather than experimenters or ingroup members.

The stereotype-performance effects exhibited by participants who believed they would be visible to the outgroup were strongly moderated by status-resistance. Participants with high status-resistance performed well under boost and poorly under threat as expected. However, participants who did not resist the stereotype performed poorly under boost and much better under threat, contrary to STB predictions.

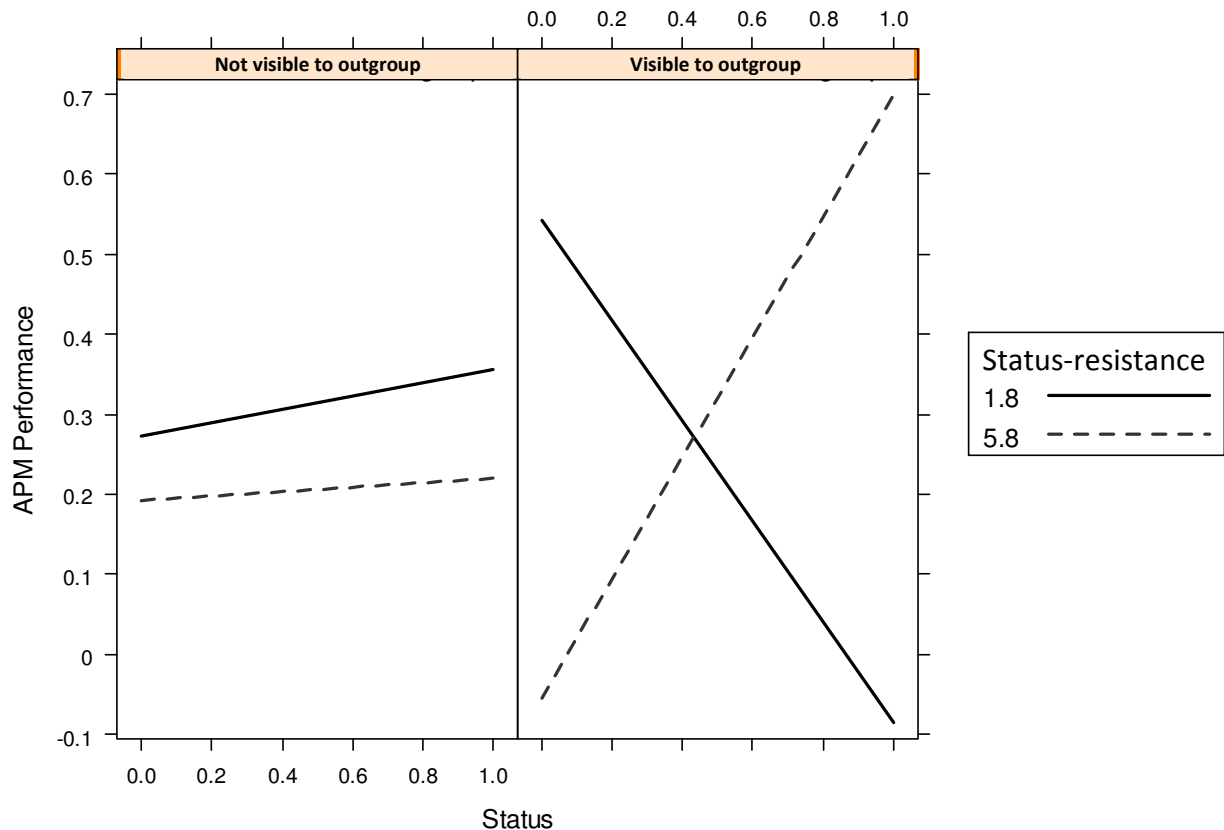


Figure 13: Interaction plot for status\*status-resistance\*visibility-to-outgroup on APM performance for Identifiability Study 3

### 6.3.6. Discussion of Identifiability Study 3

The most general outcome of this study is that it confirms the results in the stereotype threat literature showing that identifiability moderates stereotype threat and boost effects (Ambady et al., 2004; Schmader et al., 2008). This is an important finding, since the majority of STB studies are completed anonymously for ethical reasons but many performances in real-life contexts are not, which represents a substantial threat to external validity. This could at least partially explain why the powerful results obtained in laboratory studies of STB have not been clearly replicated in real-life high-stakes and non-anonymous performance contexts such as the SAT's (eg. Cullen et al., 2004).

These results also contribute to the growing portfolio of studies showing reversed or ironic STB effects in specific performance conditions (cf. Nguyen et al., 2003). Together, these ironic results challenge the conventional understanding of STB in which stereotypes are expected to produce congruent shifts in performance in a deterministic way. Specifically, it seems clear that stereotype-performance effects are real, but that the direction of the effect is not exclusively related to stereotype valence.

#### **6.3.6.1. *Identifiability and audience visibility***

The results confirmed that identifiability plays an important role in stereotype-performance effects, as suggested by the first two studies. However, it also supported the hypothesis that the audience to whom one is identifiable may be an important feature of identifiability, as predicted by the SIDE model (eg. Klein et al., 2007). Therefore, it is not simply lack of anonymity that impacts on stereotype-performance effects, but the audience to whom the performance will be visible and their relationship to the performer and the performer's group that matters.

These results support the argument that task performance in the experimental context is also a performance of identity. To put it a different way, doing well or doing badly on a performance task in a specific performance context, such as completing the APM in a STB experiment, is an interpretable act in that context to a particular audience. The performance of a task, therefore, is also a performance of identity. In this study participants oriented this performance to the audience to whom it would be visible and who they believed would judge them and/or their group on the basis of their performance.

As in the previous studies, an important driver of the effect of identifiability on performance was that participants in boost conditions were likely to perform better compared to other conditions when identifiable, but worse when they believed they were anonymous (and had low resistance to the status manipulation). They were most likely to underperform in



the boost condition when they exhibited low resistance to the manipulation and reported high levels of belief in the legitimacy of the status hierarchy, in other words, when their belief in the superiority of their group was secure.

The dominant explanation for poorer performance under boost conditions in the literature is that the experience of being positively stereotyped places participants under pressure to live up to the stereotype which interferes with their performance, in other words, they ‘choke under pressure’ (Cheryan & Bodenhausen, 2000). However, the results from this study suggest that low performance under boost is most likely under anonymous conditions. The authors of the most cited paper postulating the ‘choking under pressure’ argument (Cheryan & Bodenhausen, 2000) have confirmed that their studies were also completed under anonymous conditions (Bodenhausen, 2010). Therefore, it is possible that underperformance in their studies was similarly motivated.

The ‘choking under pressure’ hypothesis is further undermined by the fact that in the current study, participants who were under boost but high identifiability performed better than all other conditions, particularly when they believed their performance would be identifiable to other ingroup members. Although it is possible that being visible to the ingroup is relatively safe and results in decreased pressure and thereby higher performance, in line with the “choking under pressure” hypothesis, it is unlikely that identifiability [to anyone] invokes less pressure than anonymity. Therefore, according to the results in this study and the studies reported previously, the most likely reason for underperformance in the boost condition is that participants have a secure identity in the experimental context and therefore have nothing to prove and are not motivated to invest the effort required to perform well on the APM. This explanation will be further explored in the fourth and final identifiability study.

#### 6.4. Identifiability Study 4

Despite the intuitive and theoretical coherence of the argument arising from Identifiability Study 3 that participants under boost may underperform because their secure identity allows them to relax, it was still post-hoc and conjectural. The final identifiability study was therefore designed to test this specific hypothesis and to test it directly against the ‘choking under pressure’ explanation for underperformance under boost commonly advanced in the STB literature (eg. Cheryan & Bodenhausen, 2000; Smith & Johnson, 2006).

##### 6.4.1. Design

The abstraction subtest of the Shipley Institute of Living Scale (referred to as the SILS hereafter) was administered prior to any manipulations. Like the APM, it is a good general test of educative ability and generally is strongly correlated with APM performance (generally at about .72). However, in contrast to the APM, it requires pattern and sequence recognition with numbers and letters rather than images. This test therefore allows a pre-manipulation snapshot of performance on the same (or at least a similar) construct that estimates some combination of pre-test ability and motivation without invoking any practice effects. This allows statistical comparisons to see whether participants’ performance shifts *in relation to their peers*<sup>18</sup>, increasing the power and diagnosticity of the design.

Additionally, to investigate the hypothesis that participants with positive group identities (ie. under boost) were less invested in the task because they experienced less need to defend their identity with strong performance on the APM, we introduced a manipulation intended to increase personal investment in the task by modifying the reward they expected for participation. Specifically, in the condition designed to match the majority of STB research, participants were offered a flat-rate incentive of R30 regardless of their level of

---

<sup>18</sup> Since no norms are available for either test in the South African context, the SILS/APM comparison was achieved with z-scores, which necessarily made the comparison relational.

performance on the task. In the condition designed to maximize personal investment in the task, participants were told that they would get R16 just for participating and an additional R1 for every APM item correctly answered (up to a maximum of R30).

Mobbs, Hassabis, Seymour, Marchant and Weiskopf (2009) showed that performance-related monetary rewards can increase pressure to the point that performance decreases (even without a concurrent status manipulation), so if the ‘choking under pressure’ hypothesis is correct, we would expect participants to be even more likely to ‘choke under pressure’ in the boost condition when an individual monetary reward is at stake. Conversely, if the conclusion of Identifiability Study 3 is a better explanation in this context and participants in the boost condition underperform because their secure identity allows them to invest less in the task, then we would expect participants in the boost with individual reward condition to be provided with a reason for investment in the task and therefore to perform better than their counterparts in the boost with flat-rate reward condition (which is equivalent to the conditions of Identifiability Study 3).

#### **6.4.2. Design and procedure**

The study employed a 3x2x2 factorial design with status (boost, threat or control), identifiability (identifiable or anonymous), and reward (flat-rate or performance-based) on APM performance. Data was collected in batches of ten on computers running E-Prime (Psychology Software Tools, 2010), with all participants in each session assigned to the same identifiability condition to avoid accidental contagion.

#### **6.4.3. Sample**

The sample consisted of 294 students registered for degrees in the Faculty of Commerce on the Pietermaritzburg campus of the University of KwaZulu-Natal who believed they were being compared to humanities students. The results for eighteen students were dropped because they were not, it turned out, actually registered for commerce degrees,

resulting in a final sample of 279. Of these: 138 were assigned to the threat condition and 141 to boost; 139 were anonymous and 140 were identifiable; and 136 were in the flat-rate reward condition and 143 were in the performance-based reward condition. In terms of demographics: 174 participants reported their sex as female and 104 as male; 232 reported their race as black African, 6 as coloured, 28 as Indian, 10 as white and 2 as ‘other’ (see Finchilescu & Tredoux, 2010, for a discussion of the use of racial categories in South African research).

#### ***6.4.3.1. Identifiability manipulation***

In the high identifiability condition each computer was fitted with a webcam. After completing the informed consent procedure and the Shipley, the researcher went to each PC and pressed the webcam’s snapshot button. The E-Prime software then displayed the participant’s picture on-screen and asked them to confirm if it was indeed a picture of them. They were then asked to enter their name and student number.

For the low identifiability condition the webcams were removed from the venue. Instead of showing a picture of the participant, E-Prime displayed a stylized androgynous silhouette (displayed in Figure 14) and asked them to “Press ‘N’ if it is not a picture of you” (which it very clearly was not). They were then asked to “make up a fake name” and a “fake number ... that no-one else can recognize.”



*Figure 14: Placeholder image used in the low-identifiability condition*

#### **6.4.3.2. Status manipulation**

All participants were told that the “study compares the performance of Commerce and Humanities students on a test of academic ability called Raven’s Advanced Progressive Matrices.” Those assigned to the boost condition were told that “Commerce students consistently perform badly and Humanities students consistently perform very well” because “doing well in the Raven’s Advanced Progressive Matrices requires *logical pattern analysis*” which is “well suited to the skills of Commerce students” and that “Humanities students are generally better at intuitive symbolic analysis and are therefore disadvantaged in this task.” Those in the threat group received exactly the same description except they were told that Commerce students consistently perform badly because the APM requires “intuitive symbolic analysis” which is “very well suited to the skills of Humanities students.”

Participants in the control group were not reminded of their category membership as Commerce students in the experimental context, and nor were any positive or negative group stereotypes triggered.

#### **6.4.3.3. Reward manipulation**

Participants in the flat-rate reward (FRR) condition read the following instructions: “You will get R30 just for participating. You will not get additional money for every item that you get right on the Raven's Progressive Matrices task. However, please do your best on the Raven's Progressive Matrices task.”

Participants in the performance-based reward (PBR) condition read the following: “You will get R16 just for participating. You will also get an additional R1 for every item that you get right on the Raven's Progressive Matrices task. However, please do your best on the Raven's Progressive Matrices task.”

#### 6.4.3.4. *Resistance*

Status resistance was calculated as described in Chapter 5. Identifiability resistance was calculated in the same way, but using manipulated identifiability and reported identifiability to arrive at an identifiability resistance score.

#### 6.4.4. Results of study Identifiability Study 4

The basic saturated model regressing manipulated status, manipulated identifiability, and manipulated reward on APM performance resulted in a weak and insignificant model,  $F_{7, 271} = .690$ ,  $p = .680$ ,  $r^2 = .018$ , with only the effect for identifiability approaching significance,  $t = -1.823$ ,  $p = .069$ ,  $b = -.092$ .

When status-resistance was added to the model (calculated as discussed in Chapter 5) and simplified with stepwise selection on AIC using the `step()` function in R (R Development Core Team, 2010) a model emerged that was not significant,  $F_{9, 269} = 1.548$ ,  $p = .131$ , but in which the only significant effect was the interaction between status, status-resistance and reward,  $t = -2.194$ ,  $p = .029$ . Therefore, including status-resistance helped to illuminate stereotype-performance effects, but not very much.

A model regressing status, identifiability, reward and social desirability on APM performance approached significance,  $F_{15, 262} = 1.569$ ,  $p = .082$ ,  $r^2 = .082$ . Only three effects were significant, and these included the saturated four-way status\*identifiability\*reward\*social-desirability interaction,  $t = -2.070$ ,  $p = .039$ ,  $b = 1.797$ .

Since four-way interactions are difficult to plot, the data were partitioned by a median split on social-desirability and the fully saturated status\*identifiability\*reward model was fitted on the low and high social-desirability subsets independently. The model was not at all significant for the low social-desirability subset and nor were any substituent effects, but closely approached significance for the high social-desirability subset,  $F_{7, 134} = 2.071$ ,  $p = .051$ ,  $r^2 = .098$ ,  $adj.r^2 = .098$ . The significant effects were for status,  $t = -2.289$ ,  $p = .024$ ,

$b = -.836$ , and  $\text{status} \times \text{reward}$ ,  $t = 2.147$ ,  $p = .034$ ,  $b = 1.100$ . However, these were subject to the three-way  $\text{status} \times \text{identifiability} \times \text{reward}$  interaction that closely approached significance,  $t = -1.915$ ,  $p = .058$ ,  $b = -1.414$ .

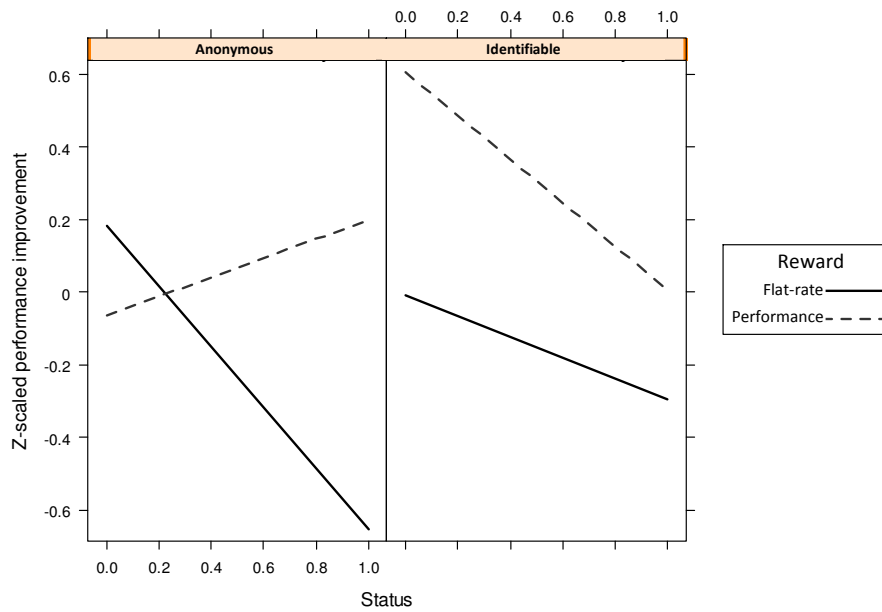


Figure 15: Interaction plot for  $\text{Status} \times \text{Identifiability} \times \text{Reward}$  for high social desirability subset in Identifiability Study 4

The interaction plot (Figure 15) shows that, as predicted, the worst performers were those in the low identifiability boost condition expecting a flat-rate reward. However, participants in the equivalent low identifiability boost condition performed much *better* if they were expecting a performance-based reward. This interaction did not occur for participants in the high identifiability condition, who generally performed better under threat than boost and better if they were expecting a performance-based reward than a flat-rate reward. The best performers overall were in the low status condition, believed they were individually identifiable, and that they would receive a performance-based reward.

#### 6.4.5. Discussion of Identifiability Study 4

##### 6.4.5.1. *Underperformance in boost conditions due to slipstreaming, not choking under pressure*

The ‘choking under pressure’ hypothesis (eg. Cheryan & Bodenhausen, 2000) argues that participants who are targets of a positive stereotype underperform because they experience increased stress due to the pressure to live up to the positive group stereotype. This study tested this hypothesis by adding a reward manipulation to the status and identifiability manipulations used in the previous studies as a means of adding further pressure to participants in the boost condition (Mobbs et al., 2009). If they were already ‘choking under pressure’ because of the high expectations attached to high-status group membership, then they should perform even worse when the additional pressure of performance-based reward was applied.

However, the previous studies in this chapter have suggested an alternative hypothesis: that anonymous participants in boost conditions underperform because their membership of a high-status group relieves them of the burden of proving themselves individually in the performance context. In other words, there is a possibility that they are ‘pulled along in the slipstream’ of their positive group identity and do not have to perform well in order to produce an acceptable identity in the performance context. This hypothesis would be supported if participants performed better when given a reason to do, either by being made individually accountable for their performance (by being individually identifiable), or by expecting a performance-based reward, despite the additional pressure that these alternative conditions are likely to invoke.

In the present study, the performance of participants who were unconcerned with the impression they made in the experimental context (that is, who scored low in social desirability) did not support either hypothesis. In fact, the performance for participants low in



social desirability was not predictable at all by status, reward or identifiability. However, for the subset of participants who scored high in social desirability, the worst performers were those in the anonymous flat-rate reward condition.

In the STB literature, the first study to report underperformance in boost conditions and attribute it to ‘choking under pressure’ was Cheryan and Bodenhausen (2000). They argued that the high expectations of membership of a group that is positively stereotyped in the performance domain can result in over-motivation that inhibits task success.

They reported in their paper that participants in their study received a flat-rate reward of \$5, and the second author has confirmed by personal correspondence that participants were anonymous (Bodenhausen, 2010). Therefore the conditions in which participants underperformed in Cheryan and Bodenhausen’s (2000) study are the same as the conditions in which participants performed the worst in the present study: high-status, low-identifiability and flat-rate reward. If Cheryan and Bodenhausen’s (ibid.) choking under pressure explanation is correct, then any condition that would have increased pressure for these low-performing participants should have depressed their performance even further.

However, in the present study the opposite occurred. The participants who should have been under most pressure – those who were under boost, were individually identifiable, and were expecting a performance-based reward, performed better than participants in any other condition. These results do not support the notion that participants in boost conditions underperform because of the weight of expectation upon them. On the contrary they provide additional support for the slipstreaming hypothesis.

The slipstreaming hypothesis, in other words, posits that underperformance under boost is likely when participants’ group identity as members of the high status group provides an identity buffer that excuses poor performance and allows them to maintain a positive

identity in the performance context regardless of their individual performance on the specific task.

Since social identity theory predicts that group identity is more likely to be a prominent component of self-concept in conditions of anonymity (Tajfel, 1978c), it follows that group identity will buffer self-concept most strongly when participants are known only as group members and not identifiable as individuals. It also follows that an intervention that rewards individual behavior will make participants more accountable for their own behavior and increase the impact of individual performance on the self-concept. Therefore the pattern of results observed in this study provides evidence that, in this performance context at least, underperformance in the boost condition was related to ‘identity slipstreaming’, not to ‘choking under pressure.’

This finding strengthens Steele’s argument (eg. Steele, 2010) that the “identity contingency” in STB situations is truly a contingency of identity in which participants actively (although not necessarily consciously) act in ways that are oriented to their identity resources and liabilities in the context. The outcome of this contingency is that they may perform poorly when there is no identity-advantage to performing well and the negative practical and identity-related consequences for poor performance are minimal.

These results do not rule out the possibility that boost might invoke ‘choking under pressure’ in some contexts, particularly when participants are facing pressure to enact “model minority” identity as discussed by Cheryan and Bodenhausen (2000). However, this experiment demonstrates that there are other identity contingencies which might lead to underperformance in the boost condition, and that slipstreaming is one of them. Since the combination of being anonymous and receiving a flat rate reward for participation regardless of performance (whether financial or course-credit-based) is pervasive in psychological research (Brase, 2009; Tomporowski, Simpson, & Hager, 1993) and is also the default for

stereotype threat research, it is possible that poor performance under boost conditions reported in the literature so far may also be related to identity slipstreaming as observed in the studies in this chapter.

### **6.5. General discussion of identifiability studies**

While overall individual and group identifiability, as expected, slightly improved performance as predicted by ‘social loafing’ theories, the impact of identifiability on stereotype-performance effects is complex. In the four studies discussed here, the most evident impact of identifiability on stereotype-performance effects was in the tendency for participants to slipstream when anonymous and under boost, and the tendency to perform better when perceived group identifiability was high, presumably to defend the virtue of the group.

#### **6.5.1. Slipstreaming instead of ‘choking under pressure’**

The results of all four identifiability studies presented here concur that, at least in the experimental situations that these participants were exposed to in the present studies, underperformance in the boost condition may be related to participants with a secure identity having nothing to prove and therefore slipstreaming in the experimental context, by which I mean putting in less effort while relying on a secure group identity to maintain a viable identity position in the experimental context. The evidence is fairly strong, since in all studies underperformance under boost was only likely when participants were individually anonymous and had nothing to gain from performing well. These results suggest that it is when participants are shielded by a secure group identity that are able to invest less in task performance. It seems that participants were being quite pragmatic about their investment in the task in relation to the impact that high or low performance would have on their own situated identity and the reputation of their group.

This is well predicted by SIT and self-categorization theory, which argues that in any given context, identity will be an optimal combination of group- and individual-level identity resources (eg. Oakes et al., 1994). Therefore, when the group is positively stereotyped in a given context, group identity provides a ready-made acceptable identity position and individual performance will become less critical to the situated production of acceptable identity in that context. Indeed, in some identity contingencies, poor performance can be a signifier of a secure identity position in a high status group, such as for the archetypical rich kid at an ivy-league university whose family endowments can be relied upon to secure their status.

Of course, this does not rule out that in certain high-pressure environments boost conditions might produce unmanageable levels of pressure on targets as argued by Cheryan and Bodenhausen, amongst others, (Cheryan & Bodenhausen, 2000), but it is also likely that the experience of succeeding less because there is less need to succeed is a very common identity contingency.

## **Chapter 7. Results III: Conflict**

Intergroup conflict is an important component of the SIT framework and has been shown to increase category awareness; increase ingroup and outgroup depersonalization effects; and modify how much group members identify with the ingroup and outgroup (eg. Tajfel & Turner, 1979b). Since category salience (Lee & Ottati, 1995; McGlone & Aronson, 2006) and group identification (Schmader, 2002a) have already been shown to be implicated in STB, it is very likely that intergroup conflict will be an important feature of targets' situated experiences of stereotype threat. Three studies therefore explored the impact of perceived intergroup conflict on task performance under STB conditions.

### **7.1. Conflict Study 1**

#### **7.1.1. Research question**

This study manipulated perceived intergroup conflict and stereotype threat or lift (STB) to investigate the effect of perceived intergroup conflict on STB.

#### **7.1.2. Sample**

Students registered for Science degrees at UKZN were purposively sampled. A total of 102 students participated in the study, and none were dropped during data cleaning procedures. Of these: 51 were assigned to threat and 51 to boost conditions; 50 were assigned to low conflict and 52 to high conflict conditions; 44 reported their sex as female and 57 as male; and 82 reported their race as black African, 5 as coloured, 4 as Indian and 11 as white.

#### **7.1.3. Design and procedure**

The study used a 2x2 factorial design, with conflict (low; high) and status (low; high) as independent variables and performance on a short form of the Raven's Advanced Progressive Matrices as the dependent variable. Conflict was manipulated using an adaptation of the Tajfel matrices (Bornstein et al., 1983; Tajfel, 1970). Participants were randomly assigned to and scheduled in groups of three. When they arrived at the venue on the science

campus they were seated in three seats facing a single computer screen. They were informed that they would be interacting electronically with a group of humanities students who were seated in a venue in the Arts campus approximately 1km away. They were informed that they would choose “points” for the humanities students that would determine that group’s payment for participation and vice versa. In the high conflict condition, the virtual outgroup consistently chose the minimum points for the Science students (and hence the minimum payment for the participants) regardless of the punitive responses that this strategy generated. Conversely, in the low conflict condition, the virtual outgroup consistently chose to allocate the maximum possible points to the group of Science student participants regardless of their response.

Status was manipulated by describing the APM task as either favouring or not favouring the stereotypical characteristics of science students (ie. as a task requiring logic and mental accuracy, invoking lift; or requiring mental flexibility and the ability to process conflicting information, invoking threat).

Participants completed a short-form of the Raven’s advanced progressive matrices immediately following the manipulations and then completed the social identity inventory collated for this programme of research.

#### **7.1.4. Key findings**

A manipulation check on reported conflict confirmed that this manipulation was generally successful, with participants in the high conflict condition reporting significantly higher perceived conflict than those in the low conflict condition,  $t(100) = -2.737, p = .007$ .

Once again the most basic regression model including just the status and conflict manipulations and their interaction on APM performance was not significant,  $F_{3,98} = 1.115, p = .347$ , and there were no effects even approaching significance. However, when status-resistance was added (including second-order interactions) the model achieved significance,

$F_{6, 95} = 2.31, p = .040, r^2 = .033, adj.r^2 = .003$ , although it accounted for a tiny proportion of the variance in performance. The only significant effect in this model was status\*status-resistance,  $t = -3.129, p = .002$ .

This interaction is plotted in Figure 16 and shows that participants with high status-resistance were likely to perform worse in the boost condition, while those with low resistance were likely to perform worse in the threat condition. This pattern is entirely consistent with the meaning of resistance: that both those who disagree that their group is good and those who agree that their group is bad are more likely to perform badly. This means that, as expected, participants with low status-resistance exhibited a conventional STB effect and those with high status-resistance displayed a reversal of the STB effect. However, while this interaction tells us something about participants' responses to the status manipulation, it says little about the impact of the conflict manipulation – if there was any impact at all.

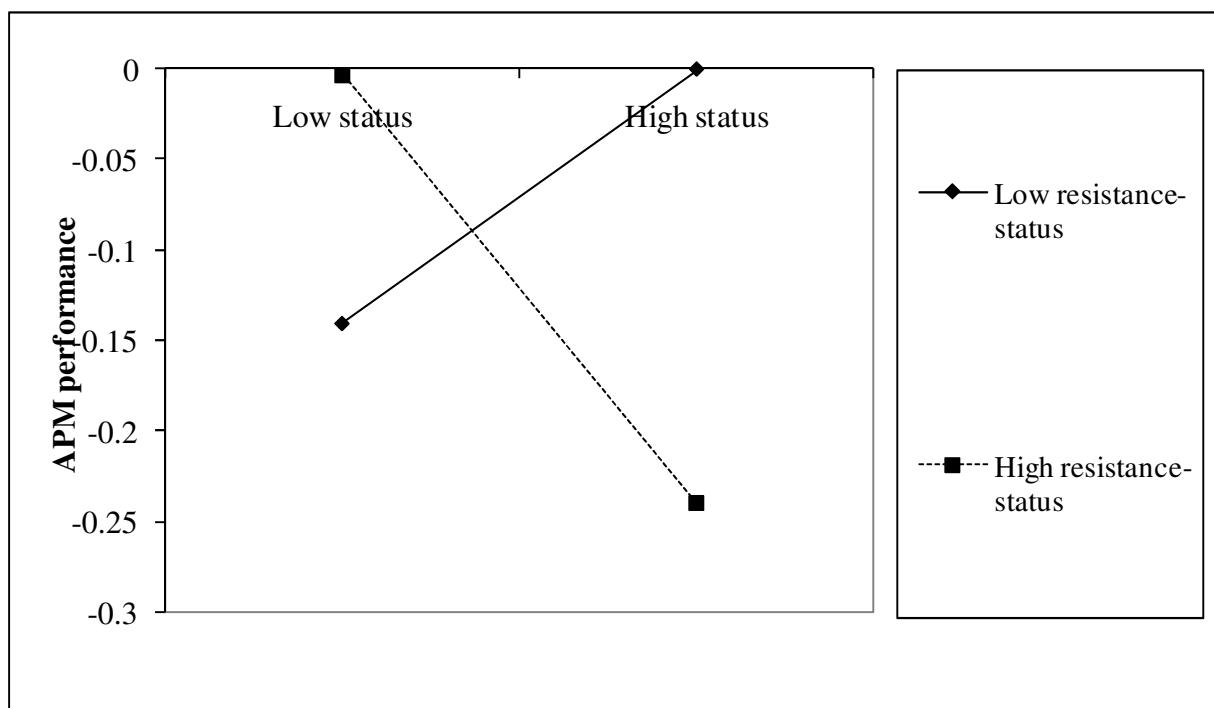


Figure 16: Interaction plot for the status \* status-resistance interaction in Conflict Study 1

#### 7.1.4.1. *Accounting for individual identifiability*

The previous chapter showed that individual identifiability or anonymity is an important feature of stereotype-performance effects. A fully saturated model regressing conflict\* status\*status-resistance\*identifiability on APM performance (with a square transformation to correct skewness) approached significance,  $F_{15, 85} = 1.7, p = .066, r^2 = .231$ . There were only two significant effects, namely, status\*status-resistance,  $t = -2.215, p = .029, b = -.053$ , and the four-way conflict\* status\*status-resistance\*identifiability interaction,  $t = -2.131, p = .036, b = -.067$ .

To interpret the significant four way interaction, fully saturated three way conflict\* status\*status-resistance models were run separately on low and high status-resistance subsets (determined by a median split on status-resistance). The model for the high-status-resistance subset was not at all significant, and neither were any constituent effects. However, the model for the low status-resistance subset was significant,  $F_{7, 42} = 2.408, p = .036, r^2 = .286$ . The status\*identifiability interaction achieved significance,  $t = -2.513, p = .016, b = -.133$ , as did the conflict\*identifiability interaction,  $t = -2.603, p = .013, b = -.145$ . However, both of these were subject to the significant higher-order conflict\* status\*status-resistance interaction,  $t = 2.785, p = .008, b = .172$ .



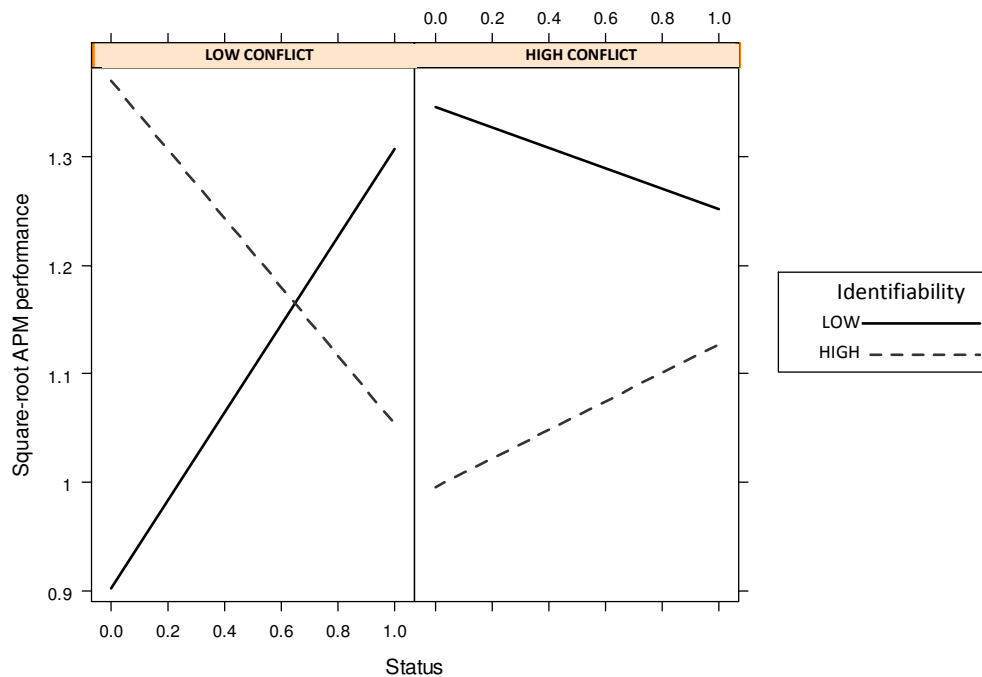


Figure 17: Interaction plot for status\*conflict\*identifiability for the low status-resistance subset of Conflict Study 1

The three-way interaction is plotted in Figure 17, and shows that stereotype-performance effects were more pronounced in low conflict conditions. The low-status condition produced the strongest stereotype threat effect for participants in the low conflict condition who also reported low identifiability. Of course, as argued in the previous chapter, for ethical reasons the vast majority of psychological experiments take place in conditions of low identifiability (anonymity) and stereotype threat research generally takes place in contexts, such as Universities, in which every effort is taken to minimize race and gender conflict. In other words, the conditions that produced the strongest threat in the present study match the conditions of the typical stereotype threat experiment.

In this study, participants who reported low identifiability were likely to exhibit conventional stereotype threat and boost effects in the low conflict condition. However, those participants reporting higher identifiability exhibited reversed stereotype-performance effects,

performing better under threat and worse under boost. The same applies to participants under threat in the high conflict condition, who performed particularly well.

### **7.1.5. Discussion of Conflict Study 1**

The results of Conflict Study 1 are complex, but they offer support for: (1) the importance of considering resistance in stereotype-performance effects; and (2) the general hypothesis that perceived intergroup conflict is an important element of stereotype-performance effects in some situations.

#### **7.1.5.1. Resistance**

The results of this study demonstrated once again that it is not (always) true that simply knowing that others hold a stereotype about your group is enough to invoke STB-related performance shifts. On the contrary, in this study participants' resistance to the status manipulation was related to their performance on the stereotyped task and performance could not be predicted without considering resistance to the stereotype manipulation alongside the manipulation itself. Specifically, those who resisted the stereotype demonstrated a reversal of the conventional STB effect, performing better under threat and worse under boost.

It is possible that this effect could simply be a case of participants knowing themselves, assuming that they are prototypical of the group and projecting this self-knowledge of their own abilities onto the ingroup as a whole in their answers to the status manipulation check. However, since the status manipulation-check by which the status-resistance score is calculated was completed *before* the APM, it is also likely that participants' orientations to the group stereotype impacted on their engagement with the task and therefore on their performance.

#### **7.1.5.2. Conflict**

The effect of the conflict manipulation was not detectable in the model until reported individual identifiability was included, at which point a four-way interaction between

conflict, status, status-resistance and identifiability emerged. Unsurprisingly, the conditions under which the negative effects of stereotype threat were most pronounced (low conflict and low identifiability) are exactly those of a typical psychology experiment. If identifiability was high but conflict was low then participants performed particularly well under threat but more poorly under boost, in contrast to the predictions of the STB literature, but congruent with the findings of the previous chapter. In high conflict conditions the effects of status on performance were much less pronounced.

### **7.1.5.3. *Incongruent effects***

In this study it was demonstrated that negative and positive stereotypes can have opposite effects on performance depending on specific structural and experiential features of the intergroup setting. In this case these features were conflict and identifiability, but there is no reason to believe that these are the only socio-structural features that might have this type of effect.

Indeed, in this study, conventional stereotype threat effects were the exception rather than the norm. A strong conventional stereotype-threat effect was only observed in low-conflict conditions when participants reported low individual identifiability. It is unlikely to be a coincidence that these are the precise conditions in the vast majority of experimental contexts in social psychology. If these results are generalizable, these findings may go some way towards accounting for the notoriously tricky nature of stereotype threat research: detecting the effect at all is often hit-or-miss; the direction and size of the effect varies dramatically across studies (Nguyen & Ryan, 2008); no stable moderators or mediators have emerged (Smith, 2004) despite the large number of studies that have been undertaken; and the effect has not generalized well to real-life testing environments (eg. Cullen et al., 2004; Cullen et al., 2006).

The solution being suggested here is not simply to add “conflict” and “identifiability” to the ever-growing list of moderators of the STB. Instead, it will be argued that stereotype threat and boost should be considered as subtypes of more general stereotype-performance effects. The nature of the variables found to be implicated show that task performance in the context of this study was closely tied in to situated identity and, as such, is intelligible in a social identity framework. Therefore, stereotype-performance effects should not be considered as ends in themselves, but in light of their identity-constructing value in social life.

#### **7.1.5.4. *Slipstreaming under boost***

This analysis provides additional support for the slipstreaming hypothesis as a way of understanding underperformance in boost conditions. In the present study although participants performed worse under boost when they were anonymous in the high conflict condition, they were most likely to do so when they reported high identifiability in the low conflict condition. These results are in contrast to the studies in the previous chapter that consistently showed that poor performance under boost was most likely when participants were anonymous.

However, the previous studies did not take perceived intergroup conflict into account. In the present study the participants experienced an overtly cooperative interaction with the outgroup prior to completing the performance task. This task was presented in all studies as a means of competitively comparing the abilities of the ingroup and outgroup. Therefore it is likely that, first, the cooperative low conflict manipulation enhanced high status participants’ sense of identity security, since “clear-cut status differences may lead to a quiescent social system” in which open competition is avoided (Tajfel & Turner, 1979a), particularly by members of the dominant group since conflict risks destabilizing the favourable status hierarchy (Moscovici & Mugny, 1985). Second, a strong performance by someone with a

substantial advantage could potentially be perceived as unkind and disaffiliative – rubbing the outgroups’ faces in the ingroup’s superiority. By way of analogy, it would be something like a first-division squash player facing a beginner and never letting them reach the ball. In other words, it is quite possible that participants in the low-conflict boost condition performed worse when they were identifiable because performing too well could have been interpretable as an antisocial act. This interpretation emphasizes the need to consider the potential meaning of a strong or weak task performance in the specific performance context.

Although these explanations are conjectural, they fit the results well and support the emerging theory of stereotype-performance effects as an outcome of situated identity performance. They are also a reminder that slipstreaming is likely to occur in conditions of secure identity and when good task performance is either unnecessary or even an identity liability in the performance context. Therefore the finding in the previous chapter that slipstreaming is most likely under anonymous conditions is only likely to hold to the extent that anonymity contributes to secure identity. Underperformance in boost conditions, therefore, cannot be predicted by a specific set of universal predictors, but requires an understanding of the social and individual meanings and consequences of task performance and the relationship of these consequences to situated identity.

## **7.2. Conflict Study 2**

This study was a replication Conflict Study 1 with minor modifications to the measures to improve the design.

### **7.2.1. Research question**

This study aimed to further investigate the effect of intergroup conflict on stereotype threat and boost (STB).

### 7.2.2. Sample

Science students registered at UKZN were purposively sampled and 122 took part in the study. No participants were dropped during data cleaning procedures. Of these: 36 were assigned to the threat condition, 36 to the boost condition and 50 to the control condition; 58 were assigned to the low conflict condition and 64 to the high conflict condition; 52 reported their sex as female and 65 as male; 90 reported their race as black African, 6 as coloured, 11 as Indian and 11 as white.

### 7.2.3. Design and procedure

The design of this study was closely modeled on the first conflict study except that participants completed the Shipley Institute of Living Scale (SILS) prior to experiencing the manipulations. The Shipley has been shown to have a high correlation to the Raven's Advanced Progressive Matrices (APM) even though it has a completely different look and feel. Therefore it was an ideal covariate measure of *a priori* ability or performance since it is extremely unlikely to produce confounds such as practice effects. Apart from the addition of the SILS, the manipulations and measures were identical to Conflict Study 1.

While a conventional manipulation check on self-reported perceptions of intergroup conflict was not significant,  $t = -.576$ ,  $p = .566$ , a factorial ANOVA on perceived conflict by assigned conflict condition and status suggested that there was an interaction between status and conflict manipulations, with those in the threat and control conditions reporting much more conflict than those in the boost condition,  $F(4, 115) = 2.212$ ,  $p = .072$ . When the sample was filtered for those in the threat condition only, the difference in reported conflict between those assigned to the low and high conflict conditions was significant,  $t = -2.15$ ,  $p = .039$ .

Following the conflict manipulation, participants were assigned to a stereotype activation or control condition using the same manipulation described in Conflict Study 1.

Specifically, participants read a description of the study in which either a positive or negative stereotype was invoked regarding the performance of Science students on the APM invoking STB (ie. as a task requiring logic and mental accuracy, invoking lift; or requiring mental flexibility and the ability to process conflicting information, invoking threat).

Participants completed a short-form of the APM immediately following the conflict and status manipulations, along with a battery of social identity measures.

#### **7.2.4. Key findings**

##### **7.2.4.1. Status and conflict**

Since there was no *a priori* reason to believe that the effects of being assigned to the threat, control, or boost groups would be linear in that order, the status assignment factor would need to be dummy-coded for inclusion in a regression model which would make interaction effects more difficult to interpret. Therefore the analysis of the base model was done using a factorial ANOVA to aid interpretation, despite being less comparable with the analysis of other studies.

A 3x2 ANOVA on APM performance by status (low; high; control) and conflict (low; high) revealed that neither status,  $F(2,116) = 1.733, p = .181$ , nor conflict,  $F(1,116) < .001, p = .995$ , had direct effects on APM performance. However, the interaction between conflict and status was significant,  $F(2,116) = 3.354, p = .038, \eta^2 = .055$ . The estimated marginal means for this model are displayed in Figure 18, below, and show that the STB effect was reversed in the low conflict condition. Indeed, the highest performers were participants under *threat* and the lowest were those under *boost* in the low conflict condition. The STB effect in the high conflict condition was less pronounced and in the expected direction.

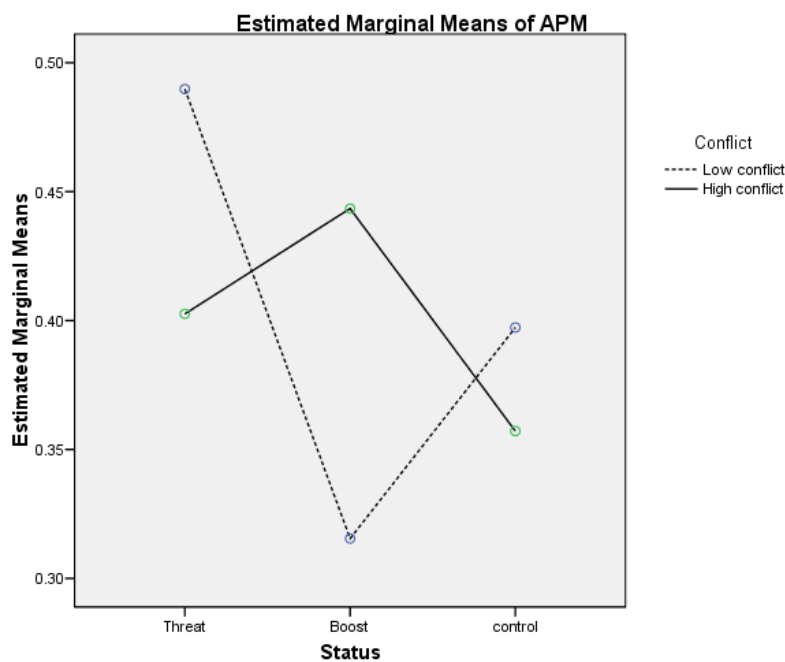


Figure 18: Estimated marginal means of APM by status and conflict

#### 7.2.4.2. Different effects of status and conflict at different baseline levels of performance

Steele and his colleagues have always argued that stereotype threat is most likely to affect the ‘vanguard,’ in other words that the top performing targets are most likely to experience stereotype threat effects (eg. Steele, 1997). Since this study incorporated the Shipley Institute of Living Scale (SILS) as a baseline measure of performance independent of, but closely related to, the dependent measure The SILS has been shown to be highly correlated to APM performance in other studies (Senior, 2001), with a median correlation of .72. In the present study the correlation between the SILS and APM results was 0.441,  $t_{120} = 5.377, p < .001$ , which – although highly significant – is lower than usual. However, it must be remembered that this relationship would have been attenuated by the status and conflict manipulations which, by definition, are designed to disrupt performance and thereby disrupt the ability to predict APM performance from SILS performance (the baseline).



Generally stereotype threat theorists argue that STB affects are generally only evident in top performers. The results of this study cast some light on why this might be the case: a regression analysis with status (dummy-coded for threat, boost and control conditions), manipulated conflict and baseline performance on the SILS yielded a highly significant model,  $F_{10, 111} = 4.065$ ,  $p < .001$ ,  $r^2 = .268$ ,  $adj.r^2 = .202$ . SILS performance was significant,  $t = 3.794$ ,  $p < .001$ ,  $b = .525$ , as was the two-way interaction between status dummy-coded for boost\*SILS performance,  $t = -2.084$ ,  $p = .039$ ,  $b = -.443$ , and the three-way interaction between status dummy-coded for threat\*conflict\*SILS performance,  $t = -1.996$ ,  $p = .048$ ,  $b = -1.126$ . Two additional interactions closely approached significance, namely status dummy-coded for boost\*conflict,  $t = 1.972$ ,  $p = .051$ ,  $b = .151$ , and status dummy-coded for threat\*conflict,  $t = 1.974$ ,  $p = .051$ ,  $b = .830$ .

While this model provides support for the general notion that habitual or baseline levels of performance are related to participants' experience of STB and the effects of STB on their performance, the specific effects are different to predictions from the stereotype-threat literature, which generally assume that moderate and poor performers are simply not subject to STB effects, or that they experience an effect of the same type but of different degree. Of key interest in understanding the relationship between baseline performance and STB effects are the status(boost)\*SILS and status(threat)\*conflict\*SILS interactions.

The status(boost)\*SILS interaction is plotted in Figure 19 and shows that participants with low baseline performance in the SILS were likely to perform at virtually the same level on the APM regardless of whether they experienced the boost manipulation or were in the control condition and therefore were exposed to no stereotype about their group at all. In contrast, for participants with high baseline performance, those who were exposed to a positive group stereotype performed significantly *worse* than their high-performing counterparts in the control condition.

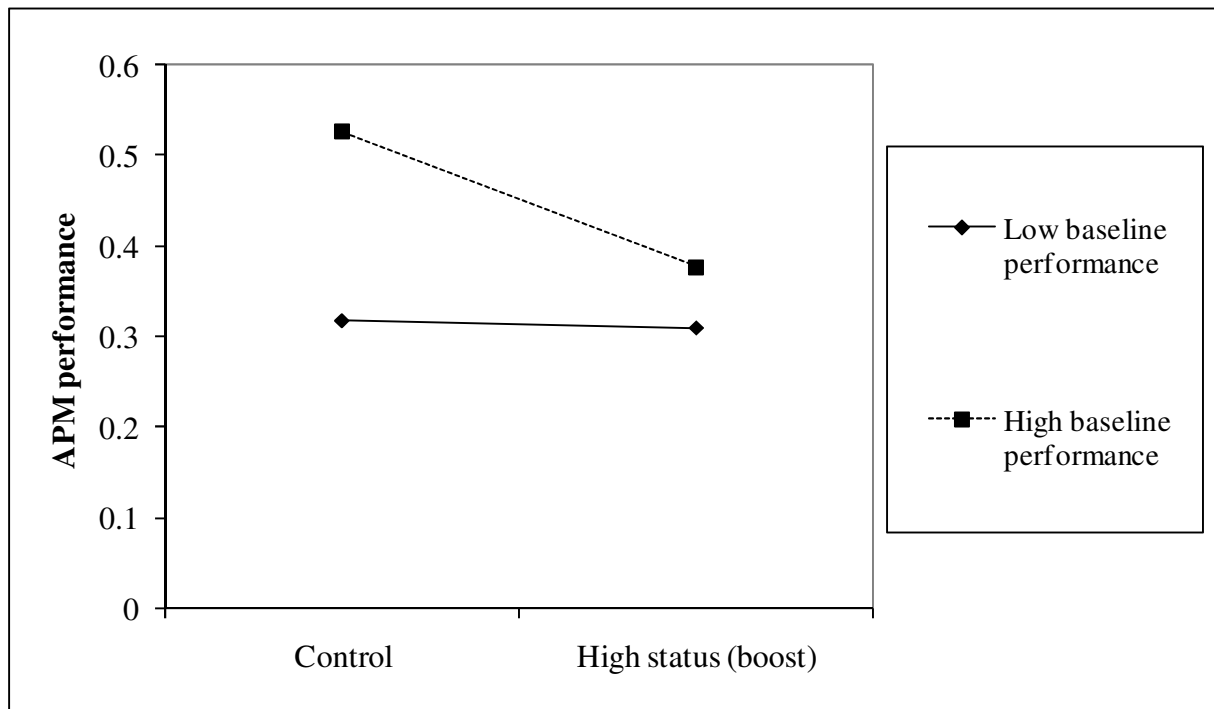


Figure 19: Interaction plot for Status(boost)\*SILS performance

The status(threat)\*conflict\*SILS interaction is plotted in Figure 20 and demonstrates that the effect of conflict and stereotype threat on APM performance was reversed for participants with low vs. high baseline performance on the SILS. Participants with low baseline SILS performed better under threat when conflict was *high*, but worse when conflict was low. Conversely, participants with high baseline SILS performance performed better under threat when conflict was *low* and worse when conflict was high. Thus, while conflict evidently changed the way that participants oriented to the performance of the APM task, this orientation depended somehow on their baseline SILS performance. It would be tempting to assume that SILS performance directly represents ability, but it is unlikely to be so simple, as will be discussed below.

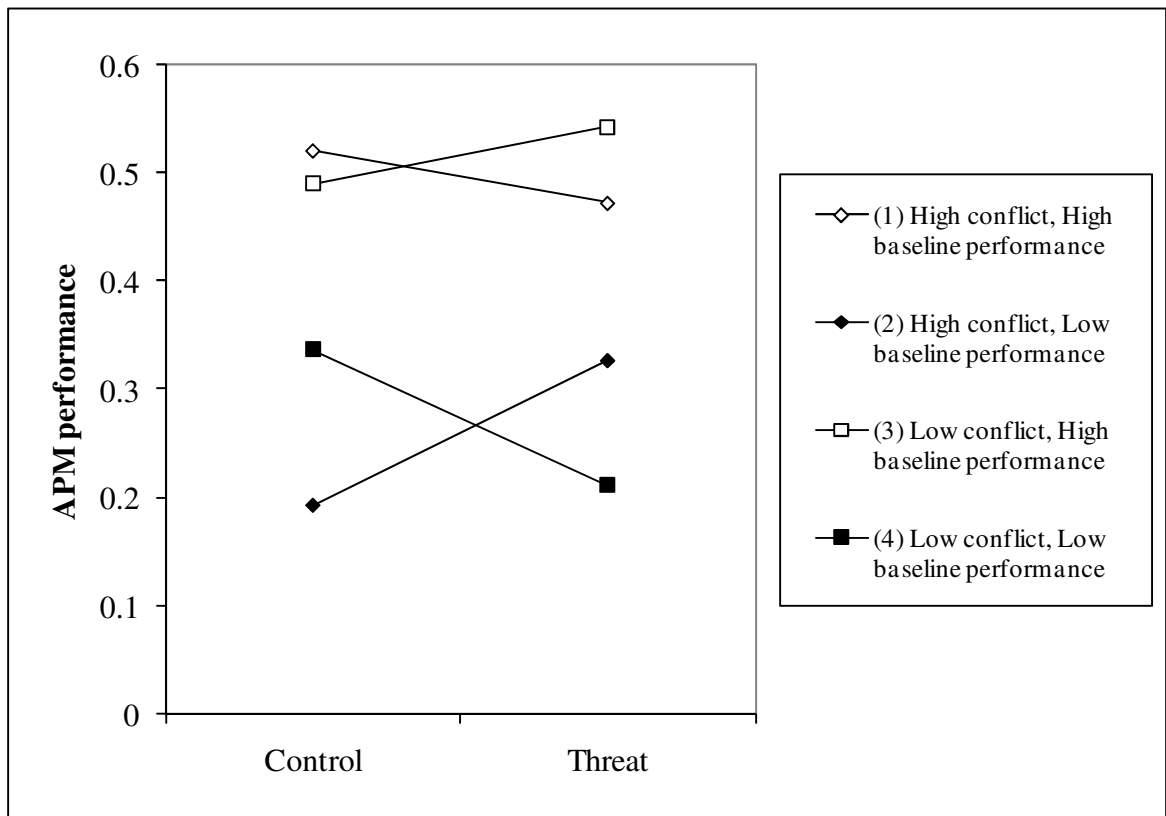


Figure 20: Interaction plot for status(threat)\*conflict\*SILS performance

### 7.2.5. Discussion of Conflict Study 2

The second study exploring the effect of conflict on the STB confirmed that perceived intergroup conflict is a potentially important feature of how participants approach task performance under stereotype threat or boost.

#### 7.2.5.1. *The importance of intergroup conflict in stereotype-performance effects*

The simplest model considering just status and conflict showed that reduced conflict resulted in a reversal of the conventional STB effect whereby participants in the threat condition performed particularly well and those in the boost condition performed worse than participants in any other cell in the ANOVA design. Taken in isolation, participants in the high conflict condition appeared to experience a conventional STB effect, performing better under boost than threat.

Once baseline performance on the SILS was included in the model, it emerged that the boost condition had most effect on participants who performed well on the baseline measure, and that the effect of the high conflict manipulation was to decrease APM performance. For participants under threat, both high and low baseline performers were affected. For low baseline performers, threat increased APM performance in low conflict conditions and inhibited it when conflict was high. For high baseline performers these effects were reversed, with threat reducing performance in high conflict conditions and increasing it in low conflict conditions.

The most general outcome of this study, then, was to confirm that intergroup conflict is not an incidental feature of stereotype-performance effects, but a feature that can result in genuine reversals of the STB effect in certain contingencies of identity. This is an important finding, since most STB studies take place in universities where generally every effort is made to eliminate intergroup conflict with respect to the categories of gender and race that are most commonly invoked in stereotype threat studies.

#### **7.2.5.2. *Slipstreaming***

These patterns of results confirm the general finding that positive stereotypes are frequently associated with performance decrements. The results also echo those of Conflict Study 1, in which participants were found to perform worse under boost in conditions of low conflict and perceived anonymity. In the present study the worst performers in the general model were those in boost and low conflict conditions. However, when conflict was high, participants under boost outperformed those under threat.

These findings provide additional support for the slipstreaming hypothesis that boost conditions can provide a secure identity position that allows participants to relax a little, or at least to invest less in the performance of the task. Low performers on the baseline measure were unaffected, since it seems that their investment was already low; but for those whose

initial performance was high, it seems likely that the experience of having a positive group stereotype affirmed gave them license to perform less well – particularly when they had also experienced a cooperative low-conflict interaction with the outgroup.

Of course, it is possible that this dip in performance was related to the additional pressure of having to represent one's positively stereotyped group and potentially failing to live up to expectations (Cheryan & Bodenhausen, 2000), but if this were the case then we would expect two additional patterns to have been observed: (1) participants who performed poorly on the baseline should also have been affected by this pressure, and possibly more so; and (2) intergroup conflict is unlikely to reduce pressure, so if the choking under pressure hypothesis were correct in this case then participants in high conflict conditions should not have performed better under boost as they did. These specific features of performance add weight to the argument that underperformance in the boost condition was related to secure identity positions releasing participants from the need to perform well as individuals rather than because of 'choking under pressure.'

#### ***7.2.5.3. Conflict affects targets' APM performance differently depending on their baseline SILS performance***

It also emerged that the impact of conflict on stereotype-performance effects varied according to participants' levels of baseline performance on the SILS, a task independent of the APM, but closely related to it. While it is tempting to jump to the conclusion that the baseline performance represents *ability*, this would be wrong, since it clearly represents a combination of factors such as task investment, motivation for completing the study, and ability that cannot easily be disentangled.

The truth is that many features have to align for participants to score well on a difficult task such as the SILS or APM: they must be motivated to perform well; they must be able to focus on the task; they must have the ability to do well at that type of task; they must

have an educational and cultural background that has naturalized such tasks and given them an analytic framework with which to understand and tackle the items, and so on. However, although a large number of features must align in the performance context for participants to do well at a task such as the APM, poor performance requires just one point of failure. In other words, there are many more reasons for someone to do badly on a task than to do well. Therefore, when a participant emerges with a poor baseline score it is equally likely to mean that they were not particularly invested in the task (for example if they were participating only for the cash incentive) as that they lacked the ability to do well.

When the measure of baseline performance (whatever it may represent) was included in the model, it emerged that conflict had opposite impacts on the stereotype-performance effect at different levels of baseline performance. High baseline performers performed better under threat than in the control condition when conflict was *low*, but low baseline performers performed better under threat than in the control condition when conflict was *high*. While this is a fascinating pattern that deserves further enquiry, the data available in this study do not allow much to be said about the potential reasons and motivations for this difference in performance in relation to conflict and status.

However, one conjectural explanation is that high performers on the baseline measure are already performing close to their ceiling and therefore are likely to be subject to regression to the mean – that is, any disruptive influence is more likely to reduce their performance than to increase it (eg. Bland & Altman, 1994). Therefore the additional emotional and motivational load involved in experiencing conflict is more likely to decrease performance than to increase it. In contrast, many participants with low baseline SILS scores may have been galvanized into action by the (simulated) experience of intergroup conflict because the intergroup competition gave them a good reason to invest in the performance task. Consequently, they performed much better under threat.

The important thing to note is that this pattern of results cannot be understood without acknowledging that performing the APM task in this particular experimental context was closely related to sociostructural features of the intergroup context. Changing participants' experience of the intergroup context (such as providing them with a cooperative or conflictual interaction with the outgroup) resulted in different reactions to the status manipulation for participants with low and high baseline performance. Therefore, the clearest conclusion that can be made from this evidence is that APM performance was oriented to, and interpretable within, the social demands of the intergroup context.

### **7.3. Conflict Study 3**

Conflict Study 3 was a replication of Conflict Study 2 using humanities students instead of science students (ie. filling in the other half of the design). A methodological advance was that participants completed all materials on computer workstations running E-Prime in a dedicated research facility rather than on paper in a temporary venue as in Conflict Study 2, increasing the realism of the conflict manipulation.

#### **7.3.1. Research question**

This study aimed to replicate the findings of Conflict Study 1 and Conflict Study 2, but using humanities students instead of science students. This increases the scope of the results by avoiding the criticism that there is something unique about science students as a group that produced the results in the first two studies.

#### **7.3.2. Design**

The design was identical to the two previous conflict studies except that humanities students were recruited instead of science students. To recap, the study employed a 2 x 2 factorial design with conflict (low; high) and status (low; high) on APM performance. Participants completed the Shipley Institute of Living Scale prior to the manipulations and other measures as a baseline test of performance, and completed a number of measures of

identity and perceptions of sociostructural features of the intergroup context, including: conflict; stability; individual identifiability; effort; category salience; intergroup closeness; permeability; group identifiability; legitimacy; ingroup identification; achievement motivation and social desirability.

The materials were completed in a laboratory of nine PC's running the E-Prime experimentation environment (Psychology Software Tools, 2010). Since the low and high conflict manipulations were visually distinctive, each batch of up to nine participants were pre-assigned to one of the conflict conditions using a predetermined schedule. Each computer in each batch was assigned to one of the status conditions using the predetermined randomization schedule. Participants were randomly assigned to a batch and a seat number within the batch as they confirmed their availability for the session.

The conflict manipulation used the same simulation developed for the previous two conflict studies, but updated for participants to engage individually, where in the previous studies participants had engaged with a single computer screen in groups of three. The program simulated an interaction between the humanities participants and a group of science students believed to be in a venue in a building on a different campus. The first part of the simulation was devoted to setting up a (fake) network connection and "selecting" a group of three participants from the larger group to work together. This group was only represented visually and no real interaction took place.

The second part of the simulation required participants to assign points to the science students which, they were told, would determine the outgroup's remuneration at the end of the study. At the same time the virtual outgroup were choosing points worth money for the actual participants. These points were chosen from a set of forced choices similar to Tajfel's matrices (Bornstein et al., 1983) and forced participants to select from three combinations of ingroup and outgroup points. These were carefully designed to correspond to intergroup



strategies: the first allocated the ingroup more points than the outgroup, even though the ingroup total was less than for other options (a win at all costs option); the second ensured fairness since both groups were awarded equal points; and the third represented cooperation, since the outgroup were awarded more points than the ingroup but the total ingroup points were higher than the first option even though lower than outgroup points. In the low conflict condition the virtual outgroup 'selected the third option in each of five rounds, demonstrating that they were willing to cooperate to achieve maximum benefit for both groups. In the high conflict condition the virtual outgroup 'selected' the first option in each of the five trials, demonstrating that they were competitively and punitively willing to sacrifice their own points just to ensure that the humanities students received less than they did.

After completing the conflict manipulation participants completed the STB manipulation check, the APM and the rest of the measures. Once they were finished they completed a written debriefing procedure, were given a chance to ask questions and given a flat-rate incentive regardless of the points awarded in the manipulation procedure.

### **7.3.3. Sample**

Humanities students were recruited on the Pietermaritzburg campus of the University of KwaZulu-Natal via pamphlets, posters, word-of-mouth and face-to-face solicitation. The sample consisted of 137 students, of whom nine were dropped because they turned out not to be humanities students after all, resulting in a final sample size of 128. Of these: 83 were female, 39 as male and 6 declined to report their sex; 102 identified themselves as black African, 3 as Indian, 8 as coloured, 8 as white, and one did not respond; 66 were assigned to the threat condition and 62 to the boost condition; and 61 were assigned to the low conflict condition and 67 to the high conflict condition.

### 7.3.4. Key findings

#### 7.3.4.1. Base model

As with most other studies so far, there was no significant direct relationship between status and performance,  $r = -0.092$ ,  $t_{128} = -1.038$ ,  $p = .3013$ . The model did not improve when conflict and status\*conflict were included,  $F_{3, 124} = 1.082$ ,  $p = .359$ ,  $\Delta r^2 = 0.017$ .

#### 7.3.4.2. Resistance

A fully saturated model regressing status, status-resistance, conflict and conflict-resistance on APM performance (with a square-root transformation to correct skewness) was highly significant,  $F_{15, 106} = 2.275$ ,  $p = .008$ ,  $r^2 = .244$ ,  $adj.r^2 = .137$ . The only effects in the model that were significant were the simple effect for status,  $t = -2.392$ ,  $p = .029$ ,  $b = -.135$ , the status\*conflict interaction,  $t = 2.286$ ,  $p = .024$ ,  $b = .016$ , and the status\*status-resistance interaction,  $t = -2.523$ ,  $p = .013$ ,  $b = -.125$ . The interaction plots for these two significant interactions are displayed in Figure 21 and Figure 22, below.

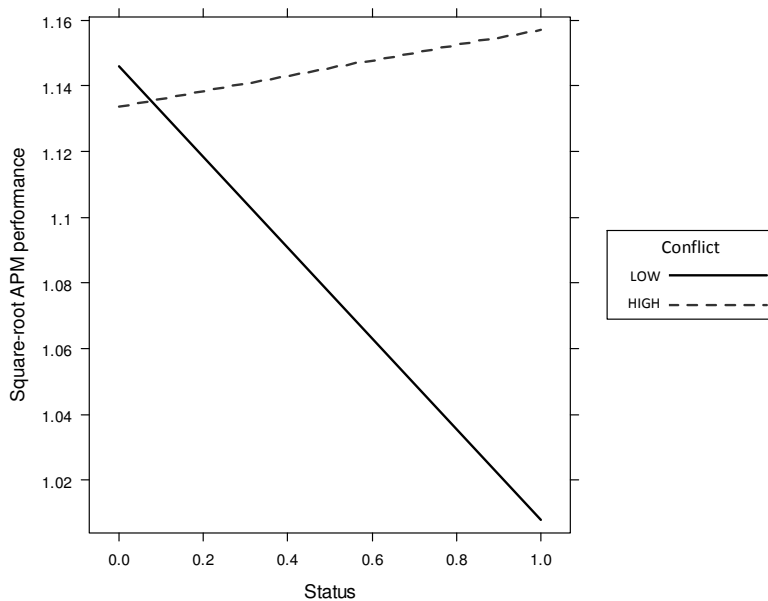


Figure 21: Interaction plot for status \* conflict

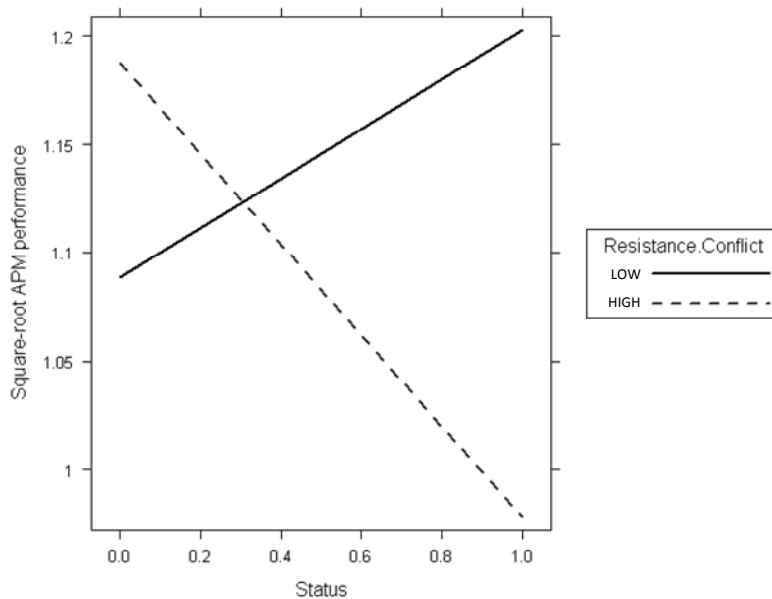


Figure 22: Interaction plot for status\*status-resistance

Figure 21 shows that, once resistance was accounted for, the most powerful stereotype-performance effect was the reduction in performance for participants in the high-status condition. However, there was also a significant interaction between status and conflict-resistance on APM performance (Figure 22) such that the stereotype-performance effects were in the expected direction for participants who did not resist the conflict manipulation and reversed for participants scoring higher on conflict-resistance. The worst performers overall were participants in the boost condition with high conflict-resistance, but similar conflict-resistance participants performed very well in threat conditions.

### 7.3.5. Discussion of Conflict Study 3

The first two conflict studies sampled science students and invoked high and low status stereotypes with humanities students as the outgroup. Conflict Study 3 sampled humanities students and invoked science students as the outgroup to explore whether findings

replicated. The results confirm several findings of the previous studies and reinforce some recurring themes in the programme of studies as a whole.

#### **7.3.5.1. *Status and conflict***

Unlike in Conflict Study 2, the model including only the status and conflict manipulation came nowhere near significance. It is possible that the more diffuse and heterogeneous identity of “humanities” students as a group compared to “science” students as a group contributes to this difference, but this is entirely conjectural.

#### **7.3.5.2. *Resistance and slipstreaming***

Confirming the results of previous studies, adding resistance variables improved the ability to model APM performance and revealed the significant impact of status. The model demonstrated that status interacted significantly with conflict, and also with status-resistance. Supporting the findings of the studies reported previously, the participants most likely to perform poorly were those under boost in the low conflict condition (when looking at the status\*conflict interaction) or those under boost who expressed high resistance to the conflict manipulation (looking at the status\*resistance.conflict interaction). Therefore, whichever way you look at it, in this study participants in the boost condition were most likely to perform poorly.

However, when high status participants were also exposed to the high-conflict manipulation, or when they did not resist the conflict manipulation, they performed particularly well. This pattern of results adds to the weight of evidence supporting the slipstreaming hypothesis that good performance under boost is most likely when identity is under threat in some way and poor performance is most likely when positive group or individual identity is secure and not in need of defense. In this case, the high status boost condition produced excellent performance when high status group members were subject to a conflictual interaction with the outgroup which they did not symbolically resist.

These results also support the tentative hypothesis developed in the previous two studies that participants in high status conditions (under boost) may perform worse as a way of smoothing-over intergroup differences and conflict. Since conflict can destabilize the intergroup status hierarchy and thereby place the high-status group's position at risk (Moscovici & Mugny, 1985), it may often be in the interests of the high-status group to minimize competitive differences such as those invoked in stereotype threat studies (Tajfel & Turner, 1979a). This is likely to be particularly true of "model minority" members, such as the Asian women in the math domain studied by Cheryan and Bodenhausen (2000), who might prefer to avoid making themselves – and their group superiority – conspicuous.

#### **7.4. General discussion of conflict studies**

SIT has shown that intergroup conflict has several effects that may be relevant to stereotype-performance effects, and three studies have confirmed that this is the case. In all of the studies, intergroup conflict was observed to modify the effects of stereotypes on performance, often resulting in reversals in the direction of STB effects. However, the effects of conflict on stereotype-performance effects were contingent on other features of participants' experience in the experimental context, such as resistance to the status manipulation and perceived individual-identifiability. For example, in one study conflict impacted performance differently depending on participants' baseline levels of performance – motivating low baseline performers but disrupting performance for those who were already performing well.

However, just as conflict was shown to have opposite effects on performance depending on other salient features of individual identity and socio-structural conditions in the performance context, status was also shown to have contingent effects on performance. In fact, stereotype-performance effects were often reversed depending on specific structural and experiential features of the intergroup setting. These results call into question, once again, the

way that mainstream stereotype-threat has been conceptualized as a unidirectional effect (Steele, 2010; Steele et al., 2002). Instead, these results confirm that STB should be considered as subtypes of more general stereotype-performance effects.

The results of the present studies demonstrate that perceived intergroup conflict is one feature that can transform the direction of stereotype performance effects. For example, for high baseline performers (traditionally the target of stereotype threat studies) conventional threat effects were observed when conflict was high, but when conflict was low threat improved performance. However, for low baseline performers threat reduced performance when conflict was low and increased it when conflict was high.

Stereotype boost conditions were also routinely associated with reduced performance. However, the specific circumstances in which high status was associated with reduced performance were consistent with the slipstreaming hypothesis rather than ‘choking under pressure’ (eg. Cheryan & Bodenhausen, 2000; Smith & Johnson, 2006). For example, reduced performance was likely for participants assigned to high-status conditions when conflict was low. In contrast, high status participants in high conflict conditions generally performed particularly well, exhibiting congruent STB effects. Since it is unlikely that increased conflict *reduced* performance pressure, the results support the slipstreaming hypothesis that participants performed worse under boost because the specific conditions provided a safe place in which to do so.

Indeed, there are convincing arguments that high performance in conditions of low conflict and high status could be an identity liability: when participants had experienced a pleasant and cooperative interaction with the outgroup then immediately performing well in the experimental context could potentially be interpreted as rubbing the noses of the low status group in their own group’s success.

These instances of slipstreaming remind us that the *meaning* of good or bad task performance in the intergroup context is critical to the specific form of stereotype-performance effects. Since intergroup conflict transforms the meaning of task-performance in specific performance contexts, it is therefore likely to be an important feature of stereotype-performance effects.

In other words, conflict is not an incidental feature of stereotype performance effects, but can be critical to the pattern of stereotype-performance effects observed in specific performance contexts. For example, the performance gap at universities could be addressed by specifically invoking intergroup conflict in certain conjunctions of other sociostructural conditions<sup>19</sup>. This controversial suggestion highlights the fact that the majority of contexts in which stereotype-threat studies have been undertaken – such as at universities – actively and explicitly avoid overt gender and race conflict, which may have influenced results of STB studies to date.

---

<sup>19</sup> Although I am not suggesting this as a serious intervention because there would be many unforeseen consequences.

## Chapter 8. Results IV: Permeability

From the earliest work in stereotype threat (eg. Aronson, Lustina, Good, Keough, & Steele, 1999; Steele, 1997) the notion of “domain identification” has been considered important to the phenomenon of stereotype threat or boost (STB) and the importance of identification with the domain or group has been empirically demonstrated to be an important mediator and moderator of STB (eg. Leyens, Desert, Croizet, & Darcis, 2000; Schmader, 2002), although its effects have been unreliably detected and the mechanisms by which identification influences STB have not been clearly articulated across the studies that have found it to be important (see Smith, 2004 for a review).

Permeability, or the extent to which intergroup boundaries are perceived as open or closed, is one of the key socio-structural variables in Social Identity Theory (SIT) that predicts how individuals are likely to position themselves in intergroup settings. Since empirical studies of permeability have demonstrated that it has a powerful influence on ingroup identification, it is highly likely that perceived permeability will be related to targets' sense of identification or disidentification with the ingroup in the STB context.

This is because perceived permeability is likely to influence status enhancement strategies: when permeability is high, then individuals are more likely to adopt *social mobility* strategies and enhance their status by working to join, or at least aspiring to join, the high status group; on the other hand, when permeability is low then it becomes difficult or impossible to transfer individually to the high status group and other strategies, such as *social change* strategies, or actively finding cognitive alternatives, become more likely (Ellemers, Spears, & Doosje, 2002; Tajfel, 1978a, 1978d).

One way of understanding ingroup identification is as an affective motivation and/or lay explanation for the status enhancement strategy that is made appropriate by specific features of the sociostructural context, such as intergroup permeability (Ellemers, 1993;



Ellemers, Spears, & Doosje, 1997; Ellemers et al., 2002; Ellemers et al., 1988; Ellemers et al., 1990; Ellemers et al., 1993; Mummendey et al., 1999). Specifically, when permeability is low (precluding individual mobility strategies), low-status group members are likely to identify more with their own group. When permeability is high (allowing the possibility of individual mobility between groups), low-status members are likely to identify less with the low-status group and more with the high status group as a psychological precursor to individual mobility between groups. Conversely, it is theorized that high status group members with insecure positions within the group are likely to identify more strongly with the ingroup when permeability is high and they risk being excluded from the high status group (Ellemers, 1993).

STB researchers have generally focused on *domain* identification (eg. Aronson et al., 1999; Keller, 2007; Lawrence et al., 2010; Smith, 2005; Smith & White, 2001; Steele, 1997) rather than *group* identification with a few exceptions (eg. Armenta, 2010; Clark et al., 2010). However, social identity theory predicts that interest and investment in a *domain* will be important only to the extent that it allows the advancement of individual or group identity.

In many cases it is possible that a domain is synonymous with a group identity and domain identification will be synonymous with group identification. For example, becoming a doctor (a member of a particular social group) requires strong interest and achievement in medicine (a domain). Indeed, these contexts where domain and group are tightly coupled are likely to be the contexts in which STB effects are most prevalent. However, in many cases domain identification is less tightly coupled with group identity (such as between the domain of Philosophy and the group humanities students). Since social identity theory predicts that domain identification will be secondary to group identification, the present studies focus on the latter.

Three studies were undertaken manipulating status and permeability that allowed the direct influence of permeability on STB to be assessed. The first two applied the same design with similar real-category samples in consecutive years. However, it is very difficult to manipulate the perceived permeability of real group categories, so the third study employed a minimal groups design in which permeability was more clearly manipulable.

## **8.1. Permeability Study 1**

### **8.1.1. Research question**

This study aimed to explore whether perceptions of permeability influenced task performance under STB conditions, or participants' experience of identity under STB. Since, to my knowledge, it was the first study ever to explore the relationships between permeability, ingroup identification and stereotype-performance effects, there were no clear or specific hypotheses beyond the very general prediction that permeability would influence ingroup identification and consequent task performance in ways consistent with SIT.

### **8.1.2. Sample**

One-hundred students participated in the study and all were registered for undergraduate degrees at Varsity College, a private tertiary college providing private tuition for degrees awarded through UNISA, South Africa's distance University. More than half (N=62) were black African, 12 were coloured, 8 were Indian and 17 were white. Just more than half were male (N=57; 57%). The design was perfectly balanced, with 50 participants each in the threat and boost conditions; and 50 each in the low and high permeability conditions.

### **8.1.3. Design**

The study used a 2x2 factorial design, with permeability (low; high) and status (low; high) as independent variables and performance on a short form of the Raven's Advanced Progressive Matrices (APM) as the dependent variable. Participants were informed that their

performance on the APM was a measure of their academic ability that would predict their future academic success and, that their institution (Varsity College) was being compared to the neighboring University of KwaZulu-Natal (UKZN), a large traditional university.

Status was manipulated by informing the low status group that small institutions produce students with lower academic ability because they *lack* “extensive academic facilities,” “large class sizes fostering independent thinking,” “large campuses resulting in greater access to resources,” and “large departments resulting in better access to lecturers.” Those in the high status group, on the other hand, were told that smaller institutions tend to produce students higher in academic ability because they have “intimate academic facilities,” “small class sizes fostering independent thinking,” “small campuses resulting in greater access to resources,” and “small departments resulting in better access to lecturers.”

The permeability manipulation was achieved by informing participants in the low permeability condition that students from small institutions such as Varsity College who transfer to postgraduate studies in large institutions such as UKZN “find it exceptionally difficult to make the transition” because they “cope very poorly with different styles of lecturing and assessment,” “are unable to develop productive relationships with lecturers,” and “adjust slowly to unfamiliar social networks.” Students in the high permeability condition were told the opposite, namely that transfer students do very well because they “cope very well with different styles of lecturing and assessment,” “are able to develop productive relationships with lecturers,” and “adjust quickly to unfamiliar social networks.” This design echoes that of Platow, Byrne, and Ryan (2005) who manipulated the relative status of students at neighboring universities in Australia with similar descriptions.

After reading the study descriptions (and thereby experiencing the relevant manipulations), participants completed a status manipulation check, completed a short form

(12 item) version of the APM and then completed a social identity inventory assessing key social identity variables.

#### 8.1.4. Key findings

##### 8.1.4.1. Status and permeability on ingroup identification

As predicted by SIT, a regression on ingroup identification by permeability and status was highly significant,  $F_{3,96} = 4.191$ ,  $p = .008$ ,  $r^2 = .116$ , with a significant main effect for permeability,  $t = -2.335$ ,  $p = .022$ ,  $b = -.674$ , and a significant interaction between status and permeability,  $t = 2.598$ ,  $p = .011$ ,  $b = 1.061$ . The interaction is displayed in Figure 23, below, and shows that ingroup identification was similar for participants assigned to low and high status groups when permeability was low, but when permeability was high then ingroup identification was lower for members of low status groups and higher for members of high status groups. This pattern is consistent with SIT predictions (eg. Ellemers et al., 1988), as discussed in the literature review (section 3.6.9.3, p. 101).

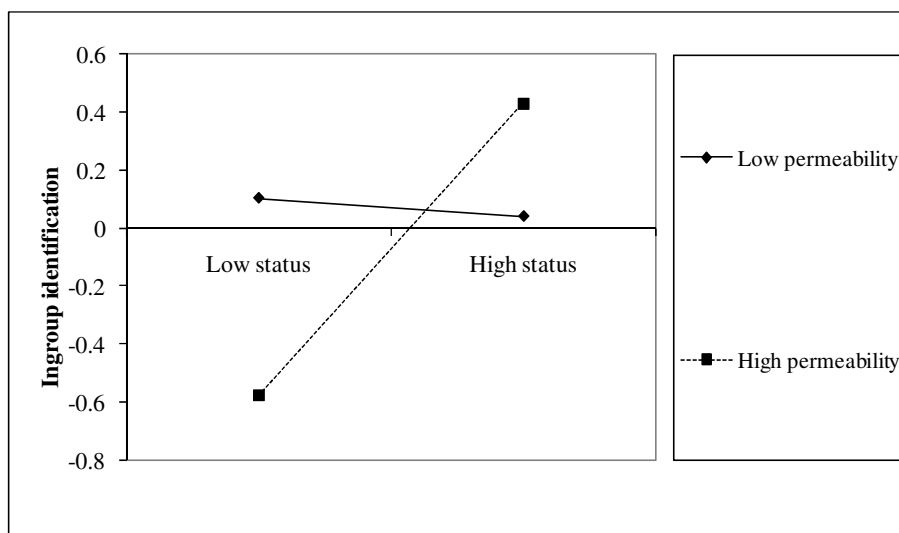


Figure 23: Interaction plot for status\*permeability on ingroup identification

#### **8.1.4.2. Status and permeability on APM performance**

Contrary to the experimental hypothesis, the model regressing status, permeability and status\*permeability on APM performance was not significant,  $F_{3, 96} = .155$ ,  $p = .926$ ,  $r^2 = .005$ . Even when ingroup identification was added to the model (in a fully saturated model regressing status, permeability and ingroup identification on APM performance) the results were not significant,  $F_{7, 92} = .324$ ,  $p = .942$ ,  $r^2 = .078$ . So, although status and permeability were shown to influence ingroup identification as predicted, these factors did not have any significant linear effect on APM performance.

#### **8.1.4.3. Status, permeability and resistance on APM performance**

A saturated model regressing status, status-resistance, permeability and permeability-resistance<sup>20</sup> on APM performance was more successful. Although the model did not achieve significance,  $F_{14, 85} = 1.338$ ,  $p = .263$ , the permeability\*permeability-resistance interaction was significant,  $t = -2.236$ ,  $p = .028$ ,  $b = -.145$ . A further two interactions approached significance, namely status-resistance\*permeability-resistance,  $t = -1.772$ ,  $p = .080$ ,  $b = -.060$ , and status-resistance\*permeability\*permeability-resistance,  $t = 1.892$ ,  $p = .062$ ,  $b = .083$ . Although this model was not significant and will not be considered further, it is noteworthy that, once again, the effect of status on APM performance (ie. STB) was more successfully modeled once participants' resistance to the experimental manipulations were acknowledged.

#### **8.1.4.4. Status, permeability, and ingroup identification**

Given the theoretical importance of ingroup identification in relation to permeability and STB, a model was fitted with the status and permeability manipulations and ingroup identification. The model (with a log transformation on APM performance and ingroup-identification and dropping three outliers) did not achieve significance,  $F_{7,87} = 1.510$ ,

---

<sup>20</sup> The calculation of resistance metrics is discussed in previous chapters

$p = .174$ . However, the theoretically critical three-way interaction between status, permeability and ingroup identification was highly significant,  $t = 2.547$ ,  $p = .013$ ,  $b = .710$ . This effect will be interpreted despite the non-significance of the omnibus  $F$  test because this interaction was the key *a priori* hypothesis in this study (as justified by Bedeian & Mossholder, 1994).

The interaction plot for this significant interaction is displayed in Figure 24 and shows that participants were most likely to perform particularly well under boost when they had low identification with their own group and were assigned to the low permeability condition.

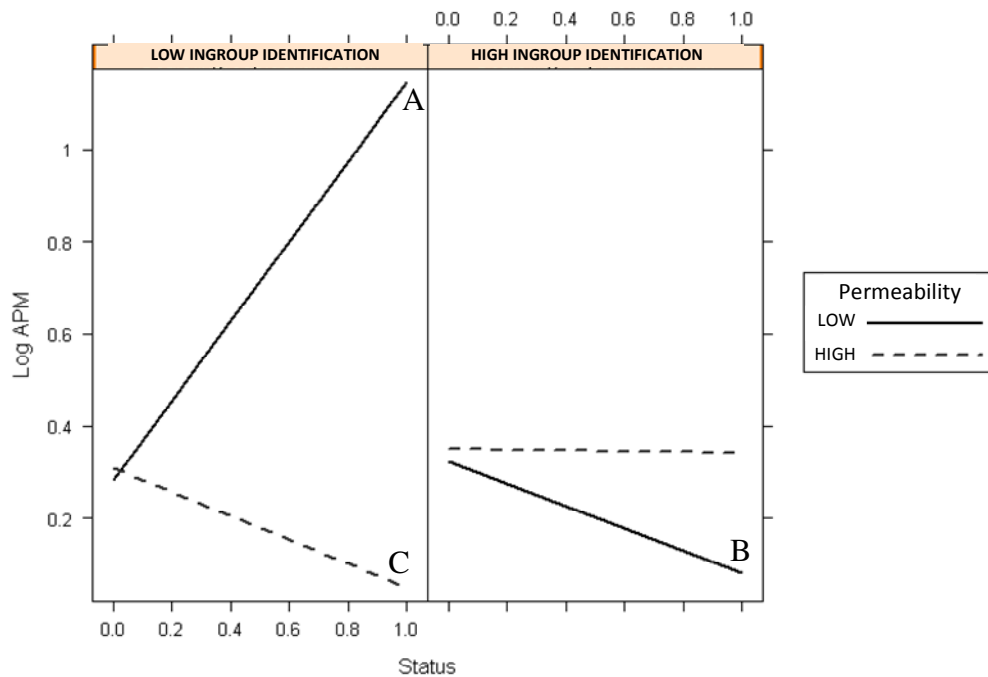


Figure 24: Interaction plot for status\*ingroup-identification\*permeability on log APM performance in Permeability Study 1

#### 8.1.4.5. Ingroup identification on APM performance

Taken alone, ingroup identification did not predict APM performance,  $F_{1,98} = .090$ ,  $p = .765$ .

#### 8.1.4.6. *Status, permeability and group differentiation*

A fully saturated model regressing status, permeability and group-differentiation on APM performance was highly significant,  $F_{7, 92} = 4.12$ ,  $p < .001$ ,  $r^2 = .239$ ,  $adj.r^2 = .181$ . The simple effect for status was significant, as were the two-way status\*permeability and status\*group-differentiation interactions, but these will not be directly interpreted because they were subject to the significant three-way status\*permeability\*group-differentiation interaction,  $t = -2.619$ ,  $p = .010$ ,  $b = -1.072$ . The interaction plot is displayed in Figure 25. In the low permeability condition stereotype-performance effects were magnified and in the expected direction for participants reporting high group differentiation and reversed for participants reporting low group differentiation. These participants reporting low group-differentiation assigned to the low permeability condition performed poorly generally, but more so when also under boost. These effects were reversed in the high permeability condition, but were small and unlikely to be significant.

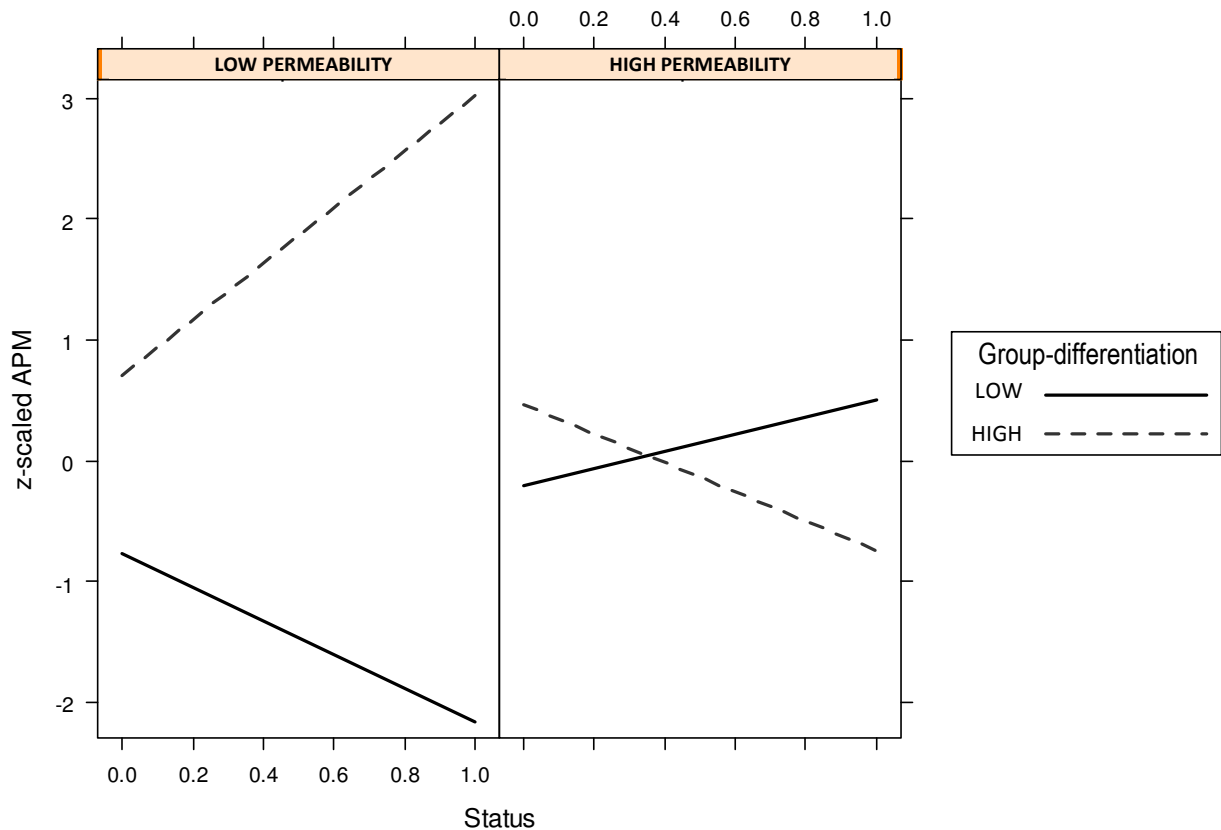


Figure 25: *Status\*permeability\*group-differentiation on z-scaled APM performance for Permeability Study 1*

### 8.1.5. Discussion of Permeability Study 1

This study aimed to explore the effects of permeability on stereotype-performance effects since: (1) SIT makes clear theoretical predictions linking permeability and ingroup identification; and (2) domain identification and ingroup identification have been amongst the most consistent mediators and moderators of the STB effect established to date.

SIT predicts that members of low status groups with enough individual ability to realistically leave their group are more likely to identify with the high status group when permeability is high. However, if permeability is low and it is impossible to gain entry to the high status group then SIT predicts that low status group members will identify more with their own group regardless of their individual ability. For high status group members, ingroup



identification is likely to be highest for members with insecure group membership who may fear exclusion (Ellemers, 1993; Ellemers et al., 1988).

This pattern was well supported by the results of the present study. As predicted, permeability and status jointly predicted ingroup identification. In high permeability conditions, low-status group members expressed lower ingroup identification (and presumably, higher identification with the outgroup) and high-status group members expressed higher ingroup identification.

Despite the fact that ingroup identification was influenced by status and permeability manipulations precisely as predicted by SIT, there were no direct effects of status and permeability on APM performance unless additional factors were considered. As in previous studies, considering participants' resistance to the manipulations improved the extent to which APM performance could be predicted, confirming that the relationship between stereotypes and performance in this context was neither automatic nor deterministic. On the contrary, APM performance could not be modeled from knowledge of participants' status or permeability manipulations without also knowing something about how they positioned themselves in relation to those manipulations (as represented by resistance) and the various groups in the intergroup context (represented by ingroup identification and group differentiation).

The fact that the status and permeability only predicted APM performance in conjunction with ingroup identification tells us something important: that there was variability in ingroup identification over and above that which was modeled by the status\*permeability relationship that contributed to participants' responses to the APM. However, ingroup identification in isolation did not predict APM performance at all. Therefore, there is something about the way that status, permeability and ingroup identification combine that is useful for understanding stereotype-performance effects.

As predicted by SIT, those with a secure position in the high status group (high ingroup identifiers in the low permeability condition) had no need to demonstrate their prototypicality and did particularly badly (labeled 'B' in Figure 24). This is interpretable as a case of slipstreaming. However, their high status counterparts who exhibited low ingroup identification and high permeability performed equally badly (labeled 'C'). The participants who did best of all, by a large margin, were those in the high-status low permeability condition who reported low ingroup identification (labeled 'A').

These patterns are not easily interpretable in terms of the standard SIT model, which would probably predict that participants assigned to the high permeability condition would perform best, particularly those in the low-status group, in order to demonstrate their prototypicality of, and suitability for, the high status group. Understanding these particular results requires a closer look at the particulars of the situation.

To effect the permeability manipulation, participants were all told that members of their group (Varsity College students) either find it easy or difficult to transfer to *postgraduate* studies at traditional universities, thus presumably invoking low or high permeability conditions. However, a critical feature of the manipulation is that (1) not all of the participants would be interested in, or eligible for, postgraduate degrees and (2) Varsity College offers a very limited portfolio of postgraduate options, requiring most students to transfer to a different institution in order to pursue postgraduate studies.

Therefore the permeability manipulation is likely to have impacted on participants differently depending on their personal abilities and academic aspirations. Ellemers and her colleagues (1988) found that individual ability beliefs interacted with permeability to predict ingroup identification, and it seems likely that the present results reflect participants own aspirations and abilities as well as their manipulated perceptions of intergroup permeability

and group status. Unfortunately, no measures of ability perceptions or postgraduate aspirations were included in the materials and it was not possible to pursue this further.

However, the large difference in APM performance between participants in the boost condition reporting low ingroup identification in low and high permeability conditions (marked A and C in Figure 24) is interpretable when the particulars of permeability in the performance context are considered. Low ingroup identification represented low affiliation to Varsity College which, for students with postgraduate aspirations, probably meant higher affiliation with the 'traditional institutions' which were the comparison group.

Students in the high permeability condition would not have had to prove their competence, since the permeability manipulation had informed them that members of their group generally found it very easy to transfer to traditional institutions. However, students in the low permeability condition were told that Varsity College students generally found it very difficult to transfer to postgraduate programmes at traditional institutions. Therefore their outstanding performance is intelligible as an act demonstrating their suitability for outgroup membership *despite* the generally low intergroup permeability. In other words, the permeability manipulation informing them that transfer was difficult (but not impossible) seems to have motivated participants in this predicament to perform particularly well to demonstrate that they, as individuals, were exceptions to the general permeability rule. If this argument is plausible, then the moderate performance of participants reporting low affiliation to Varsity College in the *low* status and low permeability condition is interpretable as a case of conventional stereotype threat, since the only systematic difference in the contingency experienced by these two groups of participants was the low or high status manipulation.

This argument is strengthened by the highly significant model regressing status, permeability and group-differentiation. Participants with postgraduate aspirations and low ingroup identification (with Varsity College) are precisely those who would be expected to

differentiate most clearly between the two institutions. It was these participants with high group differentiation in the low permeability condition who performed best overall. This pattern is congruent with the explanation that – since intergroup transfer was the only option for participants wishing to pursue postgraduate studies – those with the ability and aspiration to do so performed particularly well in the low permeability condition to demonstrate their individual ability and suitability for outgroup membership despite the low permeability generally applicable to other group members. This interpretation also explains why participants with high group differentiation did better than participants reporting low group differentiation regardless of their status assignment.

The underperformance of participants in the high status and low permeability condition who reported low group differentiation can therefore be interpreted as genuine slipstreaming, since they had neither the need nor inclination to transfer to the outgroup. They were therefore able to perform as badly as they liked without risking their positive group identity.

## **8.2. Permeability Study 2**

The second permeability study replicated the design of the first in the following year, allowing key findings to be replicated. However, the preliminary analysis of Permeability Study 1 had not yet picked up the problems with the permeability manipulation discussed above and the design was not modified to accommodate them.

### **8.2.1. Research question**

Once again, the study aimed generally to explore the relationships between permeability, ingroup identification and APM performance.

### **8.2.2. Sample**

As in the previous study, participants were recruited from Varsity College, a private education facility offering classroom-style tuition for degrees awarded by UNISA, South

Africa's distance university. Because early studies lacked power, an effort was made to increase the sample size, and 167 students were recruited from the Pietermaritzburg and Westville campuses. The design was well balanced, with 82 participants in the threat condition compared to 85 assigned to boost; and 83 participants in the low permeability condition and 84 in the high permeability condition.

Measures were put in place to ensure that participants who had participated in Permeability Study 1 at the Pietermaritzburg campus the previous year did not participate again, but this required sampling students registered for diplomas (73/79) as well as undergraduate degrees (6/79) on the Pietermaritzburg campus. All students sampled at the Westville campus were registered for undergraduate degrees. Note that the sample for Permeability Study 1 consisted entirely of students registered for degrees.

Of the final sample, half (50.3%) identified their race as black African, 6.6% as coloured, 12.6% as Indian and 28.7% as white and two participants declined to report their race. Just more than half of the sample (54.5%) reported their sex as male, 44.9% as female and 1 participant declined to report their sex<sup>21</sup>.

### **8.2.3. Design and procedure**

The study used the same design as Permeability Study 1, which was a 2 x 2 factorial design with permeability (low; high) and status (low; high) on APM performance. The manipulations were identical to the previous study (see p. 240).

### **8.2.4. Key findings**

#### **8.2.4.1. Status and permeability on ingroup identification**

Unlike the previous study, the model regressing permeability and status on ingroup identification was not significant, although it did approach significance,  $F_{3,163} = 2.236$ ,

---

<sup>21</sup> Note that in all studies demographics were reported last to avoid cross-contamination of measures and manipulations.

$p = .086$ . However, in this model the only effect approaching significance was status,  $t = 1.805, p = .073$ . In contrast, the same model on the data from the previous study was highly significant, with significant effects for permeability and status\*permeability.

However, further exploration revealed different effects for that males and females. A fully saturated model regressing status, permeability and sex on APM performance was significant,  $F_{7, 158} = 2.17, p = .040, r^2 = .0877, adj.r^2 = .047$ . The significant effects driving this model were permeability,  $t = 2.105, p = .037$ , sex,  $t = 2.277, p = .024$ , and permeability\*sex,  $t = -2.012, p = .046$ . The permeability\*sex interaction is plotted in Figure 26 and shows that women reported particularly low ingroup identification in the low permeability condition whereas men reported higher ingroup identification in the low permeability condition.

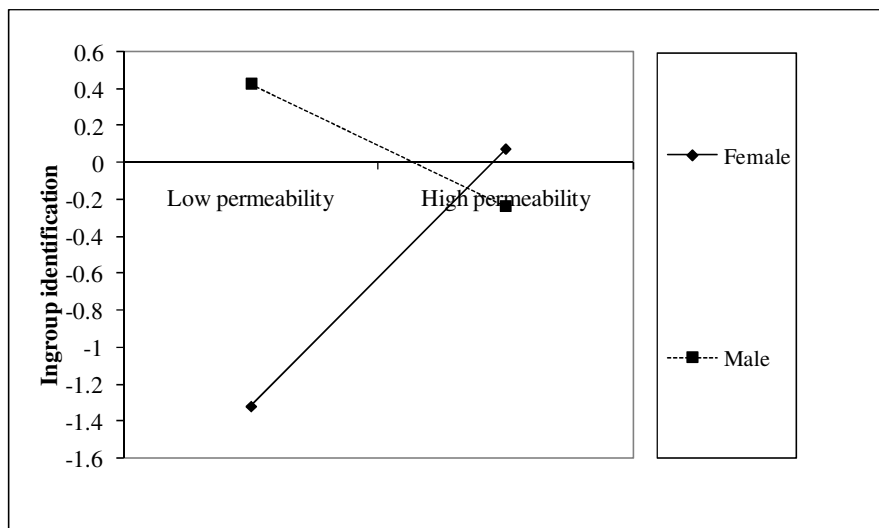


Figure 26: Interaction plot for permeability\*sex on ingroup identification

While this result is not as neat and tidy as the previous study, it still confirms that the permeability manipulation had some impact on ingroup identification, although the effect was not interpretable without also considering participants' sex. The SIT prediction that permeability would influence ingroup identification differently in different status conditions had mixed support, since the main effect for status approached significance in the 2x2 (status,

permeability) model, but dropped away when sex was included in a 2x2x2 (status, permeability, sex) model.

#### **8.2.4.2. Status and permeability on APM performance**

The omnibus test for the saturated model regressing status and permeability on APM performance was not significant,  $F_{3, 163} = 1.424$ ,  $p = .238$ , although the main effect for permeability was significant,  $t = -1.986$ ,  $p = .049$ ,  $b = -.089$ . The effect on performance was that participants in the low permeability condition performed marginally worse than participants in high permeability groups.

#### **8.2.4.3. Status, permeability and resistance on APM performance**

The model regressing status, status-resistance, permeability and permeability-resistance approached significance,  $F_{15, 150} = 1.675$ ,  $r^2 = .144$ ,  $adj.r^2 = .058$ , but no effects achieved conventional significance.

#### **8.2.4.4. Status, permeability and ingroup identification on APM performance**

A model regressing status, permeability and ingroup identification on APM performance (dropping three outliers) approached significance,  $F_{7, 156} = 1.876$ ,  $p = .078$ ,  $r^2 = .078$ ,  $adj.r^2 = .036$ . In this model, the simple effect for permeability achieved significance,  $t = -2.032$ ,  $p = .044$ ,  $b = -.310$ , as did the two-way status\*permeability interaction,  $t = 2.137$ ,  $p = .034$ ,  $b = .561$ . However, these significant lower-order effects were subject to the three-way status\*permeability\*ingroup-identification interaction which closely approached significance,  $t = -1.902$ ,  $p = .059$ ,  $b = -.332$ . The interaction plot for the three-way interaction is displayed in Figure 27, which shows that, in this study, stereotype-performance effects were only observed for participants assigned to the high permeability condition. These participants exhibited a strong conventional STB effect, performing poorly under threat and well under boost, if they reported low ingroup identification. However, participants in the high permeability condition who reported high ingroup identification exhibited very a slight

(and probably non-significant) reversal of the conventional STB effect, performing slightly better under threat than under boost.

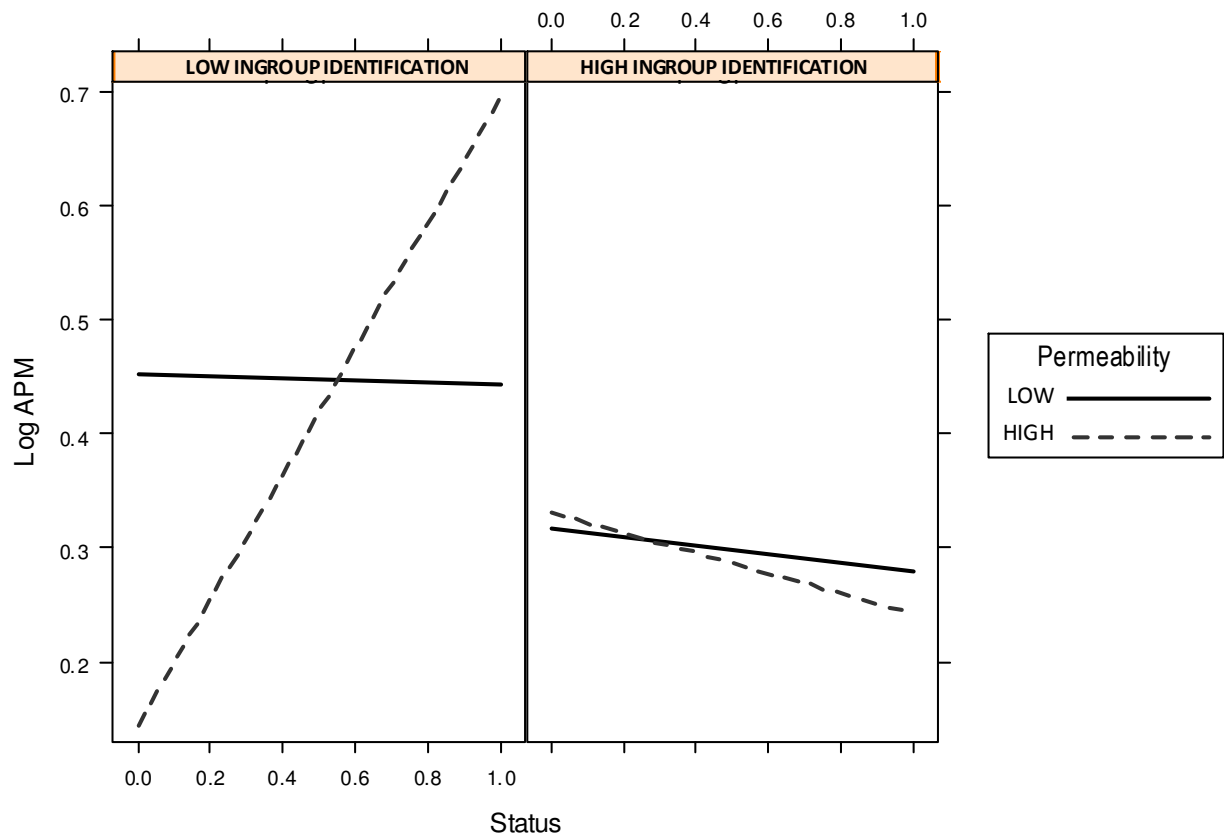


Figure 27: Interaction plot for status\*permeability\*ingroup-identification interaction for Permeability Study 2

#### 8.2.4.5. Degree or diploma?

The sample for the previous study, Permeability Study 1, consisted entirely of students registered for undergraduate degrees. However, 44% of the participants in Permeability Study 2 were registered for diploma courses. A simple correlation confirmed that Degree students (dummy-coded as '1') performed significantly better than Diploma students (dummy-coded as '2'),  $r(165) = -.189, p = .014$ . The regression analysis was therefore repeated with status, permeability, ingroup-identification and degree-type saturated to the third-order on centered APM performance (dropping one outlier). The omnibus  $F$  test



for the model approached significance,  $F_{14, 151} = 1.69$ ,  $p = .063$ ,  $r = .136$ ,  $r^2 = .055$ . Only two of the 14 effects and interactions were significant. The first was the permeability \* ingroup-identification interaction,  $t = 2.667$ ,  $p = .008$ ,  $b = .289$ . However, this was subject to the significant three-way permeability\*ingroup-identification\*degree-type interaction,  $t = -2.143$ ,  $p = .034$ ,  $b = -.139$ , plotted in Figure 28.

The interaction plot confirms that degree and diploma students responded quite differently. Diploma students tended to perform worse when ingroup identification was high regardless of their permeability assignment. However, for degree students there was a large difference between participants assigned to high and low permeability conditions, with degree students in the low permeability condition performing best overall by some margin.

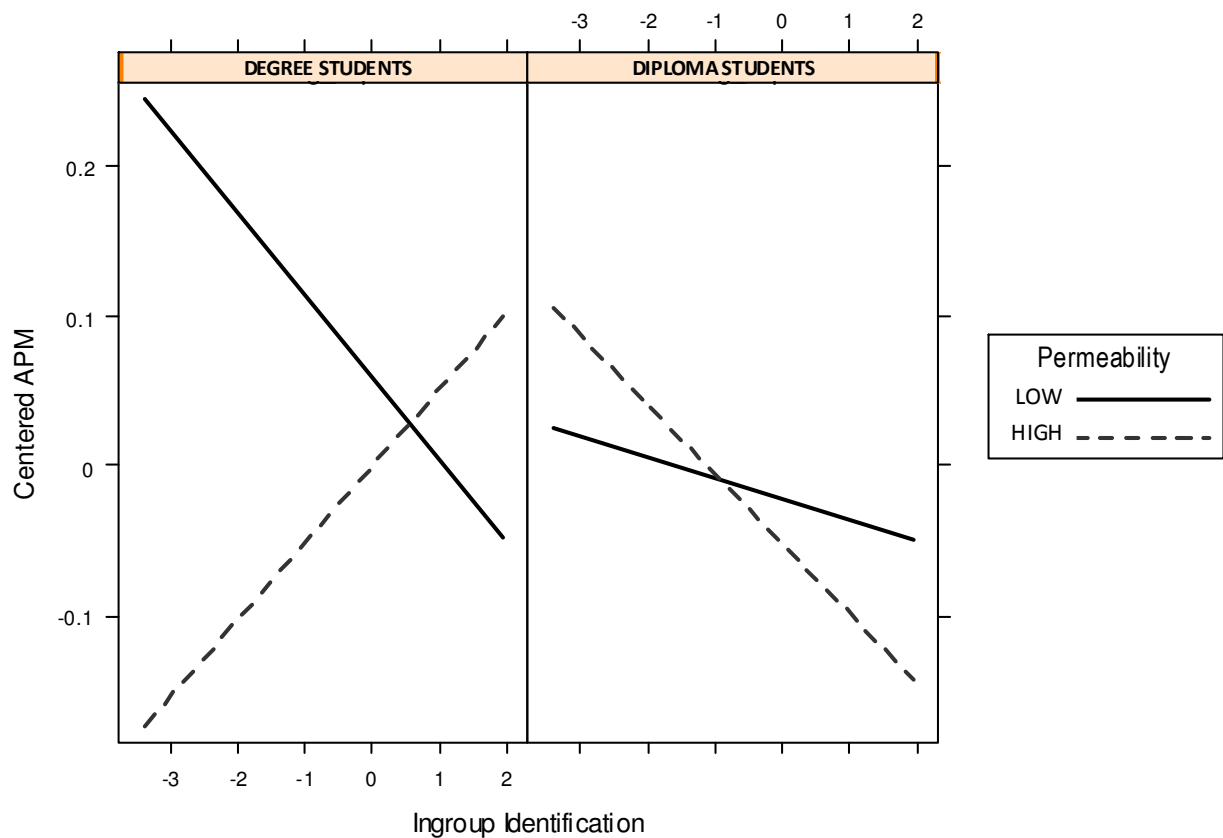


Figure 28: Interaction plot for ingroup-identification\*permeability\*registration-type on APM

### 8.2.5. Discussion of Permeability Study 2

The results of Permeability Study 2 confirmed the most general findings of Permeability Study 1. First, permeability and status did jointly influence ingroup identification. However, where in Permeability Study 1 this relationship was very close to SIT predictions, in the present study the relationship was only evident once sex was included in the model. Female participants reported higher ingroup identification under high permeability conditions, as predicted by SIT, but males exhibited the reverse. Although this pattern is not decipherable with the measures and information available to us, the results do emphasize that permeability and status do impact on ingroup identification, although they remind us that other identities (including, but not limited to sex) may moderate this effect.

As in the previous study, status, permeability and ingroup-identification significantly predicted APM performance, and the model was driven by the significant status\*permeability\*ingroup-identification interaction. As in the previous study, it was participants who expressed low ingroup-identification who showed the greatest differences in APM performance, but in the present study the key effect was reversed: according to this model the best performers were those assigned to *high* permeability and high ingroup status reporting low ingroup identification, whereas in the previous study it was the equivalent participants in the *low* permeability condition who performed the best.

Once again, this pattern is only intelligible in light of the particulars of the identity resources available to participants in the specific performance context that they experienced. In the previous study, all participants were registered for undergraduate *degrees* at Varsity College, whereas in the present study close to half of the participants were registered for *diplomas*. These participants performed worse on the APM in general, possibly reflecting lower academic ability. They were also likely to have experienced the permeability

manipulation differently, since they would have been both less eligible and, probably, less motivated to pursue postgraduate studies than degree students.

The importance of this difference was confirmed by the significant ingroup-identification \* permeability \* degree-type interaction which revealed that diploma students generally performed better at low levels of ingroup identification, regardless of permeability. However, degree students with low ingroup-identification exhibited a marked divergence in performance at different levels of permeability, with those assigned to the *high* permeability condition performing particularly badly and those assigned to the *low* permeability condition performing particularly well. Despite the differences in sample characteristics, this pattern partially replicates the result of Permeability Study 1.

Specifically, ignoring status for the moment, in Permeability Study 1 the best performers were degree students with low ingroup identification in the low permeability condition. In the present study, the top performers were precisely the same: degree students with low ingroup-identification in the low-permeability condition. Similarly, in both studies, participants in the equivalent *high* permeability condition (degree students; low ingroup identification; high permeability) were amongst the worst performers. These results contradict SIT predictions, which suggest that participants in high permeability conditions will be most motivated to demonstrate their prototypicality of, or suitability for, the high-status group.

However, the specific features of the performance context that participants experienced in both of the present studies would make the course of action predicted by SIT a sub-optimal strategy. First, unlike in previous studies of permeability from which these predictions arose (Ellemers, 1993; Ellemers et al., 1988; Ellemers et al., 1990; Ellemers et al., 1993), participants knew that their performance on the given task would have no direct relevance to their eventual permeability trajectory, in other words, that their performance on

the experimental task would have no bearing on actually gaining acceptance to postgraduate studies and becoming an outgroup member.

Second, both experiments invoked a scenario in which participants with postgraduate aspirations would, ultimately, be *required* to become members of the outgroup regardless of its relative status in order to further their studies. Even participants who believed implicitly in the superiority of Varsity College would need to transfer to a traditional university if they wanted to continue their studies at a postgraduate level. Based on the findings of Ellemers and her colleagues (Ellemers, 1993; Ellemers et al., 1988; Ellemers et al., 1990), it seems likely that this subset of participants with postgraduate aspirations would report lower ingroup-identification with Varsity College as a psychological precursor to their planned change in group membership in the future, although not all students with low ingroup identification would necessarily have postgraduate plans. Additionally, no matter how strong the wording of the permeability manipulation, participants knew the intergroup boundaries were – in reality – at least partially permeable. The permeability manipulation, therefore, did not operate between the extremes of fully open and totally closed, but across more moderate positions in-between.

The results of both studies suggest that the pressing identity challenge for participants with postgraduate aspirations in this context was to lay a claim to being the kind of student to whom the intergroup boundary would be permeable. For students in the high permeability condition, this could be achieved by doing very little, since the permeability manipulation had informed them that Varsity Colleges students generally find it very easy to transfer to the outgroup. However, participants in the low permeability condition had been informed that Varsity College students generally find it difficult to transfer. Therefore producing an identity as the type of student who could successfully transfer to a large traditional institution for participants with postgraduate aspirations (and low ingroup identification) in the low

permeability condition required an excellent performance on the APM, which is exactly what was observed in both studies.

Both the permeability and status manipulations in these two studies were clearly flawed. The permeability manipulation lacked relevance to any participants with no postgraduate aspirations – and there was no way of identifying the participants in that situation. However, the most serious problem was that participants had pre-existing knowledge of, history with, and aspirations towards the categories that were invoked. This intimate knowledge and experience necessarily influenced the ways that participants understood the manipulations and interpreted the identity contingency produced in the experimental performance context.

From a methodological point of view these were weaknesses that required *post hoc* explanation and call into question the invariance and generalizability of any effects observed. However, the way that the manipulations intersected with real, lived categories and moral careers provided theoretical purchase that demonstrated the extent to which stereotype-performance effects are truly identity *contingencies* as argued by Steele (2010). In contrast to the STL prediction that status-related stereotypes automatically and deterministically impact on task performance, the results of these two studies suggest that participants oriented their performance on the APM to strategically address individual and group identity concerns. In other words, where good or bad performance on the APM was an interpretable act, participants oriented their performance to the way that it would be interpreted.

This argument that stereotype-performance effects are related to situated identity performance does not preclude possible involuntary cognitive or affective influences on task performance, such as interference with working memory (eg. Schmader, 2010). However, it does demonstrate that cognitive and individualistic models that overlook the active and strategic features of identity production cannot fully address the phenomenon of STB either.

### **8.3. Permeability Study 3**

The key methodological problem with the first two permeability studies is one that has plagued all attempts to experimentally explore social-identity effects of permeability with real groups: that it is exceptionally difficult to manipulate participants' perceptions of the permeability of real groups that they belong to or want to belong to. The third permeability study aimed to surmount this problem by invoking stereotype threat with a minimal groups paradigm. Since minimal groups are entirely under the control of the researcher, and participants have no prior knowledge or experience of them, their perceived permeability can be much more easily manipulated.

This study was risky, because there was no evidence in the literature that stereotype threat had been successfully invoked in minimal groups settings before. Additionally, as the only study in this programme of research to use experimentally invoked (rather than real) categories, participants' active engagement with specific features of their situated identity was likely to result in different patterns of results. This is particularly true since the real groups invoked in previous studies were important to participants' real-life moral careers and personal ambitions. Those invoked in the experimental setting were bound to be much less meaningful and hold less long-term importance for participants.

#### **8.3.1. Research question**

Given the difficulties with manipulating permeability in a real intergroup setting, this study aimed to explore the effects of permeability and status on task performance in a minimal groups setting.

#### **8.3.2. Design and procedure**

As with most studies in the program of research, the design was a 2x2 factorial design with task-related status (low; high) and permeability (low; high) on APM performance.

### ***8.3.2.1. Minimal groups assignment and status manipulation***

Tajfel and his colleagues in his seminal minimal group studies with young boys in Bristol assigned participants to minimal groups by flashing dots on a screen and then asking them to estimate the number of dots they had seen (Tajfel, 1970). Regardless of their actual answers, boys in one condition were randomly assigned to be labeled “overestimators” or “underestimators,” and in the other “accurate” or “inaccurate” estimators. Since invoking STB requires that categories be closely linked to expected task performance, Tajfel’s basic strategy was adapted by collapsing these categories into “active estimators” (positive) and “reactive estimators” (negative) so that expected task performance was embedded into the value labels. These labels were chosen as they appear to be 1) plausible; 2) similar to those successfully used by Tajfel to invoke minimal groups and 3) although recognizable and comprehensible, are not likely to be categories with which participants have had previous experience.

To invoke these categories, participants were briefly shown five increasingly complex slides of dots and were asked each time to estimate how many there were. Participants assigned to the “active” estimators group were given on-screen feedback after each response that suggested that they were becoming increasingly accurate as the trial progressed, while those in the “reactive” estimators group were given feedback that suggested that they were becoming increasingly inaccurate as the slides became more complex.

After the five trials, participants were informed on-screen that their performance had demonstrated that they were either “active” or “reactive” estimators. To increase group status, active estimators were told that only 15% of the population fall into that category, and to decrease group status, reactive estimators were told that they were “below average.” Participants were informed by on-screen instructions that active estimators are “skilled at

combining multiple streams of information and arriving at the best solution” while reactive estimators tend to “over-react to feedback and lose sight of the task itself.”

### 8.3.2.2. *Permeability manipulation*

Participants randomly assigned to the *high* permeability condition were informed that reassignment would be possible on the basis of their performance on the APM, since the dot estimation task was not entirely reliable. Because their answers to the scales assessing identity and sociostructural variables depended on their status, all participants in the high permeability group were informed after completing the APM that their results were borderline and, although reassignment was not out of the question, they were not reassigned at that time. It was hoped that this instruction would maintain their status assignment and yet maintain the illusion of permeability for the duration of the session.

In the initial version of the design, participants in the *low* permeability were given the opposite information – that reassignment was not possible. However, in a pilot of the manipulations several participants noted that this statement made them suspicious and it was therefore decided to drop this instruction despite introducing a slight difference in the experience of low and high permeability participants.

### 8.3.2.3. *Measures*

The dependent measure was the short form of the Ravens Advanced Progressive Matrices. As manipulation checks, participants completed a minimal group manipulation check assessing the extent to which participants reported beliefs about active and reactive estimators congruent with their minimal group assignment, the STB manipulation check, and a measure of perceived permeability adapted for the minimal groups context. Participants also completed measures of: baseline performance (using the SILS); ingroup identification; category salience; legitimacy; stability; intergroup conflict; effort, and individual identifiability. Some of these measures were less than ideal for the minimal group context



(eg. intergroup conflict), but they were retained because of their importance in the SIT framework.

#### **8.3.2.4. Procedure**

The experiment took place in a computer-based experimental laboratory containing 12 PC's in a spatial configuration ensuring privacy. Upon arrival, participants were assigned a number from the pre-assigned randomization schedule assigning them to one of the four conditions. Once seated, they completed the informed consent procedure they began completing the materials on a PC running E-Prime (Psychology Software Tools, 2010), beginning with the SILS, followed by the dot-estimation task and manipulations, the APM and then the manipulation checks and social identity and sociostructural scales. After completing the materials they were debriefed, provided with their incentive and thanked for participating.

#### **8.3.3. Sample**

Student participants were recruited by poster and face-to-face solicitation on the Pietermaritzburg campus of the University of KwaZulu-Natal and were offered an incentive of R30 for participation. In total 151 participants participated and, of these: 73 were assigned to the threat condition and 78 to boost; 77 were assigned to the low permeability condition and 74 to the high permeability condition; 102 reported their race as black African, 13 as coloured, 20 as Indian, 15 as white and one as 'other'; 85 reported their sex as female and 66 as male.

#### **8.3.4. Key findings**

##### **8.3.4.1. Manipulation checks**

As in previous studies, participants' perceptions of own-group performance (the status manipulation check) and permeability were converted to resistance scores and entered into the main analysis rather than being used as simple manipulation checks. However, an

independent-samples *t*-test on the minimal group manipulation check was highly significant,  $t(149) = -3.580, p < .001$ , indicating that participants were distinguishing between *active* and *reactive* estimators in ways congruent with the experimental manipulations and that they did not reject the minimal groups.

#### 8.3.4.2. Status and permeability on ingroup identification

The model regressing status, permeability and status\*permeability on ingroup identification was significant,  $F_{3, 146} = 19.62, p < .001, r^2 = .287, adj.r = .272$ . As expected, the effect for status was significant,  $t = 4.254, p < .001, b = .964$ . However, the status\*permeability interaction achieved conventional significance only after three outliers were dropped,  $t = 2.354, p = .020, b = .723$ . As illustrated in Figure 29, the pattern of interaction is exactly as predicted by SIT: members of the low status group had lower ingroup identification when permeability was high and the high status group had higher ingroup identification when permeability was high.

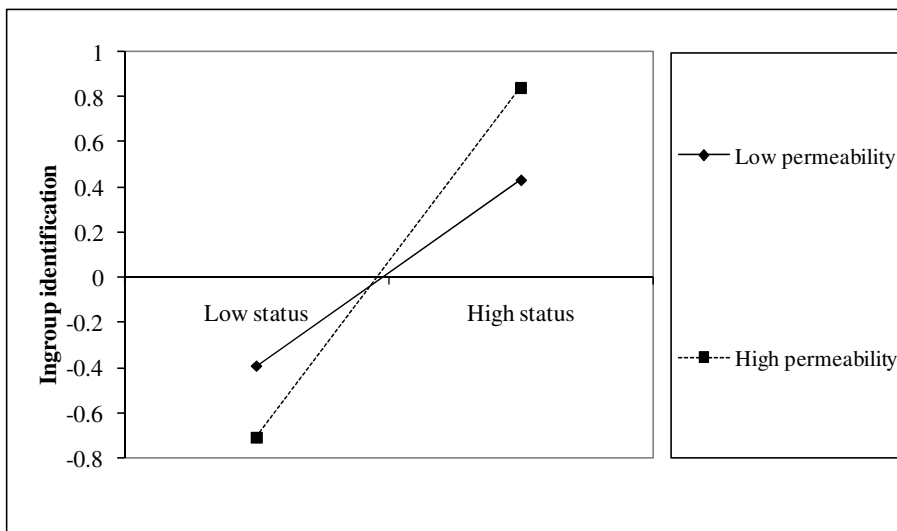


Figure 29: Profile plots for status\*permeability on ingroup identification (excluding three outliers)

To further explore the influence of status and permeability on participants' orientation towards their ingroup and outgroup, the model was repeated on the ingroup and outgroup closeness measures. The model for status, permeability and status\*permeability on *outgroup* closeness was significant,  $F_{3, 147} = 6.133, p < .001, r^2 = .111, adj.r^2 = .093$ , but the only significant effect was status,  $t = -3.606, p < .001$ , indicating that members of the low status group reported feeling higher levels of closeness to the high-status outgroup than the members high-status group reported towards the low-status outgroup. Permeability had no influence on reported outgroup closeness either as a main effect or as an interaction.

The model for status, permeability and status\*permeability on *ingroup* closeness was also significant,  $F_{3, 147} = 4.241, p = .007, r^2 = .080, adj.r^2 = .061$ , but was driven by the significant effect of permeability on ingroup closeness,  $t = -2.624, p = .001, b = -1.136$ . Specifically, members in the low-permeability condition were less likely to report feeling close to their ingroup than members of the high-permeability condition. Neither status, nor the interaction between status and permeability, had a significant effect on ingroup closeness.

#### **8.3.4.3. Status and permeability on APM performance**

As in most previous studies, the base model regressing status, permeability and status\*permeability on APM performance was not significant,  $F_{3, 147} = .018, p = .910$ .

#### **8.3.4.4. Status, permeability and resistance on APM performance**

However, in this study the model incorporating resistance was not particularly successful either. A saturated model with status, status-resistance, permeability, and permeability-resistance was regressed on APM performance and simplified with stepwise selection with BIC as the selection procedure with the step() function in R (R Development Core Team, 2010). The model as a whole was not significant,  $F_{9, 141} = .912, p = .516$  and only the status-resistance\*permeability\*permeability-resistance interaction approached significance,  $t = -1.910, p = .058$ .

#### **8.3.4.5. *Status, permeability and ingroup identification***

As shown above, the prediction from the social identity literature that ingroup identification would be related to status and permeability was upheld. Ingroup identification (and domain identification) have frequently been shown to interact with status to predict performance. However, the model regressing APM performance by status, permeability and ingroup identification was not significant.

#### **8.3.4.6. *Status, permeability, and group position measures on APM performance***

Although the reliability for the ingroup identification scale in this study was acceptable ( $\alpha = .723$ ), some of the items in the scale did not make much sense with respect to the minimal groups set up in the manipulations. For example, participants were asked to report agreement with items such as “it feels bad when people say bad things about [active/reactive] estimators” and “I feel strong ties with [active/reactive] estimators as a group.” Although a model including ingroup identification, ingroup closeness and outgroup closeness was significant with significant effects suggesting that the ingroup identification measure was contributing uniquely to the model, it was difficult to know if the items were valid given the unrealistic intergroup context inducted in the laboratory.

Therefore the ingroup and outgroup closeness measures were used as indicators of participants' position in relation to the experimental 'groups.' These were graphical measures in which the participant was represented by a small circle and the ingroup or outgroup was represented by a larger circle with the two grounded on a horizontal line (see Figure 4). Seven options were presented wherein the two circles successively became closer and ultimately completely overlapped (Schubert & Otten, 2002). Because these measures did not rely on verbal descriptions of the group and were decontextualized, they therefore did not

violate the temporally-bounded nature of the minimal groups and were therefore much more likely to be valid indicators of group position in the minimal-groups context.

A model was fitted regressing status, permeability, ingroup-closeness and outgroup-closeness on APM performance (with a square transformation and dropping one outlier) with non-significant effects removed in one step. The model very closely approached significance,  $F_{6, 143} = 2.161, p = .050, r^2 = .083, r = .044$ . The simple effects were significant for status,  $t = -2.022, p = .045, b = -.197$ , permeability,  $t = -2.114, p = .036, b = -.211$ , and outgroup-closeness,  $t = -2.041, p = .043, b = -.029$ . However, these were subject to the significant two-way interactions, namely status\*outgroup-closeness,  $t = 2.733, p = .007, b = .020$ , and permeability\*ingroup-closeness,  $t = 2.053, p = .042, b = .012$ .

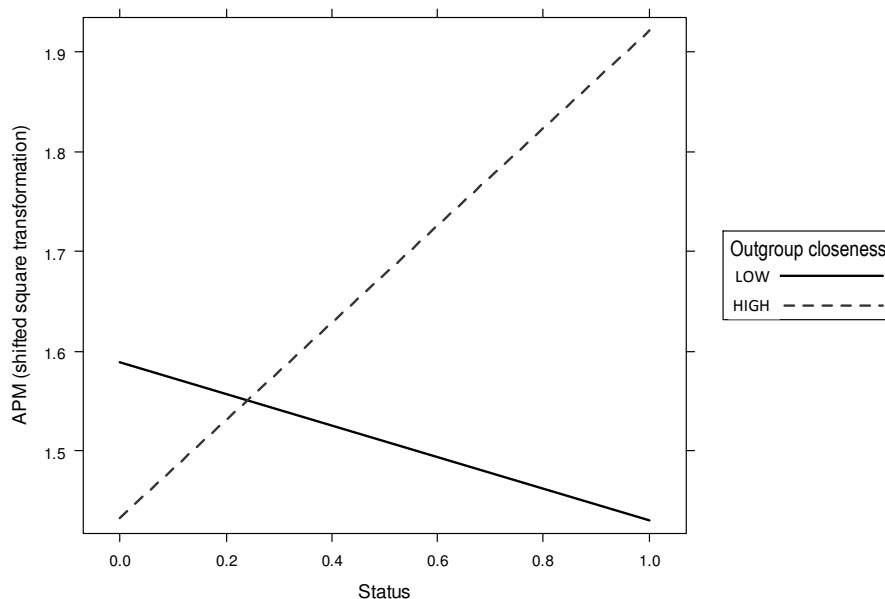


Figure 30: Interaction plot for status \* outgroup-closeness on APM for Permeability Study 3

As displayed, in Figure 30, participants expressing high outgroup closeness exhibited conventional and strong STB effects. The best performers were those in the boost condition who expressed high outgroup closeness. However, participants expressing low outgroup

closeness exhibited a mild reversed STB effect, performing worse under boost than under threat.

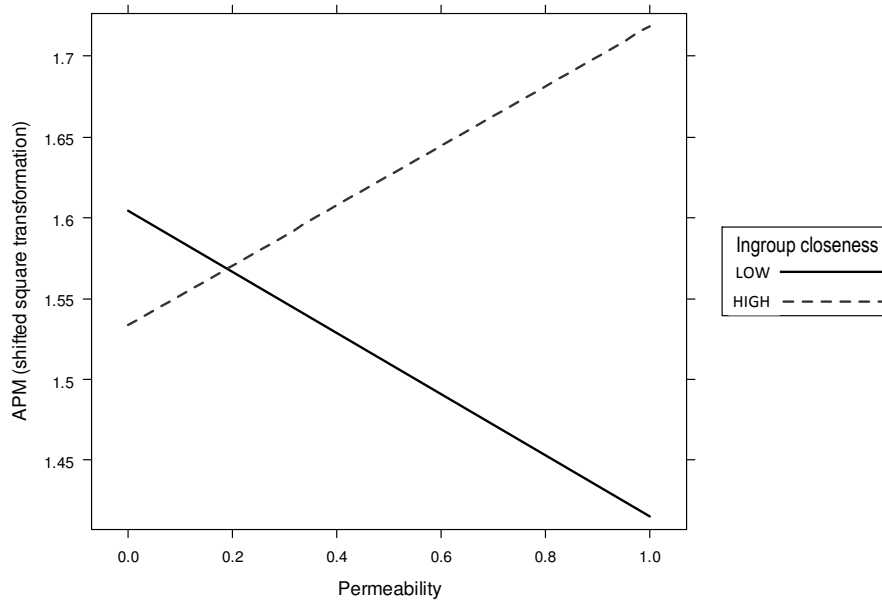


Figure 31: Interaction plot for permeability\*ingroup-closeness on APM for Permeability Study 3

The pattern displayed in Figure 31 is very similar except that it plots APM performance by *permeability* and *ingroup* closeness. The difference in performance was much greater when permeability was high than when it was low, and participants expressing low ingroup closeness performed poorly in high permeability conditions while those expressing high ingroup closeness performed very well in high permeability conditions.

### 8.3.5. Discussion of Permeability Study 3

As in the previous permeability studies, the permeability manipulation had a significant influence on ingroup identification and the interaction between status and permeability was consistent with the predictions of social identity theory (eg. Ellemers, 1993; Ellemers et al., 1988). Specifically, low-status group members reported lower levels of ingroup identification than high-status group members, and this trend was more pronounced

in high permeability conditions. However, a closer inspection of participants' social position in relation to the low and high status groups showed that perceptions of *ingroup* closeness were largely driven by permeability and perceptions of *outgroup* closeness were largely driven by status. These results are congruent with SIT predictions and demonstrate that both status and permeability perceptions impact on the way that participants orient themselves to the intergroup terrain in a typical STB performance context.

Contrary to predictions, and despite the clear impact of the manipulations on ingroup identification, there was no direct relationship between status or permeability on APM performance, either alone or in combination. Also, contrary to theoretical predictions, ingroup identification had no discernable impact on performance, despite being related to status and permeability precisely as predicted by social identity theory. However, despite the reliability of the ingroup identification scale, there were some question-marks about its suitability for this minimal groups application.

Fortunately other measures of group affiliation had been included in the materials, and significant models emerged once group closeness variables were considered. In this study, status and outgroup closeness jointly predicted APM performance while permeability did so in conjunction with ingroup closeness. The patterns observed were interpretable in SIT terms.

Looking first at the effects of status and permeability on APM performance, participants performed particularly well when they were in the boost condition and yet expressed higher than average closeness to the outgroup. This suggests that they were insecure about their place in the high-status group and performed particularly well as a way of staking their membership of the group. Work in the SIT framework has previously suggested that group members with insecure identity at the fringes of high status groups may become hyper-prototypical and enact ingroup norms and ideals far more vigorously than

high-status group members with more secure identity positions (Branscombe, Ellemers, Spears, & Doosje, 1999; Ellemers et al., 2002; Jetten, Branscombe, & Spears, 2002; Jetten, Branscombe, Spears, & McKimmie, 2003; Noel, Wann, & Branscombe, 1995). In the situation of the STB study, the only way of enacting prototypicality of the high status group was to perform well, and this is what they did.

Members of high status groups who have secure identities have the freedom to be far less prototypical, because they are not at risk of exclusion. In the present study the participants in the high status condition who perceived a large distance between themselves and the low status outgroup (that is, expressed low outgroup closeness) performed particularly badly. This could be interpreted as a case of high-status group members with secure identity not needing to perform well to maintain their positive identity – in other words, as another example of slipstreaming.

When the effects of permeability and *ingroup* closeness on APM performance were considered, participants in low permeability conditions were relatively unaffected (in contrast to the results of the previous study). In high permeability conditions, participants performed well if they reported themselves to be close to their group and poorly if they reported a distance between themselves and their group. It is difficult to interpret this interaction either in light of STB, since it does not reflect a stereotype-performance effect, or SIT, since models of permeability in the literature are closely coupled to status.

However, despite the difficulty of making a specific interpretation, these results do emphasize that status is not the only feature of identity that contribute to shifts in task performance. It is not simply whether an individual's group is positively or negatively stereotyped that counts (as argued in the STB framework), but the sociostructural features of the specific performance context (such as permeability) and the individual's location within and between the groups.



#### 8.4. General discussion of permeability studies

‘Domain identification’ and, less commonly group identification, have been amongst the most commonly explored mediators/moderators of stereotype threat in the STB literature, but studies to date have had mixed and unreliable results (Armenta, 2010; Aronson et al., 1999; Clark et al., 2010; Cullen et al., 2006; Deaux et al., 2007; Keller, 2007; Kiefer & Sekaquaptewa, 2007; Kray & Shirako, 2009; Lawrence et al., 2010; Lesko, amp, & Corpus, 2006; Leyens et al., 2000; Massey & Fischer, 2005; Osborne & Walker, 2006; Schmader, 2002a; Smith, 2004; Steele, 1997; Tagler & Teles, 2003; Twamley & Tajfel, 2009; Wout, Danso, Jackson, & Spencer, 2008).

One problem with previous studies is that ‘identification’ has generally been considered as a unitary freestanding construct. However, in SIT, identification is conceptualized as a cognitive-affective indicator of actual and aspirational positioning of an individual within an intergroup context. In this framework, group identification has been clearly shown to be sensitive to both individual identity resources, such as ability, and socio-structural features of intergroup contexts, such as permeability (eg. Ellemers, 1993; Ellemers et al., 1988).

Social identity theorists specifically critique focusing exclusively on individual-level constructs such as ingroup identification, ability, or ambition, in understanding and predicting either individual trajectories in an intergroup context or large-scale social change (eg. Tajfel, 1981a). Therefore, even in cases where identification with a domain or group mediates or moderates STB effects, its influence is only likely to be comprehensible in conjunction with the sociostructural and individual features of situated identity to which it is sensitive.

Across all three studies, ingroup identification was linked to status and permeability as predicted by SIT (Ellemers, 1993; Ellemers et al., 1988; Ellemers et al., 1990; Mummendey et al., 1999), although in Permeability Study 2 the specific patterns of ingroup

identification were unusual. As predicted, ingroup identification and other indicators of individuals' positioning within the intergroup, such as ingroup and outgroup closeness measures, most usefully modeled task performance in STB conditions in conjunction with sociostructural variables, such as permeability and group-differentiation. However, it also emerged that other variables related to individuals' specific position and aspirations within the intergroup context, such as resistance or degree-type, were critical to modeling stereotype-performance effects.

Overall, these results indicate that participants' performances were related to their membership of stereotyped groups, as predicted by STB. However, task performance was not influenced in a deterministic way simply by the knowledge that others hold a positive or negative stereotype about the group to which they had been assigned as predicted by STB (eg. Steele, 2010), but rather that task performance – at least in the experimental context – were meaningful responses to the social predicament, strategically motivated towards the production of identity in that specific context.

For example, members of the 'old money' classes in the UK can wear threadbare clothes in public and drive old cars with no threat to their status – and in fact their blatant disregard for appearances makes them more prototypical (eg. Mount, 2010). Similarly participants with secure identities in the high status experimental condition were able to perform particularly poorly without threatening their identity position. On the other hand, the participants with the most identity work to do performed the best. These participants were those who did not identify with the ingroup (and presumably had aspirations to transfer to the outgroup in the future) but were told that group boundaries were impermeable and transfer was difficult. It seems very likely that they performed particularly well to demonstrate that, despite the generally impermeable boundaries, they were suitable candidates for the high status group.

Despite their post-hoc nature, these results should prompt us to take more seriously Steele's (Steele, 2010) notion that stereotype threat is an "identity contingency." In the STB tradition, the only salient features of identity in that context are whether the participant's group is positively or negatively stereotyped on the task. However, the present results suggest – although they do not prove – that participants orient towards their identity in fine-grained ways that take into account the identity that is being applied to them by others, their reaction to these stereotypes, their reading of the sociostructural features of the performance context and their own orientation to the groups in the intergroup setting. In this framework, task performance is meaningful in identity terms and becomes a way of claiming or rejecting identity.

However, high or low task performance can be the result of many different identity orientations. For example, low performance could result from "settling in" to a secure high-status group membership which relieves the individual from the burden of high individual performance (ie. slipstreaming) or, equally likely, from accepting the inferiority of one's low status identity and consciously or unconsciously giving in.

This understanding of stereotype-performance effects as situated identity performance does not preclude finding a subset of identity contingencies in which the specific intersection of identity and sociostructural features constrain performance in ways congruent with conventional STB. However, these situations are likely to be the exception rather than the rule, and it is likely that understanding the specific sociostructural and identity-producing features of that performance will reveal ways in which STB effects can be disrupted.

The three studies exploring permeability suggest that three classes of variables will be important in understanding and predicting the strength and direction of stereotype performance effects: first, sociostructural features of the intergroup context as experienced in the specific performance situation that are generally beyond the control of the individual,

such as intergroup permeability and legitimacy. Second, variables representing participants' position within, and orientation to, the groups and identities in play in the performance context, such as ingroup-identification, or ingroup and outgroup closeness. Third, features of individual identity and self-concept such as ability and experience with similar performance tasks in the past. This layered approach to modeling stereotype-performance effects will be further explored in the following chapter.

**Chapter 9. Results V: Predicting incongruent stereotype-performance effects**

The stereotype-threat literature has demonstrated clearly that in many situations negative stereotypes can result in decreased task performance for members of minority groups and positive stereotypes can result in improved performance for members of high status groups. However, a careful reading of the literature also revealed that ironic stereotype-performance effects are also possible, in which positive stereotypes are associated with reduced performance and negative stereotypes with increased performance (see section 3.2, p. 72).

The general strategy of stereotype threat theorists for dealing with these incongruent effects has been to state an increasing number of boundary effects defining stereotype-threat situations that allow incongruent effects to simply be written-off as irrelevant (eg. Steele, 2010). For example, Aronson, Lustina, Good, Keough and Steele (1999) exposed white males to a stereotype that Asian men had superior math ability and assessed the impact on their performance. High domain identifiers performed poorly compared to a control group when exposed to stereotype threat, as hypothesized, but moderate domain identifiers performed better. In fact moderate domain identifiers under threat performed just as well as high domain identifiers in the control condition (figure reproduced here as Figure 32).

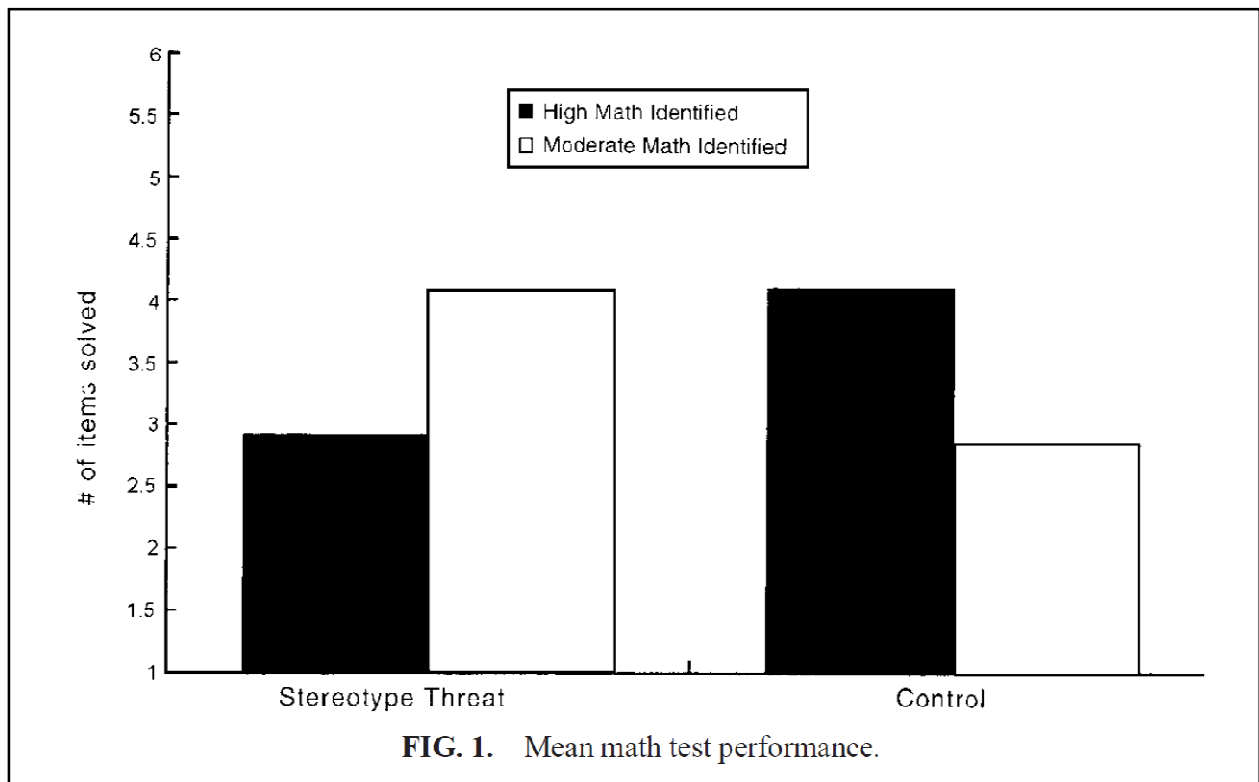


Figure 32: Figure 1 from Aronson, Lustina, Good, Keough and Steele (1999, p. 38)

Aronson and his colleagues concluded that domain identification is a necessary condition for stereotype-threat to occur. However, an equally valid conclusion is that (1) stereotype-performance effects can be congruent or incongruent with stereotype valence and (2) that both high and low domain identifiers exhibited a stereotype-performance effect, but effect-congruence was moderated by domain identification.

Without a model predicting the congruence of stereotype-performance effects, the field of stereotype-threat will be left either with an array of boundary conditions that will make the theoretical framework increasingly irrelevant to everyday task performance, or the unparsimonious prediction that both negative and positive stereotypes can both decrease or increase performance. The next stage of analysis was, therefore, to model the conditions predicting congruent or incongruent responses to threat and boost conditions.

The analyses in the previous chapters treated task-performance as the dependent variable and, although they revealed some signs of incongruent stereotype-performance

effects, these could only be inferred indirectly in complex interaction effects. To explore the congruence of the stereotype-performance effects more directly, the variables representing task performance and stereotype threat and boost were combined into one variable representing stereotype-performance-congruence.

To do so: first, the status variable was recoded with simple contrast-coding (-1 for low-status/threat; +1 for high-status/boost). Second, change in task performance was calculated by first transforming APM and SILS performance to z-scores and then calculating the difference score for each participant (APM score – SILS score). The resulting difference score represented the change in task performance following the status manipulation relative to other participants. A difference score of 0 represented performance that was roughly constant, a negative score a decrease in performance, and a positive score an increase in performance. Finally, a stereotype-performance-congruence score was calculated by multiplying the performance-difference score by the status contrast, resulting in a metric that was positive for conventional congruent stereotype-performance effects and negative for incongruent performance effects, as illustrated in Table 25. Scores near zero represented participants who were relatively unaffected by the status manipulation and who performed roughly equivalently on the SILS (prior to the status manipulation) and APM. Extreme scores represented participants who were strongly influenced by the manipulations, either congruently (positive scores) or incongruently (negative scores). This metric will hereafter be referred to as *STB congruence*.

	Status	
	Threat	Boost
Negative performance shift	status (-)	status (+)
	x	x
	performance (-)	performance (-)
	= <u>positive congruence</u>	= <u>negative congruence</u>
Positive performance shift	status (-)	status (+)
	x	x
	performance (+)	performance (+)
	= <u>negative congruence</u>	= <u>positive congruence</u>

*Table 25: Calculating the stereotype-performance-congruence score*

### 9.1. Strategy for analysis

Each study for which congruence scores could be calculated was reanalyzed with STB congruence as the dependent variable<sup>22</sup>. However, there was no reason to expect that the social and psychological conditions that best promote congruent or incongruent stereotype-performance effects should be the same under threat as they are under boost. On the contrary, it seems likely that the conditions supporting incongruent performance under threat would be different to those supporting it under boost. For example, group identifiability may have a galvanizing effect on members of the high status group because it would motivate them to demonstrate that they are “good” prototypical group members. However, poor performance would not have the same liability value for members of the low status group and therefore is less likely to predict their performance. These differences across boost and threat subsets would violate the moderated multiple regression assumption of homogeneity of error variances across conditions and would result in drastically reduced power to detect effects

<sup>22</sup> Since this method required both SILS and APM scores, several earlier studies could not be analyzed with this method. The suitable studies included Conflict Study 3, Permeability Study 2, Permeability Study 3, Identifiability Study 4, Conflict Study 2, Permeability Study 1 and Achievement Motivation Study 1.



(Aguinis, 2004). Therefore, the analysis was carried out separately for boost and threat subsets in each study despite the decrease in power and increased risk of type I error that this strategy entailed<sup>23</sup>.

Key variables were transformed<sup>24</sup> where necessary to maximize normality and minimize skewness judged on the basis of Q-Q plots. Although this procedure maximized power, it resulted in the dependent measure for each of the following models being represented in different scales which limited the interpretability of slopes and the direct comparability of coefficients across studies. However, it was still possible to interpret the direction and relative strength of each effect within and across models by comparing the interaction plots.

For theoretical reasons discussed in the method chapter, the key variables in this series of studies were identifiability, intergroup conflict and intergroup permeability. To increase the comparability of models across studies, the analysis used *reported* identifiability, conflict and permeability even when these variables had been manipulated in specific studies. Consequently, the following section of the analysis should be considered correlational rather than experimental.

The SIT variables of conflict, permeability and individual identifiability were expected to be most likely to contribute to stereotype-performance effects for theoretical reasons discussed in the literature review and method chapters and the analysis began by focusing on these. The first step in the analysis was to enter the key variables of perceived conflict, permeability and individual identifiability as simple effects in models predicting

---

<sup>23</sup> Since threat and boost subset were analyzed separately, the analysis could have used a z-scaled difference score as the dependent variable. However, the congruence metric was retained because it emphasizes that the key object of analysis here is not simply to *detect* a stereotype-performance effect, but to model the congruence of the relationship between status and effect.

<sup>24</sup> When exponential transformations were used (including log transformations) the distribution was first shifted so that the minimum value was 1.

STB congruence for each study, for the high status (boost) and low status (threat) subsets in turn. In this stage of the analysis it emerged that the most consistent predictors of STB congruence were perceived intergroup conflict for participants in boost conditions and perceived intergroup permeability for those under threat.

The second step involved dropping non-significant key variables from each model and then one-by-one including other available predictors. If a predictor improved a model (on the basis of  $r^2$  change) then it was retained, but if it did not it was excluded. Although this selection procedure added to the overall type I error of the analysis, it was not possible to test all predictors in one model for each study because of the small sample sizes.

The final step involved testing higher-order interaction effects involving variables remaining in each model by this stage. Interactions were retained if they achieved significance or accounted for variability that allowed a key simple effect to achieve significance. For the sake of brevity, only the final models for low and high-status have been reported for each study.

It should be noted that the sample sizes for most studies were not as large as they should have been to support this method of analysis. However, when the studies were designed it had been theorized that threat and boost conditions could be analyzed together. It was only after the data had been collected it was realized that the social conditions predicting congruence and incongruence might be different for threat and boost conditions and they would have to be considered separately.

There is little agreement on minimum sample sizes for multiple regression analysis. Tabachnick and Fidell (1989) suggested that the absolute minimum sample size should be five cases per predictor, but there is general agreement that this is too few for adequate power (eg. Green, 1991). Others suggest a minimum of ten (Miller & Kuncze, 1973; Weinberg & Abramowitz, 2008) or fifteen (Stevens, 1996) cases per predictor. Harris (1985)

recommended that the number of cases should exceed the number of predictors by at least 50. Knofczynski and Mundfrom (2008) calculated that a good prediction from data with a small to moderate squared population multiple correlation coefficient with three predictors requires a sample size greater than 100.

Although the sample sizes for the analyses reported in this chapter were generally uncomfortably small, all exceeded Tabachnick and Fidell's (1989) bare minimum of five cases per predictor, and only the analysis for Permeability Study 1 dipped below the minimum of ten cases per predictor recommended by Weinberg and Abramowitz (2008) and Miller and Kuncce (1973).

No samples were as large as recommended by Knofczynski and Mundfrom (2008) for good predictive generalizability, but the purpose of these studies was explanatory rather than predictive, which justifies smaller samples. Additionally, the large combined samples sizes across studies of 494 in low-status conditions and 496 for high-status conditions and the consistency of key results across these studies should allay concerns that the small sample sizes resulted in bias. Although it would have been preferable for model-building purposes to have had a smaller number of larger studies, and the overall analysis may have missed important patterns due to low power in each constituent study, it is unlikely that results observed consistently across several studies were due to statistical artifacts introduced by small sample sizes.

## **9.2. Predicting incongruent stereotype-performance effects under boost**

The pattern of significant effects and interactions for participants in high-status conditions across studies is displayed in Table 26. The most consistent predictor of STB congruence for participants in high-status conditions was perceived intergroup conflict, which was significant either as a simple effect or as a component of a significant interaction effect for all six studies in which STB congruence and conflict metrics were both available.

The next most consistent predictor was social-desirability, which was a significant component of models in five of the seven studies. Group-identifiability (which is the extent to which participants felt that they were identifiable as group members and that their group would be judged by their performance) was a significant component of models in three studies but was not included in the materials for two studies.

	C2	C3	P2	P1	P3	ID4	AM1
Conflict	*	*	*	I	I	I	X
Social-desirability			I	*	I	I	*
Group-identifiability	I	*		*	X		X
Individual-identifiability					I	I	
Ingroup Identification							*
Conflict*social-desirability			*	*		*	
Conflict*group-identifiability	*			I			
Conflict*individual-identifiability						*	
Ingroup-identifiability*social-desirability							*
Ingroup-identifiability*social-desirability				*			
Conflict*social-desirability*individual-identifiability					*		
Conflict*social-desirability * group-identifiability				*			
N	36	62	85	50	78	141	44

*Note.* C2/C3 = Conflict Studies 1 and 3; P1/P2/P3 = Permeability studies 1-3; ID4 = Identifiability study 4; AM1 = Achievement Motivation Study 1

\* = directly significant; I = significant as a component of a higher-order interaction effect; X = measure not included in study materials

*Table 26: Significant effects and interactions on STB congruence for participants in high-status conditions across studies*

### 9.2.1. Conflict Study 2

In Conflict Study 2 the saturated model regressing conflict and group identifiability on APM performance was highly significant,  $F_{3, 31} = 4.918$ ,  $p = .007$ ,  $b = -7.802$ ,  $r^2 = .323$ ,  $adj.r^2 = .257$ . The simple effect for conflict was significant,  $t = -2.507$ ,  $p = .018$ ,  $b = -7.803$ , but was subject to the higher-order two-way interaction between conflict and group identifiability,  $t = 2.712$ ,  $p = .011$ ,  $b = 1.661$ , which is plotted in Figure 33.

The plot shows that stereotype-performance effects were much more pronounced for participants expressing higher perceived intergroup conflict. Participants expressing high perceived conflict were likely to perform congruently if perceived group identifiability was high, in other words if they perceived themselves to be representatives of the group, and particularly incongruently when group identifiability was low. This pattern was reversed when reported intergroup conflict was low.

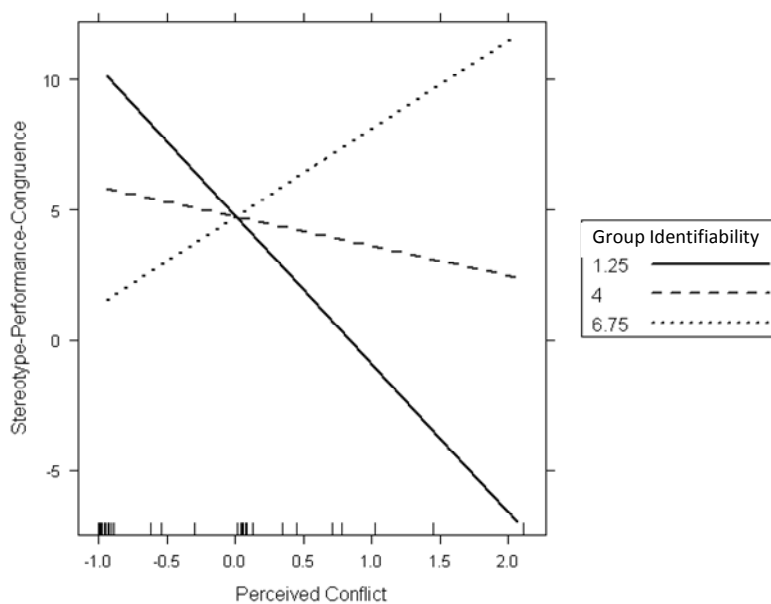


Figure 33: Interaction plot for conflict \* group-identifiability on STB congruence for Conflict Study 2

### 9.2.2. Conflict Study 3

The same model regressing perceived conflict, group-identifiability and conflict\*group-identifiability was applied to Conflict Study 3 and was significant (omitting two outliers on the basis of undue leverage),  $F_{3, 52} = 3.674$ ,  $p = .018$ ,  $r^2 = .175$ ,  $adj. r^2 = .127$ . Although the interaction effect was not significant as in the previous models, perceived conflict achieved conventional significance,  $t = -2.407$ ,  $p = .0197$ ,  $b = -.290$ , and group-identifiability was marginally significant,  $t = 1.895$ ,  $p = .064$ . Note that the effects were in

directions supporting the previous findings, namely, that perceived conflict generally reduced congruence for participants under boost and group-identifiability, or the sense of acting as a representative of the group, increased it.

### 9.2.3. Identifiability Study 4

Perceived conflict also played a role in modeling congruence for participants under boost in Identifiability Study 4 in which a model regressing conflict, individual-identifiability, social-desirability, conflict\*social-desirability and individual-identifiability\*social-desirability on STB congruence was significant,  $F_{5, 134} = 2.519$ ,  $p = .033$ . The perceived-conflict\*social-desirability interaction was significant,  $t = -2.891$ ,  $p = .004$ ,  $b = -1.262$ , and is plotted in Figure 34. The plot suggests, once again, that the stereotype-performance effect was more pronounced when participants perceived high intergroup conflict, but that this converted to congruent boost effects when social-desirability was *low* and to incongruent effects when social desirability was *high*.

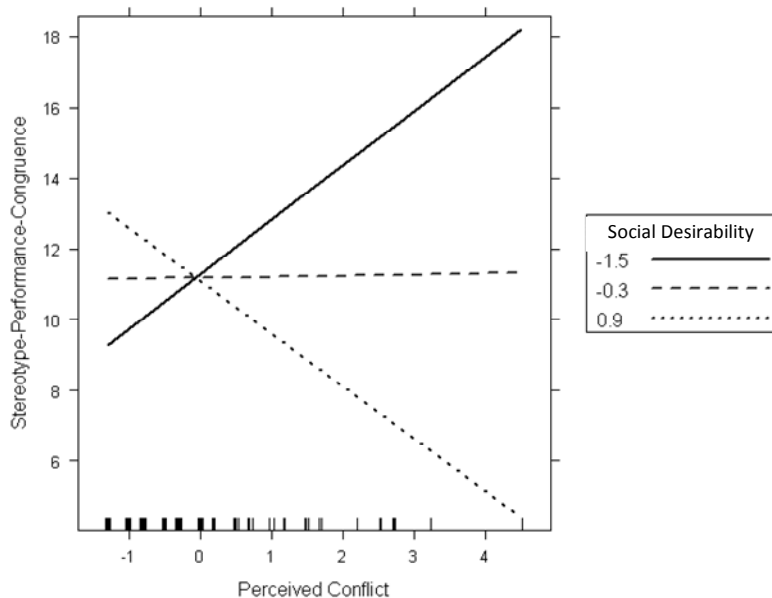


Figure 34: Interaction plot for conflict \* social-desirability on STB congruence for Identifiability Study 4

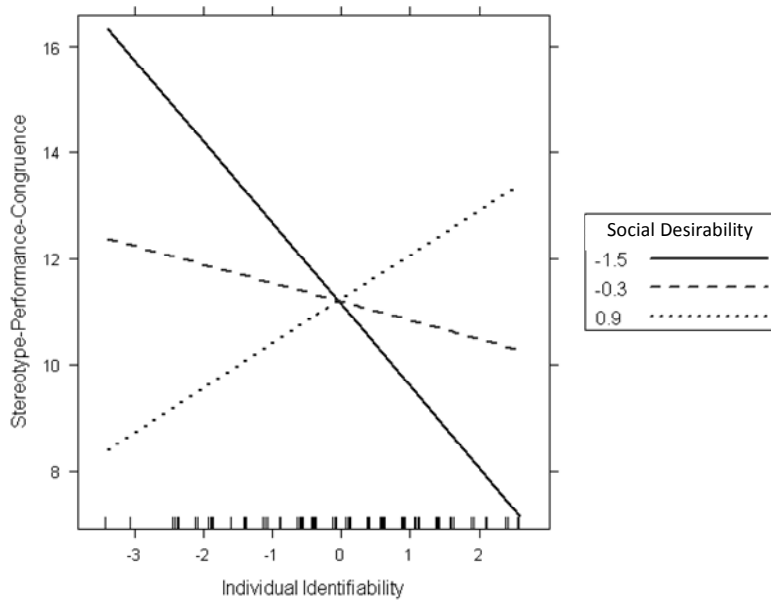


Figure 35: Interaction plot for individual-identifiability \* social-desirability on STB congruence for Identifiability Study 4

The individual-identifiability\*social-desirability interaction was also significant,  $t = 2.527$ ,  $p = .013$ ,  $b = .993$ , and is plotted in Figure 35. The plot reveals that congruent STB effects were likely under conditions of low individual identifiability and high social desirability, or high individual identifiability and low social-desirability. Conversely, incongruent effects were likely either when participants expressed high social desirability but felt relatively anonymous or expressed low social desirability and high individual identifiability.

#### 9.2.4. Permeability Study 2

A reduced version of the same model regressing perceived conflict, social-desirability and conflict\*social-desirability was applicable to Permeability Study 2,  $F_{3, 81} = 3.2$ ,  $p = .028$ ,  $r^2 = .106$ ,  $adj.r^2 = .073$ . The main effect for conflict was directly significant,  $t = -2.923$ ,



$p = .005$ , but subject to the significant two-way conflict\*social-desirability interaction,  $t = 2.123$ ,  $p = .037$ ,  $b = .124$ , which is plotted in Figure 36.

As in the previously discussed studies, the stereotype performance effect was much more pronounced – and in this case only evident – when perceived conflict was high. However, the direction of the interaction was reversed compared to Identifiability Study 4. In this study, it was participants with *high* social desirability and high conflict who performed congruently under boost and those with *low* social desirability who performed incongruently. Since in other studies (eg. Permeability Study 3, discussed below) the conflict\*social-desirability interaction was reversed at different levels of significant higher-order interactions, the unexpected direction of the interaction observed in the present study was likely to be related to unobserved higher-order interactions.

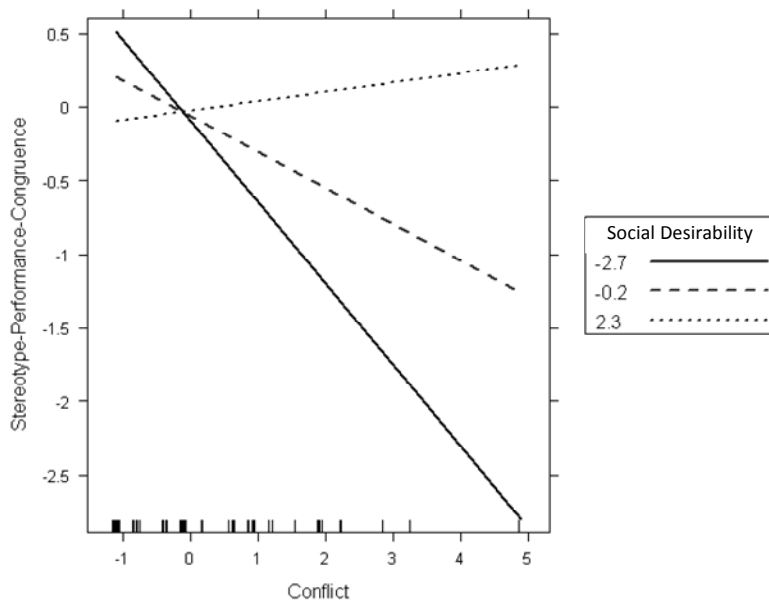


Figure 36: Interaction plot for conflict \* social-desirability on STB congruence for Permeability Study 2

### 9.2.5. Permeability Study 3

When the same model was applied to the boost subset of Permeability Study 3, regressing conflict, social-desirability and individual-identifiability on STB congruence (using logarithmic transformations on all variables to correct skewness and non-normality) the model was not significant,  $F_{7, 69} = 1.665$ ,  $p = .132$ ,  $r^2 = .145$ ,  $adj.r^2 = .058$ , although all effects but one were significant or marginally significant. The three-way conflict\*social-desirability\*individual-identifiability interaction was significant,  $t = 2.269$ ,  $p = .026$ ,  $b = 2.673$  (Figure 37).

As in previous studies, stereotype-performance effects were amplified when perceived conflict was high. When individual identifiability was high, participants performed incongruently under boost when they also expressed high social desirability. When individual identifiability was low, it was participants expressing high social desirability and high conflict who were more likely to perform congruently, while those expressing low social desirability and high conflict were more likely to perform incongruently. The relationship between identifiability, social-desirability and conflict is intelligible in the specific context of the minimal groups paradigm in which the participants' willingness to cooperate with the researcher, believe the experimental materials and perform in appropriate ways are critical features of the specific performance context.

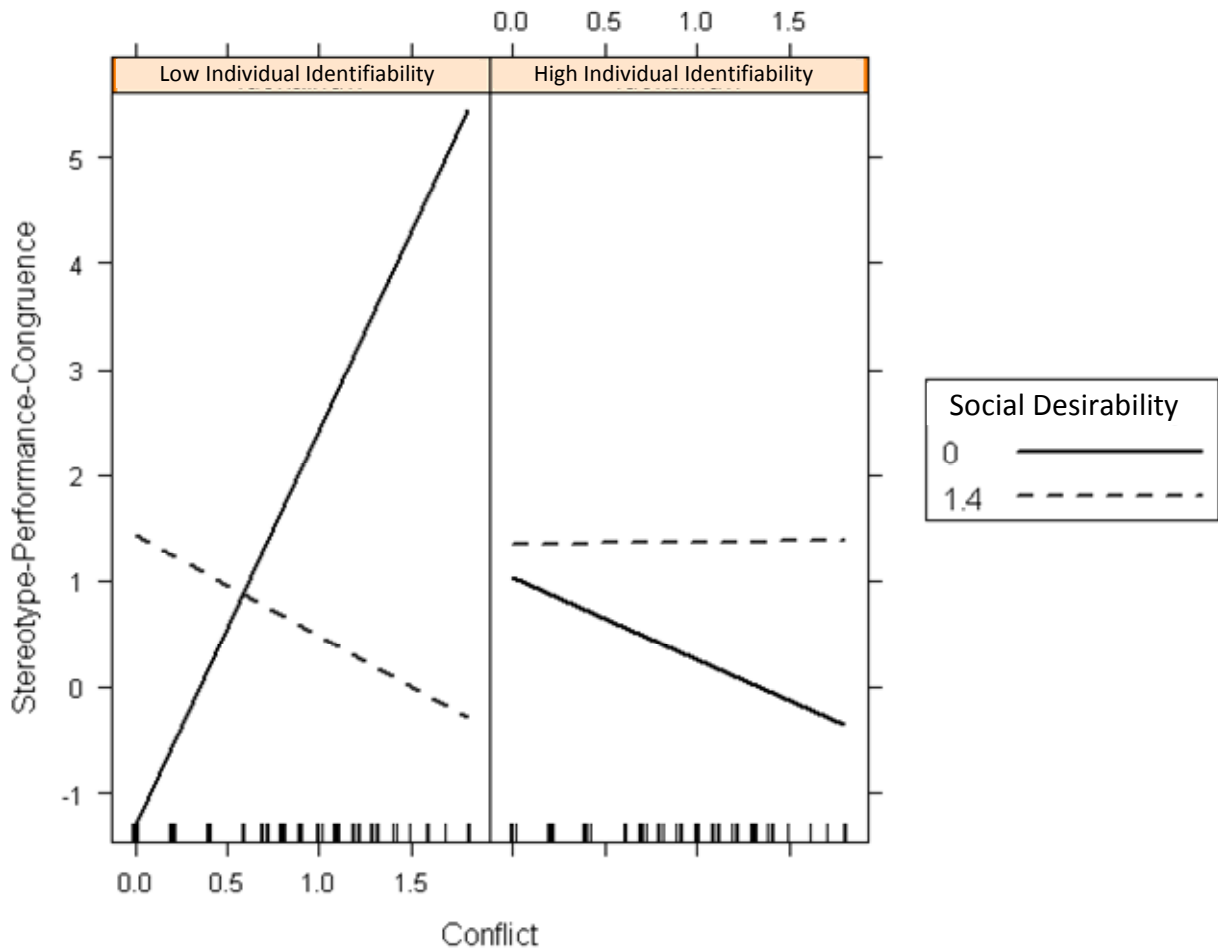


Figure 37: Interaction plot for conflict\*social-desirability\*individual-identifiability on STB congruence for Permeability Study 3

#### 9.2.6. Permeability Study 1

In Permeability Study 1 a fully saturated model regressing conflict, social-desirability, and group-identifiability (dropping one outlier) on STB congruence was significant,  $F_{7,41} = 2.312$ ,  $p = .044$ ,  $r^2 = .283$ ,  $adj.r = .161$ . All lower order effects and interactions were either significant or marginally significant (displayed in Table 27), but all were subject to the three-way interaction which very closely approached significance,  $t = -2.002$ ,  $p = .052$ ,  $b = -.079$ , plotted in Figure 38.

	<i>b</i>	Std. Error	<i>t</i>	<i>p</i>	
(Intercept)	0.589	0.207	2.845	0.007	**
Conflict	-0.589	0.325	-1.790	0.081	.
Social-desirability	0.374	0.164	2.282	0.027	*
Group identifiability	0.115	0.044	2.582	0.014	*
Conflict*social-desirability	0.389	0.197	1.978	0.055	.
Conflict * group-identifiability	0.120	0.068	1.782	0.082	.
Social-desirability * group-identifiability	-0.096	0.036	-2.690	0.010	*
Conflict * Social-desirability * group-identifiability	-0.079	0.040	-2.002	0.052	.

*Table 27: Model regressing conflict, social-desirability, and group-identifiability on STB congruence for Permeability Study 1*

The interaction plot shows that incongruent underperformance in the boost condition was most likely for participants who expressed low social desirability and low group identifiability. Conversely, congruent increases in performance under boost were most likely for participants low in social desirability who expressed high group identifiability, in other words, who believed that their group was being scrutinized and that their actions would reflect on the reputation of their group.

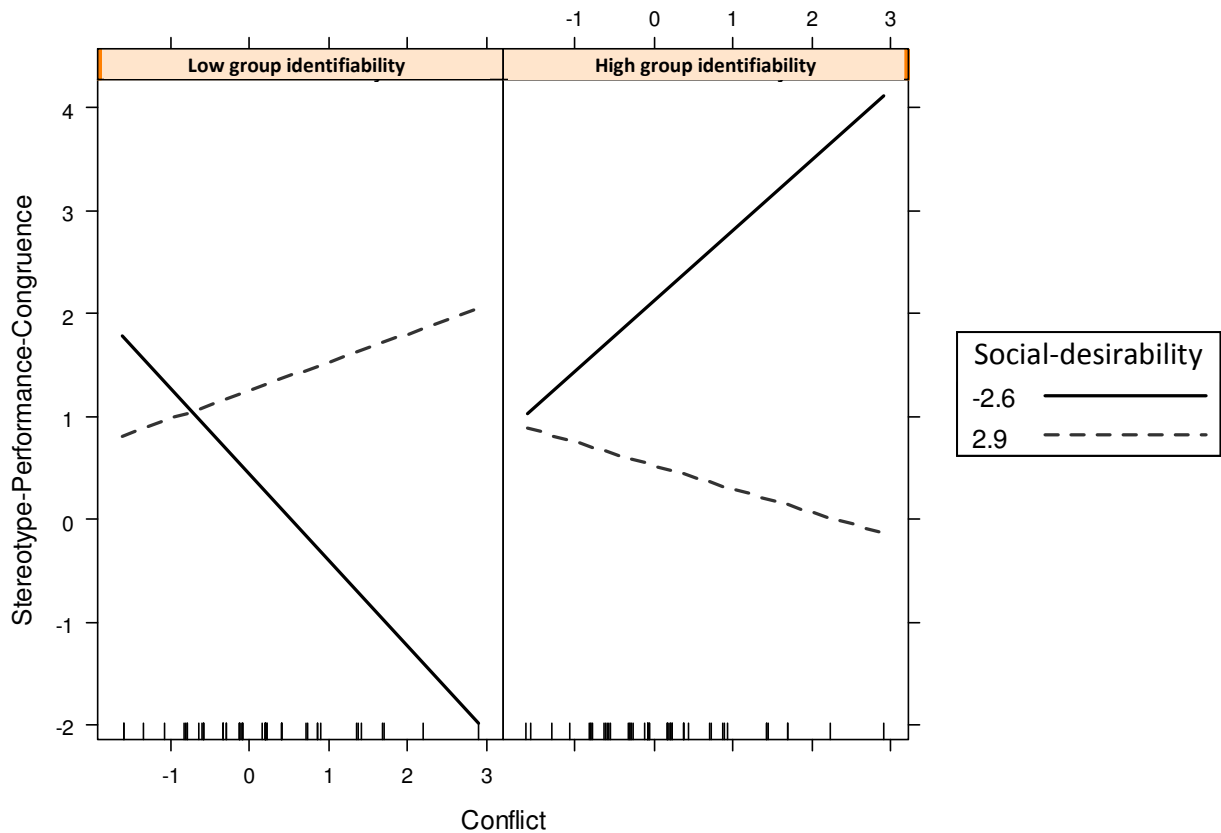


Figure 38: Interaction plot for conflict\*Social-desirability\*group-identifiability on STB congruence for Permeability Study 1

### 9.2.7. Achievement Motivation Study 1

The final study that allowed the calculation of STB congruence was Achievement Motivation Study 1 which, unfortunately, did not include conflict, group- identifiability or individual- identifiability measures and therefore could be used to test the conflict models that were applicable to the studies above. However, a significant model was fitted regressing social-desirability and ingroup-identification on STB congruence (using a square-root power transformation and dropping one outlier),  $F_{3, 29} = 3.204$ ,  $p = .038$ ,  $r^2 = .249$ ,  $adj.r^2 = .171$ . The simple effects for social-desirability,  $t = 2.945$ ,  $p = .006$ ,  $b = 6.101$ , and ingroup-identification,  $t = 3.068$ ,  $p = .005$ ,  $b = 8.538$ , were both highly significant, but subject to the

significant two-way social-desirability\*ingroup-identification interaction,  $t = -2.994$ ,  $p = .006$ ,  $b = -3.925$ , plotted in Figure 39.

The plot shows that stereotype performance effects under boost were more pronounced for participants reporting low social desirability and, for those participants, more likely to be congruent when they reported high ingroup-identification and incongruent when they reported low ingroup identification. Participants reporting high social-desirability were more likely to perform incongruently when they expressed high ingroup identification.

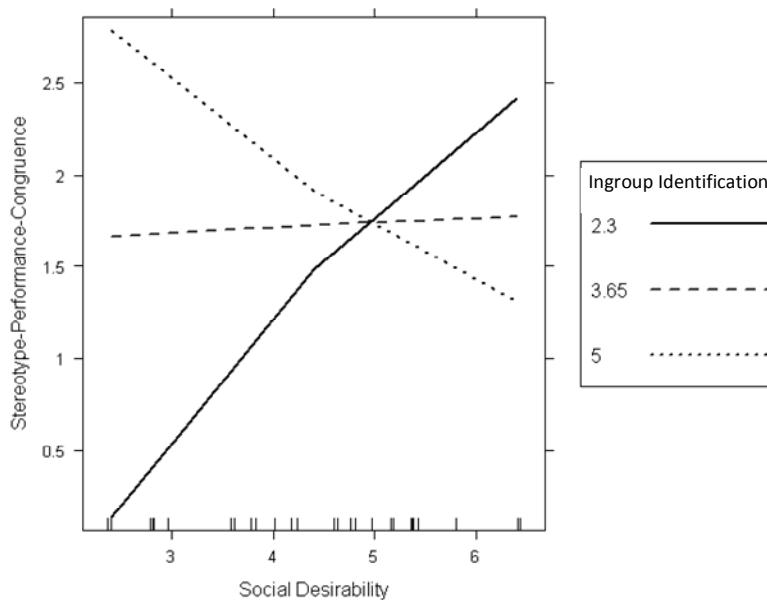


Figure 39: Interaction plot for Social-desirability\*ingroup-identification on STB congruence for Achievement Motivation Study 1

### 9.2.8. Summary of predictors of incongruent stereotype-performance effects under boost

The most consistent predictor of incongruent performance under boost was perceived intergroup conflict, which was significant as a simple effect or as a component of a significant interaction effect in all seven of the studies that allowed the analysis of stereotype-performance congruence. The next most common predictors were social desirability (5/7);

group-identifiability (5/7); individual identifiability (2/7); and ingroup identification (1/7).

This pattern of results demonstrates that the probability of reacting to a positive group stereotype with improved or reduced performance depends on (at least): how participants perceive the nature of the intergroup relations underlying the group comparison being made in the experimental context; their own individual orientation to the experiment and the extent to which they want to make a good impression in that context as individuals; and the extent to which they believe they are accountable for their performance as group members and as individuals.

### **9.3. Predicting incongruent stereotype-performance effects under threat**

As predicted, the social conditions that best predicted incongruent stereotype-performance effects under threat conditions were not the same as those resulting in incongruent performance under boost. Where the key predictor of incongruent stereotype-performance effects under boost (that is, performing worse in high status conditions) was perceived intergroup conflict, the crux for participants under threat was perceived intergroup permeability, which contributed significantly to models predicting STB congruence, either as a simple effect or as a component of an interaction effect, for each of the seven studies in which the congruence metric could be calculated. Social desirability contributed significantly in three studies and conflict, ingroup-identification and legitimacy contributed significantly to models in two studies each (see Table 28).

	C 2	C3	P2	P1	P3	ID4	AM1
Permeability	*	*	*	I	*	*	X
Social-desirability					*	*	*
Conflict		I	*				X
Ingroup-identification			I				*
Legitimacy		*	X				X
Individual-identifiability				I			X
Symbolic threat		X	X	*	X	X	X
Realistic threat	*	X	X		X	X	X
Social-desirability*ingroup-identification							*
Permeability*individual-identifiability				*			
Permeability*realistic-threat	*						
Permeability * conflict		*					
Permeability * ingroup identification			*				
Permeability*symbolic-threat*individual-identifiability				*			
N	36	66	82	50	73	138	49

*Note.* C2/C3 = Conflict Studies 1 and 3; P1/P2/P3 = Permeability studies 1 -3; ID4 = Identifiability study 4; AM1 = Achievement Motivation Study 1

\* = directly significant; I = significant as a component of a higher-order interaction effect; X = measure not included in study materials

*Table 28: Significant effects and interactions on STB congruence for participants in low-status conditions across studies*

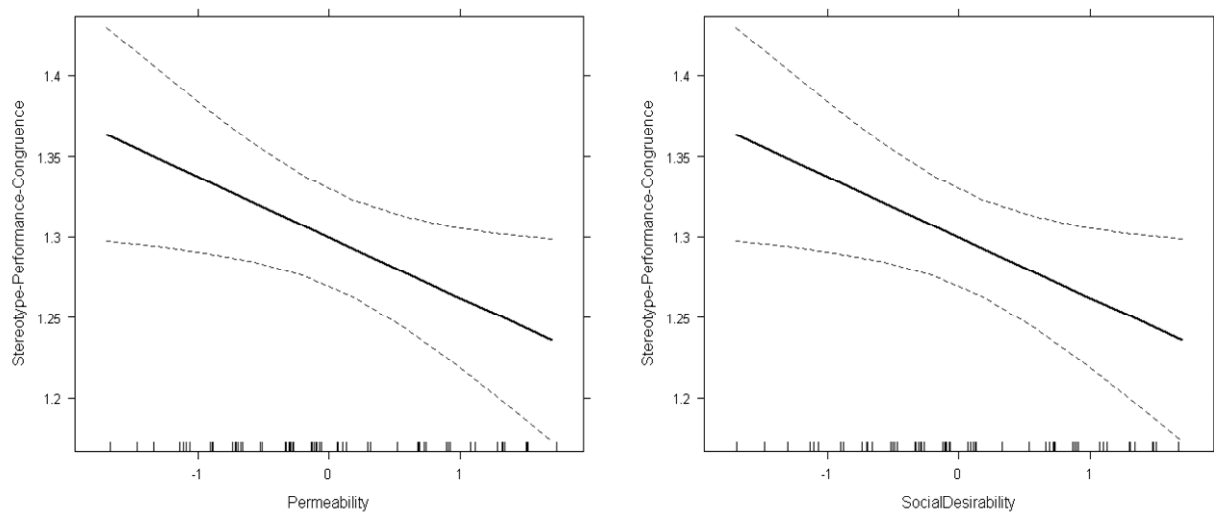
### 9.3.1. Permeability and social-desirability

In Permeability Study 3, STB congruence was significantly predicted by social-desirability, ingroup-identification and the permeability\*ingroup-identification interaction,  $F_{4,67} = 2.515, p = .050$ . The significant effects were for permeability,  $t = -2.656, p = .010, b = -.044$ , and social-desirability,  $t = -2.230, p = .029, b = -.023$ . Although the



permeability\*ingroup-identification interaction was not directly significant, the model was not significant if that interaction was excluded.

Both increased permeability and increased social desirability were associated with decreased STB congruence, (plotted in Figure 40). In other words, the participants most likely to respond to stereotype threat with congruent decreased performance as predicted by the STB literature were those who did not believe that it was possible to leave the low status group. Conversely, participants were likely to behave incongruently by performing well under threat if they expressed high permeability (in other words, the belief that they would be able to work themselves into the high status group) or high social desirability (and were therefore likely to be oriented to making a good impression in the experimental context).



*Figure 40: Effects of permeability and social desirability on STB congruence in Permeability Study 3(dotted lines represent 95% confidence intervals)*

A similar model regressing the simple effects of permeability and social desirability on STB congruence was significant for Identifiability Study 4,  $F_{2, 135} = 5.121, p = .007$ . Participants who expressed higher perceived permeability were significantly more likely to perform in opposition to the negative group stereotype, as were those with high levels of social desirability, as displayed in Figure 41.

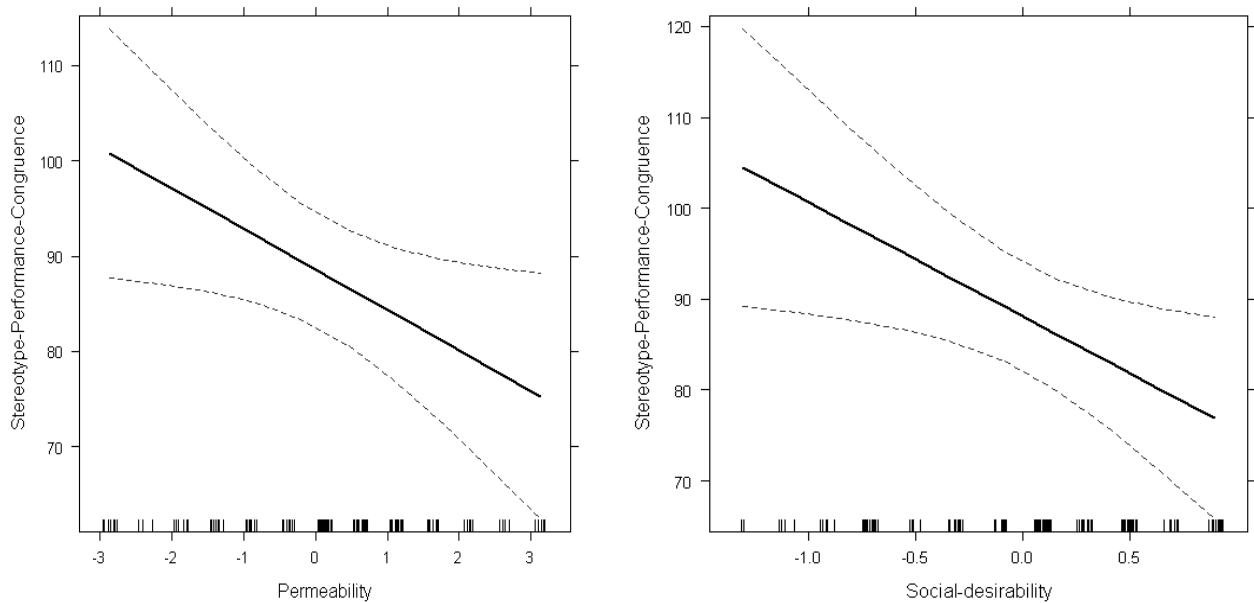


Figure 41: Effects of permeability and social desirability on STB congruence in Identifiability Study 4 (dotted lines represent 95% confidence intervals)

Although Achievement Motivation Study 1 did not include a permeability scale, it did include social desirability and ingroup identification scales. The saturated model regressing social-desirability, ingroup-identification and social-desirability\*ingroup-identification (excluding one outlier) was significant,  $F_{3, 37} = 11.86$ ,  $p < .001$ ,  $r^2 = .490$ ,  $adj.r^2 = .449$ , as were the simple effects for ingroup-identification,  $t = -3.761$ ,  $p = .001$ ,  $b = -3.028$ , and social-desirability,  $t = -4.388$ ,  $p < .001$ ,  $b = -3.167$ .

The social-desirability\*ingroup-identification interaction is plotted in Figure 42. Although the same interaction was significant in the model predicting STB congruence for participants under boost in the same study, the direction of effects was reversed in this model for participants under threat. Specifically, participants under threat were most likely to perform congruently when they either expressed low ingroup-identification with high social-desirability or high ingroup-identification with low social-desirability. Participants were most

likely to perform incongruently, in other words to perform well under threat, when they expressed high ingroup-identification with low social-desirability or low ingroup-identification with high social-desirability.

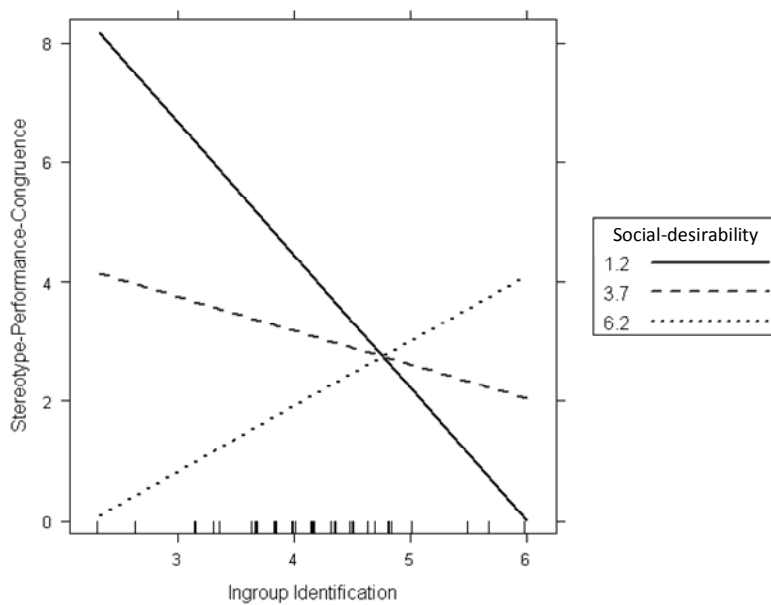


Figure 42: Interaction plot for Social-desirability\*ingroup-identification on STB congruence for Achievement Motivation Study 1

### 9.3.2. Permeability, symbolic threat and realistic threat

In two studies permeability predicted STB congruence in conjunction with symbolic or realistic threat. In Permeability Study 1, a saturated model regressing permeability, symbolic-threat and individual-identifiability on STB congruence was significant,  $F_{7, 42} = 2.553, p = .028$ . The significant effects included the simple effect for symbolic-threat,  $t = -2.197, p = .034, b = -.221$ , the two-way permeability\*individual-identifiability interaction,  $t = 2.139, p = .038, b = .531$ , and the three-way permeability\*symbolic-threat\*individual-identifiability interaction,  $t = -2.207, p = .033, b = -.135$ .

As displayed in Figure 43, the joint effects of permeability and social-desirability were reversed at different levels of individual-identifiability. When individual-identifiability

was low, participants with low social desirability were likely to perform congruently when they expressed low intergroup permeability and incongruently if expressed permeability was high. Conversely, participants with high social desirability were likely to perform congruently when permeability was high and incongruently when permeability was low.

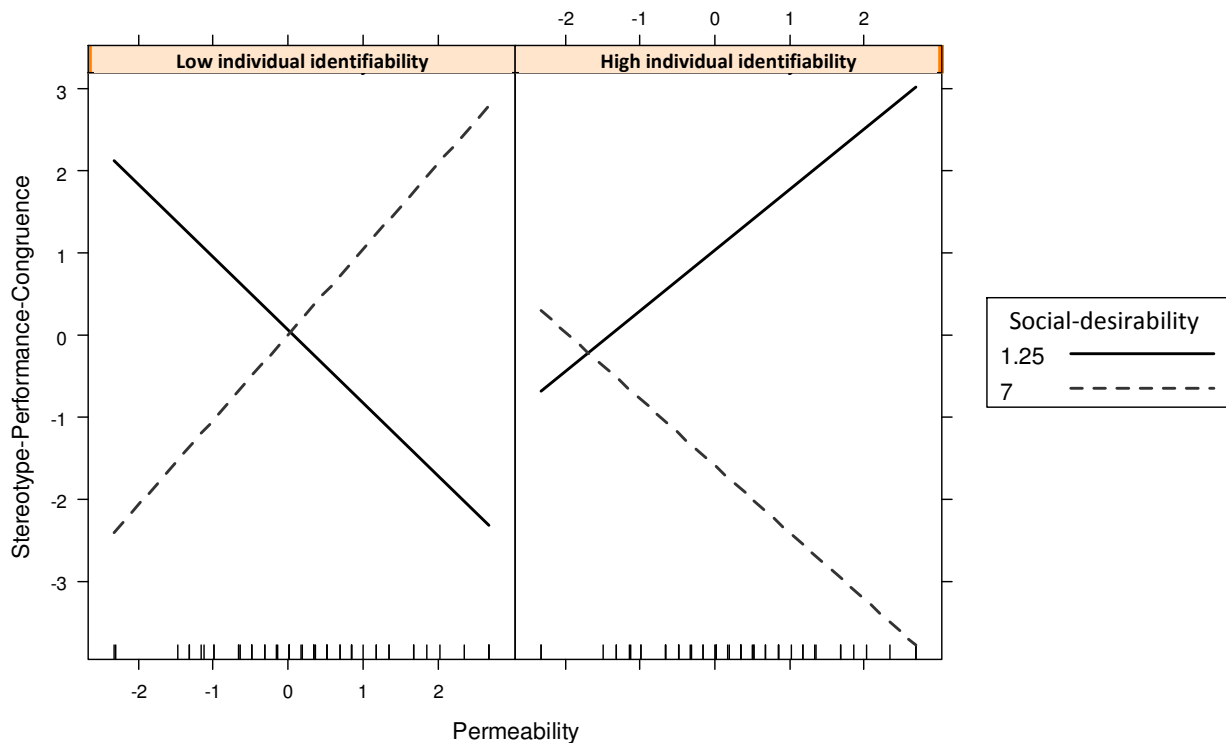


Figure 43: Interaction plot for permeability\*individual-identifiability\*sosial-desirability on STB congruence for Permeability Study 1

This pattern was roughly reversed for participants who perceived themselves to be individually identifiable. Although participants who expressed low permeability were virtually unaffected, those expressing high permeability and high individual-identifiability exhibited magnified stereotype-performance effects, with low social-desirability participants performing congruently and low social-desirability participants performing incongruently.

In Conflict Study 2 it was permeability and realistic threat (rather than symbolic threat) that best predicted STB-congruent performance. A model regressing perceived

permeability, group-identifiability, realistic-threat and the two-way permeability\*realistic-threat interaction was significant,  $F_{4,31} = 2.695$ ,  $p = .049$ ,  $r^2 = .258$ , as was the simple effect for permeability,  $t = -2.300$ ,  $p = .028$ , although the remaining effects were only marginally significant. Dropping one outlier increased the model significance,  $F_{4,30} = 3.678$ ,  $p = .015$ ,  $r^2 = .329$ ,  $adj.r^2 = .240$ , and resulted in the simple effect for realistic threat achieving significance,  $t = 2.728$ ,  $p = .011$ ,  $b = .443$ , along with the simple effect of permeability,  $t = -3.154$ ,  $p = .004$ ,  $b = -2.348$ , and the two-way permeability\*realistic-threat interaction,  $t = 2.749$ ,  $p = .010$ ,  $b = .629$ .

Both significant simple effects were subject to the significant higher order interaction, which is plotted in Figure 44 which shows that, in this study, participants who expressed low realistic-threat were likely to exhibit congruent stereotype-performance effects if they also expressed low permeability. However, they were likely to exhibit incongruent performance if they expressed high permeability. The pattern was reversed for participants expressing high realistic-threat, who were more likely to exhibit congruent performance if they expressed high permeability and incongruent performance if they expressed low permeability.

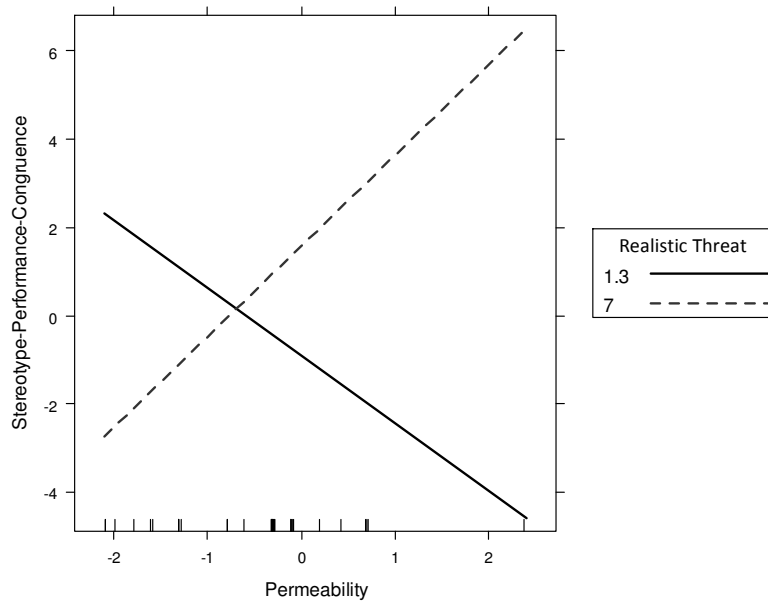


Figure 44: Interaction plot for permeability\*realistic-threat on STB congruence for Conflict Study 2

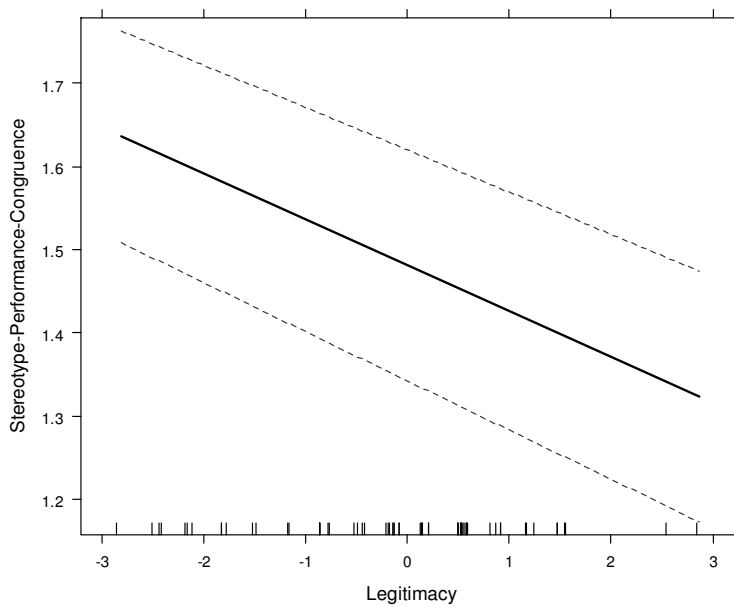
### 9.3.3. Permeability and conflict

In two studies permeability modeled STB congruence in conjunction with perceived conflict. In Conflict Study 3, a model regressing perceived conflict, legitimacy, permeability, and the two-way conflict\*permeability interaction (with all variables transformed to z-scores to correct their distributions) closely approached significance,  $F_{4,59} = 2.417$ ,  $p = .059$ ,  $r^2 = .141$ ,  $adj.r^2 = .083$ , while the effects for legitimacy,  $t = -2.389$ ,  $p = .020$ ,  $b = -.068$ , permeability,  $t = 2.021$ ,  $p = .048$ ,  $b = 0.059$ , and the two-way permeability\*conflict interaction,  $t = 2.012$ ,  $p = .049$ ,  $b = .059$ , achieved conventional significance. Only the simple effect for conflict was not significant,  $t = 1.572$ ,  $p = .121$ .

The simple effect for legitimacy can be directly interpreted, and is displayed in Figure 45. The plot shows that as legitimacy increased, so the stereotype-performance relationship became more incongruent. This means, strangely, that participants were *more* likely to challenge the negative stereotype by performing well if they expressed high belief in the

legitimacy of the status hierarchy. The relationship with permeability suggests that this may be due to participants in these conditions working to distance themselves from the low status group and demonstrate their affiliation with the high-status group.

As in the analyses of incongruent boost effects, stereotype-performance effect was magnified when participants reported high levels of intergroup conflict (see the conflict\*social-desirability interaction displayed in Figure 46). However, when conflict was high, the direction of the stereotype-performance effect was contingent on social-desirability, with participants reporting high social-desirability performing congruently and those reporting low social desirability performing incongruently and performing well under threat.



*Figure 45: Effect of legitimacy on STB congruence in Conflict Study 3(dotted lines represent 95% confidence intervals)*

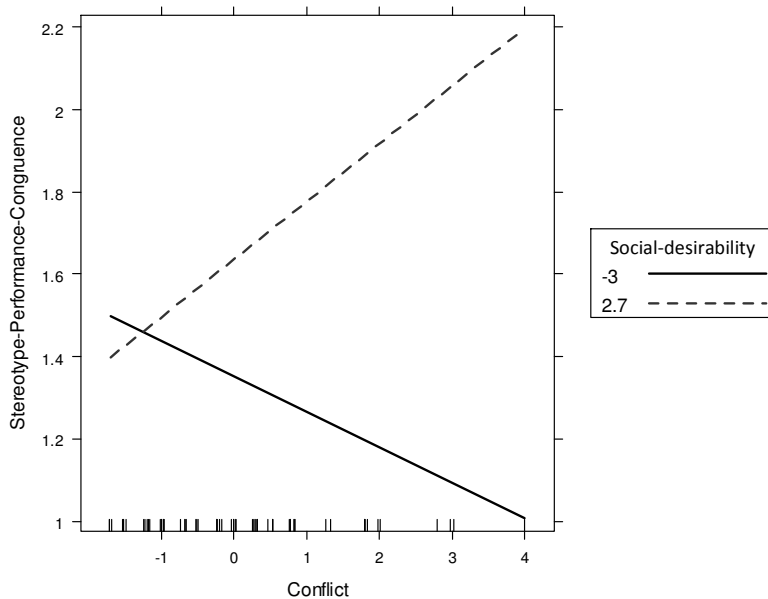


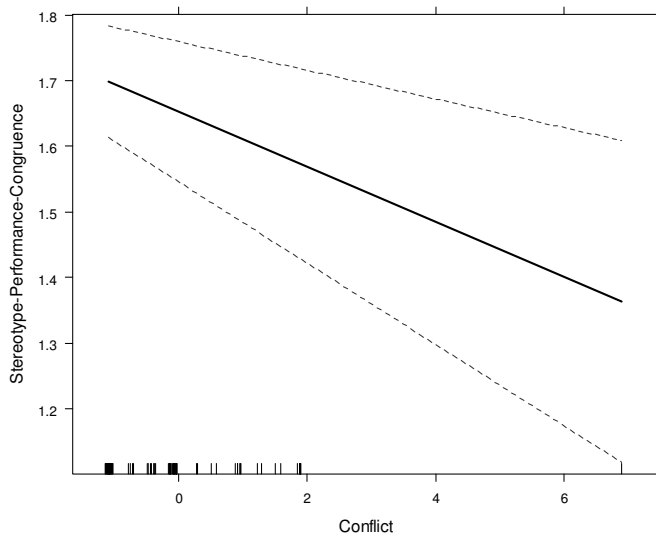
Figure 46: Interaction plot for conflict\*social-desirability on STB congruence for Conflict Study 3

In Permeability Study 2 a model regressing perceived permeability, conflict, ingroup-identification and permeability\*ingroup-identification on STB congruence was significant (applying a square-root power transformation and dropping two outliers),  $F_{4,75} = 3.264$ ,  $p = .016$ ,  $r^2 = .148$ ,  $adj.r^2 = .103$ . Two simple effects achieved significance, namely permeability,  $t = -2.118$ ,  $p = .037$ ,  $b = -2.236$ , and conflict,  $t = -2.291$ ,  $p = .025$ ,  $b = -.168$ , while the effect of ingroup-identification was marginally significant,  $t = -1.862$ ,  $p = .066$ ,  $b = -2.043$ . The two-way permeability\*ingroup-identification interaction was also significant,  $t = 2.044$ ,  $p = .045$ ,  $b = 1.046$ .

The plot for the simple effect of conflict (Figure 47) suggests that participants were more likely to perform congruently when they reported low perceived conflict and to perform less congruently – but also less predictably – when perceived conflict was high. The interaction plot for the permeability\*ingroup-identification interaction (Figure 48) shows that participants with low ingroup identification responded very differently to the low status



manipulation depending on the level of perceived permeability. Specifically, participants who reported low levels of ingroup-identification performed congruently with the status manipulation (that is, they performed poorly under threat). However, participants who reported high levels of permeability responded incongruently to the status manipulation, performing well under threat.



*Figure 47: Effect of conflict on STB congruence in Permeability Study 2(dotted lines represent 95% confidence intervals)*

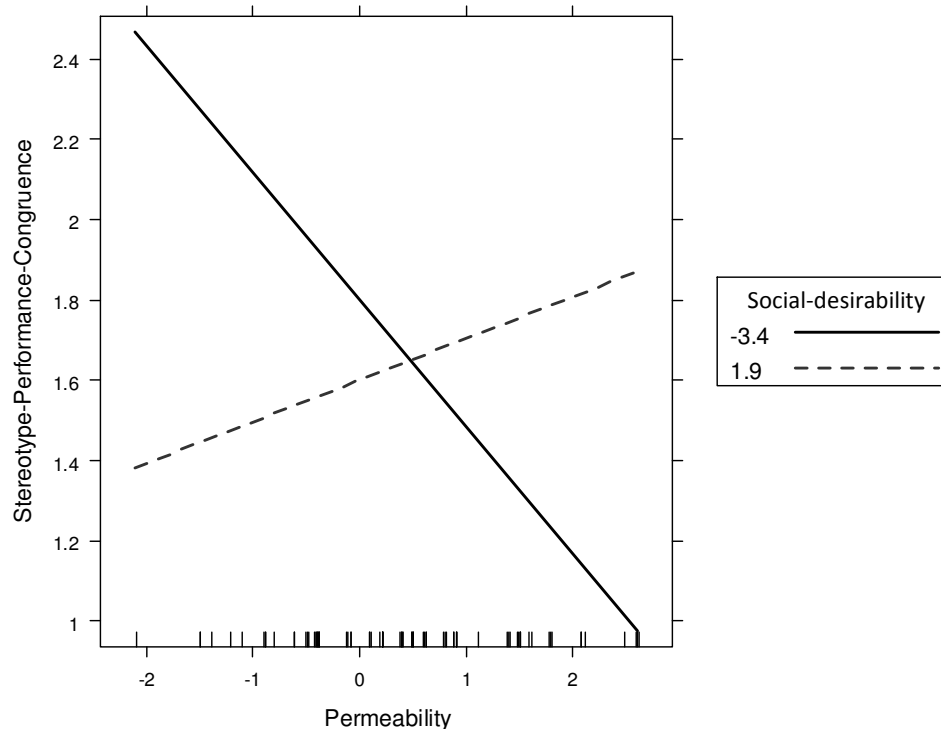


Figure 48: Interaction plot for permeability\*social-desirability on STB congruence for Permeability Study 2

#### 9.4. Discussion of the social-psychological conditions under which congruent or incongruent stereotype-performance effects become likely

The most general outcome of this analysis was that congruent and incongruent effects were common in both threat and boost conditions. This was despite the fact that many boundary conditions for stereotype threat were satisfied in the study design, specifically: (1) the APM is a difficult task; (2) in comparison to the general population, South African university students can be considered academic high achievers and part of the ‘vanguard,’ since only 9.1% of the population have any tertiary education at all and only 2.7% have undergraduate or postgraduate university degrees (Statistics South Africa, 2007); and (3) the experimental manipulations were designed to maximize the relevance of task performance to participants’ career and university success.

Critics could challenge the extent to which each of these boundary conditions was truly satisfied in this series of studies in order to defend the key tenet of stereotype threat, which is that in specific circumstances negative stereotypes are associated with reduced task performance. However, these analyses showed that conventional stereotype-threat effects frequently occurred, but that incongruent effects were also possible and likely in certain conjunctions of identity conditions.

Therefore, instead of policing the boundary conditions to ensure that analytic attention remains focused only on the congruent-threat quadrant of stereotype-performance effects, we should rather: first, accept that incongruent stereotype-performance effects are both possible and likely; and second, focus attention on understanding the social-psychological conditions under which each response-quadrant becomes most tenable.

The review of incongruent stereotype-performance effects already reported in the STB literature (p. 84) concluded that three broad levels of features would be important in modeling the congruence of stereotype-performance effects: (1) individual identity resources and habits-of-being such as impression management (Curhan & Overbeck, 2008); (2) features of the specific task-performance environment, such as whether participants are anonymous or identifiable; and (3) the socio-structural features of the intergroup context. This provides a useful framework for understanding the predictors of STB congruence for boost and threat subsets observed in the present studies.

#### **9.4.1. Boost**

Across this series of studies, the incongruence of stereotype-performance effects under boost conditions (when high status was associated with reduced performance) was most consistently predicted by perceived intergroup conflict (level 3), which was significantly implicated in all six of the studies for which the scale was available. Less consistent predictors were social desirability (level 1; four studies), group identifiability

(level 2; three studies), individual identifiability (level 2; two studies), ingroup identification (level 3; one study), or some combination of these in two- or three-way interactions.

In the three studies in which conflict was significant as a simple effect (Conflict Study 2, Conflict Study 3 and Permeability Study 2) the direct effect of conflict on STB congruence was negative. In other words, higher levels of perceived intergroup conflict were associated with incongruent stereotype-performance effects. However, there was only one study (Conflict Study 3) in which conflict was not subject to a higher-order interaction, so this result must be treated cautiously. Indeed, when conflict was considered jointly with social desirability, group identifiability and individual identifiability, the effect of conflict was generally to amplify the stereotype-performance effect, but the direction of the performance shift was usually reversed at different levels of the other predictors.

Social desirability was significant as a simple effect in two studies and, in both, was associated with congruent stereotype-performance effects. In other words, without considering higher order interactions, participants high in social desirability were more likely to perform well under boost. Conversely those who were low in social desirability, and therefore less concerned with making a positive impression in the experimental context, were more likely to perform incongruently with the activated stereotype.

Group identifiability, or the extent to which participants perceived themselves to be group representatives whose performance would reflect on their group's reputation, had a similar positive effect on STB congruence in Permeability Study 1 and Conflict Study 3. Congruent performance – performing well under boost – was more likely for those participants who expressed high group identifiability. Conversely, participants who expressed the belief that their performance would make little difference to the reputation of their group were more likely to perform poorly.

However these variables frequently interacted. Significant interactions included conflict \* group-identifiability (Conflict Study 2), conflict \* social-desirability (Permeability Study 2, Permeability Study 1, Identifiability Study 4), conflict \* individual identifiability (Identifiability Study 4), ingroup-identification \* social-desirability (Achievement Motivation Study 1), social-desirability \* ingroup-identification (Permeability Study 1), conflict \* social-desirability \* individual-identifiability (Permeability Study 3), and conflict \* social-desirability \* group-identifiability (Permeability Study 1).

These interactions were complex, but generally consistent with the social identity framework. Although the results are fragmented because they are derived from multiple low-power studies, they do allow a tentative conclusion that targets of positive stereotypes are likely to perform well, congruently with stereotype boost, when they identify strongly with their group, are concerned with making a positive impression in the performance context, believe that they are representatives of their group in the performance context and/or that their performance will reflect on the reputation of their group. However, these variables also interacted in complex ways that emphasize the importance of considering individual and sociostructural features of identity alongside the situational features of performance contexts generally considered by stereotype threat theorists.

#### **9.4.2. Threat**

For participants under threat it was perceived intergroup permeability that most consistently predicted STB congruence. In six of the studies for which measures of both permeability and STB congruence were available, permeability predicted STB congruence directly in five studies and as a component of a significant interaction in one study. In four of the five studies in which it was significant as a simple effect, perceived permeability was associated with decreased STB congruence. In other words, participants who performed better under threat were likely to express the belief that the intergroup boundaries were

permeable and that it would be possible for them to gain entry to the high status group.

However, the effect of permeability was also subject to higher-order interactions with other social identity variables.

Social-desirability negatively predicted STB congruence in three studies. In these studies participants who were concerned with creating a good impression in the experimental context were more likely to perform well in contravention of their negative group stereotype.

In two studies STB congruence was predicted by perceived intergroup conflict, which generally had a negative relationship to STB congruence. Participants who expressed higher levels of intergroup conflict in these studies were more likely to respond incongruently and perform better under threat. However, participants expressing high social desirability in conjunction with high conflict were particularly likely to succumb to threat and perform worse.

The balance of effects predicted STB congruence in only one study each. Ingroup identification (Achievement Motivation Study 1), legitimacy (Conflict Study 3), and symbolic threat (Permeability Study 1) were all negatively associated with STB congruence in their respective studies while realistic threat was positively associated with STB congruence in Conflict Study 2.

Although only permeability consistently predicted STB congruence for participants under threat across studies, the results paint a relatively consistent picture of the conditions under which incongruent responses to stereotype threat are possible or likely. Firstly, when targets of negative stereotypes believe that group boundaries are impermeable and that they have little chance of improving their own status by engaging in a social mobility strategy then they are more likely to succumb to the negative stereotype by performing worse under threat; Secondly, when participants are concerned with social desirability then they are more likely to perform well under threat, possibly as a way of distancing themselves from their low status

group. Thirdly, when they identify strongly with their low-status group, question the legitimacy of the status hierarchy, express higher levels of intergroup conflict or believe that the high-status group poses a symbolic threat to the identity of their group then they are likely to respond to threat by raising their game and performing better. These conditions are all related to the advancement of individual identity (permeability and social desirability) or concern for the reputation of the group (ingroup-identification, legitimacy, conflict and symbolic threat).

In summary, these results suggest that targets of negative stereotypes are likely to perform better if: (1) they believe that they can advance their individual identity by demonstrating that they are not prototypical members of the low status group (indicated by negative relationships between STB congruence and permeability and social-desirability); or (2) if they are motivated to act in defense of their group identity (indicated by negative relationships between STB congruence and ingroup identification, conflict, legitimacy, and symbolic threat).

### **9.5. Limitations**

Unfortunately it was decided at the start of the programme of research to undertake a large number of moderately sized studies to try to explore the effects of variety of social identity constructs on stereotype-performance effects. However, this design decision resulted in small to moderate sample sizes which then became too small for comfort when the emerging results required that analysis proceed by subsets.

Since power was generally low, it is likely that important predictors of STB congruence were missed in each study and that a clearer pattern of results would have emerged from a smaller number of larger studies. However, the fact that any significant effects emerged at all, that the results were relatively consistent across these small studies,

and that they were theoretically interpretable in social identity terms all suggest that the results obtained are theoretically meaningful.



## Chapter 10. General discussion

In their seminal review, Steele and his colleagues criticized the “dominant paradigms of psychology” for relying on “decontextualized internal processes and traits such as low self-esteem, low performance expectancies, oppositional attitudes towards mainstream achievement and the like” (Steele et al., 2002, p. 436), arguing that:

... the broadened explanatory framework developed in this research ... adds to the internal influences, an understanding of group differences in the intersubjective experience of the school, classroom, and test settings. By “intersubjective” we mean the capacity to represent what other people are thinking and likely to do and to incorporate that representation into the regulation of one’s own behavior, It is through this capacity that different groups of students come to understand that, based on their social identity and the stereotypes about it, they may be at risk of devaluation or obstruction in a school setting. And it is through this capacity that a host of external, life context influences—influences that have less to do with their internal psychology and more to do with their appraisals of how other people, structures, arrangements, practices, norms and so on are predisposed toward their social identity—come to influence their sense of belonging and performance in these settings. We hope that by exploring and documenting this intersubjective influence in academic performance, that we have both learned something useful about the problem we set out to understand and perhaps made a case for the value of an explanatory paradigm in social psychology that better incorporates the ongoing, intersubjective dimensions of social experience (p. 436)

The primary argument made in the present dissertation is that Steele and his colleagues did not go far enough in developing a framework for understanding the socially situated nature of task performance, or the active ways in which targets of stereotypes engage

with their identity resources and liabilities in the performance context. In fact, despite his stated objective to understand task performance as a situated and intersubjective experience of identity, Steele's work has systematically excluded good evidence that negative stereotypes sometimes improve performance and that positive stereotypes can also help or hinder it.

By 2002 there were already signs that positive stereotypes can increase task performance and that negative stereotypes can be associated with increased performance e.g. (Ambady et al., 2001; Kray et al., 2001; Shih et al., 1999) but this was not mentioned in Steele and colleagues' (2002) review which exclusively focused on the negative effects of negative stereotypes. By 2010 the evidence was mounting that both negative and positive stereotypes can either hinder or improve performance (eg. Crisp et al., 2009; Nguyen & Ryan, 2008; Walton & Cohen, 2003), but again, none of this evidence was acknowledged in *Whistling Vivaldi*, Steele's (2010) book.

This focus on one quadrant of stereotype-performance has not been lost by the general audience for whom the book was intended. For example, Amazon.com's editorial synopsis describes the book as follows: "... Claude M. Steele ... shares the experiments and studies that show, again and again, that exposing subjects to [negative] stereotypes ... impairs their performance in the area affected by the stereotype." (Amazon.com, 2010). A lay reader says "I was really amazed with the premise of the book: that anyone who is reminded in any way of a stereotype against them before a test *inevitably* does worst [sic] on the test" (Andres, 2010, emphasis added).

The strategic ring-fencing of stereotype threat as referring only to the negative impact of negative stereotypes is not limited to Steele's work: it is endemic in the stereotype threat literature, for example Spencer, Logel and Davies' review due to be published in the *Annual Review of Psychology* (Spencer, Logel, & Davies, 2011) define stereotype threat as "the risk

of being judged in light of ... negative stereotypes [which] can elicit a disruptive state that undermines performance” (abstract), thereby disregarding three classes of empirically confirmed stereotype performance effects: positive effects of negative stereotypes (eg. Crisp et al., 2009; Kray et al., 2004); and both positive and negative effects of positive stereotypes (eg. Smith & Johnson, 2006; Walton & Cohen, 2003).

The problem with disregarding the evidence that negative stereotypes can sometimes improve performance and ignoring positive stereotypes altogether is that instead of developing an understanding of “the ongoing, intersubjective dimensions of social experience” as Steele and colleagues planned (2002, p. 436), these ringfencing practices are resulting in increasingly deterministic intra-individual models of stereotype-performance effects in which stereotypes are thought to act on passive subjects to disrupt performance.

However, the present study provides evidence to support the argument developed in the literature review that: (a) both positive and negative stereotypes can have congruent or incongruent effects on performance and there are therefore four quadrants of stereotype-performance effects to be explored; and (b) that people can and do respond to negative and positive stereotypes actively and strategically. The problem is that, unless some explanatory mechanism or empirical regularities can be identified to predict when task performance will be congruent or incongruent these findings are as useless as a weather report announcing that it will either be raining or fine tomorrow.

The results of the present studies support the argument that it is the meaning of good or bad performance as an interpretable act in the specific context that targets of stereotypes find themselves that determine – at least partially – whether they respond with better or worse results. In other words, where the majority of STB research treats stereotypes as a stimulus and performance as the response in a stimulus-response network, the results of the present programme of research suggest that task performance can be considered one way (but not the

only way) of producing identity embedded in the specific context in which the task is performed.

If this is the case, then we must also (c) acknowledge that identity is complex and – as Steele and his colleagues suggested (Steele et al., 2002) – begin to explore the experience of STB as an identity contingency (Steele, 2010) and task performance as an act with identity-constructing meaning in that context.

### **10.1. Task performance as a socially meaningful and accountable act**

The results of these ten studies support the general assertion in the STB literature that some features of situated identity are likely to galvanize increased engagement while others are likely to result in disengagement. However, in contrast to conventional STB research, here it is argued that the impact on task performance is unlikely to be directly predictable from the valence of a simple stereotype. It is not the stereotype that directly impacts on performance (or at the very least, not *only* the stereotype), but how the stereotype constrains targets' identities in the performance context and revises the way that good or bad performance would be read as a meaningful act, either by other potential observers or by the targets themselves (cf. Klein et al., 2007).

Therefore, task-related group status (ie. the 'stereotype' in 'stereotype-threat') should be considered relative to other features of identity and intergroup relations in the performance context. Understanding which features of situated identity are likely to result in reduced performance (the goal of STB research) consequently requires understanding the specific meaning of task performance for participants in the performance context and the consequences that good or bad performance may have for participants and their groups in that context.

Once task performance is understood as an accountable act with specific identity-related meaning, it must also be acknowledged that opposite performance strategies could be

equally justifiable responses to specific social predicaments. Therefore, because high or low performance is such a deprived “message,” it is possible to think of multiple valid reasons for high or low performance in most possible identity contingencies. In other words, it is unlikely that there is a one-to-one deterministic relationship between specific identity contingencies and performance strategies. A much more valid conclusion is that participants actively engage with their identity in the specific task performance context; the meaning of high or low performance in that context; and the impact that high or low performance would have on their own fortunes or the fortunes of the group in that specific context and more generally.

Obviously, in an anonymous experimental context, the identity-related consequences of good or bad performance are very different than in a real-life examination or job-interview (Sackett, Hardison, & Cullen, 2004). Therefore the specific identity-constructing features of task performance in a given context are not incidental to the pattern of effects observed.

This argument contradicts the general positivist assumption that specific features of experimental environments not directly relevant to the researcher’s hypotheses can be ignored when generalizing to other contexts (cf. Danziger, 1994) and potentially explains why results from experimental studies of STB have not generalized to real-life performance settings particularly well (eg. Cullen et al., 2004; Cullen et al., 2006; McFarland et al., 2003; Nguyen et al., 2003; Ployhart et al., 2003; Stricker et al., 1999; Stricker & Ward, 2004). This is not to say that there are no regularities in stereotype-performance effects, but rather that – to date – it has been the particular strategic features of task performance that have been most likely to be ignored in the process of generalization. However, the results of the present programme of research suggest that it is these particular identity-constructing features of task performance that are critical to understanding stereotype-performance effects.

### **10.2. Three levels of identity structure implicated in stereotype-performance effects**

Most studies in stereotype threat have explored intra-individual mediators and moderators such as effort, anxiety, self-handicapping, evaluation-apprehension, confidence, stereotype endorsement and self-esteem (Smith, 2004). Some studies have examined variables assessing the way that participants are positioned in the performance context with respect to the high and low status groups and the powerful-others administering the test itself, such as ingroup identification, domain identification and perceptions of the test (Smith, 2004) or individual identifiability (Ambady et al., 2004). However, social identity theory predicts that a third level of socio-structural variables will also be important for understanding stereotype-performance effects, such as permeability, legitimacy and intergroup conflict. This prediction was confirmed in the present programme of research.

### **10.3. Usefulness of the SIT framework for modeling and explaining stereotype-performance effects**

Reicher (2004) has criticized the way that social identity theory has been applied in many mainstream psychology applications, arguing that:

The task of social psychology is to explain the flexibility of human beings in creating and relating to their social worlds. Social identity and self-categorization theories provide a thoroughgoing interactionist framework for achieving such a task. However, in order to do so, it is necessary to avoid reductionist misreadings of the theories that would explain human social action simply by reference to psychological processes, without examining how the play of process depends on the cultural and structural settings in which they occur. More specifically, to the extent that self-categories shape social action, flexibility is achieved through the

categories to which we belong, the others with whom we compare ourselves, and the dimensions along which such comparisons occur (p. 921)

Despite Steele and his colleague's intentions to develop an "explanatory paradigm in social psychology that better incorporates the ongoing, intersubjective dimensions of social experience" (Steele et al., 2002, p. 436), the field has relied on individualized and decontextualised notions of identity. Even where more nuanced frameworks of identity have been applied, they have generally been reduced to their most individualistic components (eg. Schmader, 2002b), although there have been some exceptions (Crisp et al., 2009; Haslam, Salvatore, Kessler, & Reicher, 2008).

The present programme of research attempted to explore the situation of STB by applying a fully interactionist social identity framework. In so doing, these studies have been able to model the identity contingency of STB in a much more nuanced way than previous studies, because the SIT framework provided theoretical scaffolding and a set of explanatory constructs by which the extra-individual and socio-structural features of the local production of identity could be understood in addition to the individual constructs generally considered in STB research.

A close reading of the social identity literature suggested that intergroup conflict, permeability, and identifiability might be critical to the way that stereotype targets might approach a performance task, so the programme of research was designed with these as key manipulated variables. However, other sociostructural and identity variables such as stability, legitimacy, ingroup identification, ingroup closeness, realistic threat and symbolic threat were also measured and deserve closer attention in future studies.

In all ten studies the combination of status manipulations (STB), SIT identity variables (such as identifiability, ingroup identification, ingroup and outgroup closeness), and

SIT sociostructural variables (including permeability, legitimacy and stability) were useful components of models of stereotype-performance effects. Indeed, in most studies, task-related status had no discernable effect on performance at all unless other SIT variables were also considered.

### **10.3.1. The impact of experimentally manipulated SIT variables**

The impact of three key social identity variables on STB was experimentally explored in this series of studies, with four studies experimentally manipulating status with identifiability, and three each manipulating status with conflict and status with permeability. The clearest findings from each of these will be discussed in turn.

#### ***10.3.1.1. Identifiability***

Increasing individual identifiability was generally related to slightly improved performance regardless of status. However, the effects of identifiability were sensitive to specific variations of the performance situation. Identifiability was also associated with reversals of the STB effect in conjunction with other identity-related variables, such as minority status.

These results confirm Ambady et al's (2004) conclusion that “individuation” reduces STB effects, but the argument for why this might be the case is quite different. Ambady et al. argued that individuation is “a protective measure against the detrimental effects of negative stereotype activation,” but here the evidence suggests that identifiability produced different forms of accountability in the experimental context and thereby transformed the meaning and accountability of good or bad performance on the dependent task. In other words, identifiability shifted the way that task performance reflected on individual and group reputation and thereby transformed the meaning of good or bad performance.

These results confirmed that threat can improve performance, if individuals believe that their performance will impact either on their own reputation or improve the reputation of



their group. These results are similar to those of Deaux, Bikmen, Gilkes, Ventuneac, Joseph, Payne, & Steele (2007) who found that Afro-Caribbean participants performed better under threat, presumably as a way of demonstrating that they and their group were not prototypical of the African-American stereotype of intellectual inferiority.

However, the results of the present studies suggested that it is not simply being identifiable or anonymous that is important, but whom one is identifiable *to*. For example, in Identifiability Study 3 participants performed particularly well under boost conditions when they believed their performance would be identifiable to other members of their ingroup rather than to outgroup members or experimenters. These results add weight to the argument that it is not *individuation* as a free-standing phenomenon that results in different performance under identifiable or anonymous conditions as argued by Ambady and her colleagues (2004), but the transformation of the meaning and reach of good or bad task performance for individual and group reputations. In other words, if task performance is an interpretable act with consequences for identity and reputation in the immediate performance context (in experiments) and more generally (in real life) then the extent to which your performance is visible to particular audiences and coupled to your reputation as an individual and/or to the reputation of your group must transform the way that you approach the performance task.

If these conclusions are correct, then it is not possible to support a deterministic law of behavior that increasing identifiability will increase individuation and thereby reduce stereotype threat effects. Instead, the effect of identifiability depends on the social meaning of high or low performance in the specific performance context, how individuals and groups will be accountable for that performance, and how the audience to whom the performance is visible will read that performance.

### 10.3.1.2. *Conflict*

Intergroup conflict was shown to moderate stereotype-performance effects in all three of the studies in which conflict was manipulated, although its influence was often only statistically detectable once other identity variables were accounted for. The actual pattern of relationships between status, conflict and performance differed across studies.

In the first conflict study, STB effects seemed less pronounced for participants assigned to low conflict conditions and more pronounced for participants in high conflict conditions. The conditions that maximized the negative effects of stereotype-threat were precisely those of the standard STB experiment (low perceived identifiability and low perceived conflict).

In the second study (Conflict Study 2), there was a direct interaction between conflict and status, with those in the low conflict condition exhibiting a reversal of the conventional STB effect and those in the high conflict condition exhibiting a conventional STB effect. However, closer investigation revealed that the effect of conflict on STB depended on levels of baseline performance on the SILS, with high baseline performers exhibiting reversed threat effects when conflict was *low* and low baseline performers performing better under threat when conflict was *high*. This pattern was closely replicated in the third conflict study (Conflict Study 3).

Of course, it has long been acknowledged in the STB literature that stereotypes have different effects on performance for targets at different levels of baseline performance. This evidence has been used to justify a boundary condition – a line in the theoretical sand demarcating stereotype threat which then allows other forms of stereotype-performance effect to simply be ignored (eg. Steele, 1997, 2010; Steele et al., 2002).

However, the argument supported by the present data is that participants enter the performance context with different motivations and understandings of performance in the

performance context. The manipulations therefore transform the meaning of good or bad task performance in different ways for participants with high or low baseline performance. In other words, the present results support the more parsimonious explanation that the opposite effects exhibited by low and high baseline performers result from the same underlying social mechanism. That mechanism, it seems, is at least partly related to the way that the meaning of, and accountability for, task performance shifts for low and high baseline performers.

The simplest conclusion, then, is that positive and negative stereotypes did not deterministically influence task performance as predicted by conventional STB models. Instead, participants' APM performance was oriented to the social demands of the intergroup context in light of their own abilities and initial orientation to the experimental context.

These results all challenge the general assumption in the stereotype literature that stereotypes impact on performance directly via individualistic cognitive processes. Instead, the results suggest that stereotypes impact on performance to the extent that they render task performance meaningful and relevant to individual and group reputation. Understanding the effect of conflict on STB, therefore, requires understanding the way that high or low conflict conditions transforms the meaning and accountability of good or bad task performance and the types of individual and group identity strategies that are validated by specific combinations of conflict and other features of situated identity.

#### ***10.3.1.3. Permeability***

As predicted, by SIT, the combination of status and permeability generally predicted ingroup identification and other measures of group position (ingroup and outgroup closeness) quite well, and in patterns congruent with SIT predictions. These group-position variables jointly predicted APM performance in conjunction with status and permeability in all three studies, but specific patterns of performance sometimes (but not always) differed from both

SIT (Ellemers, 1993; Ellemers et al., 1988; Ellemers et al., 1990) and STB (eg. Schmader, 2002a; Smith, 2004) predictions.

Previous studies in the STB genre have explored ingroup identification as a unitary freestanding construct. However, the results of the present studies support the argument made by SIT theorists that ingroup identification is a cognitive-affective indicator of actual and aspirational individual positioning in the perceived intergroup context that is sensitive to sociostructural features such as perceived permeability (Ellemers, 1993; Ellemers et al., 1988; Ellemers et al., 1990; Mummendey et al., 1999).

In the present studies ingroup identification did improve statistical models of task performance under threat or boost, but only jointly with permeability. In other words, the effect of ingroup identification was only statistically meaningful in conjunction with the permeability conditions to which it was related. This suggests that “ingroup identification” does not impact on performance under threat or boost in a simple deterministic way, but has different identity implications in low or high permeability conditions and, therefore, different effects on performance.

As with studies exploring identifiability and conflict, the pattern of results in each study were most intelligible once the particulars of the performance context were considered and task-performance was understood as a meaningful and accountable act within that context. These results confirmed that permeability, like conflict, is an important sociostructural feature of situated identity that can transform experiences of STB and stereotype-performance effects.

However, the results also suggest a reason that the empirical results of studies exploring the effects of group- and domain-identification on performance under STB conditions have been so mixed. These studies have tended to treat STB as an abstract and invariant phenomenon and ingroup identification as a stable individual trait. However, as

predicted by SIT, the present studies have demonstrated that ingroup identification is a fluid indicator of actual and aspirational positioning within an intergroup context, sensitive to socio-structural features of intergroup relations (such as permeability) and the social demands placed on people in the immediate performance context.

#### **10.4. Resistance**

STB generally assumes that the experimenter's version of the stereotype/group identity is accepted uncritically by participants and that there is a deterministic relationship between stereotypes and performance. Indeed, the argument that agreement with the stereotype is not necessary for stereotype threat to occur (eg. Steele, 1997) is a cornerstone of its appeal, because it distances STB from internalized accounts of inferiority by placing the 'responsibility' for poor performance more squarely on the context rather than on the poorly performing minority member of the negatively stereotyped group.

In this series of studies, status alone did not significantly predict performance in any individual studies. However, in all studies in which a resistance metric could be calculated, adding resistance to statistical models substantially improved the sensitivity of models to stereotype-performance effects. The first conclusion that can be drawn from this is that participants' individual- and group-level orientations to the imposed group stereotype are critical to task-performance in relation to that stereotype, at least with the relatively weak stereotypes invoked in this programme of research.

Some might argue that resistance as conceptualized here is simply a measure of the failure of a poor manipulation, and this is a fair criticism. However, if this were so, it would not be expected to contribute to models of APM performance beyond chance levels. On the contrary, the results show that the effects of status on performance are undetectable beyond chance unless resistance is also included in the model.

The empirical fact that participants in these studies frequently resisted both the status and SIT manipulations, and that the effect of STB on performance could generally only be modeled once such resistance was accounted for challenges the generally passive way that stereotype targets are considered in the STB literature; as corks bobbing on the sea of outgroup opinion. Instead, the analysis of resistance in the present studies makes a strong case that participants responded to stereotype manipulations both actively and strategically.

Obviously, some stereotypes are easier to resist than others, so resistance is likely to play a different role in studies with more powerful stereotypes. However, that is exactly why stereotypes and resistance should be considered in relation to specific features of the intergroup context. These features include – but are not limited to – sociostructural features of intergroup relations, such as task-related group status (STB), the legitimacy, permeability, and stability of the intergroup status hierarchy, and features that position the individual within the performance context, such as identifiability, ingroup identification, ingroup and outgroup closeness and so on.

However, the capacity for an individual to resist is not unlimited: it depends on individual identity resources and liabilities, such as ability, social desirability, self-esteem, achievement motivation, familiarity with the experimental situation as a genre of experience (cf. Danziger, 1994) and group level resources, such as additional negative or positive identities that confirm or undermine the status manipulation, social resources such as perceived group support, and discursive resources (such as, but not limited to cognitive alternatives) that allow the stereotype to be strategically supported or undermined.

To add to the theoretical complexity, the experience of resisting a positive stereotype (challenge) is likely to be quite different to that of resisting a negative one (denial). Therefore these two types of resistance are likely to have different relationships to situated identity and, therefore, on situated task-performance.

The results of the present studies demonstrate that group status (in the form of activated negative or positive stereotypes) is not the only identity constraint on performance. Indeed, in the present studies, with the relatively weak stereotypes invoked, other features of situated identity such as identifiability and conflict often had more direct effects on task performance than status.

### **10.5. Incongruent stereotype-performance effects**

There is mounting evidence at the fringes of the STB literature that positive stereotypes can inhibit performance and negative stereotypes can galvanize it. These incongruent effects were observed frequently in the present programme of research.

#### **10.5.1. Positive stereotypes undermining performance: slipstreaming**

*Slipstream:*

*“an assisting force regarded as drawing something along behind something else” (Oxford Dictionaries, 2010)*

*“an area of reduced air pressure and forward suction that is directly behind and caused by a rapidly moving vehicle” (Encarta, 2009)*

*Slipstreaming/Drafting:*

*“Riding closely behind another rider in the slipstream ... enables the second rider to maintain speed with less effort. A drafting rider can save as much as 25 percent of effort and be more rested at the finish of the race” (Howard, 2010, p. 169)*

While the evidence in the literature is clear that incongruent responses to negative and positive stereotypes are possible, these responses are considered unconventional and do not

enter mainstream accounts of stereotype threat (eg. Steele, 2010; Steele et al., 2002). Having said that, the ironic effect that has received most attention at the fringes of the STB literature is the phenomenon of performing poorly under boost. This is generally explained as ‘choking under pressure,’ with the argument that for ‘model minority’ groups with strong positive stereotypes in competitive domains (such as Asian American math students) being aware of the positive stereotype can increase motivation to unproductive levels and ironically interfere with optimal performance (eg. Cheryan & Bodenhausen, 2000).

In the present studies there were many instances in which targets of positive stereotypes were shown to perform significantly worse than participants in control or threat conditions. However, there was no evidence that this ‘underperformance’ was related to increased pressure. Instead, the specific features of identity in which members of the high status group were likely to perform poorly in boost conditions suggested that they are most likely to do so when their identity position was secure enough that poor performance on the experimental task was unlikely to negatively impact their individual reputation or the reputation of their group.

For example, in the first identifiability study (Identifiability Study 1), double minority participants performed very badly under boost when identifiability was low, but much better under boost in high identifiability or threat conditions. In the third identifiability study (ID3) participants were most likely to underperform in the boost condition when anonymous, but more so when they also reported low resistance to the high status manipulation and high belief in the legitimacy of the intergroup status hierarchy. The equivalent participants with low status-resistance and high legitimacy were top performers under boost when identifiability was high (presumably increasing pressure). In fact, in all four identifiability studies, low performance in the boost condition was only likely when participants were anonymous and in configurations of identity that could be described as secure.



Of course it could be argued that the more secure the identity the more pressure is felt by targets of the positive stereotype to live up to it, resulting in over-arousal and choking. To test this hypothesis the fourth identifiability study manipulated individual reward in addition to status and conflict. Since attaching financial reward to individual performance has produced enough performance pressure to induce ‘choking’ in previous studies even without a status manipulation (eg. Mobbs et al., 2009), if the choking under pressure hypothesis was correct then underperformance would have been most likely in the boost + individual-reward condition. However, if participants were simply working less hard because their identity was secure and did not need defense then they would be more likely to perform badly in the boost + flat-rate reward condition and better in the individual reward condition.

The data did not support the choking under pressure hypothesis: the only participants to perform better in the boost condition (without including additional variables in the model) were those in the individual reward condition, which should have increased pressure. Once resistance was accounted for, the only participants to perform better under boost in the flat-rate-reward condition were those who were in the high identifiability condition and believed that they were indeed identifiable. Therefore, the pattern of results in the boost condition showed (1) that performance was not universally improved or hindered in the boost condition; and (2) that features of the task performance context that should have increased *pressure*, such as increased identifiability and receiving performance-based rewards instead of a flat-rate incentive, were generally also associated with increased performance, refuting the ‘choking under pressure’ explanation in this context.

These results suggest the possibility of a second – and much more prosaic – reason for underperformance under positive stereotypes: that when participants have a secure identity in the experimental context they may have no *need* to well on the experimental task. Since in this situation they are ‘sheltered’ by their positive and secure group identity, their experience

can be compared to “slipstreaming” in that they are being “pulled along” by the positive group identity and therefore have to work less hard as individuals to maintain an acceptable identity within the experimental context.

In the conflict studies it emerged that performing too well as a member of a high-status group may even be a social liability in certain contexts. In these studies, participants in the boost condition regularly performed well when conflict was high, but poorly when conflict was low. It seems likely that performing to the best of one’s ability as a high-status group member immediately following a cooperative interaction with low-status group members could be interpreted as a disaffiliative response. Therefore, in that specific context, good performance could have been perceived as a threat to positive social identity in the performance context.

Of course, there might be situations in which being a member of a positively stereotyped group produces so much pressure that performance suffers as argued by Cheryan and Bodenhausen (2000), particularly when the group identity is insecure (with low stability or legitimacy) or the individual’s position within the high status group is insecure (with high permeability). These types of identity security would require the individual to perform well to defend their own reputation or the reputation of the group and it is easy to imagine circumstances in which that type of pressure could harm performance. However, in the present studies the evidence is strong that this type of pressure was not related to underperformance in the boost condition. Rather, it seems that boost conditions justified reduced performance because the positive group identity provided an identity buffer that relieved individuals with secure high status identity positions from the chore of proving themselves in the experimental context. When you have nothing to prove, why prove it?

However, it also seems that participants in the boost condition had a good platform for good performance when it was required (for example, to defend an insecure identity

position or to secure an individual reward). The best conclusion seems to be that participants were *pragmatic* about their investment in task performance in relation to the impact that lower or higher performance would have on their own fortunes or the fortunes of their group.

To summarize, these results open up a very simple yet novel explanation for underperformance in boost conditions that has not yet been noted in the STB literature: that slipstreaming may occur when participants have a secure positive identity and poor performance will not have negative consequences for their own or their group's reputation.

However, to avoid the mistakes of the STB literature more generally, it must be remembered that the likelihood of slipstreaming in the boost condition depends on how participants understand the meaning of good or bad performance, in other words, how good or bad performance may affect the fortunes of their group or their individual fortunes. This itself depends on cultural expectations around task performance and identity and local-cultural understandings of the task-performance context (such as an experiment run at a university).

Consequently – since the social meaning of an experiment with students at a university is different to the meaning of performance in other contexts – slipstreaming in experimental contexts is much more likely than in the high-stakes performance contexts that are generally the idealized focus of STB theory such as tests or job interviews. In other words, the patterns of results in the field of STB can only be understood and generalized once it is acknowledged that the vast majority of studies have used experimental analogues of these high-stakes settings with few real-life consequences. Since the present series of studies has demonstrated fairly convincingly that the social meaning of good or bad performance in the specific performance context is critical to predicting whether targets will perform well or badly, the assumption that the findings from anonymous experiments with few extra-

experimental consequences are directly generalizable to high-stakes testing contexts is questionable.

### **10.5.2. Negative stereotypes as challenge – stereotype threat conditions galvanizing performance**

There is emerging evidence in the STB and social identity threat literatures that negative stereotypes may sometimes invoke increased performance (see section 3.2.3). Kray et al. (2001) were “led to conclude that reactance effects might be limited to certain contexts” that were different to “noninteractive, academic context[s] in which the necessary trait or behavior is not ... easily obtained (e.g., analytic math ability)” (p. 955). However, the APM is one of the most specific tests of analytic ability available, so the occurrence of “reactance” in the present study suggests that it is not simply the type of test that facilitates these ironic effects.

These results add to the growing body of evidence that in certain circumstances stereotype threat can galvanize improved performance. However, previous explanations for this phenomenon have largely been individualistic. For example, Crisp, Bache, and Maitner’s (2009) central argument was that “there is some research which shows that, through experience, people can sometimes develop strategies that help them cope with self relevant threat,” (p. 11). Hoyt and Blascovich (Hoyt & Blascovich, 2010) found that responding positively under threat was primarily related to participant’s “leadership self-efficacy” (p. 89).

The results from the present studies suggest that individual skills and motivation are important features of resistance (demonstrated by the importance of baseline performance in predictive models), but that participants’ perceptions of sociostructural features of the intergroup setting are even more important for determining which targets of threat will resist and overcome the manipulation.

In the present studies, incongruent stereotype-performance effects under threat were most reliably predicted by perceived permeability, followed by social-desirability, intergroup conflict, ingroup identification, legitimacy, and symbolic threat. These variables interacted in complex ways that emphasized that task-related group status (ie. positive or negative stereotypes) is not a particularly meaningful predictor of task performance divorced from other individual, positional and sociostructural features of identity in a given intergroup performance context.

More specifically, targets of negative stereotypes seemed to perform better if they believed that doing so could: (1) advance an independent identity, either by demonstrating individual ability or by distancing themselves from the low status group or moving towards the high status group; or (2) effectively defend the reputation of their group.

These results echo Kray and her colleagues' findings that malleability beliefs increase the probability of performing well under threat (2007), since beliefs about the malleability of a negatively stereotyped attribute are closely related to stability and legitimacy beliefs. Kray, Reb, Galinsky and Thompson (2004) found that "blatantly telling women that they lack the attributes necessary to prevail at the distributive bargaining table produces stereotype threat in the absence of power and stereotype reactance when sufficient power is possessed" (p. 409).

The findings of the present studies strongly suggest that good performance under threat is closely related to the participant's reading of the sociostructural features of the intergroup context, their own position in that context, and the potential meaning or consequences of good or bad context for themselves or their group. In other words, when task performance is a meaningful act in an intergroup setting, it will be sensitive to the performer's position and prospects in that setting.

The empirical evidence demonstrating that people may perform worse under boost and better under threat, and the conditions in which these ironic effects were most likely to occur can best be understood if we acknowledge that participants responded strategically to produce identity in the experimental context, drawing on personal and group identity resources available to them to do so (cf. Danziger, 1994; Spears & Smith, 2001; Tajfel, 1972) and using the performance task as one of the means of positioning themselves in the context. Therefore task performance must be understood as an active production of identity as well as an outcome of the manipulations. In order to understand task performance, we must also understand its identity-constructing power with respect to socio-structural features of the context, such as permeability, stability, legitimacy, and conflict as described by social identity theory.

#### **10.6. STB as situated identity performance**

The most important theoretical finding from this series of studies is also the simplest: that to understand when and why stereotypes will influence task performance, we have to know something about what that task performance means to stereotype targets and what they think it means to people who may observe it or judge it. The lived experience of threat or boost is a complex identity experience, and task performance is only one component of that experience. However, task performance is only meaningful at the intersection of multiple social and practical features of the experimental context and the participant's own phenomenological experience of the task and context.

Asking why participants are completing a task and what it means to them requires us to acknowledge that – in most contexts relevant to STB at least – “the observers become the audience [and] actions which appear to be done on objects become gestures addressed to the audience” (Goffman, Lemert, & Branaman, 1997, p. 22). This is particularly true of experimental contexts, in which participants are aware that the observation of their

performance is the sole object of the social form in which they are engaged (Danziger, 1994), and most other performance contexts central to STB research such as educational tests and exams (eg. Steele, 1997).

The notion that the meaning of the task performance is critical to our understanding has been implicit in mainstream STB theorizing. For example, Spencer, Steele and Quinn (1999) argued “it is only when the test reflects on ability and is difficult and the test takers care about this ability that the stereotype becomes relevant and disturbing” (p. 25). However, this insight has never been taken seriously in the STB literature. Instead, performance has always been taken as an uncomplicated self-referential indicator, in the same way that cycling speed was used by Triplett (1898) to indicate social facilitation.

However, if task performance under positive or negative stereotypes is related to the meaning attached by stereotype targets and their audience to good or bad task performance and how it “reflects” on the performer as argued by Spencer, Steele and Quinn (1999), then the task performance itself must be oriented to the performer’s reading of the social meaning of the performance context, their position in it and how that position intersects with their own identity, the meaning and accountability of good or bad performance in that context, and the individual resources available to increase (or decrease) performance in response to the demands of the context.

This insight challenges the standard deterministic model of stereotype-performance effects in which positive or negative stereotypes, mediated or moderated by key variables, impact on performance directly and linearly. If ‘task-performance’ is not simply a self-referential measure, an end in itself, but an expressive act oriented to the production and defense of situated identity, then our models of stereotype-performance effects must consider how participants might orient to making meaning of performance in STB contexts.

However, this understanding of task performance raises some problems in modeling STB effects statistically. The first is that task performance is a very ambiguous symbol. Good performance could be oriented towards distancing one's self from a low status ingroup or to defending the reputation of a high-status ingroup. Poor performance could indicate that participants have disengaged from the experimental context in protest at the negative stereotypes that have been invoked, that they have a secure identity that does not require defense by good performance or that they simply lack the ability or cultural capital (such as familiarity with the requirements of the task) to perform well. There are therefore many potential readings of a social context that make either good or bad performance intelligible as a socially meaningful act.

Since the prediction of performance relies on understanding how specific features of the performance context may be 'read' by performers, the prediction of task performance which, after all is the goal of STB research, can be tricky. However, first, since performance contexts are generally recognizable cultural forms, and good or bad performance generally has a limited number of meanings within those understandable contexts, it should still be possible to predict whether performers will respond with increased or decreased performance in specific contingencies of identity. Second, understanding stereotype threat and boost as the outcome of situated identity performance introduces possibilities for new classes of construct that may allow stereotype performance effects to be successfully theoretically and empirically modeled.

For example, in this series of studies a resistance metric was developed that substantially improved the ability to detect and understand the stereotype-performance effects that occurred. 'Resistance' helps to make sense of stereotype-performance effects as outcomes of situated identity performance because, like the SIT conceptualization of ingroup identification, it represents an intersection of individual and social identity resources and



liabilities in the specific performance context. It is likely that future research will both improve the means of assessing resistance and also identify additional constructs at the situated intersection of identity resources and liabilities and specific features of performance contexts that will improve our ability to theoretically and empirically model stereotype-performance effects.

This understanding of task performance as a situated identity performance means that any deterministic model that fails to account for the active nature of performance as identity-work cannot universally model STB effects. Additionally, since neither the cultural meaning of task performance nor the intergroup relations within any given context is fixed, it is unlikely that any static predictive model of STB will be generalizable across contexts or maintain its predictive validity over time unless it can account for the active nature of task performance as a means of producing or contesting identity or the ways that the meanings of good or bad performance of the specific task in the specific context may shift across contexts and over time.

Understanding stereotype threat and boost as the outcome of situated identity performance means acknowledging that performance shifts can be best predicted once: (a) we know what meaning is attached to good or bad performance by the performer in the *specific* context in which the performance occurs; (b) we understand the specific identity constraints on, and resources available to, the task performer in that context; and (c) how good or bad performance would reposition the participant in the specific context and affect their moral careers more generally. These insights may go a long way towards explaining why the STB effect has been so difficult to reliably replicate and why so many mediators and moderators have been identified in the literature.

Instead of treating performance decrements as universal outcomes of negative stereotypes, this meta-theoretical stance will allow the development of: (1) a taxonomy of

strategic performance orientations that predict performance shifts in response to positive or negative stereotypes; and (2) the conjunctions of individual, positional and socio-structural variables in which in different strategic orientations are likely to be taken up or resisted.

### **10.7. Stereotype threat revisited**

In previous chapters it was demonstrated that being the target of a positive or negative stereotype did indeed impact on participants' performance on a stereotype-relevant task. Steele (eg. 2010) and other stereotype-threat theorists argue that stereotype threat is a *situational* phenomenon that targets have little control over. In fact, they argue that individual attempts to resist it often make it worse (Steele, 2010; Steele et al., 2002). For example, Steele and his colleagues have argued that "compared to viewing the problem of Black underachievement as rooted in something about the group or its societal conditions, [stereotype threat] uncovers a social psychological predicament of race, rife in the standardized testing situation, that is amenable to change" (Steele & Aronson, 1995, p. 810).

As a situationist account of the underperformance of minority groups in critical contexts, stereotype threat theory has played an important role in undermining pathologizing theories of minority underperformance that attribute deficits to biological, cultural or individual inferiority (eg. Allport, 1954; Johnson & Bouchard, 2007; eg. Rushton & Jensen, 2005a; Stafford, 1972) . However, the results of the present study suggest that theoretical frameworks of stereotype threat have been both too situational and not situational enough.

#### **10.7.1. Stereotype threat is not situational enough**

By not situational enough, I mean, firstly, that mainstream models of stereotype threat (eg. Steele, 1997, 2010; Steele & Aronson, 1995; Steele et al., 2002) have sidelined the numerous signs of incongruent stereotype-performance effects that have appeared in the literature. Generally, incongruent findings in the field have been dealt with by proliferating boundary conditions under which stereotype threat is likely to occur (eg. Aronson et al.,

1999), thereby drawing lines in the theoretical sand that have allowed stereotype threat theorists to ignore incongruent effects.

This practice has obscured from view situations in which negative stereotypes might improve performance or positive stereotypes might inhibit it. This narrow focus in the stereotype threat literature on the threat quadrant of stereotype-performance effects underestimates the true scope and power of the stereotype-performance situation.

Secondly, stereotype threat has been theorized in an atheoretical and individualistic framework that has resulted in a socially deprived notion of “the situation.” In most studies, “the situation” is theoretically represented only as a stereotype and a prime without consideration of the practical and socio-structural features of the specific performance context (Danziger, 1994) or the intergroup relations within it. Where such features are represented, it is usually with individualistic features of group relations such as ingroup identification (eg. Schmader, 2002a; Schmader, 2002b) or, more frequently, ‘domain identification’ (eg. Steele, 1997).

As a result, despite Steele and his colleagues’ assertions that stereotype threat is a situational phenomenon (Steele, 1997, 2010; Steele & Aronson, 1995; Steele et al., 2002), the way that it has been conceptualized and explained is highly individualistic and generally explains underperformance via individualistic cognitive mechanisms. For example, in one of his most recent papers, Carr and Steele (2009) invoked stereotype threat by asking participants to report their gender and informing them that a problem-solving task was highly diagnostic of mathematical prowess. They measured perseverance, stereotype suppression, performance expectancy, math identification, mood, “mistake activation” and motivation. With the possible exception of math identification, these measures are agnostic of the social and practical features of the performance situation and any analysis of these constructs can only result in *individual* rather than situational or social explanations of stereotype-

performance effects. This example illustrates the widespread disjunction in the stereotype threat literature between the level of conceptualization (the situation) and the level of explanation (the individual), although – to be fair – this is widespread problem in social psychology.

### **10.7.2. Stereotype threat is too situational**

However, stereotype threat is also too situational, because it fails to acknowledge the identity-constructing functions of task performance or the active and strategic features of identity performance in stereotype threat or boost contexts. The analysis of STB congruence in this chapter has demonstrated that participants can and do respond to negative stereotypes by incongruently performing better or to positive stereotypes by performing worse, and the conditions in which incongruent responses are likely suggest that participants were often responding to stereotype threat or boost conditions actively and strategically.

The most consistent findings were that negative stereotypes were likely to galvanize better performance when members of low status groups perceived intergroup boundaries to be permeable and positive stereotypes were likely to result in reduced performance when participants reported higher levels of intergroup conflict. These are fascinating findings because the vast majority of research in stereotype threat invokes categories that are highly impermeable, such as race and gender, in contexts where – in the past few decades, at least – every effort has been made to minimize explicit intergroup conflict, such as at Universities. According to the results of the present programme of research, these are the exact social conditions that would maximize the threat-boost performance gap.

This pattern of results also suggests that participants approached the performance task as an identity-constructing activity. For example, social identity theorists have demonstrated that members of low-status groups who believe that group boundaries are permeable will often identify strongly with the outgroup and behave in ways that demonstrate their

suitability for the high-status outgroup. In Tajfel's words, members of low-status groups who have the option are likely to "leave it – psychologically, objectively or both" (Tajfel, 1978d, p. 67). In the experimental context they found themselves in, participants could position themselves in relation to the ingroup and outgroup most convincingly by performing better or worse on the performance task.

In the present studies the robust tendency for participants who reported high permeability to perform better under threat fits the predictions of social identity theory very well. In conjunction with high permeability beliefs, improved performance under threat can be interpreted as a means of positioning oneself away from the low status group and towards the high-status group.

Participants also tended to perform better under threat when they believed that their performance would make a difference to their individual reputation (such as when they were high in social desirability or reported high levels of individual identifiability), or when they cared about their group and believed it was under illegitimate threat (indicated by ingroup-identification, legitimacy, individual-identifiability, symbolic threat or realistic threat). In contrast to the deterministic understanding of the stereotype-performance situation promoted by stereotype threat theorists, the present results can best be understood by considering the identity-constructing nature of "task performance" in the specific performance context of the social-psychological experiment and participants' active engagement with the identity-related resources and liabilities available to them in the context.

### **10.7.3. An interactionist understanding of stereotype-performance effects**

This invites us to consider task-performers under negative or positive stereotypes not just as passive targets, but as agents in the performance context. Tajfel argued that social psychologists often treat the world as a ping-pong table and experimental subjects as balls, expecting that the trajectories of each ball and their interactions will be fully predictable on

the basis of a small number of psychological ‘laws’. However, people are not ping-pong balls and the social world is not a table. People change themselves, they change the world and they change each-other (Tajfel, 1972).

By changing himself [sic] the individual changes the social environment; by changing it he changes himself. He [sic] must create change, resist it, adapt to it or prepare for it; most often he [sic] will do all these things at the same time” (p. 108).

Applying Tajfel’s insight to stereotype-performance effects requires three acknowledgements. First, that stereotype targets are agents who engage actively with the stereotype threat or boost situation they find themselves in. Second that the task performance itself, and its outcome, has specific social meaning and, possibly, social consequences within and beyond the specific performance context. The way that task performance positions the target within the specific context and impacts on their moral career more generally is therefore critical to understanding how participants orient to the task. Third, that the task is performed in a context that offers identity resources and liabilities to the performer that will impact on the way the task is understood, how performance can be interpretable as an identity-constructing act, and the specific task performance strategies that performers adopt<sup>25</sup>.

However, it is important that no babies are thrown out with bathwater: adopting an interactionist understanding of stereotype-performance effects does not require us to ignore existing evidence, such as that negative stereotypes may invoke cognitive processes that undermine performance (eg. Rydell et al., 2009; Schmader, 2010; Steele, 2010). There may

---

<sup>25</sup> Note that an interactionist understanding of stereotype-performance effects is agnostic about whether agency, strategy and responses are conscious or unconscious, intentional or unintentional, cognitive or affective. Indeed, recent studies in social identity have demonstrated that the subjectivity and performance of identity involves all of these (eg. Scheepers et al., 2009).

indeed be specific configurations of stereotypes, identity and context that invoke performance strategies that undermine performance in these ways, for example when negative stereotypes encourage targets to simply try harder which in turn undermines their creativity (Jamieson & Harkins, 2007; Jamieson & Harkins, 2009). Rather, adopting fully interactionist stance would encourage theorists to look not only for cognitive mechanisms that undermine performance, but to also understand the specific socio-structural features of performance contexts in which these strategies are likely to be invoked and to identify socio-structural conditions and identity-performance strategies that allow targets to overcome them.

#### **10.8. The reach of these findings**

The results of the present studies have made a case for understanding stereotype threat and boost as outcomes of situated identity performance. A corollary of this argument is that it is very likely that contexts exist in which task performance is constrained in the ways described by mainstream stereotype threat theorists. The results suggest that these contexts might be ones in which the negative stereotypes invoked have low legitimacy, high stability and consensus, and refer to groups with low permeability such as race and gender stereotypes in specific contexts such as science classrooms in the USA.

In such a context what difference would individual performance make to one's own reputation or the reputation of one's group? Perhaps if performers could be convinced that their task performance in the local context might make a positive difference to their identity in relation to those stereotypes then they might be oriented to improved performance. Indeed, this is exactly what many successful stereotype threat interventions manage to do. For example, convincing participants that the researchers believe that intelligence is malleable (thereby destabilizing the legitimacy and stability of the stereotype) has been shown to eliminate stereotype threat (eg. Aronson et al., 2002; Good et al., 2003; Steele et al., 2002; Thoman et al., 2008). Steele and Aronson (2002) argued that these interventions enabled

performers to “have greater trust in the setting” (p. 394), which can also be interpreted as enabling them to believe that good performance would be a viable way of producing positive identity in the experimental context.

Understanding stereotype-performance effects as the outcome of situated identity performance actually advances Steele's political project because it suggests that STB is even more dependent on situated and contingent features of identity than he has argued so far. However, it also suggests STB is more sensitive to intractable and powerful meta-contextual features of group identity and status (ie. stereotypes) such as stability, legitimacy and permeability that preclude easy intervention. If this is the case then reducing stereotype threat cannot be as simple as simply ensuring that stereotypes are not activated in the performance context.

However, understanding STB as the outcome of situated identity performance does suggest other more radical alternatives. For example, intentionally activating the stereotype but simultaneously invoking intergroup conflict and/or demonstrating that the categories are illegitimate, the status hierarchy is unstable or that the categories are highly permeable may all be successful to the extent (a) that the interventions convince participants that their performance will have social meaning and illocutionary force in that context (either to defend their own reputation, shift their group membership or to defend the reputation of their group) and (b) performers have confidence in their abilities to perform in ways that challenge the intergroup status hierarchy and their place in it.

Although the models of stereotypes, identity and performance did not account for a much variance in performance outcomes in the present programme of research, there are three important comments to be made here: first, even small differences in performance at the rarified levels often examined by STB theorists (eg. Ivy-League universities) can make a large difference to individual moral careers and trajectories; second, these experimental



contexts were intentionally chosen so that they were not particularly important to participants (to avoid ethical harms) and it is possible (but by no means certain) that effects would be greater for more meaningful identity contingencies; and third, these models are just a start towards understanding STB as situated identity performance. More sophisticated models based on a more sophisticated understanding of particular identity contingencies may account for more variance in performance.

### **10.9. Limitations and challenges for future research**

The present series of studies is limited in several ways. First, the stereotypes and categories invoked were very weak compared to the powerful and stable stereotypes invoked in most other stereotype threat research. Therefore it is not entirely clear to what extent these findings would generalize to more intractable categories and stereotypes. Further research will obviously be needed to assess if, and how much, these findings extend to more entrenched identities such as sex and race.

Second, given the weak stereotypes that were experimentally invoked, the effect sizes on performance were relatively weak. Although sample sizes were progressively increased during the course of the programme of research, observed power was still low for most studies. This does not cast the significant findings in doubt, since statistical methods safeguard the Type I error rate, but it does make it likely that each study lacked the power to detect important relationships that may have informed theory building.

Third, only the very basic hypothesis that the socio-structural variables described in SIT impact on stereotype-performance effects was made in the design phase. It was not possible to specify more sophisticated hypotheses because of (a) the inconsistency of results within the STB literature to date; (b) the loose relationship between constructs used in the STB literature and SIT (such as domain identification compared to group identification); and (c) the lack of research in the SIT framework about how task performance relates to identity.

Therefore, beyond the most basic finding that socio-structural variables do indeed impact on performance, many of the results and interpretations are *post-hoc* and require verification.

Therefore the results of these studies cannot be considered strictly generalizable, since given these issues of design, replications would be likely to yield slightly different configurations of significant results and specific coefficients. However, the point is well made by these studies that this general class of variables sociostructural and identity variables – previously largely ignored in the STB literature – are important features of the STB experience for participants, at least in the specific identity contingencies presented to them in this series of studies. In other words, although the specific forms of the statistical models that described performance in this series in this series of studies are unlikely to be generalizable, the analysis was able to convincingly demonstrate that each of these variables – and the SIT framework that makes sense of them – deserve further exploration in the STB literature.

## Chapter 11. Conclusion

As discussed previously, in the conclusion to their seminal review of stereotype threat research up to that point, Steele, Spencer and Aronson (2002) argued that in explaining the underperformance of minorities “the dominant paradigms of psychology put a great deal of weight on relatively decontextualized internal processes and traits such as low self-esteem, low performance expectancies, oppositional attitudes towards mainstream achievement, and the like” (p. 436). Specifically, they said:

We hope that [we have] ... made a case for the value of an explanatory paradigm in social psychology that better incorporates the ongoing, intersubjective dimensions of social experience (p.436).

However, despite this expressed intention for stereotype threat research to be characterized by social and intersubjective models and explanations, Steele and his colleagues' critique of mainstream psychology was also a good forecast of the explanatory models to be applied to STB in the coming years. The majority of stereotype threat research in practice has actually relied on constructs and explanatory models just as “internal” and “decontextualized” as those offered by the individualistic models so soundly criticized by Steele and his colleagues, including self-handicapping, anxiety, evaluation apprehension, confidence, stereotype endorsement, test perceptions, feelings about the self (such as self-esteem), effort, and working memory (Schmader & Johns, 2003; Smith, 2004).

Like Steele et al. (2002), Social Identity theorists have also vigorously critiqued the reduction of the ‘social’ in social psychology to individual processes writ large. Tajfel argued that “...the study of social stereotypes by social psychologists is a travesty of our reality unless the term ‘social’ is taken seriously as the fulcrum of our work on the subject” (Tajfel, 1981a, p. 4). The key obstacle to STB theorists taking the social seriously is that the

stereotype-threat model is embedded in an individual-cognitive paradigm that lacks a meta-theoretical framework to bridge individual and social features of the stereotype situation. Social Identity Theory provides just such a meta-theoretical framework and is eminently applicable to the situation of stereotype threat.

An unexpected connection between STB and SIT is that Steele's early work in identifying the stereotype threat phenomenon was "much influenced" (Steele, 2003a, p. 316) by the studies of Katz three decades earlier that had specifically linked the underperformance of black students in recently integrated schools to issues of identity and social comparison (Katz, 1968; Katz et al., 1965). By 1978 Tajfel had also noted Katz's work, arguing that the underperformance of black students in white schools represents "the practical implications of a comparative (and negative) self-image" (Tajfel, 1981a, p. 327). Thus Tajfel, 19 years before the publication of Steele's first paper on stereotype threat, had noticed the empirical phenomenon that STB would later unpack and suggested that it may be an important *practical* feature of the lived experience of minority group members. However, uncharacteristically, he did not theoretically expand on how these "practical implications" of negative social identity theoretically related to other features of social identity theory. It was left to Steele and other STB theorists to demonstrate just how important these practical implications of identity can be to individuals and groups.

However, more recently social identity theorists have begun to consider the importance of practical everyday behavior in the production of identity and the importance of identity in constraining situated behaviour. Of key interest here is the social identity model of deindividuation effects (Reicher & Levine, 1994; Reicher et al., 1998; Reicher et al., 1995), and particularly the work of Klein, Spears and Reicher (2007) extending the model to understand and predict how people enact identity-related behaviours as a means of claiming or defending identity in everyday interaction. This work demonstrates that the performance of

everyday tasks and behaviour is not *incidental* to identity, but is both an outcome of identity and a strategic resource for achieving, enacting, or defending social identity in practice.

Identity performance can take a variety of forms, such as physical action and the manipulation of physical appearance (including displays, symbols, signs) as well as the verbal expression of representations and attitudes viewed as normative of the in-group (e.g., stereotypes and prejudice). By *identity performance*, we mean the purposeful expression (or suppression) of behaviors relevant to those norms conventionally associated with a salient social identity (Klein et al., 2007, p. 30).

Despite Steele and his colleagues' encouragements to consider stereotype threat as a situational contingency of social identity, research in the field has all-too-often uncritically treated stereotypes as stimuli and performance shifts as simple responses without considering the strategic or identity-constructing meaning of task performance in a stereotype-threat situation. The present study has attempted to treat task performance in stereotype threat as a situated and strategic orientation to task performance and as an identity-constructing act.

To some extent this feature of task performance has already been acknowledged in the stereotype threat literature, for example, in Steele's contention that the participants with most investment in the task-related identity (as indicated by domain identification) are likely to be most affected by negative stereotypes (Steele, 1997). However, in stereotype-threat models it is usually assumed that stereotypes influence performance via deterministic, unidirectional and often cognitive mechanisms (eg. Schmader, 2010). Indeed, this is a key explanatory feature, because it is this unidirectionality that makes stereotype-threat theory non-obvious. Many laypeople hearing about stereotype threat respond by saying "but if that happened to me I'd just try harder," but STB predicts that this response is likely to result in "overtrying" and reduced performance (Steele, 2010). For example, a woman called in to an interview with Claude Steele on NPR radio and said:

CHRIS (Caller): Hi. I am a 53-year-old woman. So I grew up in the '50s and '60s, and I have to say that my personal response to anyone telling me you can't do that because you're a girl has consistently been in my life to say, well, yes I can. And I'm going to prove it to you, and I'll do it better than you do.

Mr. STEELE: Yes, it's important to realize that the causes of underperformance in our research are not from women giving up. They don't give up under that kind of pressure. They do try extra hard, but sometimes, that can not foster [sic]. That can interfere with, performance in its own right.

[11 turns omitted]

CHRIS: Yeah, yeah. No, I just wonder if I mean, to me, it's not a threat, it's a challenge. It's an opportunity. It's an invitation. (Steele & Conan, 2010)

Crisp, Bache and Maitner (2009) set out to test the hypothesis that women who, like Chris in the interview above, had responded to threat as 'a challenge', 'an opportunity', 'an invitation' and "successfully entered" (p. 1) a domain in which they were negatively stereotyped would have developed successful strategies for dealing with the negative stereotypes in the performance context. As Crisp and his colleagues (and Chris, the caller to Steele's radio interview) predicted, these women performed *better* when exposed to the conventional negative stereotypes about their group. As Crisp and his colleagues argue, this demonstrates that some people "who face stereotype threat on a daily basis ... are able to thrive in the face of negative performance expectancies" (p. 11).

A close reading of the empirical literature demonstrates that increased performance in response to negative stereotypes, while not as common as the negative congruent response, is not an uncommon outcome. Just as importantly, positive stereotypes can also impact performance either positively or negatively. By building a field around the politically

important effect – the underperformance of negatively stereotyped minorities in socially critical domains – the empirical and theoretical investigations in the STB literature have been artificially limited to exploring only the negative effects of negative stereotypes while ignoring the potential positive effects that both negative and positive stereotypes may have in certain circumstances.

Renaming such evidence as “reactance” (Hoyt & Blascovich, 2007; Kray et al., 2004) or “stereotype susceptibility” (Shih et al., 1999) allows it to be packaged as a distinct and separate phenomenon, which can then be justifiably ignored in mainstream models of stereotype threat. However, failing to integrate the empirical evidence that negative stereotypes can sometimes enhance performance into theories of STB has resulted in deterministic, individualistic and passive models of the experience of stereotype threat.

Crucially, despite some early focus on the notion of “domain identification,” the field of STB is agnostic on what performance in a given context *means*, either to the participant or the researcher. Like an intelligence test, it is considered good enough that task performance is how the participant performed on the task. Although Steele (2010) emphasized that STB is the result of an identity contingency, his model – and the field of STB in general – lacks a framework for understanding identity and therefore treats performance as an end in itself.

The present study therefore aimed to re-imagine STB in light of social identity theory. Of the two previous attempts to do so, one applied what Reicher (2004) would call a “reductionistic misreading” (p. 921) of SIT but actually had empirical data (Schmader, 2002b), while the other applied a theoretically rich reading of SIT but did not apply it to empirical data (Haslam et al., 2008). The present study therefore is the first comprehensive attempt to develop an empirically-grounded social identity model of stereotype threat and boost.

A general experimental paradigm was developed for the programme of research that, unlike the majority of studies in STB, allowed (almost) true random assignment, but with the proviso that the stereotypes invoked were relatively innocuous compared to the powerful and inescapable stereotypes invoked in most stereotype research such as race and gender (Nguyen et al., 2003; Walton & Cohen, 2003). This could be seen as a disadvantage, in that the effect size of stereotype-performance effects was likely to be smaller, or as an advantage in that it most likely allowed a broader range of potential responses. This, in turn, allowed an exploration of situated agency that would have been limited with the high-stability, high-legitimacy and low permeability categories stereotypes usually invoked in STB research such as race and gender.

Ten studies were undertaken involving about 1500 participants in the course of the programme of research, each manipulating status and one other critical factor of social identity in the SIT framework. Of these, four explored the impact of identifiability on the stereotype-performance experience; three explored the effects of conflict; and three explored the effects of permeability. In addition to the SIT-relevant manipulated factors, most studies included measures for a comprehensive battery of SIT identity and sociostructural variables that allowed correlational investigation of a broader range of SIT constructs alongside the experimental investigations.

The initial analyses were not supportive of the claim frequently made in the STB literature that simply knowing that someone holds a negative stereotype about your group is enough to invoke stereotype threat effect (eg. Steele, 2010; Steele et al., 2002). On the contrary, in most of the 10 studies, models of status on performance were not significant unless a metric of participants' resistance to the status manipulation was also included. Although it is possible that the small sample sizes and modest effect sizes resulted in inadequate power to detect conventional stereotype threat effects, this is unlikely since the



studies with the largest sample sizes were generally furthest from significance. On the contrary, the evidence is fairly compelling that performance could be best predicted by knowledge of both the status manipulation experienced by a participant and that participant's resistance to it.

This finding, and others reported in the present programme of research, challenges the model of the passive subject generally assumed in the STB literature and suggests that performance under STB conditions is an active and strategic activity oriented not only to the social features of the performance context (as argued by most SIT theorists), but also to the target's own reading of that context in the context of their total identity liabilities and resources (including their ability and alternative identities), and their strategic identity orientations in the context. For example, participants who were anonymous exhibited different responses to STB compared to participants who were identifiable; and participants receiving a performance-based reward responded differently to STB than to those receiving a flat-rate reward. These results support the argument that the identity-constructing and practical features of task-performance in a specific context contribute to participants' strategic orientation to the performance-task which, in turn, contributes to their performance. As Klein, Spears and Reicher argued:

Identity performance, like self-presentation, is by its very nature a creative act by purposive agents. It is as much a function of the way in which these agents imagine—and seek to create—the future as of the nature of the present (Klein et al., 2007, p. 38)

But, in the situation of stereotype threat, participants have very limited means of enacting social ambitions. Indeed, the key defining feature of the typical stereotype-threat

context is that the performance task represents the participants' most credible "voice" (Tajfel, 1981a) or means of challenging or yielding to the social order as it is presented to them<sup>26</sup>.

As Tajfel argued: "members of minorities ... have some problems in common; there is only a limited number of possible psychological solutions (or attempts at solutions) to these problems; the kind of solution adopted is closely related to the social conditions in which minorities live" (Tajfel, 1981a, p. 316). In the case of stereotype threat performance context, the "problems in common" are even more tightly constrained than usual, as are the "possible psychological solutions."

In other words, the present studies suggest that stereotype-performance effects should be understood as outcomes of *situated identity performance*<sup>27</sup>. This model of STB proposed here understands task performance as often – but not always – an expression of identity, not in a general sense, but as it is experienced in a specific performance context. This expression of identity is responsive to situational demands (such as the positive or negative stereotype invoked) but subject to the identity resources available to the performer in the context, such as alternative identities, individual ability and experience, social support and so on. Better or worse task performance can therefore be understood as an act of identity-production intelligible within the rules of the specific context in which the performance took place.

This subtle shift in the way that task performance is understood makes stereotype threat and boost intelligible in terms of social identity theory framework and allows the truly social and contingent features of the experience to be explored in more depth than has previously been possible. Specifically, the model proposed here suggests that three levels of

---

<sup>26</sup> Especially since participants who respond by disengaging or 'disidentifying' with the domain (Steele, 1997) or exiting the experiment are no longer considered part of the focus of the stereotype-threat research despite the fact that these might be viable identity performance strategies in the lived-context.

<sup>27</sup> Klein, Spears and Reicher (2007) use the term "social identity performance." However, I have labeled the performance "situated" since the key feature of stereotype threat is that the possible avenues for establishing identity are tightly constricted by the performance situation.

analysis will be required in order to understand and predict stereotype-performance effects: (1) individual-level variables of the type often explored in the STB literature (such as ability, self-esteem, performance expectancies, achievement motivation or anxiety); (2) variables representing the momentary and longer-term positioning of the performer with respect to the social groups and categories relevant in the domain (such as ingroup identification, ingroup and outgroup closeness or identifiability); and (3) socio-structural features of the immediate and more general social context such as conflict, permeability, stability and legitimacy.

To date STB research has exhaustively explored the first category of individual features and has incorporated some of the second category of social-positioning variables into theory and empirical investigation. However, according to SIT, it is the third category of variables that determines what identity strategies will be available to individuals and groups in a given social context. It is these sociostructural variables that are most likely to predict the performance strategies of similar participants in particular identity contingencies, and these have not been explored in the STB literature at all.

The present studies experimentally explored the impact of identifiability, conflict and permeability on stereotype-performance effects and found that all three variables had intelligible effects on performance under positive or negative stereotypes if performance was understood as a meaningful and accountable social act. The most important effect of identifiability was to shift the nature and strength of participants' accountability for their performance and thereby shift the identity-related meaning and consequences for poor performance. For example, low identifiability (anonymous) conditions facilitated *slipstreaming* for participants in boost conditions who were able to rely on their positive group identity safe in the knowledge that their individual performance would not undermine their positive social identity. However, participants under threat did not generally perform

worse when anonymous because, it seems, their low group status meant that their individual performance was still relevant to their identity despite their anonymity.

Manipulated permeability and conflict shifted performance in ways that differed across studies but, despite and because of this, were intelligible for their effects on situated identity performance. For example, participants in low status conditions (threat) often performed better after experiencing a non-cooperative and conflictual interaction with the outgroup. This suggests that they responded to the combination of negative stereotype and conflict as a challenge (Alter et al., 2010; Derks et al., 2010; Scheepers, 2009).

When the incongruent stereotype performance effects (threat improving performance and boost reducing it) were modeled across studies (using measured rather than manipulated variables), it emerged that incongruent boost effects were most consistently predicted by perceived permeability while incongruent threat effects were related to perceived conflict. These patterns of performance are intelligible when task-performance is seen as an accountable act with consequences for positioning one's self and one's group in a performance context (eg. Tajfel, 1981a).

Understanding stereotype threat and boost as outcomes of situated identity performance makes sense of the incongruent stereotype performance effects observed in the literature and in the present study. For example, participants who underperform in a boost condition could be negatively affected by the increased pressure to enact model minority status (as argued by Cheryan & Bodenhausen, 2000), or be slipstreaming because their positive social identity relieves them of any pressure to prove themselves as individuals (as observed in many of the present studies). Similarly, participants may perform poorly under threat because they believe that they will do badly and give up; believe that they need to do well to disprove the stereotype and ironically perform worse because of "overeffort"; or as a way of demonstrating their prototypicality of the negatively stereotyped group (cf. Willis,

1981). All of these identity strategies are valid and likely reasons for increased or decreased performance on the APM. However, unless the performance shifts are understood *as* identity strategies, the congruence or incongruence of stereotype-performance effects is unintelligible.

The problem with this model is that it is not easily generalizable beyond the very general statement that task performance will often be a meaningful identity-constructing response in the specific performance situation experienced by each individual. As Tajfel cautioned about applying SIT to the experiences of minority group membership:

It cannot be assumed in any discussion of these effects that facile generalizations would emerge which could be applied to all, most, or even many members of one or another minority group. All the 'effects' we shall be describing apply to *some* members of *some* minorities, and a variety of patterns can be found within any one minority. All that can be achieved is to identify some patterns which appear more important than others because they are adopted by a variety of people in a variety of groups in a variety of circumstances. The generalizations of social psychology are (thankfully) limited by the creative and boundless diversity and flexibility of human social behaviour (Tajfel, 1981a, p. 316).

The saving grace for the model is that the task-performance context in stereotype threat is generally a tightly constrained context, such as in the academic performance of women in science. To develop an intervention, according to the model, educators would need to identify the strategies and identity resources that would provide the most likely platform for successful and active resistance in the specific task-performance context. For example, the results of the present studies suggest that invoking intergroup conflict (for example, by teaching the students about sexism in the academy, ensuring they have several successful role models and providing convincing cognitive alternatives) could set-up a much more successful strategic platform for resistance than the usual strategy in universities of pretending that there

is no sex-discrimination and that everyone has equal opportunities. However, if this strategy was successful it would not be because of anything intrinsic to perceived intergroup conflict, but because of how it would orient the minority group members in the specific performance context to adopt an identity position oriented to optimal performance. In other performance contexts perceived conflict might have the opposite effect; and there may be many other identity platforms that would encourage active resistance in any particular context.

It is critical to note that saying that identity performance is active and strategic, and that task performance is the outcome of situated identity performance, is *not* to say that the quality of the performance is under the conscious or full control of the performer. First, if a target of a stereotype is acting from an inadequate identity platform in the context then (in agreement with conventional stereotype threat theorists) it is unlikely that they will be able to perform optimally simply by choosing to do so. Second, minority group members face many practical and material challenges that limit their range of responses. Third, since the optimal identity conditions may frequently be non-obvious, performers may in fact undermine their own performance by adopting less than optimal strategies (Jamieson & Harkins, 2009). Fourth, most identity resources and liabilities are social and therefore out of the full control of the performer.

A second caveat is that this model does *not* suggest that existing models of stereotype threat are wrong: indeed, there must be situated identity strategies that would result in overtrying (Jamieson & Harkins, 2009; Steele, 2010), cognitive interference (Schmader, 2010; Schmader & Johns, 2003; Schmader et al., 2008) or unmanageable anxiety (Chung et al., 2010; Osborne, 2007). However, the findings of the present studies, and the inconsistent findings in the stereotype threat literature to date (Nguyen & Ryan, 2008; Smith, 2004), suggest that these specific models are applicable to some subsets of identity performance strategies and not others.

As Tajfel argued, “Social identities are not simply means of simplifying a complex social reality. Rather, they are a means of both creating and defining one’s place in a dynamic social world. Above all ... social identity is a *guide to social action*” (Tajfel, 1972, as cited in Reicher, 1996, emphasis original). Applying this insight to stereotype-performance opens a new field of enquiry for truly exploring stereotype-performance effects as identity contingencies as originally envisaged by Steele and his colleagues (Steele et al., 2002).

This shift in focus, ironically, will increase—rather than decrease—the ‘mediators’ and ‘moderators’ that will be found to influence the stereotype-performance relationship. If the model of the subject as an active producer of identity is tenable, then ‘mediators’ and ‘moderators’ will differ across cultures for similar categories, and will even differ across stereotyped categories within a single culture, because “mediators” and “moderators” can no longer be understood as permanent, universal ‘switches’ that deterministically modify performance in any context. Instead they must be understood as social resources that are responded to by active subjects who are busy producing identity in a setting. The key to understanding stereotype-performance effects in this framework is not how features of the context *produce* stereotype threat, but how individuals actively assess their identity resources and liabilities in a particular context and use task performance as one means of positioning themselves and producing identity in that context. As Chris from Delaware said to Claude M. Steele, “it’s not a threat: it’s a challenge; it’s an opportunity; it’s an invitation. (Steele & Conan, 2010).

## References

- Abad, F., Colom, R., Rebollo, I., & Escorial, S. (2003). Sex differential item functioning in the Raven's Advanced Progressive Matrices: Evidence for bias. *Personality and Individual Differences, 36*, 1459-1470. doi: 10.1016/S0191-8869(03)00241-1
- Abrams, D., Crisp, R. J., Marques, S., Fagg, E., Bedford, L., & Provias, D. (2008). Threat inoculation: Experienced and imagined intergenerational contact prevents stereotype threat effects on older people's math performance. *Psychology and Aging, 23*(4), 934-939.
- Adams, G., Garcia, D. M., Purdie-Vaughns, V., & Steele, C. M. (2006). The detrimental effects of a suggestion of sexism in an instruction situation. *Journal of Experimental Social Psychology, 42*(5), 602-615.
- Aguinis, H. (2004). *Regression analysis for categorical moderators*: New York: The Guilford press.
- Allport, F. H. (1920). The influence of the group upon association and thought. *Journal of Experimental Psychology, 3*, 159-182.
- Allport, F. H. (1923). Editorial comment upon the effect of an audience. *Journal of Abnormal and Social Psychology, 18*, 342-344.
- Allport, G. W. (1954). *The nature of prejudice*: New York: Addison-Wesley.
- Alter, A. L., Aronson, J., Darley, J. M., Rodriguez, C., & Ruble, D. N. (2010). Rising to the threat: Reducing stereotype threat by reframing the threat as a challenge. *Journal of Experimental Social Psychology, 46*(1), 166-171. doi: 10.1016/j.jesp.2009.09.014
- Amazon.com. (2010). Editorial review: Whistling Vivaldi: And other clues to how stereotypes affect us, from <http://www.amazon.com/Whistling-Vivaldi-Stereotypes-Affect-Issues/dp/039306249X>
- 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*(3), 401-408.
- Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science, 12*(5), 385-390.
- Andreolletti, C., & Lachman, M. E. (2004). Susceptibility and resilience to memory aging stereotypes: Education matters more than age. *Experimental Aging Research, 30*(2), 129-148.
- Andres. (2010, April). Review: Whistling Vivaldi: And other clues to how stereotypes affect us (issues of our time series), from <http://www.goodreads.com/book/show/6649312-whistling-vivaldi>
- Aries, P. (1962). *Centuries of childhood*: New York: Vintage Books.
- Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or "would jesse jackson 'fail' the implicit association test?". *Psychological Inquiry, 15*(4), 257-278.
- Armenta, B. E. (2010). Stereotype boost and stereotype threat effects: The moderating role of ethnic identification. *Cultural Diversity and Ethnic Minority Psychology, 16*(1), 94-98.
- Aronson, J., Cohen, G., & McColskey, W. (2009a). Reducing stereotype threat in classrooms: A review of social-psychological intervention studies on improving the achievement of black students. *Regional Educational Laboratory Southeast*(076).



- 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., Jannone, S., McGlone, M., & Johnson-Campbell, T. (2009b). The Obama effect: An experimental test. *Journal of Experimental Social Psychology, 45*(4), 957-960.
- 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*(1), 29-46.
- Arthur, W., JR., & Day, D. V. (1994). Development of a short form for the Raven Advanced Progressive Matrices test. *Educational and Psychological Measurement, 54*(2), 394-403.
- Arthur, W., Jr., Tubre, T. C., Paul, D. S., & Sanchez-Ku, M. L. (1999). College-sample psychometric and normative data on a short form of the Raven Advanced Progressive Matrices test. *Journal of Psychoeducational Assessment, 17*(4), 354-361.
- Athanassoulis, N., & Wilson, J. (2009). When is deception in research ethical? *Clinical Ethics, 4*(1), 44-49.
- Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2004). No place for nostalgia in science: A response to arkes and tetlock. *Psychological Inquiry, 15*(4), 279-289.
- Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71*(2), 230-244.
- Baumeister, R. F., & Showers, C. J. (1986). A review of paradoxical performance effects: Choking under pressure in sports and mental tests. *European Journal of Social Psychology, 16*(4), 361-383.
- Bedeian, A. G., & Mossholder, K. W. (1994). Simple question, not so simple answer: Interpreting interaction terms in moderated multiple regression. *Journal of Management, 20*(1), 159 - 165.
- Beilock, S. L., Jellison, W. A., Rydell, R. J., McConnell, A. R., & Carr, T. H. (2006). On the causal mechanisms of stereotype threat: Can skills that don't rely heavily on working memory still be threatened? *Personality and Social Psychology Bulletin, 32*(8), 1059-1071.
- Beilock, S. L., & McConnell, A. R. (2004). Stereotype threat and sport: Can athletic performance be threatened? *Journal of Sport and Exercise Psychology, 26*(4), 597-610.
- Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General, 136*(2), 256-276.
- Bell, A. E., Spencer, S. J., Iserman, E., & Logel, C. E. R. (2003). Stereotype threat and women's performance in engineering. *Journal of Engineering Education, 92*(4), 308-312.
- Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology, 41*(2), 174-181.
- Bergeron, D. M., Block, C. J., & Echtenkamp, A. (2006). Disabling the able: Stereotype threat and women's work performance. *Human Performance, 19*(2), 133-158.
- Berglas, S., & Jones, E. E. (1978). Drug choice as a self-handicapping strategy in response to noncontingent success. *Journal of Personality and Social Psychology, 36*(4), 405-417.
- Berkowitz, L. (1984). Some effects of thoughts on anti- and prosocial influences of media events: A cognitive-neoassociation analysis. *Psychological Bulletin, 95*, 410-427.

- Bettencourt, B. A., & Hume, D. L. (2001). Status differences and in-group bias: A meta-analytic examination of the effects of status stability, status legitimacy, and group permeability. *Psychological Bulletin*, *127*(4), 520-542.
- Billig, M. (1996). *Arguing and thinking: A rhetorical approach to social psychology*: Cambridge University Press.
- Billig, M. (2001a). Humour and embarrassment: Limits of 'nice-guy' theories of social life. *Theory, Culture and Society*, *18*(5), 23-43.
- Billig, M. (2001b). Humour and hatred: The racist jokes of the ku klux klan. *Discourse and Society*, *12*, 267-289.
- Bland, J. M., & Altman, D. G. (1994). Statistics notes: Some examples of regression towards the mean. *BMJ*, *309*(6957), 780.
- Blanton, H., Christie, C., & Dye, M. (2002). Social identity versus reference frame comparisons: The moderating role of stereotype endorsement. *Journal of Experimental Social Psychology*, *38*(3), 253-267.
- Blanton, H., Crocker, J., & Miller, D. T. (2000). The effects of in-group versus out-group social comparison on self-esteem in the context of a negative stereotype. *Journal of Experimental Social Psychology*, *36*, 519-530.
- Blanz, M., & Aufderheide, B. (1999). Social categorization and category attribution: The effects of comparative and normative fit on memory and social judgment. *British Journal of Social Psychology*, *38*, 157-179(123).
- Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science*, *12*(3), 225-229.
- Blumer, H. (1958). Race prejudice as a sense of group position. *The Pacific Sociological Review*, *1*(1), 3-7.
- Bodenhausen, G. (2010, 11 February). [Personal correspondence].
- Boltwood, S. (2001). Race, miscegenation, and the victorian staging of irishness. *Victorian Literature and Culture*, *29*(02), 383-396. doi: 10.1017/S1060150301002078
- Bornstein, G., Crum, L., Wittenbraker, J., Harring, K., Insko, C. A., & Thibaut, J. (1983). On the measurement of social orientations in the minimal group paradigm. *European Journal of Social Psychology*, *13*(4), p321-350.
- Bors, D. A., & Stokes, T. L. (1998). Raven's Advanced Progressive Matrices norms for first-year university students and the development of a short form. *Educational and Psychological Measurement*, *58*(3), 382-398.
- Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of Experimental Social Psychology*, *40*(2), 247-255.
- Branscombe, N. R., Ellemers, N., Spears, R., & Doosje, B. (1999). The context and content of social identity threat. In N. Ellemers, R. Spears & B. Doosje (Eds.), *{social identity: Context, commitment, content}*. Oxford: Blackwell Publishers.
- Brase, G. L. (2009). How different types of participant payments alter task performance. *Judgment and Decision Making*, *4*(5), 419-428.
- Brodish, A. B., & Devine, P. G. (2009). The role of performance-avoidance goals and worry in mediating the relationship between stereotype threat and performance. *Journal of Experimental Social Psychology*, *45*(1), 180-185.
- Brouwers, S. A., de Vijver, F. J. R. V., & Hemert, D. A. V. (2009). Variation in Raven's progressive matrices scores across time and place. *Learning and Individual Differences*, *19*(3), 330 - 338.
- Brown, R. (2000). Social identity theory: Past achievements, current problems and future challenges. *European Journal of Social Psychology*, *30*(6), 745-778.

- Brown, R. P., & Day, E. A. (2006). The difference isn't black and white: Stereotype threat and the race gap on Raven's Advanced Progressive Matrices. *Journal of Applied Psychology, 91*(4), p979-985.
- Brown, R. P., & Lee, M. N. (2005). Stigma consciousness and the race gap in college academic achievement. *Self and Identity, 4*(2005), 149-157.
- Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology, 39*(6), 626-633.
- Burnette, J. L., Pollack, J. M., & Hoyt, C. L. (2010). Individual differences in implicit theories of leadership ability and self-efficacy: Predicting responses to stereotype threat. *Journal of Leadership Studies, 3*(4), 46-56.
- Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology, 33*(2), 267-285.
- Cadinu, M., Maass, A., Lombardo, M., & Frigerio, S. (2006). Stereotype threat: The moderating role of locus of control beliefs. *European Journal of Social Psychology, 36*(2), 183-197.
- Carpenter, W. B. (1852). *On the influence of suggestion in modifying and directing muscular movement, independently of volition*. Paper presented at the Royal Institution of Great Britain Weekly Evening Meeting.  
<http://www.sgipt.org/medppp/psymot/carp1852.htm>
- Carr, P. B., & Steele, C. M. (2009). Stereotype threat and inflexible perseverance in problem solving. *Journal of Experimental Social Psychology, 45*(4), 853-859.
- Cassidy, T., & Lynn, R. (1989). A multifactorial approach to achievement motivation: The development of a comprehensive measure. *Journal of Occupational Psychology, 62*(4), 301-312.
- Chalabaev, A., Sarrazin, P., Stone, J., & Cury, F. (2008). Do achievement goals mediate stereotype threat? An investigation on females' soccer performance. *Journal of Sport and Exercise Psychology, 30*, 143-158.
- Chang, M. J., Eagan, M. K., Lin, M. H., & Hurtado, S. (2009, April ). *Stereotype threat: Undermining the persistence of racial minority freshmen in the sciences*. Paper presented at the American Educational Research Association San Diego.
- Chasteen, A. L., Bhattacharyya, S., Horhota, M., Tam, R., & Hasher, L. (2005). How feelings of stereotype threat influence older adults' memory performance. *Experimental Aging Research, 31*, 235-260.
- Chatard, A., Selimbegovic, L., Konan, P., & Mugny, G. (2008). Performance boosts in the classroom: Stereotype endorsement and prejudice moderate stereotype lift. *Journal of Experimental Social Psychology, 44*(5), 1421-1424.
- Cherry, F. (1995). *The 'stubborn particulars' of social psychology*: London: Routledge.
- Cheryan, S., & Bodenhausen, G. V. (2000). When positive stereotypes threaten intellectual performance: The psychological hazards of model minority status. *Psychological Science, 11*(5), 399-402.
- Christensen, L. (1988). Deception in psychological research: When is its use justified? *Personality and Social Psychology Bulletin, 14*(4), 664-675.
- Chung, B. G., Ehrhart, M. G., Holcombe Ehrhart, K., Hattrup, K., & Solamon, J. (2010). Stereotype threat, state anxiety, and specific self-efficacy as predictors of promotion exam performance. *Group Organization Management, 35*(1), 77-107.

- Clark, J. K., Eno, C. A., & Guadagno, R. E. (2010). Southern discomfort: The effects of stereotype threat on the intellectual performance of us southerners. *Self and Identity*, 9(1), 1-15.
- Claude M. Steele: Award for distinguished senior career contributions to the public interest. (2003). *American Psychologist*, 58(11), 909-911.
- Cocchiara, F. K., & Quick, J. C. (2004). The negative effects of positive stereotypes: Ethnicity-related stressors and implications on organizational health. *Journal of Organizational Behavior*, 25(6), 781-785.
- Cohen, G. L., & Garcia, J. (2005). I am us: Negative stereotypes as collective threats. *Journal of Personality and Social Psychology*, 89(4), 566-582.
- Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science*, 313(5791), 1307-1310.
- Cohen, G. L., Steele, C. M., & Ross, L. D. (1999). The mentor's dilemma: Providing critical feedback across the racial divide. *Personality and Social Psychology Bulletin*, 25(10), 1302-1318.
- Cole, J. C., Michailidou, K., Jerome, L., & Sumnall, H. R. (2006). The effects of stereotype threat on cognitive function in ecstasy users. *Journal of Psychopharmacology*, 20, 518-525.
- Colom, R., Escorial, S., & Rebollo, I. (2004). Sex differences on the progressive matrices are influenced by sex differences on spatial ability. *Personality and Individual Differences*, 37(6), 1289 - 1293.
- Costarelli, S. (2007). Intergroup threat and experienced affect: The distinct roles of causal attributions, ingroup identification, and perceived legitimacy of intergroup status. *Personality and Social Psychology Bulletin*, 33(11), 1481-1491.
- Cox, E. P. (1980). The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research*, 17(4), 407-422.
- Crisp, R. J., Bache, L. M., & Maitner, A. T. (2009). Dynamics of social comparison in counter-stereotypic domains: Stereotype boost, not stereotype threat, for women engineering majors. *Social Influence*, 4(3), 171-184.
- Crisp, R. J., & Beck, S. R. (2005). Reducing intergroup bias: The moderating role of ingroup identification. *Group Processes Intergroup Relations*, 8(2), 173-185.
- Croizet, J.-C., Desert, M., Dutrevis, M., & Leyens, J.-P. (2001). Stereotype threat, social class, gender, and academic under-achievement: When our reputation catches up to us and takes over. *Social Psychology of Education*, 4(3 - 4), 295-310.
- Croizet, J.-C., Despres, G., Gauzins, M.-E., Huguet, P., Leyens, J.-P., & Meot, A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin*, 30(6), 721-731.
- Croizet, J. C., & Claire, T. (1998). Extending the concept of stereotype and threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin*, 24(6), 588-594.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4), 349-354.
- Cullen, M. J., Hardison, C. M., & Sackett, P. R. (2004). Using SAT-grade and ability-job performance relationships to test predictions derived from stereotype threat theory. *Journal of Applied Psychology*, 89(2), 220-230.
- Cullen, M. J., Waters, S. D., & Sackett, P. R. (2006). Testing stereotype threat theory predictions for math-identified and non-math-identified students by gender. *Human Performance*, 19(4), 421-440.

- Curhan, J. R., & Overbeck, J. R. (2008). Making a positive impression in a negotiation: Gender differences in response to impression motivation. *Negotiation and Conflict Management Research, 1*(2), 179-193.
- Danaher, K., & Crandall, C. S. (2008). Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology, 38*(6), 1639-1655.
- Danziger, K. (1994). *Constructing the subject: Historical origins of psychological research*: Cambridge: Cambridge University Press.
- Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin, 28*(12), 1615-1628.
- Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of Personality and Social Psychology, 88*(2), 276-287.
- Davis, C., III, Aronson, J., & Salinas, M. (2006). Shades of threat: Racial identity as a moderator of stereotype threat. *Journal of Black Psychology, 32*(4), 399-417.
- Davis, D. W., & Silver, B. D. (2003). Stereotype threat and race of interviewer effects in a survey on political knowledge. *American Journal of Political Science, 47*(1), 33-45.
- Dawes, J. (2008). Do data characteristics change according to the number of scale points used? *International Journal of Market Research, 50*(1), 61-77.
- Deaux, K., Bikmen, N., Gilkes, A., Ventuneac, A., Joseph, Y., Payne, Y. A., et al. (2007). Becoming American: Stereotype threat effects in afro-caribbean immigrant groups. *Social Psychology Quarterly, 70*(4), 384-404.
- Deci, E., & Ryan, R. (2005). Intrinsic motivation inventory (imi). Retrieved from Retrieved July 11, 2008 from <http://www.psych.rochester.edu/SDT/measures/intrins.html>
- Dee, T. S. (2009). Stereotype threat and the student-athlete. *NBER Working Papers*.
- Derks, B., Scheepers, D., Laar, C. V., & Ellemers, N. (2010). The threat vs. Challenge of car parking for women: How self- and group affirmation affect cardiovascular responses. *Journal of Experimental Social Psychology, In Press, Corrected Proof*.
- Desert, M., Preaux, M., & Jund, R. (2009). So young and already victims of stereotype threat: Socio-economic status and performance of 6 to 9 years old children on Raven's progressive matrices. *European Journal of Psychology of Education, 24*(2), 207-218.
- DeVellis, R. F. (2003). *Scale development: Theory and applications*: Thousand Oaks: Sage Publications, Inc.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5-18.
- Dixon, J., Durrheim, K., Tredoux, C., Tropp, L., Clack, B., & Eaton, L. (2010). A paradox of integration? Interracial contact, prejudice reduction, and perceptions of racial discrimination. *Journal of Social Issues, 66*(2), 401-416.
- Doosje, B., Ellemers, N., & Spears, R. (1995). Perceived intragroup variability as a function of group status and identification. *Journal of Experimental Social Psychology, 31*(5), 410-436.
- Durrheim, K., Dixon, J., Tredoux, C., Eaton, L., Quayle, M., & Clack, B. (2009). Predicting support for racial transformation policies: Intergroup threat, racial prejudice, sense of group entitlement and strength of identification. *European Journal of Social Psychology*.
- Dutrevil, M., & Croizet, J. C. (2004). Reputation of intellectual inferiority undermines memory efficiency among college students. *Current Research in Social Psychology, 10*(8).

- Eiser, J. R. (1996). Accentuation revisited. In W. P. Robinson (Ed.), *Social groups and identities: Developing the legacy of henri tajfel* (pp. 359-370): Oxford: Butterworth-Heinemann.
- Elizaga, R., & Markman, K. (2008). Peers and performance: How in-group and out-group comparisons moderate stereotype threat effects. *Current Psychology, 27*(4), 290-300.
- Ellemers, N. (1993). The influence of socio-structural variables on identity management strategies. *European Review of Social Psychology, 4*, 27-57.
- Ellemers, N., & Barreto, M. (2006). Categorization in everyday life: The effects of positive and negative categorizations on emotions and self-views. *European Journal of Social Psychology, 36*(6), 931-942.
- Ellemers, N., Kortekaas, P., & Ouwerkerk, J. W. (1999). Self-categorisation, commitment to the group and group self-esteem as related but distinct aspects of social identity. *European Journal of Social Psychology, 29*(2/3), 371-389.
- Ellemers, N., Spears, R., & Doosje, B. (1997). Sticking together or falling apart: In-group identification as a psychological determinant of group commitment versus individual mobility. *Journal of Personality and Social Psychology, 72*, 617-626.
- Ellemers, N., Spears, R., & Doosje, B. (2002). Self and social identity. *Annual Review of Psychology, 53*(1), 161-186.
- Ellemers, N., van Knippenberg, A., de Vries, N., & Wilke, H. (1988). Social identification and permeability of group boundaries. *European Journal of Social Psychology, 18*(6), 494-513.
- Ellemers, N., Van Knippenberg, A., & Wilke, H. (1990). The influence of permeability of group boundaries and stability of group status on strategies of individual mobility and social change. *British Journal of Social Psychology, 29*(3), 233-246.
- Ellemers, N., Wilke, H., & Van Knippenberg, A. (1993). Effects of the legitimacy of low group or individual status on individual and collective status-enhancement strategies. *Journal of Personality and Social Psychology, 64*(5), 766-778.
- Encarta. (2009). Slipstream, from [http://encarta.msn.com/dictionary\\_1861701865/slipstream.html](http://encarta.msn.com/dictionary_1861701865/slipstream.html)
- Essack, Z. (2008). *Stereotype activation and university access programmes: Preparing students with one hand and holding them back with the other?* Unpublished masters thesis, School of Psychology, University of KwaZulu-Natal, Pietermaritzburg, South Africa.
- Evans, C. (2009). *Impact of stereotype threat on self-reported empathetic potential*. Doctoral thesis, Spalding University, Louisville, Kentucky. Retrieved from <http://gradworks.umi.com/33/71/3371755.html>
- Fiedler, K., Messner, C., & Bluemke, M. (2006). Unresolved problems with the I, A, and the T: A logical and psychometric critique of the implicit association test (IAT). *European Review of Social Psychology, 17*, 74-147.
- Fiedler, K., & Walther, E. (2004). *Stereotyping as inductive hypothesis testing*: Hove: Psychology Press.
- Finchilescu, G., & Tredoux, C. (2010). The changing landscape of intergroup relations in south africa. *Journal of Social Issues, 66*(2), 223-236.
- Flynn, J. R. (2007). *What is intelligence?: Beyond the Flynn effect*: Cambridge University Press.
- Forbes, C. E. (2009). *Lessons learned from "a clockwork orange": How retraining implicit attitudes and stereotypes affects motivation and performance under stereotype threat*. Doctoral thesis, University of Arizona, Tuscon, Arizona. Retrieved from <http://gradworks.umi.com/33/52/3352594.html>

- Forbes, J. (2007). *Stereotype-threat and the emotional sensitivity of psychology students*. Unpublished honour's thesis, University of KwaZulu-Natal, Pietermaritzburg, South Africa.
- 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(5), 643-653.
- Förster, A., Jens, Higgins, E. T., & Werth, L. (2004). How threat from stereotype disconfirmation triggers self-defense. *Social Cognition*, 22(1), 54-74.
- Frantz, C. M., Cuddy, A. J. C., Burnett, M., Ray, H., & Hart, A. (2004). A threat in the computer: The race implicit association test as a stereotype threat experience. *Personality and Social Psychology Bulletin*, 30(12), 1611-1624.
- Fritzsche, B. A., DeRouin, R. E., & Salas, E. (2009). The effects of stereotype threat and pacing on older adults' learning outcomes. *Journal of Applied Social Psychology*, 39(11), 2737-2755.
- Fryberg, S. A., Markus, H. R., Oyserman, D., & Stone, J. M. (2008). Of warrior chiefs and Indian princesses: The psychological consequences of American Indian mascots. *Basic & Applied Social Psychology*, 30(3), 208-218.
- Gaertner, S. L., & Dovidio, J. F. (2005). Understanding and addressing contemporary racism: From aversive racism to the common ingroup identity model. *Journal of Social Issues*, 61(3), 615-639.
- Garson, D. G. (2010). Multiple regression Retrieved 21 April, 2010, from <http://faculty.chass.ncsu.edu/garson/PA765/regress.htm>
- Gates, H. L., & Steele, C. M. (2009). A conversation with Claude M. Steele. *Du Bois Review: Social Science Research on Race*, 6(02), 251-271.
- Gerber, G. (2009). Status and the gender stereotyped personality traits: Toward an integration. *Sex Roles*, 61, 297-316.
- Goff, P. A., Steele, C. M., & Davies, P. G. (2008). The space between us: Stereotype threat and distance in interracial contexts. *Journal of Personality and Social Psychology*, 94(1), 91-107.
- Goffman, E. (1961). *Asylums. Essays on the social situation of mental patients and other inmates*: Harmondsworth: Penguin.
- Goffman, E., Lemert, C. C., & Branaman, A. (1997). *The Goffman reader*: Cambridge: Blackwell.
- Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and Social Psychology Bulletin*, 28(5), 659-670.
- Good, C., Aronson, & Inzlicht, J. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology*, 24, 645-662.
- Good, C., Aronson, J., & Harder, J. A. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology*, 29(1), 17-28.
- Good, J. J., Woodzicka, J. A., & Wingfield, L. C. (2010). The effects of gender stereotypic and counter-stereotypic textbook images on science performance. *Journal of Social Psychology*, 150, 132-147.
- Govorun, O., Fuegen, K., & Payne, B. K. (2006). Stereotypes focus defensive projection. *Personality and Social Psychology Bulletin*, 32(6), 781-793.
- Green, S. B. (1991). How many subjects does it take to do a regression analysis. *Multivariate Behavioral Research*, 26(3), 499.

- Gresky, D., Eyck, L., Lord, C., & McIntyre, R. (2005). Effects of salient multiple identities on women's performance under mathematics stereotype threat. *Sex Roles, 53*(9), 703-716.
- Griffiths, L. (1998). Humour as resistance to professional dominance in community mental health teams. *Sociology of Health & Illness, 20*(6), 874-895.
- Grimm, L. R., Markman, A. B., Maddox, W. T., & Baldwin, G. C. (2009). Stereotype threat reinterpreted as a regulatory mismatch. *Journal of Personality and Social Psychology, 96*(2), 288-304.
- Gupta, V. K., & Bhawe, N. M. (2007). The influence of proactive personality and stereotype threat on women's entrepreneurial intentions. *Journal of Leadership and Organizational Studies, 13*(4), 73-85.
- Guyll, M., Madon, S., Prieto, L., & Scherr, K. C. (2010). The potential roles of self-fulfilling prophecies, stigma consciousness, and stereotype threat in linking latino/a ethnicity and educational outcomes. *Journal of Social Issues, 66*(1), 113-130.
- Halpern, D., Aronson, J., Reimer, N., Simpkins, S., Star, J., & Wentzel, K. (2007). *Encouraging girls in math and science washington, dc*: Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.
- Halpern, D. F., & Tan, U. (2001). Stereotypes and steroids: Using a psychobiosocial model to understand cognitive sex differences. *Brain and Cognition, 45*(3), 392-414.
- Harris, R. J. (1985). *A primer of multivariate statistics*: Harcourt Brace Janovich.
- Harrison, C. K., Stone, J., Shapiro, J., Yee, S., Boyd, J. A., & Rullan, V. (2009). The role of gender identities and stereotype salience with the academic performance of male and female college athletes. *Journal of Sport and Social Issues, 33*(1), 78-96.
- Haslam, S., & Turner, J. C. (1992). Context-dependent variation in social stereotyping 2: The relationship between frame of reference, self-categorization and accentuation. *European Journal of Social Psychology, 22*(3), 251-277.
- Haslam, S. A., Salvatore, J., Kessler, T., & Reicher, S. D. (2008). The social psychology of success. *Scientific American Mind, 19*(2), 24-31.
- Hausdorff, J. M., Levy, B. R., & Wei, J. Y. (1999). The power of ageism on physical function of older persons: Reversibility of age-related gait changes. *Journal of the American Geriatrics Society, 47*(11), 1346-1349.
- Hays, R. D., Hayashi, T., & Stewart, A. L. (1989). A five-item measure of socially desirable response set. *Educational and Psychological Measurement, 49*(3), 629-636.
- Henry, J. D., von Hippel, C., & Shapiro, L. (2010). Stereotype threat contributes to social difficulties in people with schizophrenia. *British Journal of Clinical Psychology, 49*, 31-41.
- Herrera, C. (2001). Ethics, deception, and 'those Milgram experiments'. *Journal of Applied Philosophy, 18*(3), 245-256.
- Hertwig, R., & Ortmann, A. (2008). Deception in experiments: Revisiting the arguments in its defense. *Ethics & Behavior, 18*(1), p59-92.
- Hess, T., Hinson, J., & Hodges, E. (2009a). Moderators of and mechanisms underlying stereotype threat effects on older adults' memory performance. *Experimental Aging Research, 35*, 153-177.
- Hess, T. M., Auman, C., Colcombe, S. J., & Rahhal, T. A. (2003). The impact of stereotype threat on age differences in memory performance. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 58*(1), 3-11.



- Hess, T. M., Emery, L., & Queen, T. L. (2009b). Task demands moderate stereotype threat effects on memory performance. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *64B*(4), 482-486.
- Hess, T. M., Hinson, J. T., & Statham, J. A. (2004). Explicit and implicit stereotype activation effects on memory: Do age and awareness moderate the impact of priming? *Psychology and Aging*, *19*(3), 495-505.
- Hilton, J. L., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, *47*, 237-271.
- Ho, A. K., & Sidanius, J. (2010). Preserving positive identities: Public and private regard for one's ingroup and susceptibility to stereotype threat. *Group Processes Intergroup Relations*, *13*, 55-67.
- Hogg, M.-A. (2001). A social identity theory of leadership. *Personality and Social Psychology Review*, *5*(3), 184-200.
- Holland, R. W., Hendriks, M., & Aarts, H. (2005). Smells like clean spirit. *Psychological Science*, *16*(9), 689-693.
- Hornsey, M. J., Spears, R., Cremers, I., & Hogg, M. A. (2003). Relations between high and low power groups: The importance of legitimacy. *Personality and Social Psychology Bulletin*, *29*(2), 216-227.
- Howard, J. (2010). *Mastering cycling*. Leeds: Human Kinetics.
- Hoyt, C. L. (2005). The role of leadership efficacy and stereotype activation in women's identification with leadership. *Journal of Leadership and Organizational Studies*, *11*(4), 2-14.
- Hoyt, C. L., & Blascovich, J. (2007). Leadership efficacy and women leaders' responses to stereotype activation. *Group Processes and Intergroup Relations*, *10*(4), 595-616.
- Hoyt, C. L., & Blascovich, J. (2010). The role of leadership self-efficacy and stereotype activation on cardiovascular, behavioral and self-report responses in the leadership domain. *The Leadership Quarterly*, *21*(1), 89-103.
- Huguet, P., Brunot, S., & Monteil, J. M. (2000). Geometry versus drawing: Changing the meaning of the task as a means to change performance. *Social Psychology of Education*, *4*(3 - 4), 219-234.
- Huguet, P., & Régner, I. (2007). Stereotype threat among schoolgirls in quasi-ordinary classroom circumstances. *Journal of Educational Psychology*, *99*(3).
- Huguet, P., & Régner, I. (2009). Counter-stereotypic beliefs in math do not protect school girls from stereotype threat. *Journal of Experimental Social Psychology*, *45*(4), 1024-1027.
- Hyde, J. S., & Kling, K. C. (2001). Women, motivation, and achievement. *Psychology of Women Quarterly*, *25*(4), 364-378.
- Inzlicht, M., Aronson, J., Good, C., & McKay, L. (2006a). A particular resiliency to threatening environments. *Journal of Experimental Social Psychology*, *42*(3), 323-336.
- Inzlicht, M., McKay, L., & Aronson, J. (2006b). Stigma as ego depletion: How being the target of prejudice affects self-control. *Psychological Science (Wiley-Blackwell)*, *17*(3), 262-269.
- Jackson, J. W. (2002). Intergroup attitudes as a function of different dimensions of group identification and perceived intergroup conflict. *Self & Identity*, *1*(1), p11-33.
- Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology*, *93*(4).

- Jamieson, J. P., & Harkins, S. G. (2009). The effect of stereotype threat on the solving of quantitative GRE problems: A mere effort interpretation. *Personality and Social Psychology Bulletin*, 35(10), 1301-1314.
- Jetten, J., Branscombe, N. R., & Spears, R. (2002). On being peripheral: Effects of identity insecurity on personal and collective self-esteem. *European Journal of Social Psychology*, 32(1), 105-123. doi: 10.1002/ejsp.64
- Jetten, J., Branscombe, N. R., Spears, R., & McKimmie, B. M. (2003). Predicting the paths of peripherals: The interaction of identification and future possibilities. *Personality and Social Psychology Bulletin*, 29(1), 130-140.
- Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *Journal of Experimental Psychology: General*, 137(4), 691-705.
- Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle. Teaching stereotype threat as a means of improving women's math performance. *Psychological Science*, 16(3), 175-179.
- Johnson, W., & Bouchard, T. J. (2007). Sex differences in mental abilities: G masks the dimensions on which they lie. *Intelligence*, 35(1), 23-39.
- Josephs, R. A., Newman, M. L., Brown, R. P., & Beer, J. M. (2003). Status, testosterone, and human intellectual performance: Stereotype threat as status concern. *Psychological Science*, 14(2), 158-163.
- Josephs, R. A., Sellers, J. G., Newman, M. L., & Mehta, P. H. (2006). The mismatch effect: When testosterone and status are at odds. *Journal of Personality and Social Psychology*, 90(6), 999-1013.
- Jost, J. T., & Burgess, D. (2000). Attitudinal ambivalence and the conflict between group and system justification motives in low status groups. *Personality and Social Psychology Bulletin*, 26(3), 293-305.
- Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *Journal of Abnormal and Social Psychology*, 28(3), 280 - 290.
- Katz, I. (1968). Factors influencing negro performance in the desegregated school. In M. Deutsch, I. Katz & A. R. Jensen (Eds.), *Social class, race, and psychological development*. New York: Holt, Rinehart, and Winston.
- Katz, I., Roberts, S. O., & Robinson, J. M. (1965). Effects of task difficulty, race of administrator, and instructions on digit-symbol performance of negroes. *Journal of Personality and Social Psychology*, 2(1), 53-59.
- Keller, J. (2002). Blatant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles*, 47, 193-198.
- Keller, J. (2007). Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' maths performance. *British Journal of Educational Psychology*, 77(2), 323-338.
- Keller, J., & Bless, H. (2007). Communicating positive and negative stereotypic expectancies: The interplay of stereotype threat and regulatory focus. In Y. Kashima, K. Fiedler & P. Freytag (Eds.), *Stereotype dynamics: Language-based approaches to stereotype formation, maintenance, and transformation* (pp. 373-395): Mahwah, NJ: Erlbaum.
- Keller, J., & Bless, H. (2008). When positive and negative expectancies disrupt performance: Regulatory focus as a catalyst. *European Journal of Social Psychology*, 38(2).

- Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. *Personality and Social Psychology Bulletin*, 29(3), 371-381.
- Kellow, T. J., & Jones, B. D. (2005). Stereotype threat in African-American high school students: An initial investigation. *Current Issues in Education [On-line]*, 8(15).
- Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology*, 43(5), 825-832.
- Klein, K. G. (1999). *Diversity and detective fiction*. Bowling Green: Popular Press.
- Klein, O., Spears, R., & Reicher, S. (2007). Social identity performance: Extending the strategic side of side. *Personality and Social Psychology Review*, 11(1), 28-45.
- Knofczynski, G. T., & Mundfrom, D. (2008). Sample sizes when using multiple linear regression for prediction. *Educational and Psychological Measurement*, 68(3), 431-442.
- Koenig, A. M., & Eagly, A. H. (2005). Stereotype threat in men on a test of social sensitivity. *Sex Roles*, 52(7 - 8), 489-496.
- Kray, L., Locke, C., & Haselhuhn, M. (2007). *In the words of Larry Summers: Gender stereotype endorsement and implicit negotiation beliefs in mixed-gender negotiations*. Paper presented at the International Association for Conflict Management 20th Annual Conference, Budapest, Hungary. <http://ssrn.com/abstract=1080637>
- Kray, L. J., Reb, J., Galinsky, A. D., & Thompson, L. (2002). Reversing the gender gap in negotiations: An exploration of stereotype regeneration. *Personality and Social Psychology Bulletin*, 87(2), 386-409.
- Kray, L. J., Reb, J., Galinsky, A. D., & Thompson, L. (2004). Stereotype reactance at the bargaining table: The effect of stereotype activation and power on claiming and creating value. *Personality and Social Psychology Bulletin*, 30(4), 399-411.
- Kray, L. J., & Shirako, A. (2009). Stereotype threat in organizations: An examination of its scope, triggers, and possible interventions.
- Kray, L. J., Thompson, L., & Galinsky, A. (2001). Battle of the sexes: Gender stereotype confirmation and reactance in negotiations. *Journal of Personality and Social Psychology*, 80(6), 942-958.
- Lawrence, J., Marks, B., & Jackson, J. (2010). Domain identification predicts black students' underperformance on moderately-difficult tests. *Motivation and Emotion*(35), 105-109.
- 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(5), 526-537.
- Lee, Y.-T., & Ottati, V. (1995). Perceived in-group homogeneity as a function of group membership salience and stereotype threat. *Personality and Social Psychology Bulletin*, 21(6), 610-619.
- Lemaine, G. r. (1974). Social differentiation and social originality. *European Journal of Social Psychology*, 4(1), 17-52.
- Lesko, A. C., amp, & Corpus, J. H. (2006). Discounting the difficult: How high math-identified women respond to stereotype threat. *Sex Roles*, 54(1-2), 113-125.
- Levy, B. (1996). Improving memory in old age through implicit self-stereotyping. *Journal of Personality and Social Psychology*, 71(6), 1092-1107.
- Leyens, J. P., Desert, M., Croizet, J. C., & Darcis, C. (2000). Stereotype threat: Are lower status and history of stigmatization preconditions of stereotype threat? *Personality and Social Psychology Bulletin*, 26(10), 1189-1199.

- Lochbaum, M., Stevenson, S., & Hilario, D. (2009). Achievement goals, thoughts about intense physical activity, and exerted effort: A mediational analysis. *Journal of Sport Behavior, 32*(1), 53-68.
- Logel, C., Iserman, E. C., Davies, P. G., Quinn, D. M., & Spencer, S. J. (2009). The perils of double consciousness: The role of thought suppression in stereotype threat. *Journal of Experimental Social Psychology, 45*(2), 299-312.
- Looby, A., & Earleywine, M. (2010). Gender moderates the impact of stereotype threat on cognitive function in cannabis users. *Addictive Behaviors, 35*(9), 834-839.
- Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence, 32*(5), 481 - 498.
- Macrae, C. N., Stangor, C., & Hewstone, M. (1996). *Stereotypes and stereotyping*: New York: Guilford Press.
- Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology, 42*(2), 236-243.
- Marx, D. M. (2009). On the role of group membership in stereotype-based performance effects. *Social and Personality Psychology Compass, 3*(1), 77-93.
- Marx, D. M., Brown, J. L., & Steele, C. M. (1999). Allport's legacy and the situational press of stereotypes. *Journal of Social Issues, 55*(3), 491-502.
- Marx, D. M., & Goff, P. A. (2005). Clearing the air: The effect of experimenter race on target's test performance and subjective experience. *British Journal of Social Psychology, 44*(13), 645-657.
- Marx, D. M., Ko, S. J., & Friedman, R. A. (2009). The "Obama effect": How a salient role model reduces race-based performance differences. *Journal of Experimental Social Psychology, 45*(4), 953-956.
- Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin, 28*(9), 1183-1193.
- Marx, D. M., & Stapel, D. A. (2006a). It's all in the timing: Measuring emotional reactions to stereotype threat before and after taking a test. *European Journal of Social Psychology, 36*(1), 1-14.
- Marx, D. M., & Stapel, D. A. (2006b). It depends on your perspective: The role of self-relevance in stereotype-based underperformance. *Journal of Experimental Social Psychology, 42*(6), 768-775.
- Marx, D. M., & Stapel, D. A. (2006c). Understanding stereotype lift: On the role of the social self. *Social Cognition, 24*, 776-791.
- Marx, D. M., Stapel, D. A., & Muller, D. (2005). We can do it: The interplay of construal orientation and social comparisons under threat. *Journal of Personality and Social Psychology, 88*(3), 432-446.
- Massey, D. S., & Fischer, M. J. (2005). Stereotype threat and academic performance: New findings from a racially diverse sample of college freshmen. *Du Bois Review: Social Science Research on Race, 2*(1), 45-67.
- Mayer, D. M., & Hanges, P. J. (2003). Understanding the stereotype threat effect with culture-free tests: An examination of its mediators and measurement. *Human Performance, 16*(3), 207-230.
- McClelland, D. C. (1973). Testing for competence rather than for intelligence. *American Psychologist, 28*(1), 1-14.
- McFarland, L. A., Lev, A. D. M., & Ziegert, J. C. (2003). An examination of stereotype threat in a motivational context. *Human Performance, 16*(3), 181-205.

- McGlone, M. S., & Aronson, J. (2006). Stereotype threat, identity salience, and spatial reasoning. *Journal of Applied Developmental Psychology, 27*(5), 486-493.
- McGlone, M. S., Aronson, J., & Kobrynowicz, D. (2006). Stereotype threat and the gender gap in political knowledge. *Psychology of Women Quarterly, 30*, 392-398(397). doi: 10.1111/j.1471-6402.2006.00314.x
- McIntyre, R. B., Lord, C. G., Gresky, D. M., Eyck, L. L. T., Frye, G. D. J., & Bond, C. F., Jr. (2005). A social impact trend in the effects of role models on alleviating women's mathematics stereotype threat. *Current Research in Social Psychology, 10*(9), 116-136.
- McIntyre, R. B., Paulson, R. M., & Lord, C. G. (2003). Alleviating women's mathematics stereotype threat through salience of group achievements. *Journal of Experimental Social Psychology, 39*(1), 83-90.
- McKay, P. F., Doverspike, D., Bowen-Hilton, D., & McKay, Q. D. (2003). The effects of demographic variables and stereotype threat on black/white differences in cognitive ability test performance. *Journal of Business and Psychology, 18*(1), 1-14.
- McKay, P. F., Doverspike, D., Bowen, H. D., & Martin, Q. D. (2002). Stereotype threat effects on the Raven Advanced Progressive Matrices scores of African Americans. *Journal of Applied Social Psychology, 32*(4), 767-787.
- McKown, C., & Strambler, M. J. (2009). Developmental antecedents and social and academic consequences of stereotype-consciousness in middle childhood. *Child Development, 80*(6), 1643-1659.
- McKown, C., & Weinstein, R. S. (2003). The development and consequences of stereotype consciousness in middle childhood. *Child Development, 74*(2), 498-515.
- Mendoza-Denton, R., Shaw-Taylor, L., Chen, S., & Chang, E. (2009). Ironic effects of explicit gender prejudice on women's test performance. *Journal of Experimental Social Psychology, 45*(1), 275-278.
- Miles, J., & Shevlin, M. (2001). *Applying regression & correlation: A guide for students and researchers*. London: Sage Publications.
- Miller, D. E., & Kunce, J. T. (1973). Prediction and statistical overkill revisited. *Measurement and evaluation in guidance, 6*(3), 157-163.
- Mobbs, D., Hassabis, D., Seymour, B., Marchant, J. L., Weiskopf, N., Dolan, R. J., et al. (2009). Choking on the money. *Psychological Science, 20*(8), 955-962.
- Morrison, K. R., Wheeler, S. C., & Smeesters, D. (2007). Significant other primes and behavior: Motivation to respond to social cues moderates pursuit of prime-induced goals. *Personality and Social Psychology Bulletin, 33*(12), 1661-1674.
- Morton, T. A., Haslam, S. A., Postmes, T., & Ryan, M. K. (2006). We value what values us: The appeal of identity-affirming science. *Political Psychology, 27*(6), 823-838.
- Moscovici, S., & Mugny, G. (1985). *Perspectives on minority influence*. Cambridge: Cambridge University Press.
- Mount, H. (2010, January 9th). The death of 'shabby chic', *The Spectator*. Retrieved from <http://www.spectator.co.uk/essays/5686593/the-death-of-shabby-chic.shtml>
- Mummendey, A., Klink, A., Mielke, R., Wenzel, M., & Blanz, M. (1999). Socio-structural characteristics of intergroup relations and identity management strategies: Results from a field study in East Germany. *European Journal of Social Psychology, 29*(2/3), p259-285.
- Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science, 18*(10), 879-885.

- Newman, M. L., Sellers, J. G., & Josephs, R. A. (2005). Testosterone, cognition, and social status. *Hormones and Behavior*, 47(2), 205-211.
- Nguyen, H.-H. D., O'Neal, A., & Ryan, A. M. (2003). Relating test-taking attitudes and skills and stereotype threat effects to the racial gap in cognitive ability test performance. *Human Performance*, 16(3), 261-293.
- Nguyen, H.-H. H., & Ryan, A. M. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *The Journal of Applied Psychology*, 93(6), 1314-1334.
- Noel, J. G., Wann, D. L., & Branscombe, N. R. (1995). Peripheral ingroup membership status and public negativity toward outgroups. *Journal of Personality and Social Psychology*, 68(1), 127-137.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Nussbaum, A. D., & Steele, C. M. (2007). Situational disengagement and persistence in the face of adversity. *Journal of Experimental Social Psychology*, 43(1), 127-134.
- O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin*, 29(6), 782-789.
- Oakes, P. J., Haslam, S. A., & Turner, J. C. (1994). *Stereotyping and social reality*. Oxford: Blackwell.
- Osborne, J. (2006). Gender, stereotype threat, and anxiety: Psychophysiological and cognitive evidence. *Electronic Journal of Research in Educational Psychology*, 4(1), 109-138.
- Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement? *Contemporary Educational Psychology*, 26, 291-310.
- Osborne, J. W. (2007). Linking stereotype threat and anxiety. *Educational Psychology*, 27(1), 135-154.
- Osborne, J. W., & Walker, C. (2006). Stereotype threat, identification with academics, and withdrawal from school: Why the most successful students of colour might be most likely to withdraw. *Educational Psychology*, 26(1).
- Oswald, D. L., & Harvey, R. D. (2000). Hostile environments, stereotype threat, and math performance among undergraduate women. *Current Psychology*, 19(4), 338-356.
- Oxford Dictionaries. (2010). Definition of slipstream.
- Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology*, 84(4), 890-904.
- Perna, L. W., & Thomas, S. L. (2006). A framework for reducing the college success gap and promoting success for all: National Postsecondary education cooperative.
- Pettit, N. C., & Lount, J., Robert B. (2010). Looking down and ramping up: The impact of status differences on effort in intergroup contexts. *Journal of Experimental Social Psychology*, 46(1), 9 - 20.
- Pinel, E. C. (1999). Stigma consciousness: The psychological legacy of social stereotypes. *Journal of Personality and Social Psychology*, 76(1), 114-128.
- Ployhart, R. E., C., Z. J., & McFarland, L. A. (2003). Understanding racial differences on cognitive ability tests in selection contexts: An integration of stereotype threat and applicant reactions research. *Human Performance*, 16, 231-259.
- Pollack, B. (1963). The validity of the Shipley-Hartford retreat test for deterioration. *Psychiatric Quarterly*, 16(1), 119-131.
- Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1), 1-15.

- Pronin, E., Steele, C. M., & Ross, L. (2004). Identity bifurcation in response to stereotype threat: Women and mathematics. *Journal of Experimental Social Psychology, 40*, 152-168.
- Psychology Software Tools. (2010). E-prime 2.0.
- Purdie-Vaughns, V., Steele, C. M., Davies, P. G., Dittmann, R., & Crosby, J. R. (2008). Social identity contingencies: How diversity cues signal threat or safety for African Americans in mainstream institutions. *Journal of Personality and Social Psychology, 94*(4), 615-630.
- Quayle, M. (2009). *The social identity inventory: Measures and scale development*.
- Quinn, D. M., Kahng, S. K., & Crocker, J. (2004). Discreditable: Stigma effects of revealing a mental illness history on test performance. *Personality and Social Psychology Bulletin, 30*(7), 803-815.
- Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues, 57*(1), 55-71.
- R Development Core Team. (2010). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Ramos, M. D. (2010). *Group identification and perceived discrimination : A study of international students in the UK*. Doctoral thesis, University of St Andrews, St Andrews, Scotland.
- Raven, J. (1989). The Raven progressive matrices: A review of national norming studies and ethnic and socioeconomic variation within the united states. *Journal of Educational Measurement, 26*(1), 1-16.
- Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology, 41*(1), 1-48.
- Reicher, S. (1996). Social identity and social change: Rethinking the context of social psychology. In W. P. Robinson (Ed.), *Social groups and identities: Developing the legacy of henri tajfel* (pp. 317-336): Oxford: Butterworth-Heinemann.
- Reicher, S. (2004). The context of social identity: Domination, resistance, and change. *Political Psychology, 25*(6), p921-945.
- Reicher, S., Hopkins, N., & Condor, S. (1997). Stereotype construction as a strategy of influence. In P. J. Oakes, S. Russell, A. Haslam & N. Ellemers (Eds.), *The social psychology of stereotyping and group life*. Malden, MA, US: Blackwell Publishers.
- Reicher, S., & Levine, M. (1994). On the consequences of deindividuation manipulations for the strategic communication of self: Identifiability and the presentation of social identity. *Eur. J. Soc. Psychol., 24*(4), 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*(1), 15-40.
- Reicher, S. D., Spears, R., & Postmes, T. (1995). A social identity model of deindividuation phenomena. *European Review of Social Psychology, 6*, 161-198.
- Roberson, L., Deitch, E. A., Brief, A. P., & Block, C. J. (2003). Stereotype threat and feedback seeking in the workplace. *Journal of Vocational Behavior, 62*(1), 176-188.
- Rorty, R. (1989). *Contingency, irony, and solidarity*: Cambridge: Cambridge University Press.
- Rosen, B. C., & D'Andrade, R. (1959). The psychosocial origins of achievement motivation. *Sociometry, 22*(3), 185-218.

- Rosenthal, H. E. S., & Crisp, R. J. (2006). Reducing stereotype threat by blurring intergroup boundaries. *Personality and Social Psychology Bulletin*, 32(4), 501-511.
- Rosenthal, H. E. S., Crisp, R. J., & Suen, M.-W. (2007). Improving performance expectancies in stereotypic domains: Task relevance and the reduction of stereotype threat. *European Journal of Social Psychology*, 37(3), 586-597.
- Rosenthal, R. (1994). Interpersonal expectancy effects: A 30-year perspective. *Current Directions in Psychological Science*, 3(6), 176-179.
- Rosenthal, R., & Jacobson, L. (1968). Self-fulfilling prophecies in the classroom: Teachers' expectations as unintended determinants of pupils' intellectual competence *Social class, race, and psychological development* (pp. 219-253). New York: Holt, Rinehart, & Winston.
- Rushton, J. P., & Jensen, A. R. (2005a). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law*, 11(2), 235-294.
- Rushton, J. P., & Jensen, A. R. (2005b). Wanted: More race realism, less moralistic fallacy. *Psychology, Public Policy, and Law*, 11(2), 328-336.
- 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.
- Rydell, R. J., & Boucher, K. L. (2010). Capitalizing on multiple social identities to prevent stereotype threat: The moderating role of self-esteem. *Personality and Social Psychology Bulletin*, 36(2), 239-250.
- Rydell, R. J., Hugenberg, K., Ray, D., & Mackie, D. M. (2007). Implicit theories about groups and stereotyping: The role of group entitativity. *Personality and Social Psychology Bulletin*, 33(4), 549-558.
- Rydell, R. J., McConnell, A. R., & Beilock, S. L. (2009). Multiple social identities and stereotype threat: Imbalance, accessibility, and working memory. *Journal of Personality and Social Psychology*, 96(5), 949-966.
- Sackett, P. R., Hardison, C. M., & Cullen, M. J. (2004). On interpreting stereotype threat as accounting for African American-white differences on cognitive tests. *American Psychologist*, 59(1), 7-13.
- Sackett, P. R., Hardison, C. M., & Cullen, M. J. (2005). On interpreting research on stereotype threat and test performance. *American Psychologist*, 60(3), 271-272.
- Sawyer, A. G., Lynch, J. G., Jr., & Brinberg, D. L. (1995). A bayesian analysis of the information value of manipulation and confounding checks in theory tests. *The Journal of Consumer Research*, 21(4), pp. 581-595.
- Sawyer, J. T. P., & Hollis-Sawyer, L. A. (2005). Predicting stereotype threat, test anxiety, and cognitive ability test performance: An examination of three models. *International Journal of Testing*, 5(3), 225-246.
- Scheepers, D. (2009). Turning social identity threat into challenge: Status stability and cardiovascular reactivity during inter-group competition. *Journal of Experimental Social Psychology*, 45(1), 228-233.
- Scheepers, D., Ellemers, N., & Sintemaartensdijk, N. (2009). Suffering from the possibility of status loss: Physiological responses to social identity threat in high status groups. *European Journal of Social Psychology*, 39(6), 1075-1092.
- Schimmel, J., Arndt, J., Banko, K. M., & Cook, A. (2004). Not all self-affirmations were created equal: The cognitive and social benefits of affirming the intrinsic (vs. Extrinsic) self. *Social Cognition*, 22(1), 75-99.
- Schmader, T. (2002a). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, 38(2), 194-201.



- Schmader, T. (2002b, February). *Stereotype threat as seen through the lens of social identity theory*. Paper presented at the 3rd Annual Meeting of the Society for Personality and Social Psychology, Savannah, GA.
- Schmader, T. (2010). Stereotype threat deconstructed. *Current Directions in Psychological Science*, 19(1), 14-18.
- Schmader, T., Forbes, C. E., Shen, Z., & Berry Mendes, W. (2009). A metacognitive perspective on the cognitive deficits experienced in intellectually threatening environments. *Personality and Social Psychology Bulletin*, 35(5), 584-596.
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85(3), 440-452.
- Schmader, T., Johns, M., & Barquissau, M. (2004). The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. *Sex Roles*, 50(11 - 12), 835-850.
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115(2), 336-356.
- Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why can't a man be more like a woman sex differences in big five personality traits across 55 cultures. *Journal of Personality and Social Psychology*, 94(1), 168 - 182.
- Schubert, T. W., & Otten, S. (2002). Overlap of self, ingroup, and outgroup: Pictorial measures of self-categorization. *Self and Identity*, 1, 353-376.
- Schwartz, S. H., & Rubel, T. (2005). Sex differences in value priorities: Cross-cultural and multimethod studies. *Journal of Personality and Social Psychology*, 89(6), 1010 - 1028.
- Seibt, B., & Förster, J. (2004). Stereotype threat and performance: How self-stereotypes influence processing by inducing regulatory foci. *Journal of Personality and Social Psychology*, 87(1), 38-56.
- Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. *Journal of Experimental Social Psychology*, 39, 68-74.
- Senior, G. (2001). Shipley institute of living scale, from <http://www.usq.edu.au/users/senior/69301/Shipley.htm>
- 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(2), 107-130.
- Shevlin, M., Miles, J. N. V., Davies, M. N. O., & Walker, S. (2000). Coefficient alpha: A useful indicator of reliability? *Personality and Individual Differences*, 28(2), 229 - 237.
- Shih, M., Ambady, N., Richeson, J. A., Fujita, K., & Gray, H. M. (2002). Stereotype performance boosts: The impact of self-relevance and the manner of stereotype activation. *Journal of Personality and Social Psychology*, 83(3), 638-647.
- Shih, M., Bonam, C., Sanchez, D., & Peck, C. (2007). The social construction of race: Biracial identity and vulnerability to stereotypes. *Cultural Diversity and Ethnic Minority Psychology*, 13(2), 125-133.
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, 10(1), 80-83.
- Shih, M., Pittinsky, T. L., & Trahan, A. (2006). Domain-specific effects of stereotypes on performance. *Self and Identity*, 5(1), 1-14.

- Smith. (2005). Investigating a measure of computer technology domain identification: A tool for understanding gender differences and stereotypes. *Educational and Psychological Measurement*, 65(2), 336-355.
- Smith, J. (2006). The interplay among stereotypes, performance-avoidance goals, and women's math performance expectations. *Sex Roles*, 54(3), 287-296.
- 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.
- Smith, J. L., & Johnson, C. S. (2006). A stereotype boost or choking under pressure? Positive gender stereotypes and men who are low in domain identification. *Basic and Applied Social Psychology*, 28(1), 51-63.
- Smith, J. L., & White, P. H. (2001). Development of the domain identification measure: A tool for investigating stereotype threat effects. *Educational and Psychological Measurement*, 61(6), 1040-1057.
- Smith, J. L., & White, P. H. (2002). An examination of implicitly activated, explicitly activated, and nullified stereotypes on mathematical performance: It's not just a woman's issue. *Sex Roles*, 47(3 - 4), 179-191.
- Smith, J. L. S. C., & White, P. H. (2007). The stereotyped task engagement process: The role of interest and achievement motivation. *Journal of Educational Psychology*, 99(1), 99-114.
- 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(5), 538-553.
- Spears, R., & Smith, H. J. (2001). Experiments as politics. *Political Psychology*, 22(2), 309-330.
- Spencer, S., Logel, C., & Davies, P. G. (2011). Stereotype threat. *Annual Review of Psychology*, 62(1).
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1), 4-28.
- Stafford, R. E. (1972). Hereditary and environmental components of quantitative reasoning. *Review of Educational Research*, 42(2), 183-201.
- Stangor, C., Carr, C., & Kiang, L. (1998). Activating stereotypes undermines task performance expectations. *Journal of Personality and Social Psychology*, 75(5), 1191-1197.
- Statistics South Africa. (2007). Statistical release p031: Community survey 2007 (revised version): Online.
- Steele, C. M. (1992). Race and the schooling of black Americans. *The Atlantic Monthly*, 269(4), 68-78.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52(6), 613-629.
- Steele, C. M. (1999). Thin ice. *Atlantic Monthly*, 284(2), 44-53.
- Steele, C. M. (2003a). Through the back door to theory. *Psychological Inquiry*, 14(3/4), 314-317.
- Steele, C. M. (2010). *Whistling Vivaldi: And other clues to how stereotypes affect us*. London: W W Norton & Company.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797-811.
- Steele, C. M., & Aronson, J. A. (2004). Stereotype threat does not live by steele and aronson (1995) alone. *American Psychologist*, 59(1), 47-48.

- Steele, C. M., & Conan, N. (2010). 'Whistling Vivaldi' and beating stereotypes. Washington, DC: National Public Radio.
- Steele, C. M., & Davies, P. (2003). Stereotype threat and employment testing: A commentary. *Human Performance*, 16(3), 311-326.
- 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). San Diego: Academic Press.
- Steele, J. (2003b). Children's gender stereotypes about math: The role of stereotype stratification. *Journal of Applied Social Psychology*, 33(12), 2587-2606.
- Stephan, W. G., Boniecki, K. A., Ybarra, O., Bettencourt, A., Ervin, K. S., Jackson, L. A., et al. (2002). The role of threats in the racial attitudes of blacks and whites. *Personality and Social Psychology Bulletin*, 28(9), 1242-1254.
- Stephan, W. G., & Stephan, C. W. (1996). Predicting prejudice. *International Journal of Intercultural Relations*, 20(3-4), 409-426.
- Stevens, J. (1996). *Applied multivariate statistics for the social sciences*: Lawrence Erlbaum Associates.
- Stone, J. (2002). Battling doubt by avoiding practice: The effects of stereotype threat on self-handicapping in white athletes. *Personality and Social Psychology Bulletin*, 28(12), 1667-1678.
- Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on black and white athletic performance. *Journal of Personality and Social Psychology*, 77(6), 1213-1227.
- Stone, J., & McWhinnie, C. (2008). Evidence that blatant versus subtle stereotype threat cues impact performance through dual processes. *Journal of Experimental Social Psychology*, 44(2), 445-452.
- Stone, J., Perry, W., & Darley, J. M. (1997). 'White men can't jump': Evidence for the perceptual confirmation of racial stereotypes following a basketball game. *Basic and Applied Social Psychology*, 19(3), 291-306.
- Stricker, L. J., Bejar, & I, I. (1999). *Test difficulty and stereotype threat on the GRE general test. GRE research report 96-06R*: Princeton: Educational Testing Service.
- Stricker, L. J., & Bejar, I. I. (2004). Test difficulty and stereotype threat on the GRE general test. *Journal of Applied Social Psychology*, 34(3), 563-597.
- Stricker, L. J., & Ward, W. C. (2004). Stereotype threat, inquiring about test takers' ethnicity and gender, and standardized test performance. *Journal of Applied Social Psychology*, 34(4), 665-693.
- Stricker, L. J., & Ward, W. C. (2008). Stereotype threat in applied settings re-examined: A reply. *Journal of Applied Social Psychology*, 38(6), 1656-1663.
- Stroessner, S., Good, C., & Webster, L. (2010). What is stereotype threat? , from <http://www.reducingstereotypethreat.org/definition.html>
- Suzuki, L., & Aronson, J. (2005). The cultural malleability of intelligence and its impact on the racial/ethnic hierarchy. *Psychology, Public Policy, and Law*, 11(2), 320-327.
- Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics*: Harper & Row.
- Tagler, M., & Teles, S. M. (2003). *Stereotype threat: Prevalence and individual differences why is there no affirmative action in Britain?* 41  
<http://dx.doi.org/10.1177/0002764298041007010>, Kansas State University. Retrieved from <http://abs.sagepub.com/cgi/content/abstract/41/7/1004>
- Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, 223, 96-102.

- Tajfel, H. (1972). Experiments in a vacuum. In J. Israel, Amp & H. Tajfel (Eds.), *The context of social psychology* (pp. 69-119). Cambridge: Cambridge University Press.
- Tajfel, H. (1978a). The achievement of group differentiation. In H. Tajfel (Ed.), *Differentiation between social groups: Studies in the social psychology of intergroup relations.*: Oxford, England: Academic Press.
- Tajfel, H. (1978b). *Differentiation between social groups*. London: Academic press.
- Tajfel, H. (1978c). 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. (1978d). Social categorization, social identity and social comparison. In H. Tajfel (Ed.), *Differentiation between social groups: Studies in the social psychology of intergroup relations*. Oxford, England: Academic Press.
- Tajfel, H. (1981a). *Human groups and social categories*. Cambridge, UK: Cambridge University Press.
- Tajfel, H. (1981b). *Human groups and social categories: Studies in social psychology (ch7: Social stereotypes and social groups)*. Cambridge: Cambridge University Press.
- Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, 33(1), 1-39.
- Tajfel, H., & Turner, J. (1979a). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations*. Monterey, CA: Brooks-Cole.
- Tajfel, H., & Turner, J. (1979b). The social psychology of intergroup relations. In W. G. Austin & S. Worchel (Eds.), (pp. 33-47). Monterey, CA: Brooks/Cole.
- Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgement. *British Journal of Psychology*, 54, 101-114.
- Thoman, D., White, P., Yamawaki, N., & Koishi, H. (2008). Variations of gender-math stereotype content affect women's vulnerability to stereotype threat. *Sex Roles*, 58(9), 702-712.
- Thompson, B. (1986). Anova versus regression analysis of ati designs: An empirical investigation. *Educational and Psychological Measurement*, 46(4), 917-928.
- Todd Maddox, W., & Markman, A. B. (2010). The motivation-cognition interface in learning and decision making. *Current Directions in Psychological Science*, 19(2), 106-110.
- Tomprowski, P. D., Simpson, R. G., & Hager, L. (1993). Method of recruiting subjects and performance on cognitive tests. *The American Journal of Psychology*, 106(4), pp. 499-521.
- Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *The American Journal of Psychology*, 9(4), 507-533.
- Twamley, E. E., & Tajfel, H. (2009). *The role of gender identity on the effects of stereotype threat: An examination of girls' math performance in a single-sex classroom experiments in a vacuum*. Macalester College  
Cambridge: Cambridge University Press. Retrieved from [http://digitalcommons.macalester.edu/psychology\\_honors/15](http://digitalcommons.macalester.edu/psychology_honors/15)
- UNITE. (2006). The international student experience report 2006, from <http://www.unite-group.co.uk/Attachments/000171/International%20Student%20Experience%20Report%202006.pdf>
- Vehkalahti, K., Puntanen, S., & Tarkkonen, L. (2006). *Estimation of reliability: A better alternative for cronbach's alpha*. Helsinki: Department of Mathematics and Statistics, University of Helsinki.

- Vehkalahti, K., Puntanen, S., & Tarkkonen, L. (2009). Implications of dimensionality on measurement reliability. In B. Schipp & W. Krämer (Eds.), *Statistical inference, econometric analysis and matrix algebra* (pp. 143-160-160): Physica-Verlag HD.
- Verkuyten, M., & Nekuee, S. (1999). Ingroup bias : The effect of self-stereotyping, identification and group threat. *European Journal of Social Psychology*, 29(2-3), 411-418.
- von Hippel, C., Issa, M., Ma, R., & Stokes, A. (2010). Stereotype threat: Antecedents and consequences for working women. *European Journal of Social Psychology*.
- von Hippel, W., von Hippel, C., Conway, L., Preacher, K. J., Schooler, J. W., & Radvansky, G. A. (2005). Coping with stereotype threat: Denial as an impression management strategy. *Journal of Personality and Social Psychology*, 89(1), 22-35.
- Voss, K. E., Stem, D. E., & Fotopoulos, S. (2000). A comment on the relationship between coefficient alpha and scale characteristics. *Marketing Letters*, 11, 177-191.
- Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, 39(5), 456-467.
- Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology*, 92(1), 82-96.
- Walton, G. M., & Spencer, S. J. (2009). Latent ability. *Psychological Science*, 20(9), 1132-1139.
- Wax, A. L. (2008). Stereotype threat: A case of overclaim syndrome? *University of Pennsylvania Law School Public Law Research Paper No. 08-14*.
- Weber, U., Mummendey, A., & Waldzus, S. (2002). Perceived legitimacy of intergroup status differences: Its prediction by relative ingroup prototypicality. *European Journal of Social Psychology*, 32(4), 449-470.
- Wei, T. E. (2009a). *Stereotype threat, gender, and math performance: Evidence from the national assessment of educational progress*. Harvard University. Cambridge, MA. Retrieved from [http://www.people.fas.harvard.edu/~twei/papers/sthreat\\_naep.pdf](http://www.people.fas.harvard.edu/~twei/papers/sthreat_naep.pdf)
- Wei, T. E. (2009b). *Under what conditions? Stereotype threat and prime attributes*. Harvard University. Cambridge, MA. Retrieved from [http://www.people.fas.harvard.edu/~twei/papers/sthreat\\_exper.pdf](http://www.people.fas.harvard.edu/~twei/papers/sthreat_exper.pdf)
- Weinberg, S. L., & Abramowitz, S. K. (2008). *Statistics using spss: An integrative approach*: Cambridge Univ Pr.
- West, S. G., Aiken, L. S., & Krull, J. L. (1996). Experimental personality designs: Analyzing categorical by continuous variable interactions. *Journal of Personality*, 64(1), 1-48.
- Wheeler, S., & Petty, R.-E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin*, Vol 127 (6), 797-826.
- Wheeler, S. C., & DeMarree, K. G. (2009). Multiple mechanisms of prime-to-behavior effects. *Social and Personality Psychology Compass*, 3(4), 566-581.
- Wicherts, J. M. (2005). Stereotype threat research and the assumptions underlying analysis of covariance. *American Psychologist*, 60(3), 267-269.
- Wicherts, J. M., & de Haan, C. (2009). *Stereotype threat and the cognitive test performance of African Americans*. Paper presented at the Program: 10th Annual Conference of the International Society for Intelligence Research, Madrid, Spain. <http://www.isironline.org/meeting/pdfs/program2009.pdf#page=68>
- Wicherts, J. M., Dolan, C. V., Carlson, J. S., & van der Maas, H. L. J. (2010). Raven's test performance of sub-Saharan Africans: Average performance, psychometric properties, and the Flynn effect. *Learning and Individual Differences, In Press, Corrected Proof*.

- Williams, K. B. (2006). The effects of stereotype threat on test performance of male and female college students. *College Student Journal, 40*(3), 679-684.
- Willis, P. E. (1981). *Learning to labor: How working class kids get working class jobs*. New York: Columbia University Press.
- Wolfe, C. T., & Spencer, S. J. (1996). Stereotypes and prejudice: Their overt and subtle influence in the classroom. (multiculturalism and diversity in higher education). *American Behavioral Scientist, 40*(2), 176-186.
- 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*(3), 792-799.
- Wraga, M., Duncan, L., Jacobs, E. C., Helt, M., & Church, J. (2006). Stereotype susceptibility narrows the gender gap in imagined self-rotation performance. *Psychonomic Bulletin & Review, 13*(5), 813-819.
- Wraga, M., Helt, M., Jacobs, E., & Sullivan, K. (2007). Neural basis of stereotype-induced shifts in women's mental rotation performance. *Social Cognitive and Affective Neuroscience, 2*(1), 12-19.
- Yeung, N. C. J., & von Hippel, C. (2008). Stereotype threat increases the likelihood that female drivers in a simulator run over jaywalkers. *Accident Analysis & Prevention, 40*(2), 667-674.
- Yopyk, D. J. A., & Prentice, D. A. (2005). Am i an athlete or a student? Identity salience and stereotype threat in student athletes. *Basic and Applied Social Psychology, 27*(4), 329-336.
- Zajonc, R. B. (1965). Social facilitation. *Science, 149*(3681), 269-274.

## Appendix A. Shipley Institute of Living Scale (SILS)

NAME .....

Complete the following. Each dash (—) calls for either a number or a letter to be filled in. Every line is a separate item. Take the items in order, but don't spend too much time on any one.

start here

- (1) 1 2 3 4 5 —
- (2) white black short long down — —
- (3) AB BC CD D —
- (4) Z Y X W V U —
- (5) 1 2 3 2 1 2 3 4 3 2 3 4 5 4 3 4 5 6 — —
- (6) NE/SW SE/NW E/W N/—
- (7) escape scape cape — — —
- (8) oh ho rat tar mood — — — —
- (9) A Z B Y C X D —
- (10) tot tot bard drab 537 — — —
- (11) mist is wasp as pint in tone — —
- (12) 57326 73265 32657 26573 — — — — —
- (13) knit in spud up both to stay — —
- (14) Scotland landscape scapegoat — — — — ee
- (15) surgeon 1234567 snore 17635 rogue — — — — —
- (16) tam tan rib rid rat raw hip — — —
- (17) tar pitch throw saloon bar rod fee tip end plank — — — — — meals
- (18) 3124 82 73 154 46 13—
- (19) lag leg pen pin big bog rob — — —
- (20) two w four r one o three—