

The Effects of Stereotype Threat on Females' Mathematical Performance: A Multi-Faceted Situational Phenomenon?

Charlotte Rebecca Pennington
Faculty of Arts & Sciences,
Department of Psychology,
Edge Hill University

This thesis is submitted to the Department of Psychology, Edge Hill
University, in partial fulfilment of the requirements for the degree of

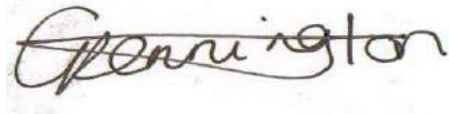
DOCTOR OF PHILOSOPHY

August 2016

I. Declaration

I declare that this thesis is my own work carried out under the normal terms of supervision. I confirm that this work has not been submitted for any comparable academic award.

Signed:

A handwritten signature in black ink, appearing to read "E. Pennington". The signature is written in a cursive style with a large initial "E" and a long, sweeping underline.

II. Acknowledgements

My sincere thanks and gratitude go to my director of studies, Professor Derek Heim. Over the past three years, you have taught me about both academic life, and life more generally, and that happiness and health are the most important things. You have been a fantastic supervisor and a true friend. I would also like to thank Dr Andrew Levy and Dr Derek Larkin for your supervision and guidance throughout this project. Whether it has been statistical guidance or chatting over a coffee, you have been instrumental in developing this final thesis. To (doctors to be) Marianne and Rowan and Drs Jin Zhou, Rebecca Monk, Adam Qureshi, Linda Kaye, Damien Litchfield and Tom Mitchell; you have helped me throughout the past three years in many ways, and I want to thank you for being such great colleagues and lifelong friends. Linda, most of this thesis would not have been written in a timely manner without our residential writing days filled with food, and fun (and bunny snuggles)!

Thank you to my beautiful friends: Lauren, Kelsie, Sarah, Sarah, Katy, Rebecca, Chelsea and Lois for providing welcome distractions and offering different perspectives on things when the 'going gets tough'. To all my friends at *PsyPAG*, we have made memories that I will cherish for the rest of my life.

Finally, thank you to my Mum, Dad, Simon, and the rest of my family. I could not have completed this without your continued love and support; we have been on this journey together.

III. Publications from Thesis

Chapter 2

Pennington, C. R., Heim, D., Levy, A., & Larkin, D. (2016). Twenty years of stereotype threat research: A review of psychological mediators. *PLoS One*, *11*(1). doi: 10.1371/journal.pone.0146487

Chapter 7

Pennington, C. R., & Heim, D. (2016). Creating a critical mass eliminates the effects of stereotype threat on women's mathematical performance. *British Journal of Educational Psychology*, *86*, 353-368. doi: 10.1111/bjep.12110

IV. Conference Presentations

Chapter 2

Pennington, C. R. (2015). *Stereotype Threat and Test Performance: Appraising Critically Two Decades of Research*. Paper presented at the Cutting Edge Conference, Edge Hill University.

Chapter 4

Pennington, C. R. (2015). *Is diminished working memory accountable for the distinct forms of stereotype threat on women's mathematical performance?* Paper presented at the Annual PsyPAG Conference (30th Anniversary), University of Glasgow.

Chapter 6

Pennington, C. R., Qureshi, A., & Heim, D. (2016). *A multi-threat approach to uncovering the mediators of stereotype threat: The role of updating, shifting and inhibition*. Poster presented at the 28th Association for Psychological Science Annual Convention (APS), Chicago, USA.

Chapter 7

Pennington, C. R. (2015). *Enhanced social identity as a stereotype threat intervention?* Paper presented at the British Psychological Society Social Section Conference, Palace Hotel Manchester.

V. Table of Contents

I. Declaration	ii
II. Acknowledgements	iii
IV. Conference Presentations	v
VI. Index of Tables	x
VII. Index of Figures	xi
VIII. Abstract	xii
1. CHAPTER 1 – Introduction	1
1.1. Females in Mathematics	2
1.1.1. Explanations for Gender Differences in Mathematics	3
1.2. An Overview of Stereotype Threat Theory	6
1.3. Research Premise	15
1.4. Overview of Thesis	17
2. CHAPTER 2 – Systematic Literature Review	20
Abstract	21
2.1. Introduction	22
2.1.1. Susceptibility to Stereotype Threat	23
2.1.2. Mediators of Stereotype Threat	24
2.1.3. Objectives of the Review	26
2.2. Methodology	26
2.2.1. Literature Search	26
2.2.3. Categorising Different Stereotype Threats	28
2.2.4. Mediators: Conceptualisations and Definitions	29
2.3. Findings	30
2.3.1. Affective/Subjective Mechanisms	42
2.3.2. Cognitive Mechanisms	49
2.3.3. Motivational Mechanisms	55
2.4. Discussion	61
2.4.1. Chapter Conclusion	65
3. CHAPTER 3 - Theoretical and Methodological Considerations	67
3.1. Chapter Overview	68
3.2. Critiques of the Stereotype Threat Literature	68

3.3.	Current Gaps in the Stereotype Threat Literature	72
3.4.	Ethical Considerations.....	78
3.5.	Measuring Mathematical Performance	79
3.6.	Approach to Data Analysis	81
4.	CHAPTER 4 – Experiment One.....	84
	Abstract.....	85
4.1.	Chapter Overview.....	86
4.2.	Introduction	86
4.3.	Method.....	90
4.3.1.	Participants	90
4.3.2.	Stereotype Threat Manipulations.....	91
4.3.3.	Measures.....	93
4.3.4.	Procedure.....	95
4.4.	Results	95
4.5.	Discussion	100
4.5.1.	Chapter Conclusion	105
5.	CHAPTER 5 – Experiments Two and Three	107
	Abstract.....	108
5.1.	Chapter Overview.....	109
5.2.	Introduction	110
5.3.	Experiment 2 Method.....	115
5.3.1.	Participants	115
5.3.2.	Stereotype Threat Manipulations.....	115
5.3.3.	Measures.....	117
5.3.4.	Procedure.....	118
5.3.5.	Data Preparation.....	119
5.4.	Results	120
5.5.	Discussion	124
5.6.	Experiment 3 Method.....	130
5.6.1.	Participants	130
5.6.2.	Stereotype Threat Manipulations.....	131
5.6.3.	Measures.....	132
5.6.4.	Procedure.....	133
5.6.5.	Data Preparation.....	134
5.7.	Results	134

5.8. Discussion	140
5.9. General Discussion.....	142
5.9.1 Chapter Conclusion	146
6. CHAPTER 6: Experiments Four and Five	149
Abstract.....	150
6.1. Chapter Overview.....	151
6.2. Introduction	152
6.3. Experiment 4 Method.....	156
6.3.1. Participants	156
6.3.2. Stereotype Threat Manipulations.....	156
6.3.3. Measures.....	158
6.3.4. Procedure.....	160
6.4. Results	162
6.5. Discussion	168
6.6. Experiment 5 Method.....	170
6.6.1. Participants	170
6.6.2. Stereotype Threat Manipulation	170
6.6.3. Measures.....	171
6.7. Results	171
6.8. Discussion	176
6.9. General Discussion.....	177
6.9.1. Chapter Conclusion	182
7. CHAPTER 7: Experiment Six.....	184
Abstract.....	185
7.1. Chapter Overview.....	186
7.2. Introduction	186
7.3. Method.....	191
7.3.1. Participants	191
7.3.2. Stereotype Threat Manipulations.....	192
7.3.3. Measures.....	192
7.3.3. Procedure.....	193
7.4. Results	194
7.5. Discussion	203
7.5.1. Chapter Conclusion	207
8. CHAPTER 8 – Thesis Discussion and Conclusions	209

8.1. General Discussion.....	210
8.1.1. Overview of Empirical Studies	211
8.1.2. Thesis Limitations and Research Implications	221
8.1.3. Thesis Conclusions	228
References.....	230
Appendices.....	281
Appendix A: Participant Information Sheet and Consent Form.	281
Appendix B. Debrief Sheet.....	283
Appendix C: Mindset Questionnaire, Experiment 6.	284

VI. Index of Tables

Table 1. <i>Six qualitatively distinct stereotype threats that emerge through the intersection of the target and source of threat. Adapted from Shapiro and Neuberg (2007, p. 113).</i>	13
Table 2. <i>Number of articles excluded in full text review, with reasons.</i>	28
Table 3. <i>Summary of stereotype threat literature examining mediational variables with key methodologies and findings.</i>	32
Table 4. <i>Experiment 2 predictions for performance on anti-saccade trials based on the mere effort and working memory accounts of stereotype threat.</i>	114
Table 5. <i>Means and corresponding standard deviations for stereotype threat manipulation checks in Experiment 2.</i>	121
Table 6. <i>Descriptive statistics for anti-saccade trials as a function of experimental condition in Experiment 2.</i>	123
Table 7. <i>Experiment 3 predictions for performance on anti-saccade trials.</i> ..	130
Table 8. <i>Means and corresponding standard deviations for stereotype threat manipulation checks in Experiment 3.</i>	136
Table 9. <i>Descriptive statistics for anti-saccade trials as a function of experimental condition in Experiment 3.</i>	138
Table 10. <i>Mean accuracy scores and corresponding standard deviations as a function of experimental condition and problem demand in Experiment 3.</i>	140
Table 11. <i>Descriptive statistics for self-report measures in Experiment 4.</i>	164
Table 12. <i>Mathematical performance and executive functioning as a function of experimental condition in Experiment 4.</i>	166
Table 13. <i>Descriptive statistics for self-report measures in Experiment 5.</i>	172
Table 14. <i>Mathematical performance and executive functioning as a function of experimental condition in Experiment 5.</i>	174
Table 15. <i>Mean arithmetic accuracy scores and corresponding deviations as a function of experimental condition, group composition, and problem presentation.</i>	200

VII. Index of Figures

<i>Figure 1.</i> Process of article inclusion (following PRISMA guidelines).....	31
<i>Figure 2.</i> Simple mediation model.....	83
<i>Figure 3.</i> Example of a horizontal and vertical high demand problem, adapted from Beilock et al. (2007).	94
<i>Figure 4.</i> Mean modular arithmetic accuracy scores (%) as a function of problem demand and orientation. Error bars represent standard deviations.....	97
<i>Figure 5.</i> Mean modular arithmetic accuracy scores (%) as a function of experimental condition and problem orientation. Error bars represent standard deviations.....	99
<i>Figure 6.</i> Mediator model: Relationship between group-as-target stereotype threat on mathematical performance through inhibition, shifting and updating. Note: * $p < .05$, ** $p < .001$	167
<i>Figure 7.</i> Mediator model: Relationship between combined stereotype threat on mathematical performance through inhibition, shifting and updating. Note: ** $p < .01$, * $p < .05$	176
<i>Figure 8.</i> Mean modular arithmetic accuracy scores (%) as a function of experimental condition, problem demand and presentation.	197
<i>Figure 9.</i> Mean modular arithmetic accuracy scores (%) as a function of experimental condition, group composition and problem presentation.	200
<i>Figure 10.</i> Mediator model: Relationship between group composition on mathematical performance through mindset. Note: * $p < .05$,** $p < .001$	202

VIII. Abstract

Background: Over the past two decades there has been an upsurge of research documenting the deleterious effects that stereotype threat exerts on females' mathematical performance. However, there is still some debate regarding the mechanisms that underpin this situational phenomenon. The current thesis argues that one reason that may have precluded finding firm evidence of mediation is the recognition of distinct stereotype threats. **Aims:**

Underpinned by social identity theory, the current thesis examines experimentally whether self-as-target and group-as-target stereotype threat influence females' mathematical performance. It aims to elucidate further whether deficits in working memory or heightened motivation mediate the stereotype threat-performance relationship. **Method: Experiment 1** – Female

participants were primed with a negative self-as-target or group-as-target stereotype and completed a modular arithmetic test to provide an initial investigation of the working memory interference account. **Experiment 2** –

Female participants were primed with a negative self- or group-relevant stereotype and completed an anti-saccade eye-tracking task to pit the working memory interference account against the mere effort motivational account.

Experiment 3 – Both the anti-saccade and modular arithmetic tasks were employed to examine whether a positive group stereotype motivated female participants to perform well or led them to 'choke under pressure'.

Experiment 4 & 5 – Female participants completed an updating, shifting and inhibition task under self-as-target, group-as-target or 'combined' stereotype threat conditions to examine whether these primes reduce general executive functioning. **Experiment 6** – Female participants were tested alone or in

groups to explore whether heightened social identity would act as a protective factor to augment their mathematical performance from self-as-target and group-as-target stereotype threat. It also examined whether stereotype threat and the group composition of the testing context influenced a fixed-ability mindset. **Results:** In line with a working memory interference account,

females who were primed with both a self-as-target and group-as-target stereotype underperformed on problems that were presented horizontally relative to vertically. Self-as-target stereotype threat appeared to have a

greater negative effect on overall performance (Experiment 1). However, these primes did not appear to influence performance on visuospatial tasks (Experiments 2 & 3). The salience of a positive group stereotype impeded females' performance on difficult maths problems consistent with theories on 'choking under pressure' (Experiment 3). Females showed reduced updating ability when they were primed concurrently with a self- and group-relevant stereotype prime, with this mediating the stereotype threat-performance relationship. This effect was not observed under conditions in which a task was deemed as solely diagnostic of personal or gender-related ability (Experiments 4 & 5). Finally, females solved more mathematical problems when they completed a maths test in single-sex groups relative to alone, suggesting that heightened in-group representation may serve to reduce stereotype threat effects. However, participants in single-sex groups appeared to endorse a weaker growth mind-set compared to those tested alone.

Conclusion: Taken together, findings suggest that females may be more susceptible to stereotype threat when both their personal and social identities are made salient in the stereotyped domain. In such situations, stereotype threat appears to diminish verbal working memory resources to bring about decrements in mathematical performance. **Original Contribution:** The empirical research presented in this thesis represents the first to examine the mechanisms that underpin the effects of different stereotype threats on females' mathematical performance.

1. CHAPTER 1 – Introduction

1.1. Females in Mathematics

The gender gap in interest, participation and performance in mathematics is well documented and hotly debated (Nosek et al., 2009; Shibley-Hyde, 2014). Whilst females¹ typically outperform males across the majority of school subjects (Mullholland, Hansen, & Kaminski, 2004; Voyer & Voyer, 2014), international comparisons reveal that males continue to achieve higher grades in mathematics in many nations (Benbow, 1988; National Science Foundation [NSF], 2013; OECD, 2015; Reilly, Neumann, & Andrews, 2014; Stoet & Geary, 2013). However, it is important to note that other research suggests that gender differences in mathematical aptitude may be small and limited to high achieving students (Ceci & Williams, 2010; Else-Quest, Shibley Hyde, & Linn, 2010; Halpern et al., 2007; Hedges & Nowell, 1995; Wai, Cacchio, Putallaz, & Makel, 2010).

Early differences in mathematics achievement have also been suggested to influence gender-maths attitudes and shape future career aspirations (Ceci, Williams, & Barnett, 2009; Hill, Corbett, & St Rose, 2010; Shapiro & Williams, 2012). In both the United Kingdom and United States of America, females represent only 25% of doctoral degree holders in mathematics (London Mathematics Society, 2013; NSF, 2013) and are less likely to enter maths-intensive careers, constituting below 20% of mathematics university faculties and 6% of Professorships (LMS, 2013; NSF, 2008; US

¹ The terms “female/male” and “women/men” are used interchangeably in this thesis. “Female/male” is appropriate when the age range is broad or ambiguous (American Psychological Association, 2010).

Department of Education, 2015). The attrition rates of females from mathematics fields are disproportionately higher compared to that of males (Beasley & Fischer, 2012), with this phenomenon being referred to as the 'leaky pipeline' (Clark-Blickenstaff, 2005; Wickware, 1997). The statistics are also disheartening for females who do enter and excel in STEM-related fields. For example, under 3% of Nobel laureates in Science are female, and only one female has received one of the top three awards in Mathematics (the Fields Medal, the Abel Prize and the Wolf Prize; Stoet & Geary, 2013). The paucity of women in mathematics is therefore a pressing societal concern (Murphy, Steele, & Gross, 2007), and has led scientists to investigate possible explanations for their underperformance in this domain.

1.1.1. Explanations for Gender Differences in Mathematics

A common and long held belief is that biological factors contribute to females' lower levels of mathematical achievement (Benbow & Stanley, 1980; 1983; Geary, 1996; Kimura, 1999; Simpkins, Davis-Kean, & Eccles, 2005; Spelke, 2005). Evolutionary theories propose that sexual selection has directly influenced mathematical development and performance (Geary, 1996). For instance, research suggests that males have developed greater visuospatial skills because these were required for successful navigation and hunting (Buss, 1995; Geary, 1995; Geary & DeSoto, 2001). On the other hand, females tend to value social relationships more than males and this may lead them to favour subjects and careers which are people oriented (Geary, 1996; 1998; 1999; c.f., also Su, Rounds, & Armstrong, 2009). These "biological traits" are seen to lay the foundation for gender differences in mathematical ability through a variety

of mechanisms including differing social roles and sex typing on children's play activities (Caplan & Caplan, 1994; Geary, 1996; 2010).

Research also suggests that prenatal exposure to sex hormones and an increase in their production during puberty may explain group differences between females and males' mathematical achievement (Collaer, Reimers, & Manning, 2007; Collins & Kimura, 1997; Geary, 2010; Kimura & Hampson, 1994). Research by Kimura and Hampson (1994) indicates that high levels of ovarian hormones (estradiol) may influence adversely quantitative reasoning ability. On the surface, there is intuitive appeal to this explanation when considering that changes in hormone production during puberty coincide with the widening of the gender-achievement gap in mathematics (Reilly et al., 2014). However, the gender-maths achievement gap has also been associated with cross-national indicators of gender equality (Else-Quest et al., 2010; Guiso, Monte, Sapienza, & Zingales, 2008), with data indicating that this gap is narrower, and sometimes disappears entirely, in more gender-equal societies (Else-Quest et al. 2010; Guiso et al., 2008). Other research, however, appears to show no association between endogenous hormone levels and performance (Halari et al., 2005; Puts et al., 2010; c.f., also Reilly et al., 2014 for review), suggesting that innate sex differences may not be the root cause of observed variations in mathematical ability.

A number of social forces, such as teacher and parental expectations of gender-subject competence and socialisation have also been suggested to underpin females' underachievement in mathematics (Bem & Lewis, 1975; Bleeker & Jacobs, 2004; Eccles, Jacobs, & Harold, 1990; Simpkins, Davis-Kean, & Eccles, 2005). In a similar vein, it has been argued that the gender

stereotype pertaining to females' perceived lower ability in mathematics might have a particularly important influence on actual performance (Spencer, Steele, & Quinn, 1999; Walton & Spencer, 2009). In support of this, research indicates that children in primary school endorse negative gender-maths stereotypes on both explicit self-report and implicit measures (Cvencek, Meltzoff, & Greenwald, 2011). Moreover, knowledge of negative gender-maths stereotypes seem to consolidate further during middle school, with research suggesting that this may have a negative influence on females' mathematical performance in school settings (Huguet & Régner, 2007).

Negative stereotypes pertaining to females' performance in mathematics appear to be pervasive, as is evidenced by the way in which multinational companies use stereotypes to market their products. For example, a public outcry led the U.S. toy company Mattel to recall a '*TeenTalk*' Barbie™ doll from the market because it said "Math is hard" (Ben-Zeev et al., 2005; Ben-Zeev, Fein, & Inzlicht, 2005). In 2015, the charity National Numeracy filed a complaint to the company L'Oreal, the French manufacturer of hair products, who featured an advert of a woman stating "Age is just a number. And maths was never my thing", asserting that this may perpetuate negative gender-maths stereotypes. The negative stereotype surrounding females' achievement and participation in mathematics thus appears to be sufficiently ingrained in Western societies and educational systems, and researchers have argued that these beliefs may primarily contribute to the gender-mathematics achievement gap (Cohen, Garcia, Apfel, & Master, 2006; Gunderson, Ramirez, Levine, & Beilock, 2012; Steele & Aronson, 1995; Steele, 1997).

1.2. An Overview of Stereotype Threat Theory

Coined by Steele and Aronson (1995), stereotype threat is a situational predicament in which members of a negatively evaluated group underperform in stereotype-salient testing environments. The main thrust of this work highlights that making individuals aware, either explicitly or implicitly, of societal stereotypes regarding their devalued group membership impairs their performance in the stereotyped domain (Nguyen & Ryan, 2008; Steele, 1997; Steele, Spencer, & Aronson, 2002). In their seminal studies, Steele and Aronson (1995) found that African American's intellectual proficiency was diminished when they perceived a verbal ability test to be indicative of race-related ability. However, African Americans performed equivalently to their Caucasian peers when the same test was presented as non-diagnostic of ability. Extending these findings, Spencer et al. (1999) found that women underperformed when they perceived a test to be confirmative of gender differences in mathematical aptitude (after controlling for pre-existing mathematical ability). Yet, they performed similarly to men when the negative gender-maths stereotype was dismissed prior to the test. These findings suggest that the mathematical ability of males and females may be relatively equal, except under circumstances where females' performance is hindered by situational cues (Steele, 1997).

Stereotype threat has been used predominantly to explain the chronic gaps in intellectual test scores between African and European Americans (Steele & Aronson, 1995; Steele, 1997) and males and females on quantitative portions of standardised tests (Spencer et al., 1999). Nevertheless, research also suggests that it is not limited to such social groups who routinely face

stigmatising attitudes. According to Steele and Aronson (1995), stereotype threat can befall anyone who is a member of a group to which a negative stereotype applies. In support of this contention, research indicates that Caucasian men, a group that typically experience relatively advantageous social statuses, underperform when they believe that their mathematical performance will be compared against that of Asian men's (Aronson et al., 1999). White men also appear to perform comparatively worse to black men when a motor task is linked ostensibly to natural athletic ability (Stone, 2002; Stone, Lynch, Sjomeling, & Darley, 1999). The theory of stereotype threat therefore suggests that individuals may be likely to underperform in testing contexts, not solely because of established factors such as poverty, socialisation or parental style (c.f., Steele, 1997), but also as a result of situational factors, such as the pervasive stereotypes that are associated with their group membership (Nguyen & Ryan, 2008). In turn, the societal attitudes held about a particular group may shape the behaviour of individual group members in a way that imperils their intellectual functioning and reinforces the stereotype further (Steele, 1997).

1.2.1. A multi-faceted situational phenomenon? Previous research has typically conceptualised stereotype threat as a singular construct, experienced similarly by targeted group members (Shapiro, 2011; Shapiro, Williams, & Hambarchyan, 2013). However, a closer look at the literature reveals that researchers have utilised diverse definitions and manipulations to elicit stereotype threat that appear to be fundamentally distinct (Shapiro & Neuberg, 2007). Steele and Aronson's (1995) original definition refers to

stereotype threat as “being at risk of confirming, as self-characteristic, a negative stereotype about one’s group” (p. 797), and therefore emphasises both the role of the self and the social group. More recent definitions have deviated from this viewpoint. For example, the majority of research has focused solely on stereotype threat as a form of social identity threat; concerns that stereotype-relevant performance will reflect adversely on the abilities of one’s group (e.g., Aronson & Inzlicht, 2004; Bosson, Haymovitz, & Pinel, 2004). Specifically, stereotype threat has been suggested to occur when “one could be seen as confirming a negative social stereotype about their in-group” (Schmader & Johns, 2003, p. 440). Other research has focused predominantly on features of the self, suggesting that stereotype threat arises when individuals apprehend that stereotype-relevant performance may be self-characteristic and a threat to self-integrity (Croizet & Claire, 1998; Kray, Thompson, & Galinsky, 2001).

These disparate definitions may underscore particularly meaningful differences in how stereotype threat operates and suggests that individuals may experience multiple, distinct forms of stereotype threat (Shapiro & Neuberg, 2007). One meaningful difference that can be observed in these varying definitions regards whether the self or the social group is the target of negative performance implications. With these distinctions in mind, the current thesis takes a social identity approach to examine situations in which the salience of a female’s devalued personal or social identity may result in mathematical performance decrements.

1.2.2. A social identity approach. Social identity theory (SIT; Tajfel, 1981; Tajfel & Turner, 1979; 1986) and self-categorisation theory (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987) posit that individuals have two sources of identity; a personal identity which defines them as idiosyncratic individuals, and a social identity which is derived from the social groups with whom they identify (Crocker & Luhtanen, 1990; Hornsey, 2008). Whereas personal identity refers to characteristics of the individual, such as *competence* or *extravert*, social identity refers to the multiple social groups with which people categorise themselves, such as *female*, *British* or *Democrat* (Crisp & Hewstone, 2007; Prati, Crisp, Meleady, & Rubini, 2016; Swann, Gómez, Conor-Seyle, Morales, & Huici, 2009). An individual's social identity serves as a reference point that enables them to compare similarities and differences with other "ingroup" relative to "outgroup" members (Turner, Oakes, Haslam, & McGarty, 1994). If an individual evaluates their role in a social group as positive, then this serves to heighten self-esteem (Hogg & Abrams, 1988; Hoelter, 1986; Stryker, 1980; Jetten et al., 2015; Schmader, 2002) and if an individual performs well in their social role, they may feel good in view of the perceived positive appraisals from others (Franks & Marolla, 1976). This quest for positive distinctiveness means that people's sense of who they are is defined in terms of their social identity, specifically in terms of 'we' rather than 'I' (Ellemers, Gilder, & Haslam, 2003; Hornsey, 2008; Tajfel & Turner, 1986). When applying a social identity approach to stereotype threat research, however, it may be questioned how people contend with a devalued social identity that does not serve to enhance self-esteem (Rydell, McConnell, & Beilock, 2009).

Tajfel (1969) reasoned that individuals come to define their sense of 'self' in terms of their group membership, and this can explain why allocation to ostensibly meaningless groups may influence behaviour. Individuals determine the superiority or inferiority of their own social group by comparing it to others (McGarty, Yzerbyt, & Spears, 2002). Distinguishing between ingroup and outgroup members allows individuals to evaluate the benefits of belonging to their social group, and helps define their place in society (Tajfel, 1981). It is also seen to help people interpret, explain, and justify their behaviour based on the characteristics of their social group (Tajfel, 1981). As such, an individual's social identity contributes to the perception of a socially structured 'self'. However, based on group categorisation, differences between the ingroup (e.g., female) and outgroup (e.g., male) can become accentuated, and intricate differences between members of the same social category (e.g., other females) may be overlooked. This can lead to the process of stereotyping in that between-group differences are perceived as large whereas within-group differences are perceived as small (Ashmore & Del Boca, 1981; Rosenthal & Crisp, 2006; 2007; McCauley, Stitt, & Segal, 1980; Tajfel, 1981). This "meta-contrast" influences a range of behaviour, such as prejudice and discrimination (Hall, Crisp, & Suen, 2009), both of which may be experienced in stereotype-salient environments (McGarty et al., 2002).

Some researchers propose that stereotype threat occurs when an individual's positive self-concept is inconsistent with the expectation that their social group should underperform in an ability domain (Rydell et al., 2009; Schmader, Johns, & Forbes, 2008). For example, most females view themselves as competent, capable and able to achieve. Nonetheless, there is

a pejorative stereotype that females have lower mathematical ability compared to males (Rydell et al., 2009). These contradictory propositions regarding the concepts of the self (i.e., I am competent, capable and able), the group (i.e., I am a female) and the ability domain (i.e., females have less mathematical aptitude) are seen to set the stage for stereotype threat because one cannot be both female and good at mathematics (Nosek, Banaji, & Greenwald, 2002; Schmader et al., 2008).

Many studies support the contribution of social identity theory to stereotype threat theory. For example, research indicates that focusing on differences between males and females can exacerbate the effects of stereotype threat on performance, whereas priming similarities between groups may alleviate these performance decrements (Crisp & Abrams, 2009; Rosenthal & Crisp, 2006; 2007). In a similar vein, women have been found to be less susceptible to stereotype threat when they are tested in same-sex relative to mixed-sex groups, suggesting that heightened ingroup representation serves to bolster their mathematical performance (Huguet & Régner, 2007; Inzlicht & Ben-Zeev, 2000).

However, SIT also suggests that self-categorisation is extremely fluid, allowing individuals to shift self-perceptions from a personal to social identity perspective dependent on contextual cues (Turner et al., 1994). Capitalising on this, researchers have proposed that individuals may be vulnerable to experiencing distinct forms of stereotype threat, which target either the self or the social group (Shapiro & Neuberg, 2007; Shapiro et al., 2013; Wout, Danso, Jackson, & Spencer, 2008). In other words, individuals may experience threats to either their personal or social identity dependent on which aspect of their

identity is more pronounced in the stereotyped domain (Schmader et al., 2008; Shapiro & Neuberg, 2007; Shapiro, et al., 2013; Wout et al., 2008). In a stereotype-salient environment, an individual's personal identity may be threatened when they compare unfavourably to other individuals (Schmader et al., 2008; Shapiro & Neuberg, 2007). Here, performance deficits may arise from a strong propositional link between the self and the ability domain (Schmader et al., 2008). In contrast, a valued social identity (i.e., being female) may be threatened when an individual's in-group compares unfavourably with the out-group (i.e., males in the domain of mathematics; Schmader, 2002). In such situations, individuals might apprehend that they will confirm the stereotype as a true representation of their social group, with a greater cognitive tension stemming from a strong association between the group and the ability domain (Schmader et al., 2008).

1.2.3. The Multi-Threat Framework. The Multi-Threat Framework (Shapiro & Neuberg, 2007) proposes six qualitatively distinct stereotype threats, which manifest from the intersection of two dimensions: the target of the stereotype (i.e., the salience of one's personal or social identity) and the source of the stereotype (i.e., the evaluative judgment of the self, the ingroup or the outgroup). See Table 1 for an overview.

Table 1.

Six qualitatively distinct stereotype threats that emerge through the intersection of the target and source of threat. Adapted from Shapiro and Neuberg (2007, p. 113).

Target of Stereotype Threat		
Source of Stereotype Threat	Personal Identity (self)	Social Identity (group)
Personal Identity (self)	<p><i>Self-as-target threat</i> <i>Being at risk of confirming, as self-characteristic, a negative stereotype about one's group.</i></p>	<p><i>Group-as-target threat</i> <i>Being at risk of confirming, or reinforcing, a negative societal stereotype pertaining to one's group-membership.</i></p>
Outgroup Members	<p><i>Own-reputation Threat (outgroup)</i> <i>Being at risk of confirming, in the minds of other group members, that a negative stereotype is true of personal ability. I will therefore be judged or treated badly by outgroup members.</i></p>	<p><i>Group-Reputation Threat (outgroup)</i> <i>Being at risk of confirming, or reinforcing, in the minds of outgroup members, that a negative societal stereotype is true of one's group membership. My group will therefore be judged or treated badly by outgroup members.</i></p>
Ingroup Members	<p><i>Own-reputation Threat (ingroup)</i> <i>Being at risk of confirming, in the minds of ingroup members, that the negative stereotypes held of my group are true of my personal ability.</i></p>	<p><i>Group-Reputation Threat (ingroup)</i> <i>Being at risk of reinforcing, in the minds of ingroup members, that the negative stereotypes held about one's group membership are confirmative about one's social group.</i></p>

Focusing on the target of stereotype threat, individuals who experience “self-as-target” stereotype threat may perceive that stereotype-consistent performance will be judged as self-characteristic of personal ability (Shapiro & Neuberg, 2007). On the other hand, people who experience “group-as-target” stereotype threat may perceive that underperformance will confirm, and thereby reinforce, a negative societal stereotype held about their ingroup (Schmader et al., 2008; Shapiro & Neuberg, 2007; Zhang, Schmader, & Hall, 2013). These types of stereotype threat are comparable from the perspective that they each result from the predicament of being a member of a devalued social group, and are prompted by the expectation that performance will be judged on the basis of a pejorative stereotype (Shapiro & Neuberg, 2007). However, researchers have theorised that these distinct experiences of stereotype threat may not only result from different eliciting conditions, but also may be moderated and mediated by somewhat different processes (Shapiro & Neuberg, 2007; Shapiro et al., 2013). As such, self-as-target and group-as-target stereotype threats may have a greater or lesser effect on females’ mathematical performance as a function of different factors that heighten their susceptibility to such threats. Furthermore, it is plausible that different mechanisms may underpin the relationship between these distinct stereotype threats and underperformance. It is with this in mind that the current thesis argues that research should recognise and distinguish between different forms of stereotype threat in order to gain a more nuanced understanding of how stereotype threat impacts performance and the mechanisms through which it may operate.

1.3. Research Premise

Considerable empirical support has been accrued for the theory of stereotype threat over the past two decades of research (c.f., Doyle & Voyer, 2016; Nguyen & Ryan, 2008; Pennington, Heim, Levy, & Larkin, 2016; Picho, Rodriguez, & Finnie, 2013 for reviews). However, when reviewing the literature, it becomes apparent that researchers have not utilised a uniform or standardised set of manipulations to evoke stereotype threat, and vary in terms of their theoretical definitions (Shapiro & Neuberg, 2007; Shapiro et al., 2013). Specifically, researchers have employed a wide range of primes that emphasise the self or the social group, yet have conceptualised this methodology as capturing the same underlying phenomenon (Shapiro & Neuberg, 2007). Other researchers have argued that this may oversimplify the stereotype threat process and consequently hinder theory development (Wout et al., 2008). Accordingly, the multi-threat framework (Shapiro & Neuberg, 2007) was developed to suggest that individuals might experience numerous different stereotype threats, which target either the self or the social group to bring about decrements in performance. Despite this framework being discussed in a number of theoretical articles (c.f., Shapiro & Neuberg, 2007; Shapiro, 2012), there exists a limited amount of empirical research that examines the effects of these distinct stereotypes on performance (c.f., Shapiro et al., 2013; Wout et al., 2008 for exceptions). Taking this into consideration, the first aim of the current thesis is to examine whether self-relevant and group-relevant stereotypes exert different (or similar) effects on females' mathematical performance. It therefore aims to make a theoretical contribution to knowledge by answering the following

questions set forth by Aronson et al. (1999), which have not been fully addressed to date:

Is stereotype threat self-threatening because it arouses a fear of being a bad ambassador of one's group to mainstream society? Or is it more simply the apprehension about appearing incompetent – for the sake of one's own reputation? Or, alternatively, is it merely the result of worrying that one might lack ability? Or is it some combination of these concerns? These are important questions that will have to await the results of future research for answers (p. 43).

There is also considerable debate regarding the underlying mechanisms of stereotype threat (c.f., Jamieson & Harkins, 2007; Smith, 2004). As will be uncovered in the following systematic literature review (Chapter 3), some researchers argue that deficits in working memory mediate the stereotype threat-performance relationship (e.g., Beilock, Rydell, & McConnell, 2007; Rydell, Van-Loo, & Boucher, 2014; Schmader & Johns, 2003), whereas others suggest that enhanced motivation accounts for this relationship (Jamieson & Harkins, 2007; 2009; Seitchik & Harkins, 2015). Furthermore, there is mixed empirical support with regard to additional affective, cognitive and motivational mechanisms that are proposed to underpin stereotype threat effects. The current thesis argues that the discrepancies between findings in the current literature may, to an extent, be a product of the heterogeneity of primes utilised to elicit stereotype threat and the different methods used to measure it. In their

theoretical review, Shapiro and Neuberg (2007) suggest that different moderating and mediating mechanisms may underpin self-as-target and group-as-target stereotype threat. However, no research has tested this empirically to date. Taking this into consideration, the second aim of the current research is to elucidate the underlying mechanisms of distinct stereotype threats. A greater understanding of the mediating mechanisms that underpin the effects of self-as-target and group-as-target stereotype threat on performance is an important endeavour, which will help researchers to understand *how* these distinct threats operate and how such effects may be reduced.

1.4. Overview of Thesis

The following chapter (Chapter 2) presents a systematic literature review of the psychological mediators that have been explored within the past twenty years of stereotype threat research. It appraises critically the underlying mechanisms of stereotype threat as a function of the type of threat primed, the population studied, and the measures utilised to examine mediation and performance outcomes. Here, it is proposed that one reason that has precluded studies from finding firm evidence of mediation is the recognition of distinct forms of stereotype threat (i.e., self-relevant and group-relevant stereotype primes). Guided by this, Chapter 3 provides an overview of the research questions asked by the current thesis and identifies methodological considerations with a view to informing the empirical chapters. The ensuing chapters then present six empirical studies. Underpinned by a working memory interference account (Beilock et al., 2007; Schmader & Johns, 2003), Chapter 4 examines whether deficits in verbal working memory may explain the effects of self- and group-

relevant stereotype threat on women's mathematical performance (Experiment 1). Chapter 5 then pits support for the working memory interference theory against the motivational 'mere effort' account of stereotype threat utilising eye-tracking methodology, which is novel within this field of research (Experiment 2). It extends this to examine whether priming a positive group-based stereotype facilitates women's mathematical performance by increasing their motivation or debilitates performance because it interferes with working memory and leads them to 'choke under pressure' (Experiment 3). Chapter 6 distinguishes between sub-components of executive functioning to examine whether updating, inhibition and shifting underpin self-as-target and group-as-target stereotype threat effects (Experiment 4). It then examines whether stereotype threat effects are more likely to emerge when implications for performance are tied to both an individual's personal and social identity (Experiment 5). Up to this point, the current thesis focuses on the debilitating effects of stereotype threat on performance and the mechanisms that may account for this relationship. The final empirical study outlined in Chapter 7 therefore investigates strategies to alleviate stereotype threat effects. Underpinned by a social identity approach, it examines whether testing women in single-sex groups may present as an effective strategy to ameliorate stereotype threat effects by heightening in-group representation (Experiment 6). It also explores the potential negative consequences of gender-segregated classroom environments by investigating whether same-sex testing influences a fixed-ability mindset because females become more cognisant of their gender. The general discussion in Chapter 8 consolidates this empirical research and provides an overall appraisal of whether stereotype threat should

be conceptualised as a singular construct or a multi-faceted situational phenomenon. It also highlights some limitations of the current thesis and discusses some potential avenues for future research.

2. CHAPTER 2 – Systematic Literature Review

Abstract

Aim: In the first of its kind, this systematic literature review appraises critically the mediating variables of stereotype threat proposed to date. **Method:** A bibliographic search was conducted across the electronic databases of PsycINFO, PsycARTICLES, Web of Knowledge, PubMed, Science Direct and Google Scholar between 1995 and 2016. The search identified 45 experiments from 38 articles and 17 unique proposed mediators that were categorised into affective ($n = 6$), cognitive ($n = 7$) and motivational mechanisms ($n = 4$). **Results:** Empirical support was accrued for mediators such as anxiety, negative thinking, and mind-wandering, which are suggested to co-opt working memory resources under stereotype threat. However, other research points to the assertion that stereotype threatened participants may be motivated to disconfirm negative stereotypes, facilitating a dominant approach which resultantly has a paradoxical effect on performance. The review also indicates that stereotype threat appears to impact diverse stigmatised groups in different ways, with no one mediator providing generalised empirical support. **Discussion:** In line with the multi-threat framework, the discussion postulates that the underlying mechanisms of stereotype threat may differ dependent on the primes utilised, the population being studied, and the measures employed to examine mediating variables and performance.

2.1. Introduction

Since the publication of Steele and Aronson's (1995) seminal research, the theory of stereotype threat has become one of the most widely researched topics in Social Psychology (Derks, Inzlicht, & Kang, 2008; Schmader, Johns, & Forbes, 2008). Over the past 20 years, Steele and Aronson's (1995) original article has gathered more than 5,000 citations² and has been referred to as a 'modern classic' (Devine & Brodich, 2003; Fiske, 2003). In stark contrast to theories of genetic intelligence (Benbow & Stanley, 1980; Spelke, 2005; c.f., Sternberg, Grigorenko, & Kidd, 2006 for debate), the theory of stereotype threat offers a situational explanation for the on-going and intractable debate regarding the source of group differences in academic aptitude (Derks et al., 2008).

More than 300 experiments have illustrated the apparent deleterious and extensive effects that stereotype threat can inflict on many different populations (Walton & Spencer, 2009). The possibility of confirming a negative stereotype about one's group has been found to contribute to underperformance on a range of diverse tasks including intelligence (Steele & Aronson, 1995; Pavlova, Weber, Simoes, & Sokolov, 2014), memory (Hess, Auman, Colcombe, & Rahhal, 2003; Levy, 1996), mental rotation (Wraga, Duncan, Jacobs, Helt, & Church, 2006), and mathematical tests (Beilock, et al., 2007; Schuster, Martiny, & Schmader, 2015; Spencer et al., 1999), golf putting (Beilock, Jellison, Rydell, McConnell, & Carr, 2006), driving (Skorich et al., 2013; Yeung & von Hippel,

²Citation reports from Google Scholar™ state that Steele and Aronson's (1995) article has been cited 5,770 times as of August 2016.

2008), inhibitory control (Pennington, Qureshi, Monk, & Heim, 2016), and childcare skills (Bosson, Haymovitz, & Pinel, 2004). Given the generality of these findings, researchers have turned their efforts to investigating the underlying mechanisms of this situational phenomenon.

2.1.1. Susceptibility to Stereotype Threat

Research has identified numerous moderators that make tasks more likely to elicit stereotype threat and individuals more prone to experience it (Shapiro & Neuberg, 2007; c.f., Nguyen & Ryan, 2008; Lamont, Swift, & Abrams, 2015; Picho et al., 2013; for meta-analyses). From a methodological perspective, stereotype threat effects tend to emerge on tasks of high difficulty and demand (Hess, Hinson, & Hodges, 2009; Keller, 2007; Neuville & Croizet, 2007). However, the extent to which a task is perceived as demanding may be moderated by individual differences in working memory (Régner et al., 2010). Additionally, stereotype threat may be more likely to occur when individuals are conscious of the stigma ascribed to their social group (Brown & Pinel, 2003; Hess et al., 2009), believe the stereotypes about their group to be true (Elizaga & Markman, 2008; Schmader, Johns, & Barquissau, 2004), and for those with low self-esteem (Rydell & Boucher, 2010) and an internal locus of control (Cadinu, Maass, Lombardo, & Frigerio, 2006). Research also indicates that individuals are more susceptible to stereotype threat when they identify strongly with their social group (Davies, Aronson, & Salinas, 2006; Kiefer & Sekaquaptewa, 2007a; Marx, Stapel, & Muller, 2005; Schmader, 2002) and value the performance domain (Appel, Kronberger, & Aronson, 2011; Aronson et al., 1999; Keller, 2007; Steele, 1997; Stone et al., 1999). However, other

research suggests that domain identification is not a prerequisite of stereotype threat effects (Keller & Dauenheimer, 2003), and has shown that heightened group membership may serve as a strategy to overcome harmful academic consequences (Oyserman, Harrison, & Bybee, 2001; Oyserman, Kimmelmeier, Fryberg, Brosh, & Hart-Johnson, 2003).

2.1.2. Mediators of Stereotype Threat

Although evidence has been accrued regarding the moderating variables that may influence the strength and direction of stereotype threat effects, research that aims to elucidate the underlying processes which account for the stereotype threat-performance relationship have produced variable results (Jamieson & Harkins, 2007; Schmader et al., 2008; Smith, 2004; Wheeler, Jarvis, & Petty, 2001). A plethora of factors have been proposed to mediate the effects of stereotype threat on performance, however, due to constraints surrounding experimental research, many of these variables have been tested in isolation (Schmader et al., 2008). Researchers have suggested that this may have contributed to the unrealistic expectation that there is a single mediator of stereotype threat (Jamieson & Harkins, 2011a; Schmader et al., 2008; Steele, Spencer, & Aronson, 2002). Overcoming this, Schmader et al. (2008) propose an integrated process model of stereotype threat, suggesting that stereotype threat heightens physiological stress responses and influences monitoring and suppression processes to deplete working memory efficiency. This provides an important contribution to the literature, signalling that multiple affective, cognitive and motivational processes may account for the effects of stereotype

threat on performance. However, the extent to which each of these variables has garnered empirical support remains unclear.

Furthermore, researchers have utilised many different manipulations and methods to instantiate stereotype threat, and it is plausible that this has precluded finding firm evidence of mediation. As discussed in Chapter 1, the majority of research has viewed stereotype threat as a situational predicament that occurs when individuals perceive their social group to be devalued by others (Crocker & Major, 1989; Steele et al., 2002), and have resultantly employed “group-as-target” primes (Shapiro & Neuberg, 2007; Shapiro et al., 2013). This research overlooks how individuals may self-stigmatise and evaluate themselves (Frey & Tropp, 2006; Leary, Terry, Allen, & Tate, 2009; van Veelan, Otten, Cadinu, & Hansen, 2016; Wheeler, Demarree, & Petty, 2007), and the conflict people may experience between their personal and social identities (Hirsh & Kang, 2015).

Researchers have also elicited stereotype threat using direct and indirect priming techniques. For example, stereotype threat has been manipulated experimentally by priming participants explicitly with a negative stereotype regarding their social group (Aronson et al., 1999; Rydell et al., 2014), activating stereotypes implicitly by asking participants to report their group membership before a stereotype-relevant task (Steele & Aronson, 1995; Quinn, Kahng, & Crocker, 2004), and manipulating the group composition of the testing environment (Inzlicht & Ben-Zeev, 2000; Sekaquaptewa & Thompson, 2003). Whilst this speaks to the robustness of this phenomenon, it is plausible that these different methods may influence performance outcomes in different ways, and that diverse experiences of stereotype threat may be

underpinned by independent mechanisms (Nguyen & Ryan, 2008; Stone & McWhinnie, 2008). Nonetheless, to date no research has taken a multi-threat approach in the investigation of mediating variables, and it remains to be assessed whether the same or different mechanisms are responsible for the effects of distinct stereotype threats on performance.

2.1.3. Objectives of the Review

The current systematic literature review distinguishes between different stereotype threat primes that target either the self or the social group to provide evidence for the existence of multiple stereotype threats that may be evoked through different pathways and accounted for by distinct mechanisms. Specifically, the purpose of this review is threefold: 1), to identify and critically examine the proposed mediators of stereotype threat; 2), to evaluate whether different mediators govern different stereotyped populations; and 3), to explore whether the effects of self-as-target or group-as-target stereotype threat on performance are the result of qualitatively distinct mediating mechanisms.

2.2. Methodology

2.2.1. Literature Search

A bibliographic search of electronic databases (PsycINFO, PsycARTICLES, Web of Knowledge, PubMed, Science Direct and Google Scholar) was conducted between the cut-off dates of 1995 (the publication year of Steele & Aronson's seminal article) and July 2016. A search string was developed by specifying the main terms of the phenomenon under investigation. Here, the

combined key words of *stereotype* and *threat* were utilised as overarching search parameters and directly paired with either one of the following terms; *mediator*, *mediating*, *mediate(s)*, *predictor*, *predicts*, *relationship* or *mechanism(s)*. Identification of relevant articles and data extraction were conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Statement (PRISMA; Moher, Liberato, Tetzlaff, & Altman, 2009). A literature search was conducted separately in each article database and records were exported to citation software, after which duplicates were removed. A call for any unpublished or 'in press' articles was sent to the European Association for Social Psychology to control for potential publication bias (Dickersin, 2005; Dickersin, Min, & Meinert, 1992; Rosenthal, 1979). Additional articles were also retrieved by reviewing the reference lists of relevant journal articles. Relevant articles were then screened by examining the title and abstract in line with eligibility criteria, after which a full text review was performed on all remaining articles (Cronin, Ryan, & Boughlan, 2008; Khan, Riet, Popay, Nixon, & Kleijnen, 2009).

2.2.2. Eligibility criteria. Studies were selected based on the following criteria: 1), researchers utilised a stereotype threat manipulation; 2), a direct mediation analysis was conducted between stereotype threat and performance; 3), researchers found evidence of moderated-mediation, and 4), the full text was available in English. Articles were excluded on the following basis: 1), performance was not the dependent variable, 2), investigations of "stereotype lift"; 3), doctorate, dissertation and review articles (to avoid duplication of included articles); and 4), moderating variables. Articles that did not find any

significant results in relation to stereotype threat effects were also excluded in order to capture reliable evidence of mediation between stereotype threat and underperformance. See Table 2 for details of excluded articles.

Table 2.

Number of articles excluded in full text review, with reasons.

Reason for exclusion	Number of articles	Percentage (%)
No direct mediation analysis	25	58.14%
No ST effects found	5	11.63%
Review papers	4	9.30%
Did not prime ST	3	6.98%
Moderators of ST	3	6.98%
No performance measure	2	4.65%
Performance not standardised	1	2.33%

2.2.3. Categorising Different Stereotype Threats

The current review distinguished between self-relevant and group-relevant stereotype primes by examining each stereotype threat manipulation. In accordance with previous research (Shapiro & Neuberg, 2007; Shapiro et al., 2013; Wout et al., 2008), self-as-target stereotype threats were categorised on the basis that participants focused on the test as a measure of personal ability.

Group-as-target threats were classified on the basis that participants perceived performance to be diagnostic of their group's ability.

2.2.4. Mediators: Conceptualisations and Definitions

Effect sizes for mediational findings are described typically through informal descriptors, such as *complete*, *perfect*, or *partial* (Hayes, 2013; Preacher & Kelley, 2011). Accordingly, the current findings are reported in terms of complete or partial mediation. Complete mediation indicates that the relationship between stereotype threat (variable X) and performance (Y) completely disappears when a mediator (M) is included as a predictor variable (Preacher & Kelley, 2011). Partial mediation refers to instances in which a significant direct effect remains between stereotype threat and performance when controlling for the mediator, suggesting that additional variables may explain this relationship further (Rucker, Preacher, Tormala, & Petty, 2011). Instances of moderated-mediation are also reported, which occurs when the strength of mediation is contingent on the level of a moderating variable (Preacher, Rucker, & Hayes, 2007).

2.3. Findings

A total of 45 experiments in 38 articles were qualitatively synthesised, uncovering a total of 17 distinct proposed mediators. See Figure 1 for process of article inclusion. These mediators were categorised into affective/subjective ($n = 6$), cognitive ($n = 7$) and motivational mechanisms ($n = 4$). The majority of included research utilised a group-as-target prime ($n = 36$, 80%) compared to a self-as-target prime ($n = 6$; 13.33%). Three studies (6.66%) were uncategorised as they employed subtle stereotype threat primes, for example, manipulating the group composition of the testing environment. Table 3 summarises the articles reviewed and details their key findings and respective methodologies.

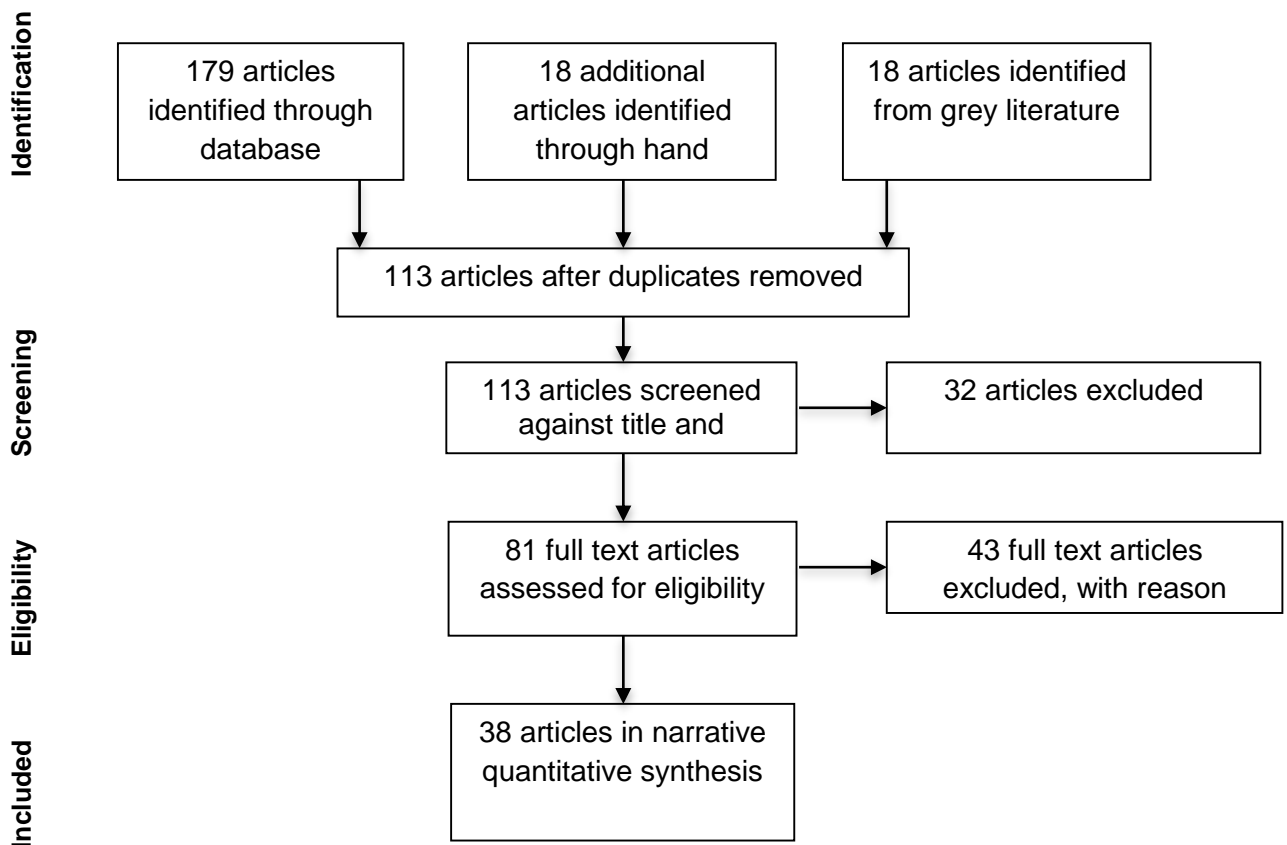


Figure 1. Process of article inclusion (following PRISMA guidelines).

Table 3.

Summary of stereotype threat literature examining mediational variables with key methodologies and findings.

Authors	Hypothesised Mediator	Mediator Method	Dependent Variable	Population	Conditions	Stereotype threat prime	Mediation findings
Appel et al. (2011), Experiment 4	Performance expectancies; Effort	Self-report expectancy scale; two-item self-report effort scale	Ability to judge encyclopaedia entries	Female STEM majors	3 conditions: 1), stereotype threat, 2), positive stereotype, 3), control	Group-as-target	None
Aronson et al. (1999), Experiment 1	Anxiety; Effort	State-trait anxiety inventory and effort questionnaire	18 Graduate Record Examination (GRE) maths questions	23 male undergraduates	2 conditions: 1), stereotype threat; 2), control condition	Group-as-target	None
Aronson et al. (1999), Experiment 2	Anxiety; Effort; Evaluation apprehension	State-trait anxiety inventory, Effort and performance expectancies questionnaire	15 GRE maths questions	75 white male undergraduates	2 conditions: 1), stereotype threat; 2), control	Group-as-target	None
Beaton et al. (2009)	Stereotype activation	Word-fragment completion task	9 GMAT and GRE questions	66 French-Canadian female undergraduates	3 conditions: 1), solo; 2), non-solo; 3), control	Group-as-target	None

Berjot et al. (2011)	Cognitive appraisals (challenge)	State primary appraisal questionnaire	Visuospatial performance (Ray figure)	92 French secondary school students (53 male)	2 conditions: 1), stereotype threat; 2), control	Group-as-target	Full
Bosson et al. (2004)	Anxiety; Evaluation apprehension	Anxiety scale observed non-verbal anxiety	Childcare (interpersonal skills)	72 male students	2 conditions: 1), stereotype threat; 2), control	Group-as-target	Full non-verbal anxiety None self-report anxiety
Brodish & Devine (2009)	Performance-avoidance goals; Anxiety	State anxiety scale and performance goals scale	20 GRE maths problems	101 female undergraduates	2 conditions: 1), stereotype threat; 2), control	Group-as-target	Full
Cadinu et al. (2003), Experiment 1	Performance expectancies	Bar graph of performance expectancies	7 difficult maths problems	95 female undergraduates	3 conditions: 1), positive stereotype, 2), negative stereotype, 3), control	Group-as-target	Partial
Cadinu et al. (2003), Experiment 2	Performance expectancies	Bar graph of performance expectancies	8 sentence-completion items	100 African-American soldiers (81 male)	4 conditions: American – 1), Negative; 2), Positive. Black - 3), Negative; 4), Positive	Group-as-target	Partial
Cadinu et al. (2005)	Negative thinking	Thought-listing sentences	7 GRE maths problems	60 female undergraduates	2 conditions: 1), stereotype threat; 2), control	Group-as-target	Full

Chalabaev et al. (2008)	Achievement goals	Achievement goals questionnaire for sports	Ability to dribble soccer ball through slalom course	51 female soccer players	3 conditions: 1), Athletic ability stereotype threat; 2), Technical ability stereotype threat; 3), control	Self-as-target	None
Chung et al. (2010)	Anxiety; Specific self-efficacy	State anxiety and self-efficacy questionnaire	Promotion performance exam	150 job applicants (134 male)	Within-participants field design	Uncategorised	Full (sequential)
Croizet et al. (2004)	Increased mental load	Heart rate variability	Raven APM cognitive test	139 college students	2 conditions: 1), stereotype threat; 2), control	Group-as-target	Full
Galdi et al. (2014)	Implicit stereotype endorsement	Implicit Association Test (IAT)	Maths test	276 first grade children (133 male)	3 conditions: 1), stereotype-consistent; 2), stereotype-inconsistent; 3), control	Group-as-target	Full

Gerstenberg et al. (2012), Experiment 3	Anxiety; Self-concept	German test anxiety scale and IAT	20 maths problems	156 female undergraduates	2 conditions: 1), subtle stereotype threat; 2), control	Group-as-target	Moderated-mediation
Hess et al. (2003)	Anxiety; Effort	Memory anxiety questionnaire; strategy use (clustered recall)	30-item free recall task	48 young (22 male) and 48 older adults (25 male)	3 conditions: 1), negative stereotype; 2), positive stereotype; 3), control	Group-as-target	Anxiety: None Effort: Full
Hess et al. (2009)	Working memory; Anxiety; Performance expectations	State anxiety scale and predicted recall task	Computation span task (maths equations), free recall task	103 older adults (52 male)	2 conditions; 1), stereotype threat; 2), control	Group-as-target	Performance expectations: Full Others: None
Jamieson & Harkins (2011a)	Effort	Coded solving techniques (conventional or unconventional)	30 GRE maths problems	76 female undergraduates	2 conditions; 1), stereotype threat; 2), control	Group-as-target	Full
Johns et al. (2008) Experiment 3	Emotion regulation; Working memory	State anxiety, re-appraisal and reading-span task	30 GRE maths problems	61 Caucasian female undergraduates	2 conditions: 1), stereotype threat; 2), control	Uncategorised	Full

Keller (2002)	Self-handicapping	2-item self-handicapping questionnaire	20 maths problems	75 German secondary school students (38 male)	2 conditions: 1), stereotype threat; 2), control	Group-as-target	Dejection: Full Others: None
Keller & Dauenheimer (2003)	Dejection; Anxiety; Self-handicapping	Anxiety and regulatory focus questionnaire	26 maths problems	74 secondary school students (39 male)	2 conditions: 1), stereotype threat; 2), control	Group-as-target	Dejection: Full Others: None
Keller & Sekaquaptewa (2008)	Individuation tendencies	24-item self-construal questionnaire	20-item spatial ability task	71 female students	2 conditions: 1), imagined solo status, 2), imagined non-solo status	Group-as-target	Partial
Laurin (2013)	Somatic and cognitive anxiety	Competitive state anxiety inventory and cognitive anxiety	Motor performance; 10 free throws	161 French high school students	3 conditions: 1), female stereotype threat; 2), male stereotype threat; 3), control	Group-as-target	Somatic anxiety: Partial Cognitive: None
Leyens et al. (2000)	Explicit stereotype endorsement	Stereotype acceptance questionnaire	Lexical decision task, valence judgment task and affective decision task	50 undergraduates (26 males)	2 conditions: 1), stereotype threat; 2), control	Group-as-target	None

Logel et al. (2009), Experiment 2	Thought suppression	Lexical decision task	20 maths problems	71 undergraduates (35 male)	2 conditions: 1), subtle stereotype threat; 2), control	Group-as-target	Partial
Mayer & Hanges (2003)	Anxiety; Cognitive interference; Self-efficacy; Evaluation apprehension	State anxiety, self-efficacy and evaluation apprehension questionnaires	Raven APM Cognitive test	60 African American and 90 White undergraduates (55 male)	2 conditions; 1), stereotype threat; 2), control	Self-as-target	None
McKown & Weinstein, (2003), Experiment 2	Anxiety; Effort; Self-appraised performance	Cognitive, physiological and affective anxiety scale.	Alphabet and word task	202 elementary school children	2 conditions: 1), stereotype threat; 2), control	Self-as-target	None
Mrazek et al. (2011), Experiment 2	Mind-wandering; Anxiety	Dundee State Stress questionnaire	30 GRE maths problems	72 female undergraduates	2 conditions: 1), stereotype threat; 2), control	Group-as-target	Full (sequential)
O'Brien & Crandall (2003)	Evaluation apprehension	Evaluation apprehension questionnaire	3 maths test: Difficult, easy and persistence tests	164 undergraduates (105 male)	2 conditions: 1), stereotype threat; 2), control	Group-as-target	None

Rosenthal et al. (2007), Experiment 2	Performance expectations	Two self-report items	10 GCSE maths problems	48 female undergraduates	4 shared characteristic conditions: 1), physical, 2), non-academic, 3), academic, 4), control	Group-as-target	Partial
Rydell et al. (2009), Experiment 2	Identity accessibility	Identity accessibility task	10 GRE maths problems	98 female undergraduates	4 conditions: 1), gender identity, 2), college identity, 3), multiple identities, 4), control	Group-as-target	Full
Rydell et al. (2009), Experiment 3	Working memory	Verbal vowel counting task	10 GRE maths problems	57 female undergraduates	4 conditions: 1), gender identity, 2), college identity, 3), multiple identities, 4), control	Group-as-target	Full
Rydell et al. (2014), Experiment 1	Updating; Shifting; Inhibition	Stroop task, letter-memory task, number-letter task	Modular maths test	168 undergraduates (93 male)	2 conditions; 1), stereotype threat; 2), control	Group-as-target	Updating: Full Others: None

Rydell et al. (2014), Experiment 2	Updating, Shifting, Inhibition	Stroop task, keep-track task, colour shape task	15 GRE Word-maths problems	90 female undergraduates	2 conditions; (1) stereotype threat; (2) control	Group-as-target	Updating: Full Others: None
Rydell et al. (2014), Experiment 3	Updating, Shifting, Inhibition	Letter-memory task, colour-shape task, anti-saccade task.	GRE Word-maths problems	82 female undergraduates	2 conditions; (1) stereotype threat; (2) control	Group-as-target	Updating: Full Others: None
Schmader & Johns (2003), Experiment 3	Working memory	Vowel-counting and operation span task	30 GRE maths problems	31 female undergraduates	2 conditions; (1) stereotype threat; (2) control	Group-as-target	Full
Seibt & Förster (2004), Experiment 2	Motivation; Expectancy	Motivation and performance expectancies questionnaire	Word-selection task	60 undergraduate students (29 male)	2 conditions; (1) stereotype threat; (2) control	Group-as-target	None
Seibt & Förster (2004), Experiment 4	Motivation; Expectancy; Mood; liking of task	Motivation, expectancies, mood and liking questionnaires	4 reasoning GRE problems and brick task	28 German undergraduates	2 conditions: (1) positive stereotype; (2) negative stereotype	Group-as-target	None

Seibt & Förster (2004), Experiment 5	Vigilance; Motivation; Expectancy; Mood; Liking of the task	Self-report eagerness and vigilance strategies, motivation and expectancy questionnaire	Analytic reasoning GRE problems and categorisation task	42 undergraduates	3 conditions; (1) positive stereotype; (2) negative stereotype; (2) control	Group-as-target	Eagerness and vigilance partially mediated
Sekaquaptewa & Thompson, (2003)	Performance expectancies	Performance expectancies questionnaire	Oral maths exam	157 undergraduates (77 male)	2 conditions; (1) stereotype threat; (2) control	Group-as-target	None
Skorich et al. (2013)	Effort	Effort measured by number of false positives on test of hazard perception	Hazard perception task	84 undergraduates (49 males)	3 conditions; (1) explicit threat; (2) categorisation threat; (3) control	Self-as-target	Full
Spencer et al. (1999) Experiment 3	Evaluation apprehension; Anxiety; Self-efficacy	State-trait anxiety index, evaluation apprehension questionnaire,	Maths portion of Graduate Management Test (GMAT)	67 undergraduates (31 male)	2 conditions; (1) stereotype threat, (2) control	Group-as-target	None
Steele & Aronson (1995) Experiment 2	Anxiety	State-trait anxiety index	Verbal GRE	20 Black and 20 White undergraduate females	2 conditions: (1) stereotype threat or control	Self-as-target	None

Stone (2002)	Self-handicapping; Anxiety	Word-fragment completion task, situational anxiety questionnaire	Athletic ability; golf-putting	38 Hispanic and 36 White undergraduates	2 conditions; (1) high stereotype threat; (2) low stereotype threat	Self-as-target	Self-handicapping: Partial Anxiety: None
Tempel & Neumann, (2014)	Anxiety	TAI-G anxiety questionnaire	8 arithmetic problems from the program for international student assessment	63 female undergraduates	2 conditions; (1) stereotype threat; (2) stereotype denial	Uncategorised	None

2.3.1. Affective/Subjective Mechanisms

Researchers have theorised that stereotype threat may stem from the fear, apprehension or anxiety of confirming a negative stereotype about one's group (Aronson & Inzlicht, 2004; Kray et al., 2001; Steele & Aronson, 1995). Consequently, many affective and subjective variables such as anxiety, individuation tendencies, evaluation apprehension, performance expectations, explicit stereotype endorsement and self-efficacy have been proposed to account for the stereotype threat-performance relationship.

2.3.1.1. Anxiety. Steele and Aronson (1995) examined whether elevated levels of anxiety underpin the effects of stereotype threat on African American's intellectual performance. Results indicated that African Americans underperformed relative to their Caucasian peers when a verbal ability test was ostensibly diagnostic of personal aptitude (i.e., a self-as-target stereotype). However, self-reported anxiety was not a significant mediator of the stereotype threat-performance relationship (Experiment 2). Extending this work, Spencer et al. (1999; Experiment 3) found that anxiety did not predict the effects of a negative group stereotype on women's mathematical performance, with further research confirming this (Aronson et al., 1999; Keller & Dauenheimer, 2003; Tempel & Neumann, 2014). Additional studies have suggested that self-reported anxiety does not influence the impact of self-as-target stereotype elicitation on African American's cognitive ability (Mayer & Hanges, 2003), Caucasian students' athletic skills (Stone, 2002), and group-as-target stereotype threat on older adults' memory recall (Hess et al., 2003; 2009).

It has also been proposed that anxiety may account for one of multiple mediators in the stereotype threat-performance relationship. For example, Chung and colleagues (2010) conducted a field study to examine the effects of stereotype threat on promotional exam performance. Results indicated that African Americans underperformed relative to Caucasians when they perceived a written knowledge test to be diagnostic of race-related ability. Furthermore, self-reported state anxiety and specific self-efficacy sequentially mediated the influence of stereotype threat on performance. This finding is supported by Mrazek et al. (2011) who found that anxiety and mind-wandering sequentially mediated the effects of stereotype threat on women's mathematical ability. Laurin (2013) also found that self-reported somatic anxiety partially mediated the effects of group-as-target stereotype threat on women's motor performance, suggesting that additional variables may explain this relationship. Nevertheless, it is viable to question whether this finding is comparable to other studies because stereotype threat facilitated performance.

In some contexts, it therefore seems that stereotype threat can lead to heightened levels of anxiety. However, the mixed results regarding anxiety as a potential mediator of stereotype threat may be indicative of various boundary conditions (i.e., moderators) that enhance stereotype susceptibility. Consistent with this claim, Gerstenberg, Imhoff and Schmitt (2012; Experiment 3) found that the impact of stereotype threat was moderated by women's self-concept of their mathematical ability. Specifically, female participants who reported a fragile maths self-concept solved fewer maths problems under group-as-target stereotype threat relative to those with a high concept of their mathematical ability. This susceptibility was mediated by increased anxiety. This moderated-

mediation suggests that women with a low self-concept in the domain of mathematics may be more vulnerable to group-relevant stereotype threat, with anxiety underpinning the effect of stereotype threat on mathematical performance.

Steele and Aronson (1995) suggest that anxiety might be relatively difficult to detect via self-report measures. Acknowledging this, Bosson et al. (2004) examined whether physiological anxiety mediated the effects of stereotype threat on homosexual males' performance on an interpersonal task. Results indicated that men who were reminded of their stigmatised homosexual identity before interacting with young children exhibited poorer childcare abilities compared to men who were not reminded of this identity. Moreover, mediational results suggested that physiological anxiety, but not self-reported anxiety, mediated the stereotype threat-performance relationship. This research was one of the first to suggest that the underlying mechanisms of stereotype threat may be best detected using indirect measures because people may not be able to reliably self-report on their experience of stereotype threat on explicit measures. Nevertheless, other research has found that physiological anxiety does not mediate the effects of group-as-target stereotype threat on older adults' memory recall (Hess et al., 2009) and self-as-target threat on children's writing ability (McKown & Weinstein, 2003). Overall, there are mixed findings in relation to anxiety as a possible mediator of stereotype threat effects, with 11 experiments resulting in null findings.

2.3.1.2. Individuation tendencies. In their seminal article, Steele and Aronson (1995) propose that stereotype threat might occur when individuals perceive a negative societal stereotype to be a true representation of their

personal ability. Based on this, Keller and Sekaquaptewa (2008) examined whether gender-based stereotypes (i.e., group-as-target threat) influence women to individuate their personal identity (the self) from their social identity (female). Female participants were assigned randomly to two conditions in which they anticipated that they would complete a spatial ability task in a group of females or males. They then completed a self-construal scale that measured levels of individualism and collectivism. Results indicated that participants appeared to underperform on a spatial ability test when they perceived that they were the single in-group representative (female) in a group of males. Moreover, stereotype threat was partially mediated by individuation tendencies in that gender-based threats influenced women to disassociate their self from the group to lessen the applicability of the stereotype. The authors suggest that this increased level of self-focused attention under solo status conditions is likely related to increased levels of anxiety.

2.3.1.3. Evaluation apprehension. Steele and Aronson (1995) also suggest that individuals might apprehend that they will confirm a negative stereotype in the eyes of out-group members. Testing this assertion, Mayer and Hanges (2003) found that African Americans reported higher levels of evaluation apprehension compared to Caucasian participants when a test was presented as diagnostic of personal cognitive ability. However, evaluation apprehension was not found to mediate the impact of this self-as-target stereotype on performance. Additional studies have found that evaluation apprehension does not mediate the effects of group-as-target stereotype threat on women's mathematical performance (O'Brien & Crandall, 2003; Spencer et al., 1999). Research to date therefore suggests that self-reported evaluation

apprehension may not mediate the effects of self or group-relevant stereotypes on performance outcomes.

2.3.1.4. Performance expectations. Under stereotype threat, individuals may evaluate the subjective likelihood of success depending on their personal resources. As these personal resources are anchored typically to group-level expectations, it is argued that in-group threatening information (i.e., women are poor at maths) may reduce personal expectancies to achieve, leading to diminished performance (Cadinu, Maas, Frigerio, Impagliazzo, & Latinotti, 2003). Testing this prediction, Cadinu et al. (2003; Experiment 1) found that women solved fewer maths problems when they were primed with a negative group-based stereotype relative to those who received a positive or no stereotype. Furthermore, performance expectancies partially mediated the effect of group-as-target threat on mathematical performance, suggesting that negative information was associated with lower expectancies. A second experiment indicated that performance expectancies partially mediated the effects of group-as-target threat on black participants' verbal ability. Research by Rosenthal, Crisp and Mein-Woei (2007; Experiment 2) also found that performance expectancies partially mediated the effects of self-based stereotypes on women's mathematical performance. However, rather than decreasing performance expectations, women who generated shared characteristics under stereotype threat reported higher predictions for performance relative to a control condition, which enhanced their mathematical performance. This research appears to suggest that the salience of a negative self-relevant stereotype may heighten the perceived differences between in-groups and out-groups to reduce performance. Furthermore, promoting shared

characteristics between in-group and out-group members may reduce stereotype threat effects.

Research has extended this work to examine the role of performance expectancies in diverse stigmatised populations. For example, Hess et al. (2009) found evidence of moderated-mediation for the effects of group-as-target stereotype threat on older adults' memory recall. Here, the degree to which performance expectancies mediated stereotype threat effects was moderated by participants' education. Specifically, elderly individuals with higher levels of education appeared to show greater susceptibility to stereotype threat. These findings suggest that lowered performance expectations may account for the effects of stereotype threat on performance, especially among individuals who identify strongly with the ability domain. Conversely, Appel et al. (2011) found that performance expectancies did not mediate the effects of group-based stereotype threat among highly identified women in the domains of science, technology, engineering and mathematics. To some extent, performance expectancies seem to contribute towards the negative effect that stereotype threat exerts on women's math performance and older adults' memory. However, it seems that additional research is required to elucidate whether performance expectancies mediate stereotype threat effects in understudied populations (i.e., the elderly), and for individuals experiencing self-as-target stereotype threat.

Other research suggests that stereotype threat can be activated through subtle cues in the environment rather than explicit stereotype activation (Sekaquaptewa & Thompson, 2003; Stone & McWhinnie, 2008). It is therefore plausible that expectancies regarding performance may be undermined when

in-group members are required to perform a stereotype-relevant task in front of out-group members. Testing this hypothesis, Sekaquaptewa and Thompson (2003) examined the interactive effects of solo status and stereotype threat on women's mathematical performance. Results revealed that women underperformed when they completed a quantitative examination in the presence of men (solo status) and under conditions of stereotype threat. Whilst performance expectancies appeared to partially mediate the relationship between group composition and mathematical ability, they did not mediate the effects of stereotype threat on performance. The authors posit that such findings may suggest that individuals are not consciously aware of how negative societal stereotypes may influence negatively their performance.

2.3.1.5. Explicit stereotype endorsement. Research has examined whether targeted individuals' personal endorsement of negative stereotypes is associated with underperformance. For example, Leyens and colleagues (2000) found that men underperformed on a decision making task when they were told that they were not as apt as women in processing affective information. Against predictions, stereotype endorsement was not found to be a significant intermediary between stereotype threat and performance. Other studies also indicate that stereotype endorsement may not be an underlying mechanism of the effects of self-as-target (Steele & Aronson, 1995) and group-as-target stereotype threat on women's mathematical aptitude (Beaton, Tougas, Rinfret, Huard, & Delisle, 2009; Spencer et al., 1999).

2.3.1.6. Self-efficacy. Research suggests that self-efficacy can have a significant impact on an individual's motivation and performance (Bandura, 1986; Maddux, 1992; Schunk, 1989), and may be influenced by environmental

cues (Bandura, 2006). Accordingly, it has been proposed that the situational salience of a negative stereotype may reduce an individual's self-efficacy. As previously mentioned, Chung et al. (2010) found that state anxiety and self-efficacy accounted for deficits in African American's performance on a job promotion exam. However, additional studies have indicated that self-efficacy does not mediate the effects of self-as-target threat on African American's cognitive ability (Mayer & Hanges, 2003) and both self-as-target and group-as-target threat on women's mathematical performance (Spencer et al., 1999; Steele & Aronson, 1995).

2.3.2. Cognitive Mechanisms

Much research suggests that affective and subjective factors underpin the harmful effects that stereotype threat exerts on performance (Schmader & Johns, 2003). However, researchers argue that stereotype threat may operate through a multi-dimensional process of affective, cognitive and motivational mechanisms (Schmader et al., 2008). Indeed, converging evidence suggests that stereotype threat may also influence performance detriments through its demands on cognitive processes (Johns, Inzlicht, & Schmader, 2008; Schmader et al., 2008; Schmader & Johns, 2003). Specifically, research has examined whether working memory, cognitive load, thought suppression, mind-wandering, negative thinking, cognitive appraisals and implicit stereotype endorsement mediate stereotype threat effects.

2.3.2.1. Working memory. Schmader and Johns (2003) propose that performance-evaluative situations might reduce working memory capacity

because stereotype-related thoughts consume valuable cognitive resources. In three studies, they examined whether working memory accounted for the influence of a group-as-target threat on women's and Latino American's mathematical ability. Findings indicated that both female and Latino American participants solved fewer mathematical problems compared to participants in a non-threat control condition. Reduced working memory capacity, measured by an operation span task (Turner & Engle, 1989), mediated the adverse effects of stereotype threat on mathematical performance. Supporting this, Rydell et al. (2009; Experiment 3) found that working memory mediated the effects of a group-as-target stereotype threat on women's mathematical performance.

Further research has also examined how stereotype threat may operate simultaneously through cognitive and emotional processes. Across four experiments, Johns et al. (2008) found that stereotype threat was accountable for deficits in women's verbal, intellectual and mathematical ability. Moreover, emotion regulation (characterised as response-focused coping) mediated the effects of group-as-target stereotype threat on performance by depleting executive resources.

Rydell et al. (2014) acknowledged that executive functioning is made up of more cognitive processes than the construct of working memory. Here the authors predicted that updating (i.e., the ability to maintain and update information in the face of interference) would mediate stereotype threat, whereas inhibition (i.e., the ability to inhibit a dominant response) and shifting (i.e., people's ability to switch between tasks) may not underpin this effect. Results indicated that women who were primed explicitly with a group-as-target stereotype displayed reduced mathematical performance compared to those in

a control condition. Consistent with predictions, only updating mediated the stereotype threat-performance relationship. These results suggest that the verbal ruminations associated with a negative stereotype may interfere with women's ability to maintain and update the calculations needed to solve difficult mathematical problems. The extent to which updating accounts for stereotype threat effects in diverse populations, however, is less straightforward. For example, Hess et al. (2009) found that working memory, measured by a computational span task, did not predict the relationship between group-based stereotype threat and older participants' memory performance.

2.3.2.2. Cognitive load. There is some evidence to suggest that stereotype threat depletes performance by placing higher demands on mental resources (Rydell et al., 2014; Schmader & Johns, 2003). These demands may exert additional peripheral activity (i.e., emotional regulation) that can interfere with task performance (Johns et al., 2008). In order to provide additional support for this notion, Croizet et al. (2004) examined whether increased mental load, measured by participants' heart rate, mediated the effects of stereotype threat on Psychology majors' cognitive ability. Psychology majors were primed that they had lower intelligence compared to Science majors. Results indicated that this group-as-target stereotype threat undermined Psychology majors' cognitive ability by triggering a psychophysiological mental load. Moreover, this increased mental load mediated the effects of stereotype threat on cognitive performance.

2.3.2.3. Thought suppression. Research suggests that individuals who experience stereotype threat may be aware that their performance will be evaluated in terms of a negative stereotype and, resultantly, engage in efforts

to disprove it (Croizet, et al. 2004; Logel, Iserman, Davies, Quinn, & Spencer, 2009; Steele & Aronson, 1995). This combination of awareness and avoidance may lead to attempts to suppress negative thoughts that tax the cognitive resources needed to perform successfully. With this in mind, Logel et al. (2009; Experiment 2) examined whether stereotype threat influences stereotypical thought suppression. Specifically, it was predicted that female participants would respond slower to gender-stereotypical words on a lexical decision task before completing a difficult maths test (suppression), but respond more quickly when they completed this task after the maths test (post-suppression rebound). Results indicated that women under stereotype threat solved fewer mathematical problems relative to men. In line with predictions, women tended to suppress stereotypical words when the lexical decision task was administered before the maths test, but showed post-suppression rebound of stereotype-relevant words when this task was completed afterwards. Mediation analyses appeared to indicate that pre-test thought suppression partially mediated the effects of stereotype threat on performance. These findings suggest that the experience of stereotype threat may lead females to suppress negative gender-related thoughts. However, this has a paradoxical effect on performance, resulting in females reinforcing the very stereotypes that they are trying to disprove (Logel et al., 2009).

2.3.2.4. Mind-wandering. Previous research suggests that the anticipation of a stereotype-laden test may produce a greater proportion of task-related thoughts and worries (Logel et al., 2009; Rydell et al., 2014). Less research has examined the role of thoughts unrelated to the task in hand as a potential mediator of the stereotype threat-performance relationship. Directly

testing this notion, Mrazek et al. (2011; Experiment 2) found that a group-as-target stereotype threat hampered women's mathematical performance in comparison to those in a non-threat control condition. Although self-report measures of mind-wandering resulted in null findings, indirect measures indicated that women under stereotype threat showed a marked decrease in attention. Mediation analyses showed that stereotype threat heightened anxiety, which in turn, increased mind-wandering and contributed to the observed impairments in mathematical performance. In contrast to these findings, other studies have found no indication that task irrelevant thoughts mediate the effects of group-as-target stereotype threat on women's mathematical performance (Beilock et al., 2007) and African American participants' cognitive ability (Mayer & Hanges, 2003).

2.3.2.5. Negative thinking. Research by Schmader and Johns (2003) suggests that the performance deficits observed under stereotype threat may be influenced by intrusive thoughts. Other research (Mrazek et al., 2011) has included post-experimental measures of cognitive interference to assess the activation of distracting thoughts under stereotype threat. However, the content of these measures are predetermined by the experimenter and do not allow participants to report spontaneously on their experiences under stereotype threat. Overcoming these issues, Cadinu and colleagues (2005) asked women to list their current thoughts whilst taking a difficult maths test under conditions of stereotype threat. Results indicated that female participants underperformed when they perceived a mathematical test to be diagnostic of gender differences. Moreover, participants in the stereotype threat condition listed more negative thoughts relative to those in the control condition, with intrusive

thoughts mediating the relationship between stereotype threat and poor maths performance. It therefore seems that negative performance-related thoughts may consume working memory resources to impede performance.

2.3.2.6. Cognitive appraisal. Research also suggests that individuals may engage in coping strategies to offset the performance implications of a negative stereotype. One indicator of coping is cognitive appraisal, whereby individuals evaluate the significance of a situation, as well as their ability to control it (Lazarus & Folkman, 1984). It is proposed that individuals may exert more effort on a task when a situation presents as a challenge, but may disengage from the task if they evaluate the situation as a threat (Drach-Zahavy & Erez, 2002; White, 2008). Taking this into consideration, Berjot, Roland-Levy and Girault-Lidvan (2011) examined whether targeted members might be more likely to perceive a negative stereotype as a threat to their group identity rather than as a challenge to disprove it. Results indicated that North African secondary school students underperformed on a visuospatial task when they were primed to perceive that French students possess superior perceptual-motor skills. Contrary to predictions, threat appraisal did not mediate the relation between stereotype threat and performance. Rather, perceiving the situation as a challenge significantly mediated the stereotype threat-performance relationship. These findings suggest that individuals may strive to confront, rather than avoid, intellectual challenges and modify the stereotype held by members of a relevant out-group in a favourable direction (Cohen & Garcia, 2005).

2.3.2.7. Implicit stereotype endorsement. Research suggests that situational cues that present as a threat may increase the activation of

automatic associations between a stereotyped concept (i.e., female), negative attributes (i.e., bad), and the performance domain (i.e., maths; Nosek et al., 2002). Implicit measures are able to detect automatic associations between such concepts and stereotypical attributes that may not be captured reliably by self-reports (Galdi, Cadinu, & Tomasetto, 2014). In a study of 240 six-year old children, Galdi et al. (2014) examined whether implicit stereotype endorsement accounted for the effects of stereotype threat on girls' mathematical performance. Consistent with the notion that automatic associations can precede conscious beliefs, results indicated that girls acquired implicit maths-gender stereotypes before they emerged at an explicit level. Specifically, girls showed stereotype-consistent automatic associations between the terms 'boy-mathematics' and 'girl-language', which appeared to account for their lower performance under stereotype threat.

2.3.3. Motivational Mechanisms

Most of the initial work on the underlying mechanisms of stereotype threat has focused on affective and cognitive processes. In more recent years, researchers have argued that instead of interfering with working memory, stereotype threat may motivate individuals to disconfirm the stereotype, with this having a paradoxical effect of harming performance (Jamieson & Harkins, 2007; 2009; 2011a). To this end, research has elucidated the potential role of effort, self-handicapping, dejection, vigilance, and achievement goals.

2.3.3.1. Effort/motivation. Underpinned by the "mere effort model", Jamieson and Harkins (2011a) examined whether motivation plays a proximal

role in the effect of stereotype threat on women's mathematical performance. Here it was predicted that stereotype threat would lead participants to use a conventional problem solving approach (i.e., use known equations to compute an answer), which would facilitate performance on 'solve' problems, but hamper performance on 'comparison' problems. Supporting this hypothesis, results indicated that stereotype threat debilitated performance on comparison problems because participants employed the dominant, but incorrect, solution approach. Furthermore, this incorrect solving approach mediated the effect of stereotype threat on comparison problem performance. These findings may suggest that stereotype threat motivates participants to perform well, which increases activation of a dominant response to the task. However, this dominant approach does not always guarantee success, resulting in participants under stereotype threat performing worse than those in the control condition.

Other researchers have argued that stereotype threat may have different effects on effort dependent on the prime utilised (Skorich et al., 2013). For example, Skorich et al. (2013) examined whether effort mediated the effects of implicit and explicit stereotypes on provisional drivers' performance on a hazard perception test. Participants in the implicit prime condition ticked their driving status (provisional, licensed) on a questionnaire, whereas participants in the explicit prime condition were provided with stereotypes relating to the driving ability of provisional license holders. Results seemed to reveal that participants detected more hazards when they were primed with an explicit stereotype relative to an implicit stereotype. Mediation analyses indicated that whilst increased effort underpinned the effects of an implicit stereotype on

performance, decreased effort mediated the effects of an explicit stereotype prime. Additional research has also indicated that reduced effort mediates the effects of an explicit stereotype on older adults' memory recall (Hess et al., 2003). These findings suggest that implicit stereotype threat primes may lead to increased effort because participants aim to disprove the stereotype, whereas explicit stereotype threat primes may lead to decreased effort as participants self-handicap themselves (Skorich et al. 2013). Other studies utilising self-reported measures of effort have resulted in non-significant findings (Aronson et al., 1999; Experiment 1 & 2; Keller & Dauenheimer, 2003; Experiment 4, McKown & Weinstein, 2003; Experiment 2, Seibt & Förster, 2004, Experiment 2, 4 & 5), suggesting that self-reports may be vulnerable to self-presentational motives, which may influence mediational results.

2.3.3.2. Self-handicapping. Individuals may engage in self-handicapping strategies to proactively reduce the applicability of a negative stereotype to their performance (Keller, 2002; Stone, 2002). Here it is theorised that people attempt to influence attributions for performance by erecting barriers to their success. Investigating this notion, Stone (2002) examined whether self-handicapping mediated the effects of stereotype threat on white athletes' sporting performance. Self-handicapping was measured by the total amount of stereotype-relevant words completed on a word-fragment task. Results indicated that white athletes practiced less when they perceived their ability on a golf-putting task to be diagnostic of personal ability, thereby confirming a negative stereotype relating to comparatively poorer white athleticism. These athletes were also more likely to complete the term 'awkward' on a word fragment completion test compared to the control condition. Mediational

analyses seemed to reveal that the greater accessibility of this term partially mediated the effects of stereotype threat on psychological disengagement and performance. The authors suggest that stereotype threat increased the accessibility of thoughts related to poor athleticism to inhibit athletes' practice efforts. However, a limitation of this research is that analyses were based on single-item measures (i.e., the completion of the word 'awkward') rather than total of completed words on the word-fragment test.

Keller (2002) also tested the hypothesis that the salience of a negative stereotype influences self-handicapping. Results indicated that women who were primed with a group-as-target stereotype underperformed on a mathematical test relative to their control group counterparts. Furthermore, they expressed stronger tendencies to search for external explanations for their weak performance with this mediating the relationship between stereotype threat and performance. Despite these preliminary findings, Keller and Dauenheimer (2003) were unable to provide support for the notion that self-reported self-handicapping is a significant intermediary between stereotype threat and women's mathematical underperformance. Whilst there is some evidence that self-handicapping may account for the effects of stereotype threat on performance across diverse populations, such as women in mathematics and sports athletes, it therefore seems that additional research is required to provide additional support for this variable.

2.3.3.3. Dejection. Research examining performance expectations suggests that stereotype threat may be mediated by goals set by the participants. Extending this work, Keller and Dauenheimer (2003) hypothesised that female participants may make more errors on a mathematical test due to

an overly motivated approach strategy. Results indicated that women underperformed when a maths test was framed as diagnostic of gender differences (group-as-target threat). Moreover, their experiences of dejection were found to mediate the relationship between stereotype threat and performance. The authors suggest that individuals may engage in a promotion focus of self-regulation because they are motivated to disconfirm the negative stereotype. However, according to the researchers, feelings of failure may elicit an emotional response that resultantly determines underperformance.

2.3.3.4. Vigilance. In contrast to Keller and Dauenheimer (2003), Seibt and Förster (2004, Experiment 5) propose that under stereotype threat, targeted individuals engage in avoidance and vigilance strategies. They predicted that positive stereotypes should induce a promotion focus, leading to explorative and creative processing, whereas negative stereotypes should induce a prevention focussed state of vigilance, with participants avoiding errors. Across five experiments, male and female participants were primed with a negative group-as-target stereotype, which stated that women have better verbal abilities than men. However, rather than showing a stereotype threat effect, results indicated a speed-accuracy trade off with male participants completing an analytical task slower but more accurately than their counterparts in a non-threat control condition. Furthermore, this prevention focus of vigilance was found to partially mediate the effects of stereotype threat on men's analytical abilities (Experiment 5). The authors suggest that the salience of a negative group stereotype may elicit a vigilant, risk-averse processing style that diminishes creativity and speed whilst bolstering analytic thinking and accuracy.

2.3.3.5. Achievement goals. Achievement goals theory (Elliot & Church, 1997) posits that participants will evaluate their role in a particular achievement context and endorse either performance-focused or performance-avoidance goals. In situations where the chances of success are low, individuals may engage in performance-avoidance goals, corresponding to a desire to avoid confirming a negative stereotype. With this in mind, Chalabaev, Sarrazin, Stone and Cury (2008) examined whether performance avoidance goals mediated the effects of stereotype threat on women's sporting performance. The impact of two self-as-target stereotypes (i.e., poor athletic and soccer ability) on performance was assessed relative to a control condition. Results indicated that women in the athletic ability condition performed more poorly on a dribbling task, but not in the soccer ability condition. Although these participants endorsed a performance-avoidance goal, this did not appear to mediate the relationship between stereotype threat and soccer performance.

Highlighting the possible interplay between affective, cognitive and motivational mechanisms, Brodish and Devine (2009) proffered a multi-mediator model, proposing that anxiety and performance-avoidance goals may mediate the effects of group-as-target stereotype threat on women's mathematical performance. Achievement goals were measured in terms of the extent to which participants endorsed performance-avoidant (the desire to avoid performing poorly) or approach goals (trying to outperform others). Results indicated that women under stereotype threat solved fewer mathematical problems relative to those in a control condition. Mediation analyses seemed to reveal that performance avoidance goals and anxiety sequentially mediated women's mathematical performance. This may suggest

that negative expectations for performance may motivate women to avoid failure, which in turn, appears to heighten anxiety and influences underperformance.

2.4. Discussion

This systematic literature review evaluated empirical support for the proposed mediators of stereotype threat over the past twenty years of research. Through the lens of the multi-threat framework (Shapiro & Neuberg, 2007), it distinguished between self-relevant and group-relevant stereotype primes to examine the extent to which these have different effects on performance, are mediated by distinct mechanisms, and imperil diverse stereotyped populations.

On the whole, findings indicate that experiences of stereotype threat may increase individuals' feelings of anxiety, negative thinking and mind-wandering, which deplete the working memory resources required for successful task execution (e.g., Beilock et al., 2007; Schmader & Johns, 2003; Rydell et al., 2014). Specifically, the salience of a negative stereotype may influence adverse thoughts and heighten situational performance pressure and resultantly distract targeted individuals from the task at hand (Logel et al., 2009; Mrazek et al., 2011). Other research, however, suggests that individuals may be motivated to disconfirm negative stereotypes and, as a consequence, engage in efforts to suppress stereotypical thoughts that are inconsistent with task goals (e.g., Hess et al., 2003; Jamieson & Harkins, 2011a, but also c.f., Jamieson & Harkins, 2009). It therefore appears that researchers are still in disagreement with regards to the underlying mechanisms that mediate the

stereotype threat-performance relationship, with several proposed mechanisms resulting in varying degrees of empirical support.

Many different primes and manipulations have been utilised to elicit stereotype threat and this may have precluded finding firm evidence of mediation (see Table 3 findings, p. 32). For example, some researchers have employed blatant/direct stereotype threat manipulations by informing participants explicitly of a negative stereotype related to performance (e.g. Spencer et al., 1999; Steele & Aronson, 1995). Others have evoked more subtle stereotype threats by placing stigmatised group members in situations where they have minority status (e.g., Keller & Sekaquaptewa, 2008; Sekaquaptewa & Thompson, 2003). It is therefore plausible that different mechanisms may mediate the effects of blatant and subtle stereotype threat effects on performance (Chalabaev et al., 2008; Skorich et al., 2013; Stone & McWhinnie, 2008). Providing evidence consistent with this claim, Sekaquaptewa and Thompson (2003) found that performance expectancies partially mediated the effects of solo status, but not stereotype threat on performance. These results suggest that women may make comparative judgments about their expected performance when they are required to undertake an exam in the presence of outgroup members, but may not consciously recognise how a negative stereotype can impair performance directly.

Additional research suggests that working memory may mediate the effects of subtle stereotype threat on performance because individuals attend to situational cues that heighten the salience of a discredited identity (Croizet et al., 2004; Schmader & Johns, 2003). Alternatively, motivation may mediate the effects of blatant stereotype threat because individuals strive to disprove

the negative stereotype (Keller & Dauenheimer, 2003; Skorich et al., 2013; Seibt & Förster, 2004; Stone & McWhinnie, 2008). However, this is contradicted by other research, which appears to show that working memory underpins the effects of an explicit gender-related prime on women's mathematical performance (Rydell et al., 2007; 2009; Van Loo et al., 2014). Although stereotype threat effects seem to be robust (Nguyen & Ryan, 2008), it is possible that different stereotype threat manipulations diverge in the nature, the focus, and the intensity of threat they produce and may be accounted for by different underlying mechanisms (Shapiro & Neuberg, 2007).

In a similar vein, previous research has viewed stereotype threat typically as a singular construct, which may have led to the unrealistic expectation that moderators and mediators may be stable across groups and domains (Barber, 2016; Shapiro & Neuberg, 2007). The current review highlights that 80% of the articles employed a group-as-target stereotype threat prime. Here stereotype threat is manipulated to highlight that stereotype-consistent performance may confirm, or reinforce, a negative societal stereotype as being a true representation of one's social group (Steele et al., 2002). This has led to a relative neglect of situations in which individuals may anticipate that their performance may be indicative of personal ability. It is therefore conceivable that self-as-target and group-as-target manipulations may have distinct effects on performance and may be mediated by different mechanisms (Shapiro & Neuberg, 2007; Wout et al., 2008). In this context it should be noted that research to date has not examined systematically whether distinct stereotype threat primes are mediated by different mechanisms.

Recent research suggests that stereotype threat may be a self-concept threat, rather than a group-reputation threat (Barber, 2016), and further support for this notion is presented in the forthcoming chapters of this thesis. As such, stereotype threat-related performance deficits may be more likely to emerge when an individual's personal identity is tied to a negative group-related stereotype. Moreover, research appears to indicate that individuals may dissociate their sense of self from the negatively stereotyped domain when a group-based stereotype threat is primed (Keller & Sekaquaptewa, 2008). However, this may be more unlikely when an individual experiences self-as-target stereotype threat because their personal ability is explicitly tied to a negative stereotype that governs their ingroup. In such situations, the activation of a group-based stereotype may set into motion mechanisms that reflect a protective orientation of self-regulation, whereas self-relevant knowledge may heighten self-consciousness. In order to gain a more nuanced account of stereotype threat, future research would benefit from recognising the distinct forms of stereotype threat, and elucidating whether performance decrements are more likely to emerge when an individual's personal or social identity is made salient in the stereotyped domain.

Research has begun to suggest that different groups may also be more susceptible to certain types of stereotype threat (Pavlova et al., 2014; Shapiro & Neuberg, 2007; Shapiro, 2011). For example, research indicates that populations that tend to have low group identification, such as individuals with a mental illness, are more susceptible to self-as-target threats (Shapiro, 2011). Conversely, populations with high group identification, such as individuals of a certain ethnicity or religion, appear more likely to experience group-as-target

threats (Shapiro, 2011). Whilst this highlights the role of moderating variables that may heighten individuals' susceptibility to stereotype threat, it also suggests that individuals might experience stereotype threat in different ways, dependent on their stigmatised identity. This could explain why some variables (e.g., anxiety, self-handicapping) that have been found to mediate the effects of stereotype threat on some groups have not emerged in other populations.

Finally, it appears that diverse mediators may account for the effects of stereotype threat on different performance outcomes. Whilst working memory appears to be implicated in tasks that require controlled processing, it may not be required for tasks that rely more on automatic processes (Beilock et al., 2006; 2007; Stone & McWhinnie, 2008; Rydell et al., 2014). In line with this notion, Beilock et al. (2006) found that experts' golf putting skills were harmed under stereotype threat when attention was allocated to automatic processes that do not heavily rely on working memory. This suggests that well-learned skills may be hampered by attempts to bring performance back under step-by-step control. Conversely, skills such as difficult mathematical problem solving appear to involve heavy processing demands and may be harmed when working memory resources are expended by exposure to a negative stereotype. As such, distinct mechanisms may underpin different threat-related performance outcomes.

2.4.1. Chapter Conclusion

Two decades of research appear to demonstrate the harmful effects that stereotype threat can exert on a wide range of populations in a broad array of performance domains. However, findings with regards to the mediators that

underpin these effects are equivocal. This may be a consequence of the heterogeneity of primes used to instantiate stereotype threat and the methods used to measure mediation and performance. To this end, the current systematic review suggests that additional research is required to examine the influence that distinct stereotype threats exert on performance, and to elucidate further whether different mediators may underpin these effects.

3. CHAPTER 3 - Theoretical and Methodological Considerations

3.1. Chapter Overview

Chapter 3 of this thesis provides an overview of the current debate within the stereotype threat literature and the issues uncovered by the systematic review (Chapter 2). It starts by briefly exploring why the theory of stereotype threat was developed, and aims to alleviate some concerns regarding this theory that are raised in the wider psychological literature. Following on from this, it engages critically with some apparent conceptual issues of stereotype threat theory, which underpin the research questions presented by this thesis. Finally, methodological issues are considered, with a view to inform the following six empirical studies presented in this thesis.

3.2. Critiques of the Stereotype Threat Literature

The theory of stereotype threat was devised to provide a social psychological explanation for the chronic achievement gaps between African American and Caucasian students, and between women and men on quantitative portions of standardised tests (c.f., Spencer et al., 1999; Steele & Aronson, 1995; Steele, 1997). The theoretical assumptions underpinning this situational phenomenon assert that people tend to be very sensitive to cues in their environment that signal a discredited social identity (Murphy, Steele, & Gross, 2007; Purdie-Vaughns, Steele, Davies, Ditlman, & Randall-Crosby, 2008; Steele et al. 2002). In turn, this heightened vigilance is seen to distract individuals from the task at hand and lead to underperformance (Schmader et al., 2008; Seibt & Förster, 2004). Nevertheless, the theory of stereotype threat has come under considerable scrutiny in recent years (c.f., Sacket, Hardison, & Cullen, 2004;

Stoet & Geary, 2013; Streets & Major, 2014), which has led to a recent review that aims to resolve some proposed theoretical and methodological issues in the literature (Spencer, Logel, & Davies, 2016).

First, some researchers have expressed concern that stereotype threat is uncritically regarded as a “primary causal” explanation for the gender-achievement gap in mathematics (Stoet & Geary, 2012, p. 99). In their critical review, Stoet and Geary (2012) highlight that only 50% of studies that used experimental manipulations consistent with Spencer et al.’s (1999) seminal paper were able to replicate the finding that women underperform in comparison to men when they are primed with a negative societal stereotype. Furthermore, they found that studies which controlled for participants’ pre-existing mathematical ability found greater effects than those that did not adjust for this. Stoet and Geary (2012) therefore conclude that the state of the current literature does not support the enthusiasm for stereotype threat theory as a primary explanation for gender differences in mathematical performance. However, it is important to note that stereotype threat researchers have firmly stated that this theory should not be used as a sole explanation for women’s underperformance in mathematics (Spencer et al., 2016; Walton & Spencer, 2009). Rather, it was developed to rival biological theories of innate differences in mathematical ability (c.f., Benbow & Stanley, 1990; 1983; Spelke, 2005), proposing that situations themselves might bring about apparent group differences in performance (Inzlicht & Schmader; 2010; Steele, 1997; Walton & Spencer, 2009).

In support of this, the effects of stereotype threat have been observed in performance domains, in which achievement gaps between groups have not

been widely observed. For example, the salience of a negative stereotype has been found to inhibit females' video gaming performance (Kaye & Pennington, 2016), older adults' memory recall (Hess et al., 2003; 2009) and mathematical ability (Abrams et al., 2008), ecstasy users' cognitive functioning (Cole, Michailidou, Jerome, & Sumnall, 2006), females' chess performance (Rothgerber & Wolsiefer, 2013), and homosexual men's childcare skills (Bosson et al., 2004). As such, it appears that stereotype threat effects are not restricted to specific populations (e.g., women and African Americans) or tasks (e.g., standardised tests), but rather appear to impede the performance of any group to which a negative stereotype generally applies (Spencer et al., 2016).

Presenting as a further issue, researchers have argued that stereotype threat effects tend to be confined to the laboratory and do not necessarily generalise to real-world testing situations (Cullen, Waters, & Sacket, 2006; Sacket et al., 2004, 2008; Stricker & Ward, 2004). Aiming to replicate the findings from Steele and Aronson's (1995) classic research, Stricker and Ward (2004) manipulated whether students reported their ethnicity or gender on a demographic questionnaire *before* or *after* a standardised test. Findings from this field study indicated that enquiring about race and gender at the start of the test did not elicit stereotype threat effects that were "statistically or practically significant" (Stricker & Ward, 2004, p. 685). According to Stricker and Ward (2004) stereotype threat might only be potent within laboratory studies because researchers use manipulations to ensure that stereotypes are particularly salient. However, Danaher and Crandall (2008) re-analysed Stricker and Ward's (2004) findings and concluded that women, in particular, performed better on the maths test when they were asked to indicate their gender after the

test relative to before. Although this effect was small, they argued that it might have notable practical implications. Specifically, it was contended that this simple and inexpensive change could increase the number of women in America receiving mathematics credit by more than 4,700 every year (Danaher & Crandall, 2008). In line with Steele and Aronson (1995), these findings suggest that simply asking test takers to report their gender in a stereotyped domain may signal to females that their performance will be evaluated in line with gender-related expectations (however, c.f., Stricker & Ward, 2008 for a reply).

Other research also provides support for the applied efficacy of stereotype threat (Good, Aronson, & Harder, 2008; Keller, 2007; Huguet & Régner, 2007). Huguet and Régner (2007) found that girls in secondary school displayed performance decrements when they were led to believe that their ability to recall aspects of a complex geometry figure was diagnostic of mathematical ability. Female secondary students also appear to perform worse on a difficult mathematical test when they are told that it reveals gender differences (Keller, 2007). A series of meta-analyses indicate further that experiences of stereotype threat may explain between 50-82% of the gender gap on the SAT-maths test (Spencer et al., 2016). Taken together, there is relatively strong evidence to suggest that stereotype threat may account for one of a multitude of factors that impede women's mathematical performance in real life testing environments.

However, even in field studies, participants are typically assigned to either a stereotype threat condition, whereby they are primed with a negative stereotype regarding their group membership, or a "non-threat" control

condition (Nguyen & Ryan, 2008). The performance between the two conditions is then compared, with the prediction that individuals will perform worse in stereotype-salient contexts compared to non-threatening contexts (Nguyen & Ryan, 2008; Steele, 1997; Steele et al. 2002). This has led some researchers to question why participants in the control condition (who also belong to the stigmatised group) do not experience stereotype threat, particularly if this phenomenon can explain real world achievement (c.f., Spencer et al., 2016). In order to overcome this potential issue, researchers typically inform participants in the control condition that their performance is non-diagnostic of ability (c.f., Spencer et al., 1999; Steele & Aronson, 1995), and that the task measures other factors relating to performance, such as working memory ability (c.f., Schmader & Johns, 2003). Such manipulations appear to be successful because they lower the situational performance pressure that participants in the control condition experience, allowing them to perform optimally (Spencer et al., 2016).

3.3. Current Gaps in the Stereotype Threat Literature

3.3.1. Conceptual/theoretical issues. A review of the literature, presented in Chapter 2, highlighted that researchers have employed a variety of different manipulations to instantiate stereotype threat within experimental procedures, and appear to conceptualise stereotype threat in many ways. Specifically, it appears that the majority of previous research has utilised a group-as-target stereotype threat prime, which for example, is designed to lead women to believe that their mathematical performance will be diagnostic of gender-related ability. Less research has examined whether performance

decrements occur when women are primed that their performance will be diagnostic of personal ability (i.e., a self-as-target threat). Bridging this gap in the literature, the current thesis examines whether self- and group-relevant stereotypes have simultaneous negative effects on performance, or whether these manipulations impact performance outcomes differently. In other words, the empirical studies in this thesis examine whether self-as-target or group-as-target stereotype threats have a greater detrimental impact on women's mathematical performance. This thesis therefore aims to make an original, theoretical contribution to knowledge by exploring whether the 'self' and the 'social group' are distinct constructs in the conceptualisation of stereotype threat (i.e., a multi-faceted phenomenon), or whether stereotype effects occur when an individual's personal and social identity are viewed as functionally equivalent (i.e., a singular construct which occurs when the concepts of the 'self' and the 'social' group are activated in stereotype-salient situations).

In addition to the conceptualisation of stereotype threat, the systematic literature review (Chapter 2) revealed that there is considerable debate regarding the underlying mechanisms of this situational phenomenon. Empirical support has been accrued for the working memory interference and mere effort accounts of stereotype threat effects, but these theories proffer opposing explanations (c.f., Jamieson & Harkins, 2007; 2009; Schmader et al., 2008). The working memory interference account suggests that the salience of a negative societal stereotype influences verbal ruminations and worries that may reduce the working memory resources required to successfully solve difficult mathematical problems (Beilock, 2008; Beilock et al., 2007; Rydell et al., 2009; 2014). Conversely, the mere effort account argues that individuals

may be motivated to disprove the negative stereotype, which facilitates the most dominant response on a given task (Jamieson & Harkins, 2007; 2009). If this dominant response is correct, then performance is facilitated, however if the dominant response is incorrect, or participants are not given the time to correct for incorrect responses, then performance is debilitated (Jamieson & Harkins, 2007; 2009; Seitchik & Harkins, 2015). As such, it remains to be ascertained whether working memory or mere effort can explain the effects of stereotype threat on performance.

Moreover, to date no research has examined whether these explanations account for the effects of distinct stereotype threats on performance. From a working memory interference perspective (Beilock et al., 2007; Johns et al., 2008; Rydell et al., 2014; Schmader & Johns, 2003), it is plausible that manipulations that target the self (i.e., self-as-target) may have a greater effect on performance because this heightens self-consciousness and leads people to 'choke under pressure'. Conversely, from a mere effort perspective (Jamieson & Harkins, 2007; 2009; 2011a), it may be that targeted individuals are more motivated to disprove a negative gender-related stereotype when it is explicitly tied to personal performance (self-as-target stereotype threat), with this increased motivation hampering performance.

Although research has made theoretical advances to suggest that individuals can experience multiple stereotype threats, and has begun to elucidate the underlying mechanisms that account for the stereotype threat-performance relationship, these two lines of research tend to be studied in parallel, rather than in conjunction with each other. However, if females in the domain of mathematics are at risk of experiencing diverse stereotype threats,

which may have different effects on performance, then it is important to identify the mechanisms that underpin these effects. In the first of its kind, the second aim of the current thesis is to examine whether deficits in working memory or enhanced motivation underpin the effects of self-as-target and group-as-target stereotype threat on women's mathematical performance. The current thesis is therefore underpinned by the following three research questions, which have not received due attention in the literature to date:

Research Question 1: Does self-as-target and group-as-target stereotype threat influence negatively women's mathematical performance?

Research Question 2: Does self-as-target stereotype threat have a greater detrimental impact compared to group-as-target stereotype threat because it heightens self-consciousness?

Research Question 3: Do distinct mechanisms govern the effects of self-as-target and group-as-target stereotype threat on women's mathematical performance?

3.3.2. Methodological issues. In addition to the heterogeneity of primes employed to instantiate stereotype threat, the systematic literature review presented in Chapter 2 pointed to a number of methodological issues that may have precluded finding firm evidence for the mediators of stereotype threat. Here it was highlighted that researchers have tended to utilise explicit self-report measures in their efforts to uncover the mediating variables which

may underpin the stereotype threat-performance relationship. However, it has long been argued that individuals have limited access to higher order mental processes (Nisbett & Wilson, 1977; Schwarz, 1999), such as those involved in the evaluation and initiation of behaviour (Mandler, 2004; Miller, 1962). Resultantly, participants under stereotype threat may be unable to observe and explicitly report the operations of their own mind (Bosson et al., 2004; Kiefer & Sekaquaptewa, 2007b; Nisbett & Wilson, 1977; Wegner, 2002; Wheeler et al., 2001). Consistent with this assertion, Bosson et al. (2004) found that although stereotype threat heightened individuals' physiological anxiety, they did not report an explicit awareness of increased anxiety on self-report measures. This may suggest that participants are mindful of the impression they make to others and engage in self-presentational behaviours in an effort to appear inoculated to negative stereotypes (Bosson et al., 2004). Support for this notion comes from research which suggests that stereotype threatened participants tend not to explicitly endorse stereotypes (Beaton et al., 2009; Bosson et al., 2004; Kiefer & Sekaquaptewa, 2007b; Leyens et al., 2000; Spencer et al., 1999; Steele & Aronson, 1995), and are more likely to claim impediments to justify poor performance (Aronson et al., 1999; Keller, 2002; Spencer et al., 1999). Overcoming the issues associated with self-report methodology in this domain, the current thesis utilises indirect measures to elucidate the underpinning mechanisms of the stereotype threat-performance relationship (i.e., cognitive tasks of working memory ability and eye-tracking).

Research has also suggested that order effects (i.e., the order in which test instruments are administered) may present as an issue when investigating stereotype threat effects (c.f., Brodish & Devine, 2009; Logel et al., 2009). For

example, Brodish and Devine (2009) found that women reported higher levels of anxiety when they completed a questionnaire before a mathematical test compared to afterwards. This suggests that pre-test anxiety ratings may have reflected participants' apprehension towards the upcoming evaluative test, with this apprehension diminishing once the test was completed. Research by Logel and colleagues (2009) provides support for this notion, indicating that women who completed a lexical decision task after a maths test were quicker to respond to stereotype-relevant words compared to women who subsequently completed the task. These results exhibit the variability in individuals' emotions under stereotype threat and suggest that they may be unable to retrospectively report on their feelings once the experience of threat has passed. This highlights the importance of counterbalancing test instruments in the investigation of stereotype threat, purporting that the order in which test materials are administered may influence mediational findings. With this in mind, the measures utilised in the current thesis were counterbalanced and randomised between participants.

Finally, the systematic literature review (Chapter 2) highlighted that, in some studies, individuals assigned to the control condition may have also experienced stereotype threat, which may have prevented finding reliable evidence of mediation. For instance, Chalabaev et al. (2008) primed stereotype threat by presenting a soccer ability test as a diagnostic indicator of personal factors related to athletic ability. However, participants in the control condition were informed that the aim of the test was to examine psychological factors in athletic ability. Consequently, these participants may have been apprehensive about their performance being evaluated. Research has also manipulated the

salience of stereotype threat by stating that gender differences in mathematical performance are equal (Sekaquaptewa & Thompson, 2003). However, other research has utilised this prime within control conditions (e.g., Croizet et al., 2004; Jamieson & Harkins, 2011a; Kiefer & Sekaquaptewa, 2007b), underpinned by the rationale that describing a test as 'fair' or non-diagnostic of ability eliminates stereotype threat (Spencer et al., 2016; Steele & Davies, 2003). It therefore appears that, in some instances, researchers may have inadvertently induced stereotype threat. This outlines the importance of employing a control condition in which individuals are not made aware of any negative stereotypes and are told that the test is non-diagnostic of ability. In line with Steele and Davies (2003) recommendations, a control condition is employed in each empirical study presented in this thesis, in which participants are told that the tasks are non-diagnostic of ability.

3.4. Ethical Considerations

The Departmental and Faculty Ethics Committees at Edge Hill University ethically approved all of the experiments presented in this thesis. Guided by the British Psychological Society's (BPS) ethical requirements, all participants were provided with an information sheet, which gave them an overview of the study requirements and provided informed consent to take part (see Appendix A). They were assigned randomly to the stereotype threat and control conditions. In each experiment presented in this thesis, participants were recruited for a study that examined ostensibly factors related to problem solving. They remained blind to the true experimental aim in accordance with research that indicates that awareness of the concept of stereotype threat may eliminate its

effects (Johns, Schmader, & Martens, 2005). This decision was also made to control for demand characteristics, whereby participants infer the aims of the study and behave accordingly.

Participants signed up to each study via an online participation website, and studies were also advertised through e-mail and posters situated around the university. They were compensated with course credits or a small monetary reward for their time (£3 for 30-minute tasks and £5 for 45-minute tasks), and were able to withdraw from participating at any given time throughout the experiment, without any penalty applied. After completion of the experiment, participants were given a period of four weeks to withdraw their data by contacting the lead researcher. They were provided with both a verbal and written debrief which explained the true aims and predictions of the experiment. Participants who were assigned to the stereotype threat conditions were told that the negative stereotype that they had heard was not a true reflection of their personal or social group's ability (See Appendix B). All participants were given the opportunity to ask any questions regarding the research and were thanked for their time.

3.5. Measuring Mathematical Performance

Modular arithmetic (Gauss, 1801, as cited in Beilock & Carr, 2005) was utilised as a test bed to examine women's mathematical performance. This task uses standard mathematical operations such as subtraction and division but presents them in a novel way (Bellinger, DeCaro, & Ralston, 2015). Across the current experiments, this computerised task was administered via E-Prime experimental software. Participants were presented with a set of varying

problems such as $'43 = 16 \pmod{3}'$ and were instructed to subtract the middle number from the first number (e.g., $43 - 16$) and then divide their answer by the modular number in the brackets (e.g., $27/3$). Participants then responded 'true' if their final answer resulted in a whole number and 'false' if this dividend resulted in a decimal number. Modular arithmetic is typically taught at the highest levels of mathematics education in school settings, particularly in the U.S., and is therefore an advantageous laboratory task because many undergraduate students will not have been exposed to such mathematical operations (Beilock & Carr, 2005). Accordingly, this particular task was deemed suitable for use in the current thesis because it may be able to control for practice effects and task familiarity to a greater extent compared to using standardised national tests such as the Scholastic Assessment Test (SAT) and the General Certificate for Secondary Education (GCSE). A breadth of recent research has employed modular arithmetic to examine the role of working memory in mathematical cognition under conditions of mathematics anxiety, choking under pressure and stereotype threat (Beilock & Carr, 2005; Beilock et al., 2007; Rydell et al., 2014), advocating the use of this measure.

Nevertheless, because this task requires a dichotomous response (true vs. false), procedures were put in place to exclude any participants who might have guessed or pressed any response key to get through the task quickly. For example, in three of the five experiments that utilised modular arithmetic in this thesis, 50% of the problems presented to participants were 'true' and the other 50% were 'false' (unknown to participants). In line with previous research (c.f., DeCaro, Rotar, Kendra, & Beilock, 2010), participants who scored below chance (50%) on this task were excluded from data analysis. The other two

experiments did not use equal numbers of “true” and “false” problems and therefore outliers ($\pm 3 SD$) were screened in accordance with procedures outlined by Tabachnick and Fidell (2013). In such instances, the deviating scores were replaced with the next highest or lowest extreme score. Details regarding the numbers of excluded participants and detected outliers in each experimental study are outlined in their respective chapters.

3.6. Approach to Data Analysis

Data were checked to ensure that they met assumptions of normality and homogeneity of variance (c.f., Tabachnick & Fidell, 2013). In instances where normality was violated, log transformations were performed. Such instances are reported specifically in each experimental study. Bonferroni-corrected pairwise comparisons were conducted to elucidate significant main effects and interactions. Two-tailed hypothesis testing was used and significant alpha levels were accepted below the conventional level of $p < .05$. Effect sizes are reported as partial eta squared and Cohen’s d -scores (c.f., Cohen, 1992).

3.6.1 Approach to mediation analysis. Mediation analysis is a statistical method which allows researchers to identify the variables that underpin the observed relationship between an independent variable (X) and a dependent variable (Y) via the inclusion of a third intermediary variable (M) (Hayes, 2013). Figure 2 presents a simple mediation model. Here the simple relationship between the independent variable (X) and the dependent variable (Y) is referred to typically as the total effect (denoted in the model as path c). The indirect effect is the relationship between the independent variable (X)

and the dependent variable (Y) when the mediator (M) is included in the model (denoted as path c'). Complete mediation is said to be present if the confidence intervals for the indirect effect does not include zero (Shrout & Bolger, 2002). Partial mediation is said to occur when the indirect path between the independent (Y) and dependent variable (X) is reduced in absolute size but is still significantly different from zero when the mediator is introduced (MacKinnon, Fairchild, & Fritz, 2007).

Within the current thesis, mediation analysis was conducted using ordinary least squares path analysis (PROCESS macro; Hayes, 2013). Recent research suggests that path analysis provides more statistical power compared to traditional analyses, such as the causal steps analysis proposed by Baron and Kenny (1986), which is no longer recommended (Hayes, 2009; 2013; Fritz & MacKinnon, 2007; Zhao, Lynch, & Chen, 2010). Experimental conditions were dummy coded ($k - 1$, Hayes, 2013; Hayes & Preacher, 2014), with the reference group changed to examine each stereotype threat prime (Cohen, Cohen, West, & Aiken, 2003). Path coefficients are presented in unstandardised form, which is recommended when the independent variable is multi-categorical (Hayes, 2013).

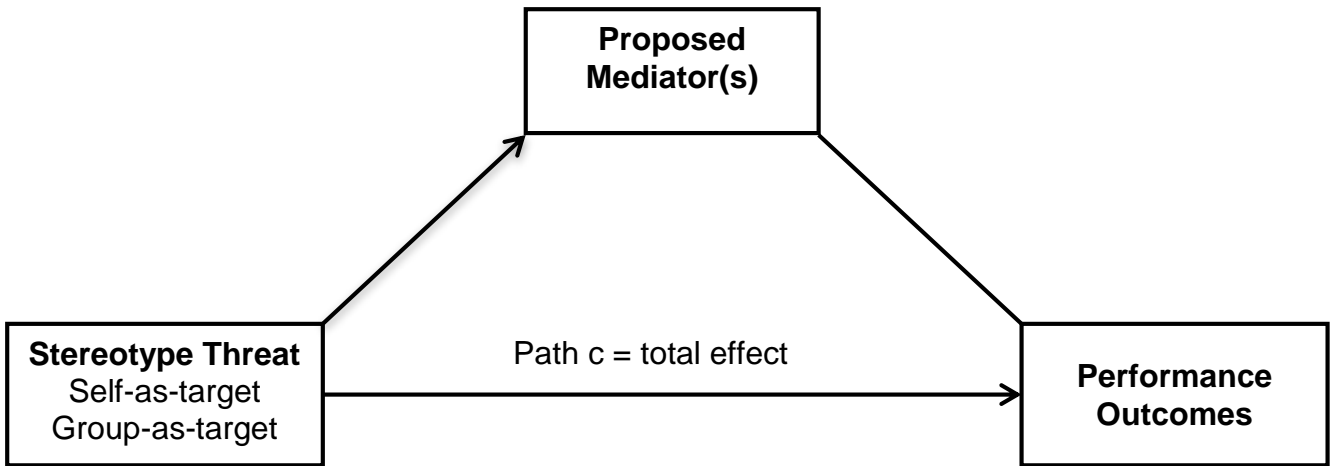


Figure 2. Simple mediation model.

4. CHAPTER 4 – Experiment One

Different Threats, Different Effects? The Influence of Self- and Group-Based Stereotype Threat on Women's Mathematical Performance

Abstract

Aim: The current study employs the multi-threat framework to examine whether self-relevant and group-relevant stereotypes exert different effects on women's mathematical performance. It elucidates further whether deficits in working memory underpin these effects. **Method:** Eighty-four female participants were assigned randomly and equally to a self-as-target or group-as-target stereotype threat condition or to a control condition. They completed a modular arithmetic test with working memory demand manipulated through problem difficulty and spatial orientation. **Results:** Findings indicate that participants under self-as-target and group-as-target stereotype threat solved fewer horizontally oriented problems relative to the control condition. Furthermore, self-as-target stereotype threat appeared to have the greatest detrimental impact on performance with participants solving fewer vertically oriented problems relative to the control. **Conclusion:** These findings suggest that ruminations garnered from negative self- and group-based stereotypes may co-opt the verbal working memory resources required to successfully solve horizontally presented problems. Moreover, negative self-relevant stereotypes may generate greater self-focused attention with associated performance decrements.

4.1. Chapter Overview

The systematic literature review in Chapter 2 highlighted that researchers have predominantly utilised group-relevant manipulations to induce stereotype threat-performance effects. However, such primes overlook the role of the self in stereotype threatening situations, presenting an avenue for additional research. In addition, the review highlighted the on-going debate regarding the mediating processes that are proposed to underpin stereotype threat effects, particularly with regard to whether deficits in working memory or enhanced motivation account for these observed performance decrements. Underpinned by the multi-threat framework (Shapiro & Neuberg, 2007), the current study sets out to examine the impact of self- and group-relevant stereotypes on women's mathematical performance. In line with previous research (Beilock et al., 2007; Rydell et al., 2014; Schmader & Johns, 2003), it also aims to elucidate whether mathematical problems that rely heavily on verbal working memory resources are more susceptible to failure under stereotype threat.

4.2. Introduction

Working memory captures variations in general executive resources critical for coping with acute stressors and functions to maintain the accessibility of task-related goals, control attention, and minimise the influence of intrusive thoughts whilst completing resource-demanding tasks (Kane, Conway, Hambrick, & Engle, 2007; Kane & Engle, 2003; Rosen & Engle, 1998). Working memory is thought to play a critical role in mathematical problem solving (Ashcraft & Krause, 2007; Cragg & Gillmore, 2014), and research indicates that working memory capacity may be depleted in high-pressure situations, particularly

when an individual strives to perform well (Beilock & Carr, 2005). Accordingly, it has been hypothesised that the situational salience of a negative gender-maths stereotype may reduce women's working memory resources by interfering with their ability to control attention and disrupting the accessibility of task-related goals (Beilock et al., 2007; Rydell et al., 2009; Schmader & Johns, 2003). Directly examining this notion, Schmader and Johns (2003) found that women's working memory capacity was depleted when they were primed with a negative stereotype pertaining to their social group's mathematical ability. These results were the first to suggest that the added burden of a negative stereotype may impede working memory capacity, with this mediating the deleterious effects of stereotype threat on women's mathematical performance.

By now it has become apparent that working memory may not be a unitary construct, but instead consists of multiple separate sub-systems (Baddeley & Logie, 1999; Friedman & Miyake, 2000; Shah & Miyake, 1996). It is theorised that working memory is made up of four components; a limited capacity central executive, a phonological loop for storing verbal information, a visual-spatial sketchpad for storing visual images, and an episodic buffer which integrates this information (c.f., Baddeley, 1986; 2000). The verbal-visuospatial distinction has been utilised to explain differences in mathematical task demands as a function of verbal and visuospatial processing requirements (DeStefano & LeFevre, 2004; Trbovich & LeFevre, 2003). It has been suggested that mathematical problems which are presented in a horizontal format require the verbal maintenance of intermediate steps in memory and appear to rely more heavily on the phonological component of the working

memory system (Trbovich & LeFevre, 2003). Conversely, vertically oriented problems are likely to require spatial processing and seem to rely on visuospatial resources (Logie, 1995; Trbovich & LeFevre, 2003). Underpinned by this model, researchers have begun to delineate the precise components of working memory that account for stereotype threat effects (Beilock et al., 2007; Rydell et al., 2014). For example, Beilock et al. (2007) found that women solved fewer horizontally oriented difficult problems compared to vertically oriented problems when they perceived a mental arithmetic test to be diagnostic of gender-related ability. Indeed, pressure-induced worries may place higher demands on the phonological aspect of working memory, which is theorised to support inner speech and thinking in the computation of complex mathematical problems (Ashcraft & Kirk, 2001; Carlson, 1997; Miyake & Shah, 1999). These results suggest that individuals' ruminations about conforming to a pejorative stereotype may debilitate their ability to focus attention on task-relevant information in the face of interference, with performance decrements most pronounced on tasks that rely heavily on verbal working memory resources (Beilock et al., 2007; Rydell et al., 2014).

However, this prior research has employed exclusively group-relevant primes to examine the effects of stereotype threat on working memory (e.g., Beilock et al., 2007; Rydell et al., 2014; Schmader & Johns, 2003). Recent research, however, suggests that stereotype threat can operate through multiple pathways, which target either the self or the social group (Shapiro & Neuberg, 2007; Shapiro et al., 2013; Wout et al., 2008). From this perspective, individuals may experience self-as-target stereotype threat when they perceive that their performance is self-characteristic of personal ability, but may

experience group-as-target threat when they apprehend performance to be confirmative of their group's abilities (Shapiro & Neuberg, 2007). Nonetheless, research has yet to examine whether these distinct forms of stereotype threat place different demands on working memory and whether self-as-target stereotype threat may have a greater detrimental impact on performance.

It is plausible that women's mathematical performance is impaired under conditions designed to increase self-focused attention. In line with this suggestion, research indicates that performance pressure may increase as a function of the personally felt importance of a situation, and that self-consciousness may have a paradoxical effect on task efficiency (Baumeister, 1984; Schmader, Croft, & Whitehead, 2013; Van-Loo, Boucher, Rydell, & Rydell, 2013). The applicability of a negative group stereotype to an individual's personal ability may therefore lead to self-doubt and trepidation that performance may be consistent with expectations (c.f., Baumeister & Showers, 1986). In such situations, the desire to perform optimally may influence individuals to 'choke under pressure', leading to performance decrements (c.f., Beilock & Carr, 2001; 2005; Beilock et al., 2004). Resultantly, self-relevant stereotype may disrupt further the working memory resources necessary to solve mathematical problems because attention is drawn to one's own abilities. Such situational pressure may be more disruptive because it increases attention directed towards the self (i.e., self-consciousness) that may result in excess worry (Baumeister, 1984; Beilock & Carr, 2005).

The current study therefore aims to examine the effects of self-as-target and group-as-target stereotype threat on women's mathematical performance. In accordance with previous research (Beilock et al., 2007), it was predicted

that participants primed with a negative self-relevant and group-relevant stereotype would solve fewer mathematical problems relative to the control condition. Specifically, these performance deficits should be more pronounced for horizontal relative to vertical oriented problems because these rely more heavily on verbal working memory resources (c.f., Beilock et al., 2007). Expanding on previous research (Baumeister, 1984; Beilock et al., 2004; Beilock & Carr, 2005), it was also predicted that self-as-target stereotype threat might have a greater detrimental effect on performance because it generates self-focused attention, which serves to heighten situational performance pressure.

4.3. Method

4.3.1. Participants

Eighty-four female participants ($M_{age} = 21.60$, $SD = 5.43$; 81% university students; 90.5% White British) participated in return for course credit or £3 payment. Using a random number generator, they were assigned equally and randomly to one of three experimental conditions; (1) self-as-target stereotype threat, (2) group-as-target stereotype threat, and (3) a non-threat control. Sample size per condition ($n = 28$) is comparable to that employed in previous stereotype threat studies (c.f., Nadler & Clark, 2011; Nguyen & Ryan, 2008 for meta-analyses).

4.3.1.1. Domain identification and perceived mathematical ability. In order to control for similar levels of perceived mathematical ability and domain

identification in the sample, participants were asked to report their level of mathematical ability (“I am good at maths”) and the degree to which they valued the importance of this (“It is important to me that I am good at maths”; Beilock et al., 2007; Steele, 1997). Responses were recorded on a Likert scale anchored between 1 (Strongly Disagree) and 9 (Strongly Agree). Participants self-reported maths skills and domain identification ($M = 5.87$, $SD = 1.71$) were above average, and did not significantly differ as a function of experimental condition, $p > .05$ ³. Previous research has suggested that moderately domain identified individuals are most susceptible to stereotype threat (Nguyen & Ryan, 2008).

4.3.2. Stereotype Threat Manipulations

In accordance with the multi-threat framework (Shapiro & Neuberg, 2007), participants were primed with either a self-as-target or group-as-target stereotype threat.

Self-as-target stereotype threat. Female participants assigned to the self-as-target condition were primed with a negative gender-related stereotype that was linked explicitly to their personal ability. The following information, designed to heighten the salience of their personal identity, was provided:

³ Moderation analyses also indicated that perceived mathematical ability and domain identification did not moderate the effects of stereotype threat.

“There is a negative stereotype that females have less mathematical aptitude comparative to males. You are a female and this maths exam is therefore diagnostic of **your personal** mathematical ability”.

Group-as-target stereotype threat. Participants in the group-as-target condition were primed explicitly that their performance would be diagnostic of gender-related ability. Here they were provided with information that heightened the salience of their social identity:

“There is a negative stereotype that females have less mathematical aptitude comparative to males. This maths exam is therefore diagnostic of **females’** mathematical ability”.

In both of these conditions, participants were first told about the negative stereotype concerning women in mathematics. This was based on Shapiro and Neuberg’s (2007) proposition that individuals may only be susceptible to self-as-target and group-as-target stereotype threat when they recognise that they belong to a negatively stereotyped group.

Control condition. In line with prior research (e.g., Rydell, Rydell, & Boucher, 2010a; Schmader & Johns, 2003; Steele & Davies, 2003), participants in the control condition were informed that the experiment was investigating the role of working memory and was not diagnostic of mathematical ability.

4.3.3. Measures

Modular arithmetic test. Participants completed a modular arithmetic test (c.f., Beilock & Carr, 2005; Beilock et al., 2007). This computerised task was administered via E-Prime experimental software and required participants to judge the validity of 50 problems, including 2 practice problems. Here participants were taught to answer questions such as ' $43 = 16 \pmod{3}$ ', by subtracting the middle number from the first number (e.g., $43 - 16$) and then dividing their answer by the number in brackets (e.g., $27/3$). Participants responded 'true' if the division resulted in a whole number and responded 'false' if the division resulted in a decimal number.

In line with Beilock et al. (2007), working memory demand (WM; high, intermediate or low) was manipulated by function of operation (borrow or carry operations; Lee & Kang, 2002) and presentational format (horizontal vs. vertical; Trbovich & LeFevre, 2003). Sixteen problems were considered low in working memory demand, requiring a single no borrow subtraction operation (e.g., $8 = 2 \pmod{2}$). Sixteen problems were considered intermediate, requiring a double digit no borrow operation (e.g., $76 = 62 \pmod{14}$). Sixteen problems were considered high demand, requiring a double digit borrow operation (e.g., $62 = 47 \pmod{7}$). Both high and intermediate demand problems are theorised to be more difficult compared to low working memory problems. This is because larger arithmetic problems (double-digits) are practiced less frequently and, as a result, are stored at lower levels of strength in long-term memory (c.f., Ashcraft & Krause, 2007; LeFevre, Sadesky, & Bisanz, 1996; Siegler & Shrager, 1984). Half of the problems were presented horizontally and half were presented vertically. Previous research suggests that the subtraction procedure

of modular arithmetic places the greatest demands on working memory (Beilock & Carr, 2005; Beilock, Kulp, Holt, & Carr, 2004). Therefore, horizontal and vertical problem orientation was altered at this point of the modular arithmetic problem. See Figure 3 for a schematic example.

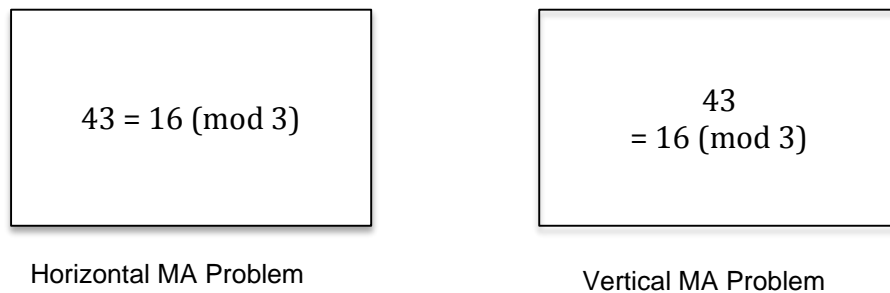


Figure 3. Example of a horizontal and vertical high demand problem, adapted from Beilock et al. (2007).

Participants were instructed to judge modular arithmetic problems as quickly and as accurately as possible (Beilock & Carr, 2005), pressing the computer keys 'Z' and 'M' to answer 'true' or 'false' respectively. Each problem appeared immediately and remained on the screen until participants had provided an answer. Accuracy scores were computed by dividing the number of problems answered correctly by the total number of problems. The percentage of correct scores was used as the dependent variable, with lower scores indicating lower modular arithmetic accuracy (Beilock & Carr, 2005; Beilock et al., 2007; Rydell et al., 2010a).

4.3.4. Procedure

After being allocated randomly and equally to an experimental condition, participants completed two self-report questions that measured their perceived mathematical ability and the importance that they attributed to these skills. Participants were then seated individually at a computer and were provided with “additional task information” which featured the experimental prime. They were taught how to solve modular arithmetic problems through on-screen task instructions, before completing two practice problems and a block of 48 test problems. Performance feedback was not provided and participants were not allowed to correct for mistakes. At the end of the experiment, participants were thanked for their participation and were given both a verbal and written debrief which emphasised that the negative stereotypes they had heard were not a true reflection of their ability.

4.4. Results

Modular arithmetic accuracy was examined in a 3 (Condition: self-as-target, group-as-target, control) x 3 (WM: high, intermediate, low) x 2 (Orientation: horizontal, vertical) Analysis of Variance (ANOVA). Experimental condition was analysed as a between-participants factor. Working memory demand and problem orientation were analysed as within-participant factors. Accuracy outliers ($\pm 3 SD$, $n = 5$) were treated in line with procedures outlined by Tabachnick and Fidell (2013). Accuracy scores were positively skewed; however, log-transformations did not influence the obtained results. For clarity of interpretation, the raw means are reported.

Problem Demand and Presentation

There was a significant main effect of problem demand, $F(2, 162) = 35.48$, $p < .001$, $\eta_p^2 = .31$. In line with predictions, pairwise comparisons indicated that performance accuracy declined for high ($M = .77$, $SD = .19$) relative to low demand problems ($M = .88$, $SD = .16$), $p < .001$, $d = - .63$. Accuracy also declined for intermediate ($M = .73$, $SD = .20$) compared to low demand problems ($M = .88$, $SD = .16$), $p < .001$, $d = - .83$. There was no significant difference in performance accuracy between high and intermediate working memory demand problems, $p > .05$.

There was a significant main effect of problem orientation, $F(1, 81) = 26.96$, $p < .001$, $\eta_p^2 = .25$. Accuracy was significantly lower for horizontally oriented problems ($M = .77$, $SD = .16$) compared to vertically oriented problems ($M = .82$, $SD = .16$), $p < .001$, $d = - .31$. This was qualified by a significant two-way interaction between problem demand and orientation, $F(2, 162) = 8.98$, $p < .001$, $\eta_p^2 = .10$. Accuracy was significantly lower when participants performed horizontally oriented high ($M = .73$, $SD = .21$) and intermediate demand problems ($M = .68$, $SD = .22$) relative to low demand problems ($M = .89$, $SD = .18$), $p < .001$, $d = - .82$ and $- 1.04$, respectively. Furthermore, vertically oriented high ($M = .81$, $SD = .21$) and intermediate demand problems ($M = .78$, $SD = .23$) were solved less efficiently compared to vertically oriented low demand problems ($M = .88$, $SD = .18$), $F(2, 68) = 13.86$, $p < .001$, $d = - .36$ and $- .48$, respectively. Accuracy appeared to decrease when high demand problems were presented horizontally ($M = .73$, $SD = .21$) compared to vertically ($M = .81$, $SD = .21$) $p < .001$, $d = - .38$, and when intermediate problems were

presented horizontally ($M = .68$, $SD = .22$) compared to vertically ($M = .78$, $SD = .23$, respectively), $p < .001$, $d = - .44$. Accuracy for low working memory problems did not differ as a function of problem orientation, $p > .05$. See Figure 4 for two-way interaction.

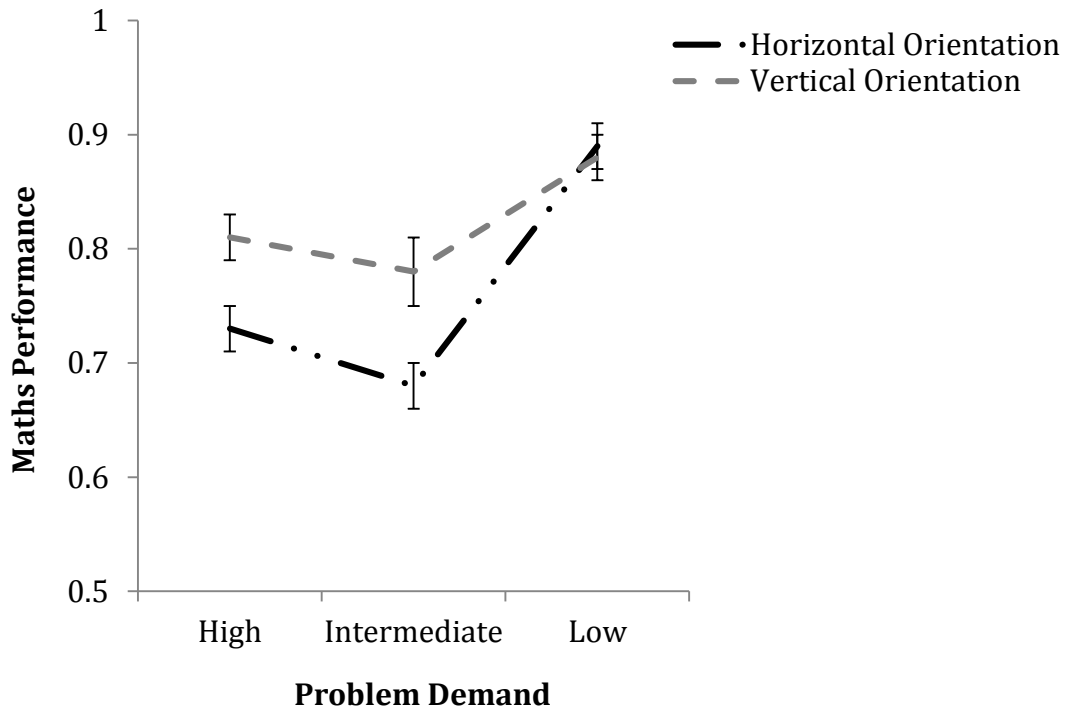


Figure 4. Mean modular arithmetic accuracy scores (%) as a function of problem demand and orientation. Error bars represent standard deviations.

Stereotype Threat

In terms of overall accuracy, there was a significant main effect of experimental condition, $F(2, 81) = 8.32, p = .001, \eta_p^2 = .17$. Pairwise comparisons indicated that participants in the self-as-target stereotype condition solved fewer problems ($M = .71, SD = .18$) compared to participants in the control condition ($M = .87, SD = .09$), $p < .001, d = -1.12$. Group-as-target stereotype threat did not appear to have a significant effect on overall accuracy and there was no significant difference between the self-as-target and group-as-target conditions, $p > .05$.

A two-way interaction between experimental condition and problem orientation was also obtained, $F(2, 81) = 3.61, p < .05, \eta_p^2 = .08$ (see Figure 5). Pairwise comparisons indicated that participants solved fewer horizontally oriented problems under self-as-target stereotype threat ($M = .67, SD = .18$) compared to control participants ($M = .86, SD = .09$), $p < .001, d = -1.34$. They also solved fewer vertically oriented problems ($M = .75, SD = .19$) compared to participants in the control condition ($M = .88, SD = .13$), $p < .05, d = -.80$. In contrast to the main effect, this interaction also revealed that participants primed with a group-as-target stereotype solved fewer horizontally oriented problems ($M = .76, SD = .16$) relative to control participants ($M = .86, SD = .09$), $p < .05, d = -.77$. Participants in the self-as-target condition solved fewer horizontal ($M = .67, SD = .18$) compared to vertically oriented problems ($M = .75, SD = .19$), $p < .01, d = -.43$. Participants in the group-as-target condition also solved fewer horizontal problems ($M = .76, SD = .16$) compared to vertical problems ($M = .84, SD = .14$), $p < .001, d = -.53$. See Figure 5 for two-way

interaction. Participants' accuracy in the non-threat control condition did not significantly differ as a function of problem orientation, $p > .05$. There was no significant difference between the self-as-target and group-as-target conditions with regards to horizontal problem accuracy and vertical problem accuracy $p > .05$. There was no significant interaction between experimental condition and problem demand, $F(4, 162) = .48, p > .05, \eta_p^2 = .01$. Participants' response times did not differ significantly as a function of experimental condition, $F(2, 81) = .56, p > .05, \eta_p^2 = .01$, suggesting that their accuracy scores were not due to a speed-accuracy trade off.

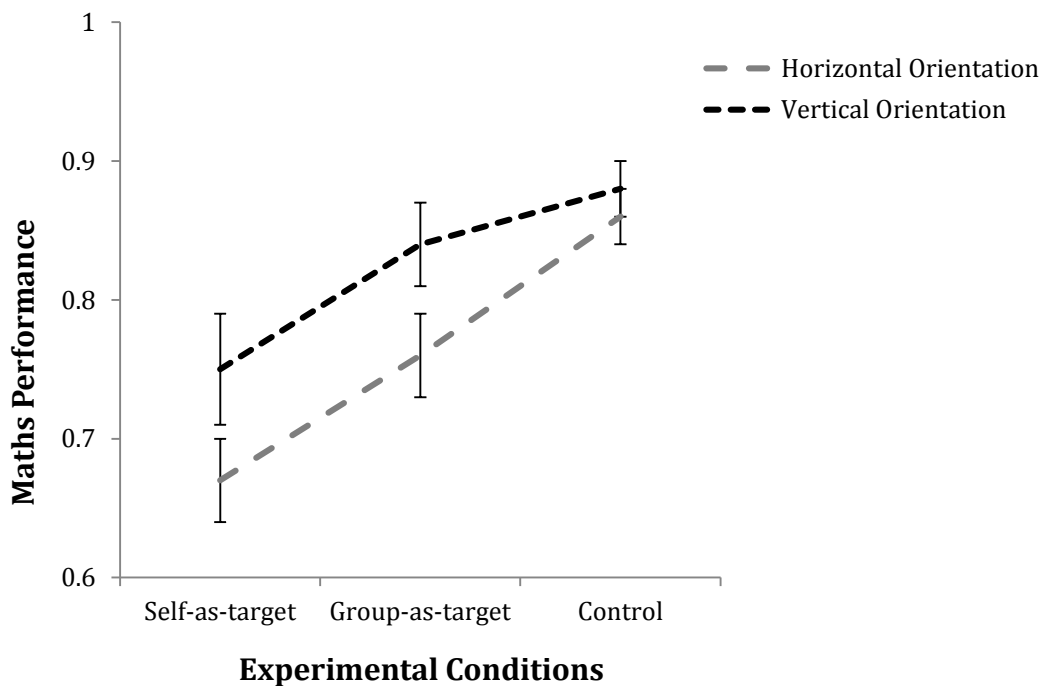


Figure 5. Mean modular arithmetic accuracy scores (%) as a function of experimental condition and problem orientation. Error bars represent standard deviations.

4.5. Discussion

The current study distinguished between self-as-target and group-as-target stereotype threat to examine the impact that these different primes exert on women's mathematical performance. It elucidated further whether performance deficits were more likely to occur for problems which are hypothesised to rely more heavily on verbal, relative to visuospatial, working memory. Findings indicate that participants solved fewer horizontally oriented problems compared to vertically oriented problems in both the self-as-target and group-as-target stereotype threat conditions. Participants in the self-as-target condition also solved fewer vertically oriented problems relative to the control condition. These findings suggest that performance may be harmed to a greater extent when individuals apprehend that a negative societal stereotype may be a true representation of their personal ability. This may generate greater self-focused attention, leading individuals to 'choke under pressure' (Beilock & Carr, 2001; 2005; Beilock et al., 2004; Van-Loo et al., 2013).

Participants primed with a negative self-relevant stereotype appeared to underperform on both horizontally and vertically oriented problems, whereas participants primed with a negative group-relevant stereotype underperformed on horizontal problems only. This finding is consistent with previous research suggesting that performance is hampered in high-pressured situations, particularly when excess attention is directed towards the self (Baumeister, 1984; Beilock et al., 2004). In relation to the current results, the applicability of a negative group stereotype to an individual's personal ability may create a distracting environment that shifts attention to task-irrelevant cues, such as

worries about the situation and its consequences (Beilock & Carr, 2001; 2005). This may lead individuals to monitor their performance and actively regulate negative thoughts and feelings, which co-opt the working memory resources required to solve mathematical problems (Schmader et al., 2008).

Female participants in both the self-as-target and group-as-target stereotype threat conditions solved fewer horizontally oriented mathematical problems compared to vertically oriented problems. Horizontally oriented problems are theorised to rely heavily on phonological aspects of the working memory system (Trbovich & LeFevre, 2003), and this finding may suggest that stereotype threat impacts performance by inducing intrusive thoughts and worries (Beilock et al., 2007; Rydell et al., 2009). Nevertheless, the current research also found that participants under self-as-target stereotype threat underperformed on vertically oriented problems. This is in contrast to the results presented by Beilock et al. (2007), who found that only horizontally oriented mathematical problems were susceptible to stereotype threat effects. However, Beilock et al. (2007) primed participants with a group-as-target stereotype, and did not investigate the possible impact of self-as-target stereotype threat on women's mathematical performance. Under self-as-target stereotype threat participants may experience a higher degree of performance pressure because their personal ability is linked to a negative stereotype. In turn, this added pressure to perform well, and to disconfirm the stereotype, may also interfere with the visuospatial component of working memory, particularly if one concocts visual images of feared consequences (Beilock, 2008; Shackman et al., 2006). In support of this contention, research seems to demonstrate the negative effects that stereotype threat can exert on spatial mental workload (i.e., spatial

orientation; Hausmann, 2014; Moè & Pazzaglia, 2012), and has indicated that women's heightened anxiety in mathematics may consume both verbal and visuospatial components of working memory (Ganley & Visilyeva, 2014).

The current findings highlight the importance of delineating between distinct stereotype threats, and investigating the impact that these may exert on mathematical performance. However, there are a number of potential limitations that should be acknowledged when interpreting the current findings. The results highlight that a self-as-target stereotype influenced performance adversely for both horizontal and vertical problems, whereas a group-as-target stereotype only impacted performance on horizontal problems. However, there was no significant difference between the self-as-target and group-as-target conditions on overall performance accuracy. In contrast to the multi-threat framework (Shapiro & Neuberg, 2007), it may therefore be asserted that these stereotype threats do not have a *distinct* effect on performance. That is, a negative self-relevant stereotype threat did not have a greater detrimental impact on performance relative to a group-relevant stereotype. From this perspective, it could be argued that women may experience simultaneously both self-as-target and group-as-target stereotype threat. Put more simply, women may apprehend that their alleged poor mathematical performance will be self-characteristic of personal ability, and thereby a true representation of their group's ability (c.f., Steele & Aronson, 1995). In support of this, research suggests that an individual's personal and social identity can become fused in so far that they value the outcomes of the group as their own (Swann, Jetten, Gómez, Whitehouse, & Bastain, 2012), and may regard their personal and social identity as functionally equivalent (Swann et al., 2009).

An alternative explanation is that distinct experiences of stereotype threat may be moderated by different factors (Shapiro & Neuberg, 2007; Wout et al., 2008). For example, research indicates that women who are highly identified with their gender underperform when they are primed with a group-as-target stereotype threat, whereas both high and low identified women may underperform in self-as-target stereotype threat conditions (Wout et al., 2008). In the current study, female participants were moderately identified with the mathematics domain and it could therefore be argued that they may have been more susceptible to self-as-target stereotype threat relative to group-as-target stereotype threat. However, sensitivity analyses indicated that domain identification did not moderate the effects of either of these stereotype threats on performance, lessening concern for this potential issue.

A key limitation of the current study, however, is that it did not employ a manipulation check to examine whether participants endorsed the negative stereotype presented to them, and attributed this to their personal or social identity. Future research would therefore benefit from examining the effectiveness of stereotype threat manipulations, and exploring whether stereotype endorsement is a factor that heightens females' susceptibility to self-as-target and group-as-target threats.

The current findings support a working memory account of stereotype threat. It advances this theory by elucidating how this mechanism may explain the impact of self- and group-relevant stereotypes on women's mathematical performance. However, it is viable to question whether working memory moderates or mediates stereotype threat effects. Although working memory is often represented as a general cognitive construct, it can also be

conceptualised as an individual-difference variable, with some individuals displaying higher working memory capacity than others (Beilock, 2008; c.f., Unsworth & Engle, 2007 for an overview). In line with this, R gner et al. (2010) found that individuals who had lower working memory capacity solved fewer problems under stereotype threat compared to those with higher working memory. In relation to the current study, it could be questioned whether participants underperformed due to the effects of stereotype threat on working memory resources, or because participants in these conditions have lower working memory capacity in relation to those in the control condition. Nonetheless, this issue was controlled, to some extent, by randomly assigning participants to an experimental condition. Furthermore, the current study employed an experimental approach by manipulating mathematical problems as a function of working memory demand and problem orientation. This was in line with recommendations which propose that researchers may benefit from adopting designs where the hypothesised mediator is manipulated, due to difficulties in justifying stringent causal assumptions in designs in which the mediator is measured (Bullock, Green, & Ha, 2010; Spencer, Zanna, & Fong, 2005).

In addition, research suggests that performance decrements under stereotype threat may only transpire on difficult problems (Keller, 2007; O'Brien & Crandall, 2003; Spencer et al., 1999). Horizontally orientated problems do not follow a typical solving convention and are considered to be more difficult (Trbovich & LeFevre, 2003). Conversely, vertically oriented problems appear to be less difficult because this format activates a representation of the problem in the visuospatial sketchpad, which is similar to how people solve problems

with pencil-and-paper (Trbovich & LeFevre, 2003). Resultantly, participants may have solved fewer horizontally oriented problems relative to vertically oriented problems not due to their taxation on verbal working memory resources *per se*, but due to their greater difficulty. Nevertheless, other researchers have used additional tasks to provide support for the suggestion that stereotype threat taxes executive functioning to bring about decrements in women's mathematical aptitude (Rydell et al., 2014), and have also implicated this finding in working memory tasks that are not associated with mathematical performance (Schmader & Johns, 2003). The current results are therefore consistent with previous research suggesting that stereotype threat may interfere with performance by reducing targeted individuals' working memory capacity (Beilock et al., 2007; Rydell et al., 2014; Schmader & Johns, 2003).

4.5.1. Chapter Conclusion

The current study distinguished between self-as-target and group-as-target stereotype threats to explore the impact these have on women's mathematical performance. Based on previous research (Beilock et al., 2007), it also examined whether mathematical problems that rely heavily on verbal working memory resources are more susceptible to failure under stereotype threat. Findings indicate that participants primed with self-as-target stereotype threat solved fewer problems compared to participants in the control condition. This may suggest that situational performance pressure is heightened when a negative societal stereotype is attributed to one's personal ability, resulting in an overall detrimental effect on performance. Participants in the self-as-target and group-as-target conditions solved fewer horizontally oriented problems

compared to participants in the control condition. They also solved fewer horizontally relative to vertically oriented problems. Consistent with previous research (Beilock et al., 2007; Rydell et al., 2014), these findings may suggest that the salience of negative stereotypes may influence verbal ruminations, which deplete the verbal working memory resources required to solve mathematical problems.

However, an alternative theory suggests that individuals may be motivated to disconfirm a negative stereotype, and as a result, may exhibit diminished inhibitory control when required to solve mathematical problems (Jamieson & Harkins, 2007; Seitchik & Harkins, 2015). For example, research testing this theory suggests that stereotype threatened participants may respond quicker than participants not under stereotype threat, and will also correct for their answers if they realise that they are incorrect. In Experiment 4, there was no significant difference in the time it took participants to answer each mathematical problem as a function of experimental condition. This may indicate that participants engaged in a similar amount of effort when completing the task. However, given that they were not able to correct for any errors within the modular arithmetic task, a motivational account of stereotype threat effects cannot be ruled out as an explanation for the observed performance deficits. Experiments 2 and 3 therefore employed an anti-saccade eye-tracking task to pit support for the working memory interference account of stereotype threat (Beilock et al., 2007; Schmader & Johns, 2003; Rydell et al., 2014) against the mere effort motivational account (Jamieson & Harkins, 2007; 2009; Seitchik & Harkins, 2015).

5. CHAPTER 5 – Experiments Two and Three

Comparisons of the Mere Effort and Working Memory Accounts Across Distinct Stereotype Threats

Abstract

Aim: The current research examines whether deficits in working memory or heightened motivation underpin the effects of distinct stereotype threats on women's mathematical performance. **Method:** In Experiment 2 of this thesis, female participants were assigned randomly to a negative self-relevant or group-relevant stereotype threat condition or to a control condition. They completed an anti-saccade eye-tracking task, in which they were instructed to look directly towards (pro-saccade) and away from a peripheral target (anti-saccade). In Experiment 3, participants were assigned randomly to a negative or positive group stereotype condition or to a control condition. They completed an anti-saccade eye-tracking task and a modular arithmetic task. **Results:** Results from Experiment 2 indicate that participants in the self-as-target and group-as-target conditions launched marginally fewer corrective saccades relative to the control condition. However, there was no significant differences in the time it took participants to launch correct and corrective saccades towards and away from the target. Findings from Experiment 3 indicate that participants solved fewer difficult mathematical problems when they were primed with a positive group-relevant stereotype, but a negative group stereotype did not appear to impact performance. Furthermore, there were no significant differences in visuospatial performance on the anti-saccade task as a function of experimental condition. **Conclusion:** The current experiments were unable to provide support for the mere effort or working memory interference accounts of stereotype threat. The discussion focuses on potential explanations for these findings, with regard to task difficulty and stereotype endorsement.

5.1. Chapter Overview

Through the lens of Baddeley's (1986, 2000) multi-component model, researchers have proposed that the verbal ruminations garnered from a negative stereotype may reduce verbal working memory resources to bring about detriments in mathematical performance (Beilock et al., 2007; Schmader & Johns, 2003). Providing initial support for a working memory interference account, findings from Experiment 1 indicate that women solved fewer horizontally oriented mathematical problems relative to vertically oriented problems when they were primed with a self-as-target or group-as-target stereotype. Indeed, previous research has shown that horizontally oriented problems rely more heavily on phonological aspects of working memory, whereas vertically oriented problems place demands on the visuospatial component of this system (Beilock et al., 2007; Trbovich & LeFevre, 2003). These results may therefore suggest that stereotype-relevant worries operate like a resource-demanding secondary task that taxes the phonological component of working memory to diminish performance. However, other researchers have argued that motivation, and not deficits in working memory, underpin the stereotype threat-performance relationship (Jamieson & Harkins, 2007; Jamieson & Harkins, 2009; Seitchik & Harkins, 2015). Experiments 2 and 3 of the current thesis therefore pit the working memory interference account against the mere effort motivational account to examine whether these two theories can explain the effects of distinct stereotype threats on women's mathematical performance.

5.2. Introduction

The mere effort account (Jamieson & Harkins, 2007) proposes that the potential for evaluation facilitates the dominant response on a stereotype-relevant task. When the dominant response is correct, performance remains unharmed, but when this dominant response is incorrect, performance is debilitated. Of central importance to this theory, however, is the proposition that stereotype threatened females will correct for their performance if they recognise that their response is incorrect and are given the opportunity to correct it (Jamieson & Harkins, 2007; McFall, Jamieson, & Harkins, 2009).

One task that allows for the comparison of a working memory and mere effort account of stereotype threat is the anti-saccade eye-tracking task, specifically because optimal performance requires working memory resources to generate volitional saccades to a peripheral target (pro-saccade trials) and inhibit the tendency to look towards it (anti-saccade trials) (Jamieson & Harkins, 2007; Munoz, & Everling, 2004; Unsworth, Schrock, & Engle, 2004). The mere effort account predicts that participants under stereotype threat should look in the wrong direction towards the target more often than participants in the control condition because increased motivation facilitates the dominant response. It predicts further that this heightened motivation will influence stereotype threatened participants to launch quicker correct saccades (eye movements directed correctly towards the target) and corrective saccades (eye movements directed towards the target following the emission of an incorrect response) compared to participants who are not subject to evaluation (c.f., Jamieson & Harkins, 2007; Experiment 3). This contrasts with a working memory

interference account, which would predict that participants under stereotype threat would launch slower correct saccades and be less likely to correct for incorrect responses because of diminished working memory capacity (Jamieson & Harkins, 2007). Although working memory is implicated in both the mere effort and working memory interference account of stereotype threat, the mere effort account proposes that participants are quicker to launch saccades because the dominant response has been potentiated, not because individuals lack the working memory capacity necessary to inhibit this response (Jamieson & Harkins, 2007).

Providing support for their theory, Jamieson and Harkins (2007; Experiment 3) found that participants primed with stereotype threat launched more reflexive saccades on a greater percentage of anti-saccade trials (i.e., incorrect responses) and appeared to launch corrective saccades more quickly than those in the control condition. There was also a tendency for participants under stereotype threat to launch reflexive eye movements more quickly on pro-saccade trials. In support of these initial findings, McFall et al. (2009) found that although participants who were subject to evaluation launched more incorrect saccades, they were quicker to produce correct and corrective saccades compared to control participants. These findings suggest that participants may be motivated to disprove negative stereotypes, which results in them launching faster saccades and correcting for any incorrect responses.

However, Schmader et al. (2008) disagree with the assertion that increased motivation to correct errors is incompatible with a working memory interference explanation of stereotype threat. They argue that despite appearing motivated to correct for their mistakes, stereotype threatened

participants in Jamieson and Harkins' (2007) study continued to produce incorrect responses. This suggests that their ability to inhibit the dominant response and produce goal-directed behaviour was diminished; a finding consistent with impaired working memory (Kane & Engle, 2003; Mitchell, Macrae, & Gilchrist, 2002).

There is therefore considerable debate regarding whether working memory deficits or motivation (i.e., mere effort) underpin the effects of stereotype threat on women's mathematical performance. Moreover, no research has examined whether these different explanations account for the effects of distinct stereotypes on performance. In their theoretical review, Shapiro and Neuberg (2007) speculate that different mechanisms may underpin self-as-target and group-as-target stereotype threat; but such assertion has not been tested empirically. Bridging this gap in the literature, the aim of Experiment 2 was to pit support for the working memory interference account against the mere effort account when females are primed with distinct stereotype threats.

In the current study, female participants were primed with a negative self- or group-relevant stereotype and completed an anti-saccade eye-tracking task in which they had to generate volitional saccades towards (pro-saccade) and away from a peripheral target (anti-saccade). Both the mere effort and working memory account predict that participants primed with a negative group-relevant stereotype should launch more incorrect saccades towards the target on anti-saccade trials relative to control participants. However, it is argued that these two theories make different predictions regarding correct and corrective anti-saccades (Jamieson & Harkins, 2007). In line with the mere effort account,

it would be predicted that participants under group-as-target stereotype threat would be more likely to correct for any incorrect responses and launch both correct and corrective saccades more quickly compared to participants in the control condition. Conversely, the working memory interference account predicts that participants under group-as-target stereotype threat would be less likely to correct for incorrect responses. This theory also predicts that stereotype threatened participants would launch correct saccades slower, rather than faster, because the salience of a negative group stereotype taxes the working resources necessary to potentiate these responses (c.f., Jamieson & Harkins, 2007).

Nevertheless, research is yet to examine whether the salience of a negative self-as-target stereotype heightens motivation or diminishes working memory capacity. From a mere effort perspective (Jamieson & Harkins, 2007), it is plausible that female participants may be motivated to disconfirm a negative gender-related stereotype when it is explicitly tied to personal ability. Accordingly, participants primed with self-as-target stereotype threat may launch correct and corrective saccades more quickly compared to participants in both the group-as-target stereotype threat and control conditions. They may also be more likely to correct for incorrect responses compared to participants in the group-as-target condition. Conversely, from a working memory perspective (Beilock et al., 2007; Rydell et al., 2014; Schmader & Johns, 2003), the applicability of a negative societal stereotype to one's personal performance may heighten self-consciousness, thereby depleting working memory capacity. Underpinned by this account, it would be predicted that participants primed with a self-as-target stereotype would launch slower correct and corrective

saccades compared to participants in both the group-as-target stereotype threat and control conditions. See Table 4 for overview of experimental predictions.

Table 4.

Experiment 2 predictions for performance on anti-saccade trials based on the mere effort and working memory accounts of stereotype threat.

	Group-as-target stereotype threat (GAT)		Self-as-target stereotype threat (SAT)	
	Mere effort	Working memory	Mere effort	Working memory
Correct %	GAT fewer than control	GAT fewer than control	SAT fewer than control	SAT fewer than control
Correct RT	GAT quicker than control	GAT slower than control	SAT quicker than GAT and control	SAT slower than GAT and control
Corrective %	GAT more than control	GAT fewer than control	SAT more than GAT and control	SAT fewer than GAT and control
Corrective RT	GAT quicker than control	GAT slower than control	SAT quicker than GAT and control	SAT slower than GAT and control

5.3. Experiment 2 Method

5.3.1. Participants

Sixty-four females ($M_{age} = 22$ years, $SD = 5.53$; 95.3% university students) participated in exchange for course credit or £3 remuneration. They were assigned randomly, using a random number generator, to one of three experimental conditions: 1), self-as-target stereotype threat; 2), group-as-target stereotype threat; and 3), a non-threat control. Decisions regarding sample size were based on Jamieson and Harkins' study (2007; Experiment 3) and recruitment was stopped once a sample of 60 participants had been met through online sign ups to the study. In accordance with previous research (Beilock et al., 2007; Steele, 1997), participants answered two self-report questions to measure their perceived mathematical ability (i.e., "I am good at math") and their domain identification (i.e., "It is important to me that I am good at math"). Responses were recorded on a Likert scale anchored between 1 (Strongly Disagree) and 9 (Strongly Agree). There were no significant differences in participants' perceived mathematical ability, $F(2, 55) = 2.40$, $p > .05$, $\eta_p^2 = .08$, or domain identification as a function of experimental condition, $F(2, 56) = .42$, $p > .05$, $\eta_p^2 < .01$.

5.3.2. Stereotype Threat Manipulations

Group-as-target stereotype threat. Participants in the group-as-target stereotype threat condition were provided with written information which stated that the task they were about to complete was diagnostic of gender-related ability (c.f., Aronson et al., 1999; Shapiro & Neuberg, 2007). This

particular prime was adapted from that used by Jamieson and Harkins (2007, p. 548) and led participants to believe that the anti-saccade task is a measure of visuospatial capacity that is indicative of mathematical ability:

The eye-tracking task that you are about to complete is a test of visuospatial capacity. This measure is closely linked to maths ability. As you may know, there has been some controversy about whether there are gender differences in maths and spatial ability. Previous research has demonstrated that gender differences exist on visuospatial and mathematical tasks. Specifically, females are shown to perform less accurately compared to males. The task that you are about to complete will therefore provide a measure of the differences between **male and females** visuospatial and mathematical ability.

Self-as-target stereotype threat. Participants in the self-as-target stereotype threat were also primed with a negative gender-related stereotype⁴ but instead were told that the task they were about to undertake was diagnostic of personal ability (c.f., Shapiro & Neuberg, 2007). Specifically, they received the following written information:

The eye-tracking task that you are about to complete is a test of your visuospatial capacity. This measure is closely linked to your maths ability. As you may know, there has been some controversy about whether there are gender differences in maths and spatial ability. Previous research has

⁴ Research suggests that participants should be knowledgeable of a negative stereotype in order to be susceptible to stereotype threat effects (Shapiro & Neuberg, 2007). Accordingly, both the self-as-target and group-as-target prime included reference to gender differences in visuospatial and mathematical performance.

demonstrated that gender differences exist on visuospatial and mathematical tasks. Specifically, females are shown to perform less accurately compared to males. The task that you are about to complete will therefore provide a measure of **your personal** visuospatial and mathematical ability.

Given the similarities between the two experimental manipulations, the researcher also verbally primed participants that the task they were about to undertake was diagnostic of personal (self-as-target) or gender-related ability (group-as-target) before they commenced the task.

Control condition. Participants in the control condition were informed that the anti-saccade task was non-diagnostic of ability (c.f., Steele & Davies, 2003) and that the experiment was investigating the role of working memory (Schmader & Johns, 2003).

5.3.3. Measures

Anti-saccade eye-tracking task. All participants reported normal or corrected-to-normal vision. Participants' eye movements were recorded using an EyeLink 1000 head-mounted eye-tracker, with a sampling rate of 1,000 Hz. The experimental tasks were designed using Experiment Builder (SR Research Ltd) and participants' heads were stabilised by a chin rest situated 57cm from the computer monitor. Both the anti-saccade and pro-saccade task consisted of 84 trials, split into 4 blocks (including 4 practice trials). Each trial started with a fixation cross that appeared on the screen and was presented randomly for 800-1000ms. A target then appeared 8° from the fixation point and was

presented randomly on the right or left hand side of the computer screen for 1000ms. The targets consisted of a square (neutral stimuli) or a number (numerical stimuli), which were presented randomly and equally across trials. These two target-types were selected because previous research which investigates inhibition from a mere effort account has utilised neutral stimuli (Jamieson & Harkins, 2007; Experiment 3), whereas numerical stimuli has been used to elucidate the working memory interference account (Rydell et al., 2014; Experiment 3). It was therefore deemed appropriate to examine any potential different effects that the type of stimulus may exert on inhibition. Each target was exactly the same size (1.4°) to ensure that this did not influence inhibitory control (c.f., Roberts, Hager, & Heron, 1994).

5.3.4. Procedure

Participants were recruited for a study that ostensibly examined factors relating to problem solving and were assigned randomly and equally to one of three experimental conditions; self-as-target stereotype threat, group-as-target stereotype threat, and a control condition. Participants were seated in front of a computer and their eye movements were validated using a 9-point calibration system. Before they commenced with the anti-saccade eye-tracking task, they were provided with “additional task information”, which corresponded to their experimental condition. On-screen instructions explained how to respond to anti-saccade and pro-saccade trials. During anti-saccade trials participants were instructed to look directly away from the target, to its mirror position, as quickly and as accurately as possible. During pro-saccade trials participants were asked to look directly towards the target. Participants completed 4 blocks

of 80 pro-saccade trials and 4 blocks of 80 anti-saccade trials, with task order (i.e., pro-saccade or anti-saccade first) counterbalanced between participants. Appropriate breaks were provided throughout to minimise fatigue. After the task, participants responded to two questions in order to evaluate the effectiveness of the stereotype threat manipulations. In accordance with Jamieson and Harkins (2007), participants were asked: “To what extent are there gender differences in visuospatial performance?” (1 = gender differences, 10 = no differences) and “Who do you believe performs better on this task?” (1 = females, 10 = males). Participants were then provided with a verbal and written debrief and were thanked for their time.

5.3.5. Data Preparation

In line with Jamieson and Harkins (2007), filters were used prior to data analysis to ensure that eye movements recorded by the eye tracker represented responses to the stimulus presented. The four practice trials were removed from any analyses, resulting in a total of 160 trials for each participant. Eye movements were categorised as ‘valid’ if participants’ initial eye position did not vary by more than 2.82° (50 pixels) from the central fixation cross. Eye movements more than 2.82° were considered as invalid and were removed from analyses. A total of 3% of pro-saccade and 3% of anti-saccade trials were excluded using this criterion. Eye movements were classed as anticipatory if participants initiated saccades in less than 80 milliseconds (ms) (c.f., Crevit & Vandierendonck, 2005; Jamieson & Harkins, 2007) and saccades beginning at 1,000ms or greater were excluded because they could not have been initiated in response to the target. This criterion resulted in the exclusion of another 3%

of anti-saccade trials and 6% of pro-saccade trials. As a total, 9% of pro-saccade and 6% of anti-saccade trials were removed from the analysis. Data from four participants were excluded from the overall analysis because of invalid centre starts and calibration error on the anti-saccade task.

5.4. Results

Stereotype Threat Manipulation Check

A Multivariate Analysis of Variance (MANOVA) was conducted to evaluate the effectiveness of the stereotype threat primes. There was no significant difference in the extent to which participants believed that there were gender differences in visuospatial performance as a function of the three experimental conditions, $F(2, 59) = 1.81, p > .05, \eta_p^2 = .06$. Participants in the group-as-target stereotype threat condition ($M = 6.32, SD = 1.86$) were more likely to endorse that men outperformed women on the anti-saccade task relative to the control condition ($M = 4.95, SD = 1.54$), $F(2, 59) = 4.95, p < .05, \eta_p^2 = .14$. However, there was no significant difference between responses of participants in the self-as-target condition compared to the control condition, $p > .05$. See Table 5 for descriptive statistics.

Table 5.

Means and corresponding standard deviations for stereotype threat manipulation checks in Experiment 2.

	Self-as-target	Group-as-target	Control
To what extent are there gender differences in visuospatial performance?	5.95 (1.99)	6.00 (2.25)	4.95 (1.54)
Who do you think performs better on this task?	5.90 (2.10)	6.32 (1.86) _a	4.53 (1.65) _a

Note: Rows with a common sub-script differ significantly at $p < .05$.

Anti-saccade Task

Two separate repeated measure ANOVAs were conducted on correct saccades and corresponding saccadic reaction times (SRT) as a function of trial type (pro, anti). There was a significant main effect of accuracy, $F(1, 58) = 35.09$, $p < .001$, $\eta_p^2 = .38$. Pairwise comparisons indicated that participants responded more accurately on pro-saccade ($M = .99$, $SD = .02$) relative to anti-saccade trials ($M = .84$, $SD = .20$), $p < .001$, $d = 1.10$. There was also a significant main effect of response time, $F(1, 59) = 205.04$, $p < .001$, $\eta_p^2 = .78$. Participants responded significantly faster on pro-saccade ($M = 182.12$, $SD = 24.38$) relative to anti-saccade trials ($M = 243.36$, $SD = 33.39$), $p < .001$, $d = -2.09$.

A series of between-participant ANOVAs were then conducted on participants' performance on the pro- and anti-saccade trials as a function of

experimental condition. There were no significant differences on any of the dependent variables as a function of the stimulus used (i.e., number vs. shape) and therefore this variable was collapsed within all subsequent analyses.

Pro-saccade trials.

Correct responses (%). There was no significant difference between correct responses on pro-saccade trials as a function of experimental condition, $F(2, 57) = 1.54, p > .05, \eta_p^2 = .05$. All participants responded with above 98% accuracy.

Saccadic response time (SRT). There was no significant difference in the time it took participants to launch saccades towards the target as a function of experimental condition, $F(2, 57) = .66, p > .05, \eta_p^2 = .02$.

Anti-saccade trials.

Correct responses (%). There was no significant difference between correct responses on anti-saccade trials as a function of experimental condition, $F(2, 57) = .04, p > .05, \eta_p^2 = .001$.

Correct responses (SRT). There was no significant difference in the time it took participants to launch saccades away from the target as a function of experimental condition, $p > .05, F(2, 57) = .43, p > .05, \eta_p^2 = .02$.

Corrective responses (%). There was a significant main effect of condition on corrective responses, $F(2, 57) = 3.57, p = .04, \eta_p^2 = .11$. However,

pairwise comparisons indicated that these differences were marginally significant. Specifically, participants in the self-as-target ($M = .54$, $SD = .32$) were marginally less likely to correct for incorrect responses relative to those in the control condition ($M = .75$, $SD = .24$), $p = .07$, $d = .74$. Participants under group-as-target ($M = .55$, $SD = .28$) were also marginally less likely to correct for incorrect responses relative to those in the control condition ($M = .75$, $SD = .24$), $p = .08$, $d = .77$.

Corrective responses (SRT). There was no significant difference in the time it took participants to launch a corrective saccade away from the peripheral target as a function of experimental condition, $F(2, 53) = .30$, $p > .05$, $\eta_p^2 = .01$. See Table 6 for descriptive statistics.

Table 6.

Descriptive statistics for anti-saccade trials as a function of experimental condition in Experiment 2.

	Self-as-target	Group-as-target	Control
	Mean (SD)		
Correct %	.81 (.17)	.81 (.16)	.82 (.21)
Correct SRT	245.99 (38.53)	237.87 (32.28)	246.64 (29.61)
Corrective %	.54 (.32) _a	.55 (.28) _b	.75 (.24) _{ab}
Corrective SRT	401.14 (86.35)	392.08 (78.06)	379.98 (87.51)

Note: % = percentage correct, RT = reaction time.

5.5. Discussion

Experiment 2 utilised the anti-saccade eye-tracking task to examine support for the mere effort account and the working memory interference theory of stereotype threat-performance effects. In line with the mere effort account (Jamieson & Harkins, 2007; 2009; Seitchik & Harkins, 2015), it was predicted that female participants who were primed with a negative group stereotype would launch quicker correct and corrective saccades and correct for any incorrect responses on a greater percentage of trials compared to control participants. It was also predicted that participants under self-as-target stereotype threat would launch quicker correct and corrective saccades relative to both the group-as-target and control conditions. This was based on the proposition that the applicability of a negative group stereotype to females' personal ability may motivate them to disconfirm the stereotype. Conversely, and in line with a working memory interference account (c.f., Beilock et al., 2007; Schmader & Johns, 2003; Rydell et al., 2014), it was predicted that participants subject to self-as-target and group-as-target stereotype threat would launch slower correct and corrective saccades compared to the control condition and launch fewer corrective saccades. This is because the salience of a negative stereotype may co-opt working memory resources (Jamieson & Harkins, 2007).

Findings indicate that participants in both the self-as-target and group-as-target stereotype threat conditions launched marginally fewer corrective saccades relative to participants in the control condition on anti-saccade trials. These findings appear to be consistent with a working memory interference

account of stereotype threat (Beilock et al., 2007; Rydell et al., 2014), suggesting that female participants may have been unaware that they had initiated an incorrect response and were therefore less likely to launch corrective saccades relative to the control condition. In contrast to the hypotheses proffered by both the working memory interference and mere effort account, however, participants in both stereotype threat conditions did not launch fewer correct saccades compared to the control condition. Furthermore, there was no significant difference in the time it took participants to launch correct and corrective anti-saccades as a function of experimental condition.

These results contrast with findings reported by Jamieson and Harkins (2007; Experiment 3) who found that participants primed with a negative group-relevant stereotype launched quicker correct and corrective anti-saccades relative to participants in a control condition. This is despite the current study using approximately the same amount of participants in each condition compared to Jamieson and Harkins (2007) (current, $n = 20$, Jamieson & Harkins', 2007, $n = \sim 18$). However, it appears that participants in the current study were less likely to endorse the stereotype threat manipulation compared to those in Jamieson and Harkins' (2007) study, and this may explain the discrepant findings. In line with Jamieson and Harkins (2007), participants were primed that the anti-saccade eye-tracking task was a test of visuospatial capacity, which is closely linked to mathematical ability. Although visuospatial ability is theorised to be related to mathematical proficiency (c.f., Tosto et al., 2014), the anti-saccade task is a relatively simple task which is predominantly used as a measure of inhibitory control (Munoz & Everling, 2004). As such, participants may not have endorsed this particular task to be a valid measure

related to mathematical ability, which may explain why both the self-as-target and group-as-target primes did not influence anti-saccade performance.

In a similar vein, both the current study and the original by Jamieson and Harkins (2007) did not examine women's mathematical performance. Resultantly, participants may not have associated their performance on the anti-saccade task as a diagnostic indicator of their mathematical ability (as the stereotype threat prime implied). This may have weakened the effectiveness of the manipulation and may explain the null findings. Due to this issue, the current study (and that of Jamieson & Harkins, 2007) is not able to directly associate performance on the anti-saccade task to the effects of stereotype threat on women's mathematical performance. With a view to overcome these potential issues, Experiment 3 aimed to corroborate the findings presented in Experiment 2, utilising both the anti-saccade eye-tracking task and a test of modular arithmetic to examine the effects of a negative group-stereotype on women's mathematical performance.

Recent research suggests that heightened motivation underpins the effects of stereotype threat on females' mathematical performance (Jamieson & Harkins, 2009; Seitchik & Harkins, 2015). For example, Jamieson and Harkins (2009) found that stereotype threatened participants were more likely to adopt a conventional solving approach, which facilitated performance on 'solve' problems but hampered performance on 'comparison' problems. Extending this work, Seitchik and Harkins (2015) examined support for the mere effort and working memory account of stereotype threat effects. Here they identified a dominant response for solving horizontal problems (i.e., the method of adjustment) and taught one group of women how to solve modular arithmetic

problems based on this method. When they were not told how to solve the problems using this strategy, females under stereotype threat solved fewer problems correctly compared to those in the control condition. However, when told which approach to use to solve the problems, stereotype threatened participants solved more problems correctly than control participants. These findings therefore suggest that females under stereotype threat may have difficulty controlling their tendency to utilise easy-to-adopt but incorrect problem-solving approaches. However, within both of these studies, participants were provided with paper-and-pencil to solve problems (c.f., Jamieson & Harkins, 2009; Seitchik & Harkins, 2015), as well as additional paper to show their calculations (c.f., Jamieson & Harkins, 2009). Allowing individuals to solve problems in this way lessens the demands placed on working memory resources and thus limits the extent to which this working memory interference account of stereotype threat can be elucidated (Raghubar, Barnes, & Hecht, 2010; Trbovich & LeFevre, 2003).

On the surface, it may seem that only negative stereotypes diminish performance. However, research has also revealed that a positively stereotyped social identity can influence performance decrements (Baumeister, Hamilton, & Tice, 1985; Cheryan & Bodenhausen, 2000). For example, Cheryan and Bodenhausen (2000) found that Asian American females underperformed on a mathematical test when they were primed with a positive group stereotype relative to a positive personal stereotype. This is consistent with research which suggests that high expectations for personal success may facilitate performance (Baumeister et al., 1985; Rosenthal & Jacobson, 1968), whereas high group-based expectations often lead to diminished performance

(Brown & Josephs, 1999). Additional research also indicates that highly identified male mathematics students underperform when they are primed with both a positive gender and student identity compared to when they are primed with one of these positive social identities alone (Rosenthal & Crisp, 2007). These findings suggest that when individuals are primed with positive group stereotypes they may experience apprehension about positively representing their social group, with such high expectations leading them to 'choke under pressure' (Cheryan & Bodenhausen, 2000; Rosenthal & Crisp, 2007). However, the opposite effects have been found on tests of spatial ability, with research demonstrating that women do better on spatial ability tasks when they are primed with a positive gender-related stereotype (Moè, 2009; Wraga, Duncan, Jacobs, Helt, & Church, 2006). Indeed, it appears that further research is required to elucidate the impact that positive stereotypes exert on visuospatial and mathematical performance.

Experiment 3 therefore examined the effects that a positive and negative group stereotype exerts on women's visuospatial and mathematical performance. The mere effort account (Jamieson & Harkins, 2007) predicts that the salience of a positive group-relevant stereotype threat motivates participants to perform well. Accordingly, it would be hypothesised that participants primed with a positive stereotype would be more likely to look in the wrong direction towards a peripheral target relative to participants in the control condition because motivation facilitates the dominant response (c.f., Jamieson & Harkins, 2007; 2009; Seitchik & Harkins, 2015). Furthermore, they should launch quicker correct and corrective saccades compared to control participants, and correct for any incorrect responses on a greater proportion of

anti-saccade trials, in a bid to confirm the positive stereotype.

However, based on predictions garnered from a working memory interference account (Beilock & Carr, 2005; c.f., also Jamieson & Harkins, 2007), it would be hypothesised that the salience of a positive stereotype may exacerbate situational performance pressure and, resultantly, diminish working memory capacity. In turn, this theory predicts that participants primed with a positive stereotype would launch more incorrect saccades, and correct for these incorrect responses slower and less often compared to participants in the control condition. See Table 7 for experimental predictions.

With regard to mathematical performance, it was predicted that women would solve fewer difficult problems when they were primed with a negative group stereotype (Beilock et al., 2007; Rydell et al., 2014). Furthermore, it was hypothesised that a positive group stereotype threat might facilitate women's performance on simple problems because they are motivated to perform well, but diminish their performance on difficult problems because this heightened expectation for success influences them to 'choke under pressure' (Beilock & Carr, 2005; Cheryan & Bodenhausen, 2000; Rosenthal & Crisp, 2007).

Table 7.

Experiment 3 predictions for performance on anti-saccade trials.

	Negative group stereotype		Positive group stereotype	
	(ST)		(PS)	
	Mere effort	Working memory	Mere effort	Working memory
Correct %	ST fewer than control	ST fewer than control	PS fewer than control	PS fewer than control
Correct RT	ST quicker than control	ST slower than control	PS quicker than control	PS slower than control
Corrective %	ST more than control	ST fewer than control	PS more than control	PS fewer than control
Corrective RT	ST quicker than control	ST slower than control	PS quicker than control	PS slower than control

5.6. Experiment 3 Method

5.6.1. Participants

Sixty females ($M_{age} = 21$ years, $SD = 5.87$) participated in exchange for course credits or £3 by way of remuneration. They were assigned randomly and equally to one of three conditions: 1), negative group stereotype; 2), positive group stereotype; and 3), a non-threat control condition. Decisions regarding sample size were based on Jamieson and Harkins' study (2007; Experiment 3), and recruitment stopped once a sample of sixty participants had been reached.

Participants self-reported mathematical ability, $F(2, 52) = .23, p > .05, \eta_p^2 < .01$, and domain identification did not significantly differ as a function of experimental condition, $F(2, 52) = 1.62, p > .05, \eta_p^2 = .06$.

5.6.2. Stereotype Threat Manipulations

Negative stereotype threat manipulation. Participants in the negative stereotype condition received the same prime that was implemented in Experiment 2.

Positive stereotype threat manipulation. Participants assigned to the positive stereotype condition received the same information as those in the negative stereotype threat condition. However, rather than highlighting a negative gender-related stereotype, they were informed that women typically outperform men on tests of visuospatial and mathematical ability. Specifically, they were provided with the following written information:

The eye-tracking task that you are about to complete is a test of visuospatial capacity. This measure is closely linked to maths ability. As you may know, there has been some controversy about whether there are gender differences in maths and spatial ability. Previous research has demonstrated that gender differences exist on visuospatial and mathematical tasks. Specifically, females have been found to outperform males. The tasks that you are about to complete will therefore provide a measure of the differences between male and females visuospatial and mathematical ability.

In an equivalent procedure to Experiment 2, the researcher also verbally reiterated the stereotype threat primes to participants before they commenced with each task.

Control condition. Participants in the control condition were informed that the task was measuring factors related to working memory and was non-diagnostic of ability (Steele & Davies, 2003).

5.6.3. Measures

Anti-saccade task. Participants completed the same anti-saccade eye-tracking task that was employed in Experiment 2.

Modular arithmetic task. A modular arithmetic task, similar to that employed in Experiment 1, was employed and was presented using E-Prime experimental software (See Method Section, Experiment 1 for additional task information). Participants completed 64 modular arithmetic problems, including 2 practice problems, and were told to respond as quickly as possible without sacrificing accuracy, indicating their responses using the 'Z' or 'M' key on a standard computer keyboard. Problem difficulty was manipulated for each problem, with 32 problems considered simple (low WM), requiring a single-digit no borrow subtraction operation (e.g., $7 = 2 \pmod{5}$) and 32 problems considered difficult (high WM), requiring a double-digit borrow subtraction (e.g., $43 = 16 \pmod{3}$). Accuracy scores were calculated by dividing the number of problems answered correctly by the total number of problems, with lower scores indicating lower accuracy.

5.6.4. Procedure

Participants were informed that the experiment was investigating factors relating to problem solving and would complete the anti-saccade eye-tracking task and a test of arithmetic. Before commencing with each of the tasks, participants were given written information, which corresponded to their experimental condition (see “Stereotype threat manipulation”). Participants completed the same anti-saccade eye-tracking task that was employed in Experiment 2. They were taught how to solve modular arithmetic problems with on-screen instructions and progressed through these at their own pace. After completing two practice problems, they were required to solve a total of 64 test problems. The order of the anti-saccade and modular arithmetic tasks were counterbalanced between participants and there was no significant difference in performance as a function of task order, $p > .05$. At the end of each task, participants completed two questions to examine the effectiveness of the stereotype threat manipulations. After the anti-saccade eye-tracking task, participants were asked: “To what extent are there gender differences in visuospatial performance?” (1 = gender differences, 10 = no differences) and “Who do you believe performs better on this task?” (1 = females, 10 = males). Following the modular arithmetic task, participants’ were asked to respond to the following two questions: “To what extent are there gender differences in mathematical performance?” (1 = gender differences, 10 = no differences), and “Who do you believe performs better on this task?” (1 = females, 10 = males). Upon completion of the experiment, participants were thanked for their participation and received both a verbal and written debrief.

5.6.5. Data Preparation

Data were prepared using equivalent procedures from Experiment 2. Eye movements more than 2.82° were considered as invalid and were removed from analyses. A total of 4% of pro-saccade and 5% of anti-saccade trials were excluded using this criterion. An additional 6% of pro-saccade trials and 3% of anti-saccade trials were excluded because participants initiated saccades less than 80ms or greater than 1,000ms. As a total, 10% of pro-saccade and 8% of anti-saccade trials were excluded. Eye-tracking data from three participants were removed due to invalid centre starts and calibration error. Mathematical accuracy data from four participants were excluded from analyses because they responded to the problems with below 50% accuracy (c.f., Beilock & DeCaro, 2007; DeCaro, Rotar, Kendra, & Beilock, 2010). Overall, seven participants were excluded from analyses due to invalid eye-tracking data or modular arithmetic accuracy.

5.7. Results

Stereotype Threat Manipulation Check

A MANOVA was conducted to examine the effectiveness of the stereotype manipulations. Findings indicated that participants in the positive stereotype condition were more likely to endorse gender differences in visuospatial performance ($M = 6.12$, $SD = 3.41$) relative to participants in the control condition ($M = 3.56$, $SD = 2.55$), $F(2,48) = 3.99$, $p < .05$, $\eta_p^2 = .14$. However, there was no significant difference between the negative stereotype and control condition on this measure, $p > .05$. Participants in the negative stereotype

condition were more likely to endorse that men outperformed women on the anti-saccade task ($M = 6.31$, $SD = 1.49$) relative to both participants in the positive stereotype ($M = 4.41$, $SD = 2.83$) and control conditions ($M = 4.44$, $SD = 1.46$), $F(2,48) = 4.73$, $p < .05$, $\eta_p^2 = .16$. Participants in the negative stereotype condition appeared to endorse more strongly that there were gender differences in mathematical performance ($M = 6.88$, $SD = 2.00$) relative to participants in the control condition ($M = 4.28$, $SD = 2.32$), $F(2,48) = 3.55$, $p < .05$, $\eta_p^2 = .13$. They were also more likely to report that men outperformed women on this task ($M = 7.25$, $SD = 1.18$) relative to participants in the control condition ($M = 5.33$, $SD = 1.68$), $F(2,48) = 3.85$, $p < .05$, $\eta_p^2 = .14$. However, participants in the positive stereotype condition did not differ significantly compared to either the negative stereotype or control condition on these two latter measures. See Table 8 for descriptive statistics.

Table 8.

Means and corresponding standard deviations for stereotype threat manipulation checks in Experiment 3.

	Negative	Positive	Control
	Mean (SD)		
To what extent are there gender differences in visuospatial performance?	5.75 (2.70)	6.12 (3.41) _a	3.56 (2.55) _a
Who do you think performs better on this task?	6.31 (1.49) _{bc}	4.41 (2.83) _b	4.44 (1.46) _c
To what extent are there gender differences in mathematical performance?	6.88 (2.00) _d	5.82 (3.89)	4.28 (2.32) _d
Who do you think performs better on this task?	7.25 (1.18) _e	5.94 (2.86)	5.33 (1.68) _e

Note: Rows with a common sub-script differ significantly, $p < .05$.

Anti-saccade Task

Two separate repeated measure ANOVAs were conducted on correct saccades and corresponding saccadic reaction time (SRT) as a function of trial type (pro, anti). There was a significant main effect for trial-type accuracy, $F(1, 58) = 45.96, p < .001, \eta_p^2 = .44$. Pairwise comparisons indicated that participants responded more accurately on pro-saccade ($M = .98, SD = .07$) relative to anti-

saccade trials ($M = .80$, $SD = .22$), $p < .001$, $d = 1.11$. There was also a significant main effect of reaction time, $F(1, 56) = 92.98$, $p < .001$, $\eta_p^2 = .62$. Participants were significantly faster to respond on pro-saccade ($M = 177.35$, $SD = 26.83$) relative to anti-saccade trials ($M = 251.38$, $SD = 49.27$), $p < .001$, $d = -1.87$.

A series of between-participants one-way ANOVAs were conducted on participants' performance on the pro- and anti-saccade trials. Consistent with Experiment 1, participants' accuracy and response times did not significantly differ as a function of the stimulus used (i.e., number vs. shape), and consequently this variable was collapsed for all subsequent analyses.

Pro-saccade trials.

Correct responses (%). There was no significant differences between correct responses on pro-saccade trials as a function of experimental condition, $F(2, 56) = 1.05$, $p > .05$, $\eta_p^2 = .04$. All participants responded with above 96% accuracy.

SRT. There was no significant difference in the time it took participants to launch saccades towards the target as a function of experimental condition, $F(2, 54) = .26$, $p > .05$, $\eta_p^2 = .05$.

Anti-saccade trials.

Correct responses (%). There was no significant difference between the percentage of correct responses on anti-saccade trials as a function of experimental condition, $F(2, 56) = .45$, $p > .05$, $\eta_p^2 = .02$.

Correct responses (SRT). There was no significant difference in the time it took participants to launch correct saccades away from the target as a function of experimental condition, $F(2, 54), p > .05, \eta_p^2 = .02$.

Corrective responses (%). There was no significant difference between the percentage of corrective saccades launched as a function of experimental condition, $F(2, 54) = .33, p > .05, \eta_p^2 = .01$.

Corrective responses (SRT). There was no significant difference in the time it took participants to correct for any incorrect responses as a function of experimental condition, $F(2, 54) = .001, p > .05, \eta_p^2 < .001$. See Table 9 for descriptive statistics.

Table 9.

Descriptive statistics for anti-saccade trials as a function of experimental condition in Experiment 3.

	Positive group stereotype	Negative group stereotype	Control
	Mean (SD)		
Correct %	.84 (.13)	.78 (.27)	.78 (.24)
SRT Correct	261.60 (67.87)	246.32 (33.71)	246.21 (40.54)
Corrective %	.55 (.32)	.62 (.23)	.59 (.27)
SRT Corrective	353.71 (130.47)	352.50 (89.75)	354.23 (94.58)

Note: % = percentage correct, RT = reaction time.

Modular Arithmetic

Problem demand. There was a significant main effect of problem difficulty on accuracy scores, $F(1, 51) = 145.58, p < .001, \eta_p^2 = .74$. Pairwise comparisons indicated that participants solved fewer difficult ($M = .78, SD = .11$) compared to simple problems ($M = .97, SD = .08$), $p < .001, d = -1.98$.

Stereotype threat. There was no significant main effect of experimental condition on accuracy scores, $F(2, 51) = 4.46, p > .05, \eta_p^2 = .09$. There was a significant interaction between experimental condition and problem demand, $F(2, 51) = 3.51, p < .05, \eta_p^2 = .12$. However, pairwise comparisons indicated that these differences were marginally significant. Specifically, participants in the positive stereotype condition solved marginally fewer difficult problems ($M = .74, SD = .11$) relative to participants in the control condition ($M = .83, SD = .10$), $p = .058, d = -.86$. All other pairwise comparisons were non-significant, $p > .05$. Reaction times did not differ as a function of experimental condition, $p > .05$. See Table 10 for descriptive statistics⁵.

⁵ Mediation results are not presented in Experiment 3 because there was no effect of stereotype threat on anti-saccade performance (the proposed mediator) or mathematical performance (the dependent variable). When mediation analyses are conducted there are no significant paths for any variables.

Table 10.

Mean accuracy scores and corresponding standard deviations as a function of experimental condition and problem demand in Experiment 3.

	Positive group stereotype	Negative group stereotype	Control
	Mean (SD)		
Difficult problems	.74 (.11)	.77 (.11)	.83 (.10)
Simple problems	.85 (.08)	.85 (.10)	.90 (.05)
Overall Accuracy			

5.8. Discussion

Experiment 3 examined the impact of a positive and negative group-relevant stereotype on women's mathematical performance. Utilising the anti-saccade eye-tracking task, it also explored whether enhanced motivation or deficits in working memory may account for these effects. Findings indicate that participants under group-as-target stereotype threat did not appear to launch significantly more corrective saccades on the anti-saccade task compared to the control condition. This is in contrast to the findings presented in Experiment 2, which indicate that participants who were primed with a negative stereotype launched marginally fewer corrective saccades compared to participants in the control condition. Furthermore, participants in the group-as-target stereotype condition did not differ in their ability to initiate correct saccades or the time it took to launch these saccades, which contradicts the findings reported by

Jamieson and Harkins (2007; Experiment 3).

Participants primed with a negative group-relevant stereotype did not appear to underperform on a mathematical test relative to participants in the control condition. This is inconsistent with previous research (c.f., Beilock et al., 2007; Rydell et al., 2014), which shows that the salience of a negative gender-maths stereotype reduces women's mathematical performance. However, inspection of the means indicate that findings were going in the predicted direction, with participants in the negative group stereotype condition solving fewer problems compared to those in the control condition. It is therefore plausible that Experiment 3 had insufficient statistical power to detect a significant finding between groups. Specifically, a total of 7 participants had to be excluded from Experiment 3 due to invalid eye-tracking or modular arithmetic data and this resulted in a lower sample size per condition. Although this sample size is comparable to that of Jamieson and Harkins' (2007; Experiment 3), it is recommended that future research in this area should use larger sample sizes to reliably establish whether negative group-as-target stereotype threat reduces females' visuospatial and mathematical performance.

Results also revealed that participants who were primed with a positive stereotype solved marginally fewer difficult problems relative to the control condition. In line with previous research (Cheryan & Bodenhausen, 2000; Rosenthal & Crisp, 2007), this may suggest that the salience of a positive group stereotype leads individuals to 'choke under pressure', with performance decrements more likely to emerge on problems that rely heavily on working memory. The general discussion focuses on possible explanations for the

findings of Experiment 2 and 3 of the current thesis with regards to the difficulty of the anti-saccade eye-tracking task and the impact that stereotype endorsement may have on performance outcomes.

5.9. General Discussion

Although representing contrasting theories, considerable empirical support has been accrued for the working memory and mere effort explanations of stereotype threat. Despite this, research has not examined whether these two explanations account for the effects of distinct stereotype threats on performance. Experiment 2 examined whether the salience of a negative self- and group-relevant stereotype motivated participants to disprove the negative stereotype (Jamieson & Harkins, 2007) or diminished their working memory capacity to bring about decrements in anti-saccade performance (Beilock et al., 2007; Schmader & Johns, 2003; Rydell et al., 2014). Findings indicate that participants under self-as-target and group-as-target stereotype threat launched marginally fewer corrective saccades relative to participants in the control condition. However, findings from Experiment 3 were unable to corroborate these findings, revealing no significant differences in participants' visuospatial performance as a function of experimental condition.

A closer inspection of the data across both experiments appears to reveal that in Experiment 2, participants in the control condition were more likely to correct for their responses ($M = .75\%$) relative to control participants in Experiment 3 ($M = .59\%$). In comparison, participants who were primed with a negative group stereotype launched corrective saccades on approximately

55% of trials in both experiments. This may explain why a significant difference was found between the stereotype threat conditions and the control condition in Experiment 2, but not Experiment 3, because the former group of control participants were more likely to correct for their incorrect responses. Taking this into consideration, the results of Experiment 2 and 3 are therefore unable to provide support for a working memory or mere effort account of stereotype threat-performance effects due to the null findings obtained.

Experiment 3 examined the influence that a positive and negative group-relevant stereotype had on women's mathematical and visuospatial performance. Findings indicate that a negative group stereotype did not significantly reduce women's mathematical performance compared to the control condition. However, participants in the positive stereotype condition appeared to solve marginally fewer difficult problems when they were primed with a positive group-relevant stereotype relative to the control condition. These results are consistent with previous research (Cheryan & Bodenhausen, 2000; Rosenthal & Crisp, 2007), which suggests that the salience of a positive stereotype may heighten situational performance pressure and influence targeted individuals to 'choke under pressure'. However, caution needs to be taken when interpreting the current study findings, particularly because the difference between the positive stereotype and control conditions did not reach a conventional level of statistical significance (albeit showing a large effect size). Furthermore, no significant differences were found in relation to participants' visuospatial performance under a positive and negative group stereotype. In sum, findings across these two experiments contrast with those reported by Jamieson and Harkins (2007; Experiment 3), who found that

participants under stereotype threat launched more corrective saccades and were quicker to launch correct and corrective saccades, which they suggest is due to heightened motivation to disprove the negative stereotype.

One explanation for the discrepant findings is that the anti-saccade task utilised in Jamieson and Harkins' (2007) study was more difficult than that employed in the current study. For example, in the current study participants were instructed to look directly away (anti-saccade trials) and towards (pro-saccade trials) numerical and neutral targets that appeared on the left or the right of the screen. However, the task employed by Jamieson and Harkins (2007) employed a flashing cue, which was presented before an arrow target appeared on the opposite side of the screen. Participants then had to indicate which direction the target was pointing. As such, the task employed in Jamieson and Harkins' (2007) study may have been more cognitively taxing, which could explain why stereotype threatened participants launched fewer correct saccades compared to the control condition. In a similar vein, participants in the current study completed blocks of anti-saccade and pro-saccade trials separately, and this could have led to familiarisation with the task requirements, lessening the demands placed on working memory. However, this does not explain why the latencies of participants' correct and corrective saccades were not influenced by stereotype threat. Given that the current task was simpler, it could be argued that differences in saccadic response times between experimental conditions should have been detected, particularly if the potential for evaluation motivates participants to disconfirm a negative stereotype. Future research is therefore recommended to explore the mere effort and working memory account of stereotype threat, taking into consideration the difficulty of

the task. This could be achieved through the use of mixed blocks of pro- and anti-saccade trials.

Furthermore, it appears that participants in the current study may have been less susceptible to the stereotype primes employed in comparison to participants in Jamieson and Harkins' study (2007; Experiment 3). In Experiment 2, there were no significant differences in the degree to which participants endorsed gender differences in visuospatial performance as a function of experimental condition. In Experiment 3, although there was a significant difference on this measure between participants in the negative stereotype condition and control condition, inspection of the means indicates that stereotype threatened participants reported a neutral response (at the mid-point of the Likert scale), suggesting that they may have been unsure whether to endorse gender differences or not. This may have masked any potential differences in visuospatial performance within the current study because participants in the stereotype threat conditions may not have endorsed the stereotype that they were primed with. Future research is therefore warranted to examine whether stereotype endorsement moderates the effects of stereotype threat on anti-saccade performance.

There has been considerable debate regarding whether working memory or motivation underpins the effects of stereotype threat on performance. Whilst some researchers argue that these two theories make opposing predictions for task performance under stereotype threat (e.g., Jamieson & Harkins, 2007; Seitchik & Harkins, 2015), others have argued that these two mechanisms may operate concurrently in high-pressured situations (e.g., Schmader et al., 2008). For example, Schmader et al. (2008) theorise

that individuals may be motivated to disconfirm a negative stereotype regarding their group membership, and this may lead to heightened vigilance and monitoring processes that tax working memory. In contrast, Jamieson and Harkins (2007) argue that the mere effort account is incompatible with the working memory interference account. They propose that stereotype threatened individuals launch more correct and corrective saccades because the dominant response has been facilitated, and suggest that this requires an intact central executive. Conversely, they suggest that the working memory account theorises that participants do not launch correct and corrective saccades faster than control participants because the central executive has been compromised. One problem with integrating these two theories is that Jamieson and Harkins' (2007) findings cannot distinguish between the overproduction of a dominant response (consistent with a motivational theory) or the failed inhibition of a dominant response, which would result from diminished working memory and goal neglect (c.f., Schmader et al., 2008; Smith, Jostmann, Galinsky, & van Dijk, 2008). Given the discrepancies between these two accounts of stereotype threat, and the findings from the current studies, it is clear that additional research is required to elucidate whether working memory deficits and motivation may operate simultaneously under stereotype threat.

5.9.1 Chapter Conclusion

Experiment 2 and 3 of this thesis examined whether deficits in working memory or heightened motivation (mere effort) may account for the performance-impinging effects of stereotype threat. Findings from Experiment 2 indicate that

women launched marginally fewer saccades under self-as-target and group-as-target stereotype threat relative to the control condition. However, there were no apparent differences in the time it took stereotype threatened participants to launch correct and corrective saccades relative to the control condition. Findings from Experiment 3 indicate that participants solved marginally fewer difficult mathematical problems when they were primed with a positive stereotype, consistent with theories of choking under pressure (Cheryan & Bodenhausen, 2000; Rosenthal & Crisp, 2007). Nevertheless, there was no significant difference in visuospatial performance as a function of experimental condition. Possible explanations for these findings include the difficulty of the anti-saccade task utilised in the current study and the possibility that stereotype endorsement may moderate performance outcomes. Additional research that employs other tasks to examine the separate components of executive functioning may be fruitful to explore whether deficits in working memory or motivation can account for stereotype threat effects.

With the above in mind, the following two experiments of this thesis employ separate tasks to examine the working memory interference account and the mere effort account of stereotype threat. Specifically, Experiment 4 and 5 examine whether the executive functions of inhibition, updating and switching account for the effects of distinct stereotype threats on women's mathematical performance. These studies also examine whether stereotype threat effects are more likely to emerge when an individual perceives that their performance will be diagnostic of personal or gender-related ability (i.e., self-as-target vs. group-as-target stereotype threat; Experiment 4), or whether performance is diminished when both an individual's personal and social identity is made

salient in the stereotyped domain (termed “combined stereotype threat”; Experiment 5).

6. CHAPTER 6: Experiments Four and Five

The Effects of Different Stereotype Threats Manipulations on Women's Mathematical Performance and Executive Functioning

Abstract

Aim: Building on the work of Rydell et al. (2014) and Shapiro et al. (2013), the current experiments examine whether deficits in executive functioning underpin the effects of distinct stereotype threats on women's mathematical performance. **Method:** In Experiment 4, female participants were assigned randomly and equally to a self-as-target or group-as-target stereotype threat condition or to a non-threat control condition. In Experiment 5, female participants were assigned to a "combined" self- and group-as-target stereotype threat condition or to a control condition. All participants completed a modular arithmetic test and three executive functioning tasks that measured updating, inhibition and shifting. **Results:** Findings from Experiment 4 indicate that women's performance was stifled when they perceived a mathematical test to be diagnostic of gender-related ability, but it did not appear to suffer when the test was regarded solely as diagnostic of personal ability. Updating ability did not appear to mediate the effects of group-as-target stereotype threat on mathematical performance. Findings from Experiment 5 indicate that under "combined" stereotype threat participants solved fewer mathematical problems relative to a control, with reduced updating ability mediating this relationship. **Conclusion:** These findings suggest that individuals may be more susceptible to the negative effects of stereotype threat when both their personal and social identities are made salient in the ability domain. In such situations, the salience of a negative stereotype may tax the verbal working memory resources required to solve difficult mathematical problems.

6.1. Chapter Overview

Findings from Experiments 2 and 3 were unable to provide support for the effects of a self-as-target and group-as-target stereotype on women's visuospatial and mathematical performance. Resultantly, neither the mere effort nor the working memory interference explanation of the stereotype threat-performance relationship could be elucidated. It was suggested that the difficulty of the task used in these experiments might have been one reason that precluded finding significant results. Experiments 4 and 5 therefore utilise separate and more difficult tasks of executive functioning (i.e., inhibition, shifting, and updating) to examine the extent to which these are implicated in the effects of diverse stereotype threats on women's mathematical performance (c.f., Rydell et al., 2014). Specifically, tasks that measure updating ability may reveal the extent to which stereotype threat depletes verbal working memory resources (c.f., Rydell et al., 2014; Schmader & Johns, 2003), whereas tasks that measure inhibitory control may reveal how stereotype threat leads to heightened motivation (and therefore quicker responding, c.f., McFall et al., 2009).

Moreover, at this point in the thesis it was acknowledged that the self-as-target manipulation utilised in Experiments 1 and 2 also made reference to the social group. Specifically, female participants were primed with a negative stereotype pertaining to their social group's ability (i.e., "there is a negative stereotype that females are comparatively bad at maths compared to males") and were then told that the task was diagnostic of personal ability. This particular prime was developed in line with the conditions set out in Shapiro and Neuberg's (2007) multi-threat framework, which suggest that participants need

to recognise that they belong to a negatively stereotyped social group in order to be susceptible to self-as-target stereotype threat. However, it could be argued that this manipulation is not distinct in comparison to the group-as-target stereotype threat manipulation because an individual's social identity is also made salient. Accordingly, Experiment 4 aimed to examine whether performance decrements ensue when female participants are merely informed that their performance will be diagnostic of personal ability, without referring to the negative stereotype that governs their social group. Here a self-as-target and group-as-target stereotype threat manipulation was taken directly from extant research (Shapiro et al., 2013) to examine whether these have a negative effect on women's mathematical performance.

6.2. Introduction

Empirical evidence seems to suggest that negative performance-related thoughts deplete working memory to bring about decrements in women's mathematical performance (Beilock et al. 2007; Schmader & Johns, 2003; Rydell et al., 2014). To re-cap, research suggests that women exhibit reduced working memory capacity when they are primed with a negative gender-maths stereotype (Beilock et al., 2007; Schmader & Johns, 2003), and also report more negative thoughts and worries relating to the task (Beilock et al., 2007; Cadinu et al., 2005). These findings converge to suggest that stereotype threat may saturate verbal working memory resources with internal worries about one's ability, resultantly leading to performance deficits (Beilock, 2008). Consistent with this reasoning, Schmader et al. (2008) propose that stereotype threat leads individuals to actively monitor their performance and suppress

negative thoughts and emotions, which resultantly is seen to diminish their ability to control attention during complex cognitive tasks.

More recent research has suggested that stereotype threat may influence more general executive functioning, in addition to the construct of working memory. Underpinned by this rationale, Rydell et al. (2014) examined whether stereotype threat interferes with women's ability to inhibit dominant behaviours (i.e., inhibition), shift attention from one task to another (i.e., shifting), and hold information in working memory (i.e., updating). Findings indicated that women showed poorer updating ability and inhibition when they were primed with the negative stereotype that women are generally worse at maths compared to men. However, no effects were found for shifting. Furthermore, whereas inhibition partially accounted for the effects of stereotype threat on women's mathematical performance, updating completely mediated this effect. These findings suggest that stereotype threat evokes negative thoughts and feelings that interfere with individuals' ability to regulate attention and coordinate information processing to solve mathematical problems.

In contrast to the working memory interference account (Beilock et al., 2007; Schmader & Johns, 2003; Rydell et al., 2014), research suggests that motivation is the core process of stereotype threat effects (c.f., Jamieson & Harkins, 2007; Seitchik & Harkins, 2015). The mere effort account (Jamieson & Harkins, 2007) posits that individuals are motivated to disconfirm negative gender-related stereotypes, which resultantly harms performance (Jamieson & Harkins, 2007). Specifically, this increased drive increases the emission of a dominant response, which is suggested to facilitate performance on simple tasks, but hinder performance on difficult tasks (Jamieson & Harkins, 2007;

McFall et al., 2009; Seitchik & Harkins, 2015). In accordance with this theory, research indicates that stereotype threatened individuals show better inhibitory control on the classic Stroop interference task (McFall et al., 2009), the Remote Associates Task (Harkins, 2006), and the anti-saccade task (Jamieson & Harkins, 2007; McFall et al., 2009; Experiment 4). However, this only occurs when stereotype threatened participants are given enough time to respond and to recognise that they have made an incorrect response⁶. These findings are therefore inconsistent with the working memory interference account of stereotype threat, which predicts that individuals will show reduced inhibition under stereotype threat because reduced working memory interferes with their ability to deploy inhibitory processes (c.f., Rydell et al., 2014; Schmader et al., 2008).

Utilising the same tasks employed by Rydell et al. (2014), Experiment 4 examined whether the executive functions of updating, inhibition and shifting mediate the effects of self-as-target and group-as-target stereotype threat on women's mathematical performance. In accordance with Shapiro et al. (2013), women were informed that an upcoming mathematical test was either diagnostic of gender-related ability (group-as-target) or personal ability (self-as-target stereotype threat). Through the lens of the working memory interference

⁶ The mere effort account proposes that stereotype threat arouses participants and this facilitates the most dominant response on a task. On the Stroop task, this refers to reading the word, rather than naming the colour of the ink. This theory proposes that if participants are not provided with enough time, then they will underperform on the Stroop interference task, because the dominant response leads them to read the word, rather than naming the colour. However, if they are given enough time, the motivation to be correct will yield a faster reaction time for participants under stereotype threat compared to those in a control condition (c.f., Jamieson & Harkins, 2009; McFall et al., 2009).

account (Rydell et al., 2013), it was predicted that a negative group-as-target stereotype would diminish women's mathematical performance. Consistent with Shapiro et al.'s (2013) findings, it was also predicted that women under self-as-target stereotype threat would solve fewer mathematical problems because the relevance of a pejorative gender stereotype to their personal identity may serve to exacerbate situational performance pressure (c.f., Baumeister, 1984; Schmader et al., 2013; Van-Loo et al., 2013).

It was also predicted that the salience of a negative self- or group-relevant stereotype would reduce women's updating ability, with this mediating the relationship between stereotype threat and underperformance. This is theorised to occur because the verbal working memory resources that are required to update complex mathematical computations in one's mind are expanded on processing negative information regarding the pejorative stereotype (Rydell et al., 2014). However, in line with Rydell et al. (2014) it was also predicted that inhibition and switching would not explain the negative relationship between group-as-target stereotype threat and mathematical performance. Although stereotype threat appears to interfere with women's ability to utilise correct problem solving approaches (c.f., Carr & Steele, 2009), Rydell et al. (2014) argue that it is unlikely that the ability to suppress a prepotent response would uniquely account for the effects of stereotype threat on poor mathematical performance. Instead they theorise that negative gender-related stereotypes threat may interfere with women's ability to update task-relevant information in their mind and block out intrusive thoughts and worries, which distracts them from the task at hand.

6.3. Experiment 4 Method

6.3.1. Participants

Ninety-six female participants ($M_{age} = 21.80$, $SD = 5.62$; 91.7% White British; 95.8% university students) participated in return for partial course credit or £5 remuneration. They were assigned randomly to a self-as-target ($n = 32$) or group-as-target stereotype threat condition ($n = 30$) or to a control condition ($n = 34$). Decisions regarding sample size were made in line with Rydell et al. (2014), of which this particular study is based. An additional 16 participants were originally recruited but were excluded because they had accuracy scores below 50% on the modular arithmetic task (c.f., DeCaro et al., 2010). An additional three participant's data was excluded because they responded with an average reaction times less than 2,000 ms, which was deemed too quick for participants to judge and answer difficult mathematical problems reliably and may be indicative of guessing. Only female participants were recruited based on previous research indicating that males are not susceptible to stereotype threat effects in this task (c.f., Rydell et al., 2014; Experiment 2).

6.3.2. Stereotype Threat Manipulations

The current research employed two distinct stereotype threat primes, derived from previous literature, which are suggested to target either the self or the social group (Shapiro et al., 2013).

Self-as-target stereotype threat. Participants in the self-as-target stereotype threat condition were informed that an upcoming maths test was

diagnostic of personal ability. Specifically, they were provided with the following written information:

In today's session, we want to get a measure of your mathematical ability by having you take a maths test. Your performance on this test will be used to help us establish your personal mathematical ability.

Group-as-target stereotype threat. Participants in the group-as-target stereotype threat condition were told that an upcoming maths test was diagnostic of gender-related ability. Specifically, they were primed with the following information:

In today's session, we want to get a measure of mathematical ability for women and men by having you take a maths test. Your performance on this test will be used to help us establish mathematical performance norms for women and men.

Control Condition. Participants in the non-threat control condition were informed that the test was a non-diagnostic problem-solving task (c.f., Schmader & Johns, 2003; Steele & Davies, 2003). They received the following information:

In today's session, we want to measure different factors that may be involved in problem solving. The test you are about to undertake is non-diagnostic of ability.

6.3.3. Measures

The study utilised the same measures as employed by Rydell et al. (2014; Experiment 1). Participants completed three executive functioning tasks, which measured the executive functions of inhibition, updating and shifting (c.f., Miyake et al., 2000). All tasks were presented using E-Prime software (Schneider, Eschman, & Zuccolotto, 2002) and the order of each task was randomised for each participant.

Inhibition. To measure inhibition, participants completed the Stroop interference task (Stroop, 1935). The stimuli consisted of three colour words “blue”, “red” and “orange” which were presented in congruent or incongruent ink. A fixation cross was presented in the middle of the computer screen for 500ms, which was then replaced with the presentation of each word. The word remained on the screen until participants pressed a keyboard button to indicate the colour of the ink (marked by coloured stickers). There were a total of 96 trials, with 48 congruent and 48 incongruent trials. During congruent trials the colour word appeared in the colour that matched its semantic meaning (e.g., BLUE presented in blue ink). During incongruent trials the colour word appeared in a colour that did not match its semantic meaning (e.g., BLUE presented in red ink). After eliminating incorrect responses (i.e., incorrect or RT < 300 ms or > 2,000 ms), a measure of inhibition was calculated by subtracting participants’ average reaction time for the incongruent trials from the congruent trials. This scoring procedure ensured that higher scores indicated better inhibitory control (Rydell et al., 2014).

Updating. The letter-memory task (Morris & Jones, 1990) was employed to measure updating. Participants were presented with a sequence of letters, which each appeared in the centre of the computer screen for 2,500ms. Participants were required to rehearse the letter sequence and recall the last three letters presented by typing them into a response box when prompted by the computer. There were 10 trials, with five consisting of seven-letter sequences and five consisting of nine-letter sequences⁷. A score of updating ability was computed by dividing the number of correctly recalled letter triads by 10, with a greater proportion of correctly recalled triads indicating greater updating capacity.

Shifting. Participants completed the number-letter task (Rogers & Monsell, 1995) to measure shifting. Across 128 trials, participants were presented with a horizontal line on the computer screen. A cue (a 'box') was presented above (64 trials) or below (64 trials) the line for 150 ms. Subsequently, a letter and a number were presented in the box, either above or below the horizontal line. If the box appeared above the line, participants were required to indicate whether the number was odd or even by pressing one of two response keys ("c" and "m" respectively). If the box appeared below the line, participants had to select whether the letter was a vowel or a consonant, using the same computer keys. There were 64 no-switch trials, in which the box was presented in the same location of consecutive trials, and 64 switch trials, in which the box changed location (from above to below the horizontal line).

⁷ The original Rydell et al. (2014) study utilised triads of 3, 5, and 9. The corresponding author informed us that we should utilise 7 and 9-letter triads only because their recent research indicates greater differences between experimental conditions when participants were required to solve difficult sequences (personal communication).

Accuracy and corresponding reaction times were recorded for no-switch and switch trials. A shifting score was calculated by subtracting participants' average correct response latencies for switch trials from no-switch trials, with higher scores indicating better shifting ability.

Mathematical performance. After receiving instructions of how to solve two practice problems, participants completed a total of 36 modular arithmetic problems. All problems used large numbers and required a borrow operation (i.e., subtraction), making them difficult (Beilock et al., 2007). Half of the problems were presented horizontally and half were presented vertically (Beilock et al. 2007; Trbovich & LeFevre, 2003). Half of the problems were true and the other half used false correlates (e.g., $51 = 19 \pmod{4}$ vs. $51 = 19 \pmod{5}$). Participants were instructed to judge the problems as quickly and as accurately as possible, pressing the 'Z' and 'M' keys on a computer keyboard for 'true' and 'false', respectively. Each problem appeared immediately and remained on the screen until a response had been recorded. Accuracy was tabulated by dividing the total number of correctly solved problems by the total amount of problems, with higher scores indicating better performance.

6.3.4. Procedure

A blind experimental design was utilised in that participants were recruited for a study that ostensibly examined factors relating to problem solving. They were allocated randomly to either the self-as-target or group-as-target stereotype threat condition or to a control condition. In order to control for similar levels of perceived mathematical ability and domain identification, participants responded to two self-report questions; "I am good at maths" and "It is important

to me that I am good at maths” on a Likert scale anchored between 1 (Strongly Disagree) and 9 (Strongly Agree) (Beilock et al., 2007; Spencer et al., 1999; Steele, 1997). Participants were then seated at a computer and were provided with on-screen instructions which instructed them how to solve the modular arithmetic problems. After participants had completed two practice problems, the experimenter provided participants with ‘further study information’, which corresponded to their experimental condition (see “Stereotype Threat Manipulations”). Subsequently, participants completed a set of 18 modular arithmetic problems, and the inhibition, shifting and updating tasks. They were then primed with the stereotype threat manipulation again before completing a final set of 18 modular arithmetic problems. The decision was taken to split the 36 mathematical problems between the start and end of the working memory tasks to ensure that participants perceived the working memory tasks to be related to mathematical performance (c.f., Jamieson & Harkins, 2011b) and to ensure that task order did not influence the findings (c.f., Brodish & Devine, 2009; Logel et al., 2009). At the end of the experiment, all participants completed a three-item measure of threat concern (Marx, 2012), which was adapted to match each of the stereotype threat manipulations employed in the current research. To examine the effectiveness of the self-as-target prime, participants responded to the following question: “I was concerned that the researcher was examining my personal mathematical ability”. In line with the group-as-target stereotype threat, they were asked “I was concerned that the researcher was examining women’s mathematical ability to compare it to men’s mathematical ability”. Finally, to examine the extent to which participants may have experienced both self-as-target and group-as-target stereotype threat,

they were asked: “I was concerned that my mathematical ability would be used as a general indicator of women’s mathematical ability”. Participants were asked to complete these measures contingent upon the stereotype threat manipulation they were presented with. They responded on a 5-point Likert Scale, anchored between ‘Strongly Disagree’ and ‘Strongly Agree’. Participants were then given a verbal and written debrief which explained that any negative stereotypes that they had heard were not a true representation of their ability.

6.4. Results

Perceived Mathematical Ability and Domain Identification

A one-way ANOVA indicated that participants in the self-as-target condition ($M = 5.03$, $SD = 1.84$) perceived that their mathematical ability was lower relative to those in the control condition ($M = 6.35$, $SD = 1.82$), $F(2, 93) = 4.83$, $p = .01$, $\eta_p^2 = .09$. There was no significant difference between the group-as-target and control condition, $p > .05$. An additional one-way ANOVA indicated that there was no significant difference in domain identification as a function of experimental condition, $F(2, 93) = 2.20$, $p > .05$, $\eta_p^2 = .05$. Accordingly, participants’ perceived mathematical ability was entered as a covariate when examining the effects of stereotype threat on mathematical performance.

Manipulation Check for Stereotype Threat

Participants’ responses to each of the three manipulation checks were analysed by means of a MANOVA. Participants in both the self-as-target ($M = 3.38$, $SD = 1.07$) and group-as-target stereotype threat conditions ($M = 2.67$,

$SD = 1.09$) reported a higher level of concern that their mathematical performance would be diagnostic of personal ability relative to those in the control conditions ($M = 2.00$, $SD = 1.04$), $F(2, 93) = 13.65$, $p < .001$, $\eta_p^2 = .23$. Participants in the self-as-target condition also reported a higher level of concern regarding their personal ability relative to participants in the group-as-target stereotype threat condition, $p < .05$. Participants under group-as-target stereotype threat ($M = 3.43$, $SD = .97$) reported a higher level of concern that their performance would be diagnostic of gender-related ability relative to both participants in the self-as-target ($M = 2.34$, $SD = 1.07$) and control conditions ($M = 2.06$, $SD = .95$), $F(2, 93) = 16.53$, $p < .001$, $\eta_p^2 = .26$. Participants in the group-as-target stereotype threat condition also reported a higher level of concern that their performance would be diagnostic of both personal and gender-related ability ($M = 2.90$, $SD = 1.18$) relative to those in the control condition ($M = 2.12$, $SD = 1.07$), $F(2, 93) = 4.88$, $p < .05$, $\eta_p^2 = .07$. All other pairwise comparisons were non-significant, $p > .05$. These findings suggest that the stereotype threat manipulations were successful. See Table 11 for descriptive statistics.

Table 11.

Descriptive statistics for self-report measures in Experiment 4.

	Self-as-target	Group-as-target	Control
	Mean (SD)		
Perceived Mathematical Ability	5.03 (1.84) _a	5.60 (1.50)	6.35 (1.82) _a
Domain identification	5.56 (1.64)	5.60 (1.48)	6.35 (1.98)
Concern for personal identity	3.38 (1.07) _{bc}	2.67 (1.09) _{bd}	2.00 (1.04) _{cd}
Concern for social identity	2.34 (1.07) _e	3.43 (.97) _{ef}	2.06 (.95) _f
Concern for personal and social identity	2.47 (1.19)	2.90 (1.18) _g	2.12 (1.07) _g

Note: Means in a row with different sub-scripts differ significantly different at $p < .05$.

Mathematical Performance

Accuracy scores were analysed in a 3 (Condition: Self-as-target, group-as-target, control) x 2 (Orientation: Horizontal, vertical) mixed-factorial Analysis of Covariance (ANCOVA). There was a significant main effect of condition on mathematical performance, $F(2, 92) = 7.15$, $p = .001$, $\eta_p^2 = .14$. Bonferroni-corrected pairwise comparisons indicated that participants in the group-as-target condition solved fewer mathematical problems ($M = .70$, $SD = .15$) relative to the control condition ($M = .83$, $SD = .12$), $p = .001$, $d = -.96$. Self-as-

target stereotype threat had no significant effect on performance in comparison to the control condition, ($p > .05$), and there was no significant difference between the self-as-target and group-as-target conditions ($p > .05$). There was no significant effect of problem orientation, $p > .05$. A one-way ANOVA revealed no significant differences in participants' reaction times to solve the problems, $F(2, 93) = .63, p > .05, \eta_p^2 = .01$. This suggests that differences in mathematical accuracy scores were not due to a speed-accuracy trade off.

Executive Functioning

Inhibition. A one-way ANOVA indicated that there was no significant difference in participants' inhibitory control as a function of experimental condition, $F(2, 93) = 2.41, p > .05, \eta_p^2 = .05$.

Shifting. A one-way ANOVA indicated that there was no significant difference in participants' shifting ability as a function of experimental condition, $F(2, 93) = .39, p > .05, \eta_p^2 = .008$.

Updating. Participants' accuracy scores for the letter-memory task were submitted to a 3 (Condition: Self-as-target, group-as-target, control) x 2 (Sequence length: 7, 9 strings) mixed-factorial ANOVA. There was no significant main effect of condition, $F(2, 93) = 1.10, p > .05, \eta_p^2 = .02$, and no significant interaction between condition and sequence length, $F(2, 93) = 2.52, p > .05, \eta_p^2 = .05$. See Table 12 for descriptive statistics for all tasks.

Table 12.

Mathematical performance and executive functioning as a function of experimental condition in Experiment 4.

	Self-as-target	Group-as-target	Control
	Mean (SD)		
Maths Performance (%)	.77 (.12)	.70 (.15) _a	.83 (.12) _a
Maths RT	10671.13 (3775.12)	9544.88 (5204.01)	10176.56 (4217.49)
Inhibition (RT)	-67.19 (62.35)	-49.71 (75.17)	-92.88 (95.41)
Shifting (RT)	-275.63 (447.15)	-377.45 (660.55)	-264.70 (554.64)
Updating (ACC)	.78 (.16)	.76 (.18)	.82 (.16)

Note: % = percentage correct, RT = reaction time.

Mediation Analysis

A parallel multi-mediational analysis was conducted using ordinary least squares path analysis (PROCESS; Hayes, 2013). Given that the current experiment employed a multi-categorical independent variable, experimental conditions were dummy coded ($k - 1$; Hayes, 2013; Hayes & Preacher, 2014). Results indicate that participants in the group-as-target stereotype threat condition showed better inhibition compared to the control condition, $a^1 = 43.17$,

$p < .05$. They also showed poorer updating ability compared to the control condition, $a^3 = - .12$, $p < .05$. However, inhibition ($b^1 = .0001$) and updating ($b^3 = .07$) did not predict mathematical performance, $p > .05$ (See Figure 6). A bias-corrected confidence interval for the indirect effects, based on 5,000 bootstrap samples, indicated that inhibition, $\beta = .004$ (LLCI = $- .009$, ULCI = $.03$), updating, $\beta = - .008$ (LLCI = $- .04$, ULCI = $.005$), and shifting, $\beta = - .001$ (LLCI = $- .02$, ULCI = $.004$) did not significantly mediate the effects of group-as-target stereotype threat on women’s mathematical performance.

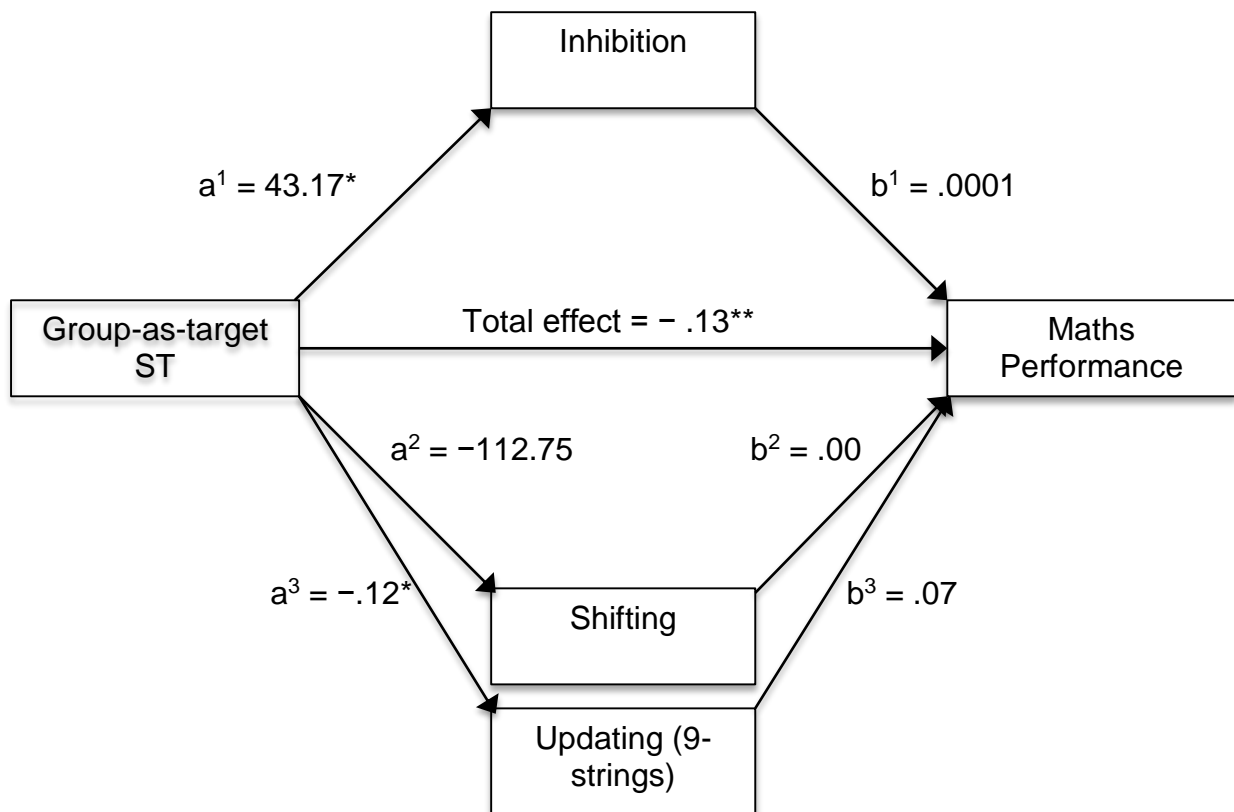


Figure 6. Mediator model: Relationship between group-as-target stereotype threat on mathematical performance through inhibition, shifting and updating. Note: * $p < .05$, ** $p < .001$.

6.5. Discussion

Experiment 4 examined whether self-as-target and group-as-target stereotype threat had a negative impact on women's mathematical performance. It also investigated whether the executive functions of inhibition, shifting and updating mediated these effects. Findings indicate that women solved fewer mathematical problems when they were primed with a negative group stereotype. In contrast with Rydell et al. (2014), stereotype threat did not influence executive functioning, and updating ability did not mediate the relationship between group-as-target stereotype threat and women's mathematical performance. Furthermore, stereotype threatened participants did not show reduced inhibitory control, which is inconsistent with previous research (McFall et al., 2009).

Findings also indicate that female participants did not appear to underperform under self-as-target stereotype threat relative to those in the control condition. The current research employed the self-as-target stereotype threat utilised by Shapiro et al. (2013), and a closer look at this prime may explain this particular finding. Here participants were not presented with a negative stereotype, but were merely informed that their performance would be diagnostic of personal ability. This is in contrast to the conditions proposed by Shapiro and Neuberg (2007), which suggest that individuals should be aware of the negative stereotype to experience both self-as-target and group-as-target stereotype threat. From this perspective, it is plausible that participants under self-as-target stereotype threat may not have experienced the same evaluative pressure compared to those under group-as-target stereotype threat because

their performance was not associated with a negative stereotype. Indeed, this assertion is consistent with Steele and Aronson's (1995) original definition, which theorises that stereotype threat may operate through concerns about how a negative societal stereotype will reflect on both an individual's personal ability and that of their valued social group.

From this perspective, it could be argued that stereotype threat may be more likely to operate through concerns for both the self and the social group. For example, a woman may apprehend that her mathematical performance will be diagnostic of personal ability (self-as-target) and therefore confirm the negative reputation that her social group lacks a valued ability (group-as-target). Providing support for this notion, research suggests that experiences of stereotype threat involve the activation of three core concepts; the concept of one's ingroup, the self-concept, the concept of the ability domain in question (Schmader et al., 2008). Here stereotype threat is said to arise from a state of cognitive imbalance in which an individual's positive self-concept is inconsistent with the expectation that their group should underperform in a stereotype-relevant domain. This claim is supported by research which suggests that an individual's personal and social identity can become fused in so far that individuals value the outcomes of the group as their own (Bilewicz & Kofta, 2011; Swann et al., 2009; Swann et al., 2012). Experiment 5 therefore examines whether stereotype threat effects are more likely to emerge when women are concurrently primed with a self-as-target and group-as-target stereotype threats. In other words, Experiment 5 aims to provide empirical evidence to evaluate whether women are vulnerable to experiencing multiple, qualitatively distinct stereotype threats, which target the self or the social group

(Shapiro & Neuberg, 2007; Shapiro et al., 2013), or whether stereotype threat effects are more likely to occur when both an individual's personal and social identity are highlighted in the stereotyped domain (Schmader et al., 2008). In line with Rydell et al.'s (2014) recent findings, it also aimed to examine further whether reduced updating ability accounts for the effects of stereotype threat on women's mathematical performance.

6.6. Experiment 5 Method

6.6.1. Participants

Sixty-five females ($M_{age} = 24.15$, $SD = 8.28$; 92.7% White British, 95.6% university students) participated in exchange for partial course credit or £5 as a way of remuneration. They were assigned randomly and equally to either a 'combined' stereotype threat ($n = 31$) or a control condition ($n = 34$). In line with Experiment 4, decisions regarding sample size were based on Rydell et al. (2014). An additional seven participants were originally recruited but then excluded because they responded with less than 50% accuracy on the maths test. There was no significant differences in participants' self-reported mathematical ability, or domain identification as a function of experimental condition, $p > .05$.

6.6.2. Stereotype Threat Manipulation

Combined self and group stereotype threat. Based on research which suggests that individuals can experience identity fusion (Swann et al., 2009; 2012), participants were primed simultaneously with both a self-as-target

and group-as-target stereotype threat. Here participants were given the following information that linked mathematical performance to both their personal and social group's ability:

In today's session, we want to get a measure of mathematical ability for women and men by having you take a maths test. Your performance on this test will be used to help us establish your personal mathematical ability. It will also be used to establish the performance norms for women and men.

Participants in the control condition were given the same information as in Experiment 4, which informed them that the task was a non-diagnostic problem-solving task.

6.6.3. Measures

The performance measures and procedure were identical to Experiment 1.

6.7. Results

Stereotype Threat Manipulation

A MANOVA was conducted on participants' responses to each of the three manipulation checks. Results indicate that participants in both the combined stereotype threat and control condition did not significantly differ with regard to their concerns that their mathematical performance would be diagnostic of personal ability, $F(1, 63) = .83, p > .05, \eta_p^2 = .01$. Participants in the combined stereotype threat condition ($M = 2.97, SD = 1.06$) reported a higher level of

concern that their mathematical ability would be diagnostic of gender-related ability relative to the control condition ($M = 2.17$, $SD = .92$), $F(1, 65) = 10.80$, $p < .01$, $\eta_p^2 = .14$. Participants in the combined stereotype threat condition ($M = 2.97$, $SD = 1.06$) also reported a higher level of concern that their mathematical performance would be diagnostic of both personal and gender-related ability relative to the control condition ($M = 1.92$, $SD = .73$), $F(1, 65) = 14.88$, $p < .001$, $\eta_p^2 = .19$. These findings suggest that the stereotype threat manipulation was successful. See Table 13 for descriptive statistics.

Table 13.

Descriptive statistics for self-report measures in Experiment 5.

	Combined ST	Control
	Mean (SD)	
Perceived Maths Ability	5.45 (1.71)	5.50 (1.26)
Domain identification	5.81 (1.56)	5.62 (1.10)
Concern for personal identity	2.52 (1.12)	2.29 (.84)
Concern for social identity	3.00 (1.06) _a	2.15 (.93) _a
Concern for personal and social identity	3.03 (1.35) _b	1.94 (.74) _b

Note: Means in a row with different sub-scripts differ significantly different at $p < .05$.

Mathematical Performance

Accuracy scores were analysed in a 2 (Condition; Combined stereotype threat, control) x 2 (Problem orientation: Horizontal, vertical) mixed-factorial ANOVA. There was a significant main effect of experimental condition, $F(1, 63) = 6.37$, $p < .05$, $\eta_p^2 = .09$. Pairwise comparisons indicate that participants in the combined stereotype threat condition solved fewer mathematical problems ($M = .76$, $SD = .10$) relative to the control condition ($M = .83$, $SD = .12$), $p < .05$, $d = -.63$. There was no significant effect of problem orientation, $F(1, 63) = .35$, $p > .05$, $\eta_p^2 < .01$, and no significant interaction between experimental condition and problem orientation, $F(1, 63) = .35$, $p > .05$, $\eta_p^2 < .01$. Additionally, there was no significant difference in reaction time as a function of experimental condition, $F(1, 63) = 3.55$, $p > .05$, $\eta_p^2 = .05$, suggesting that these results were not indicative of a speed-accuracy trade off.

Executive Functioning

Inhibition. There was no significant difference in participants' inhibitory control as a function of experimental condition, $F(1, 63) = .08$, $p > .05$, $\eta_p^2 = .001$.

Shifting. There was no significant differences in participants' shifting ability as a function of experimental condition, $F(1, 63) = .87$, $p > .05$, $\eta_p^2 = .01$.

Updating. Participants' updating accuracy was submitted to a 2 (Condition: Combined stereotype threat, control) x 2 (Sequence length: 7, 9

strings) mixed-factorial ANOVA. There was a significant main effect of experimental condition on updating accuracy, $F(1, 63) = 7.70, p < .01, \eta_p^2 = .11$. Pairwise comparisons indicated that participants in the combined stereotype threat condition had lower updating ability ($M = .76, SD = .17$) relative to the control condition ($M = .86, SD = .14$), $p < .01, d = -.64$. There was no significant interaction between experimental condition and sequence length, $F(1, 63) = .76, p > .05, \eta_p^2 = .001$. See Table 14 for descriptive statistics.

Table 14.

Mathematical performance and executive functioning as a function of experimental condition in Experiment 5.

	Combined stereotype threat	Control
	Mean (SD)	
Maths Performance (%)	.76 (.10) _a	.83 (.12) _a
Maths RT	11320.04 (3345.52)	9783.33 (3230.13)
Inhibition (RT)	-76.49 (72.60)	-83.78 (128.18)
Shifting (RT)	-290.51 (496.10)	-183.89 (427.03)
Updating (ACC)	.76 (.17) _b	.86 (.14) _b

Note: % = percentage correct, RT = reaction time.

Mediation Analysis

A path analysis was conducted to examine whether the executive functions of inhibition, updating and shifting mediated the effects of combined

stereotype threat on women's mathematical performance. Mediation findings indicate that participants in the combined stereotype threat condition displayed significantly poorer updating ability compared to the control condition ($a^3 = -.11$), with this predicting mathematical performance ($b^3 = .19$), $p < .05$. A bias-corrected confidence interval for the indirect effects, based on 5,000 bootstrap samples, did not include zero, indicating that updating mediated the effects of combined stereotype threat on mathematical performance, $\beta = -.02$, (LLCI = $-.05$, ULCI = $-.003$). Furthermore, inhibition, $\beta = .0003$ (LLCI = $-.007$, ULCI = $.006$), and shifting, $\beta = -.002$ (LLCI = $-.003$, ULCI = $.02$) did not significantly mediate the effects of combined stereotype threat on women's mathematical performance. See Figure 7 for multi-mediator model.

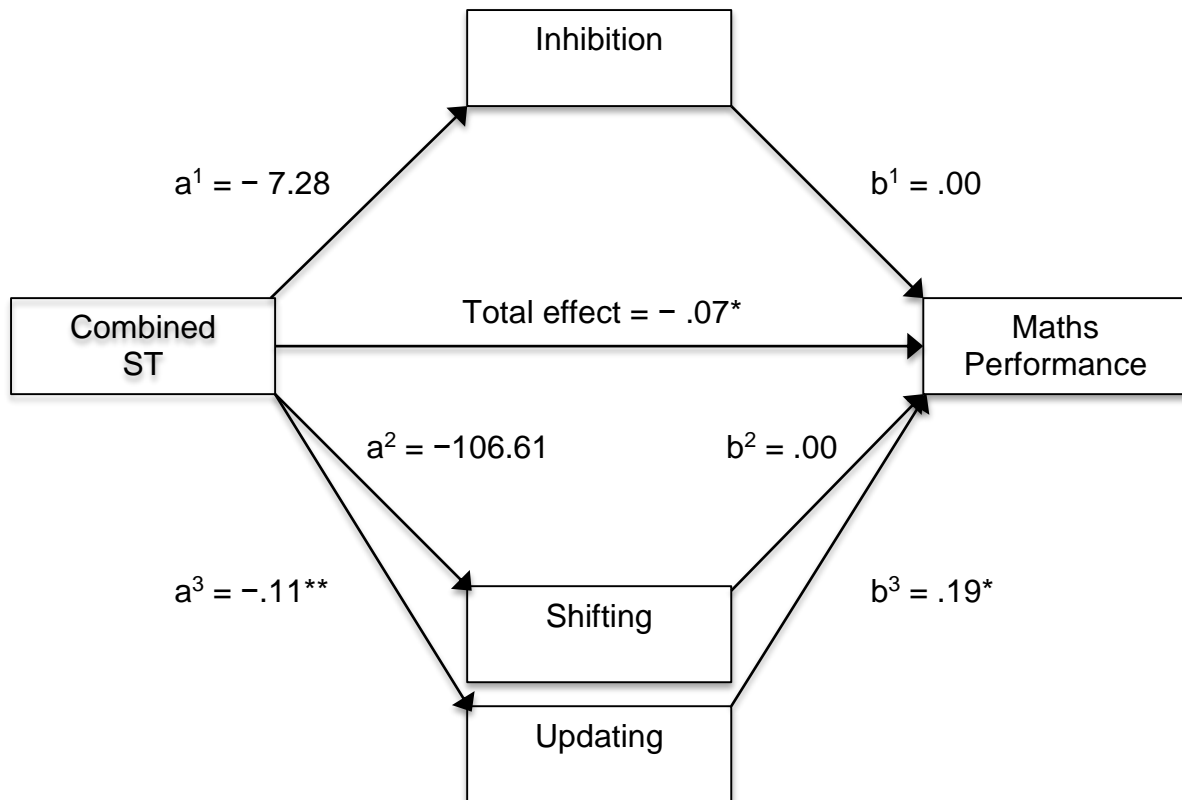


Figure 7. Mediator model: Relationship between combined stereotype threat on mathematical performance through inhibition, shifting and updating. Note: ** $p < .01$, * $p < .05$.

6.8. Discussion

Experiment 5 examined whether women's mathematical performance and updating ability was harmed when they were primed concurrently with a negative self- and group-as-target stereotype threat. Findings indicate that women in the combined stereotype threat condition solved fewer mathematical problems compared to those in the control condition. In line with Rydell et al. (2014), participants in the stereotype threat condition also showed reduced updating ability relative to control participants, with this predicting their poorer mathematical performance. Seemingly in contrast to the multi-threat framework

(Shapiro & Neuberg, 2007; Shapiro et al., 2013), these findings suggest that performance decrements may be more likely to occur when individuals perceive a negative stereotype to be a true representation of their personal aptitude, and therefore characteristic of their social groups ability (i.e., other females). Such assertion is supported by Schmader et al.'s (2008) integrated process model of stereotype threat, which theorises that stereotype threat may create a cognitive imbalance between an individual's positive self-concept, their negative group identity, and expectations for performance.

6.9. General Discussion

Although considerable inroads have been made in elucidating how stereotype threat may operate, the literature still lacks a direct empirical examination of whether self and group-relevant stereotypes hinder performance in distinct ways and the mechanisms that underpin these effects. Combining previous work (Rydell et al., 2014; Shapiro & Neuberg, 2007; Shapiro et al., 2013), Experiment 4 and 5 examined whether the executive functions of inhibition, updating and shifting were implicated in the effects of distinct stereotype threats on women's mathematical performance. Findings from Experiment 4 indicate that women solved fewer problems when they were primed with a negative group-as-target stereotype threat relative to the control condition. This is consistent with previous research which suggests that women underperform when they are primed with negative gender-related stereotypes pertaining to their perceived lower ability in mathematics (Beilock et al., 2007; Rydell et al., 2014; Spencer et al., 1999). As such, women under group-as-target stereotype threat may experience evaluative pressure because they

apprehend that their performance will confirm, and thereby reinforce, a negative societal stereotype as a true representation of their social group (Shapiro & Neuberg, 2007).

Against predictions, however, women in the self-as-target stereotype threat did not underperform in relation to the control condition. In accordance with the manipulations employed by Shapiro et al. (2013), women in the self-as-target stereotype threat condition were primed that their mathematical performance would be diagnostic of personal ability. However, in their theoretical review, Shapiro and Neuberg (2007) suggest that in order to experience self-as-target stereotype threat, women should recognise that they belong to a negatively stereotyped group and that stereotype-relevant actions are linked to oneself. Research also indicates that individuals may be more susceptible to stereotype threat when they are knowledgeable and endorse negative societal stereotypes (Elizaga & Markman, 2008; Schmader, Johns, & Barquissau, 2004). Therefore, participants under self-as-target stereotype threat may have not experienced the same evaluative pressure relative to those in the group-as-target condition because they were not made aware of any negative societal stereotypes regarding their mathematical ability. This assertion may be supported further when evaluating the manipulation checks employed in the current study. Although participants in the self-as-target stereotype threat condition reported concerns that their mathematical ability may be utilised as an indicator of their personal ability, they did not report concerns that their personal ability may be diagnostic of women's ability in general. This may suggest that for self-as-target stereotype threat effects to emerge, women should be knowledgeable of the negative gender-maths

stereotype and the implications this may have for both themselves and their social group.

The findings from Experiment 4 paved the way to examine whether women show poorer mathematical performance and reduced updating ability when they are led to perceive that their personal mathematical ability will be tied to gender-related expectations; a notion consistent with Steele and Aronson's (1995) original definition of stereotype threat. Consistent with these predictions, findings from Experiment 5 indicate that women solved fewer mathematical problems when they were presented concurrently with a negative self- and group-relevant stereotype relative to those in a control condition. Evaluations of the manipulation check seemed to show that women under 'combined' stereotype threat reported greater concerns that their personal mathematical ability may be used to evaluate the performance norms of women relative to the control condition. Indeed, previous research indicates that individuals may view their personal and social identities as functionally equivalent (c.f., Swann et al., 2009; Swann et al., 2012), and that this may particularly be the case in stereotype-salient testing environments (Schmader et al., 2008). The current findings therefore suggest that women may be more likely to underperform when they apprehend that their performance will be viewed as self-characteristic of their personal ability, and reinforce negative gender-related stereotypes as a true representation of their social group (Steele & Aronson, 1995).

The second aim of the current study was to examine whether reduced updating ability accounted for the effects of distinct stereotype threats on

women's mathematical performance. Findings from Experiment 4 indicated that participants in the group-as-target stereotype threat condition solved fewer mathematical problems relative to participants in the control condition. However, in contrast to predictions, participants in the group-as-target stereotype threat condition did not show significantly reduced updating ability relative to the control condition. One explanation for this finding may be that, although the group-as-target stereotype threat prime influenced deficits in mathematical performance, it may not have been strong enough to elicit decrements in updating ability. For example, Rydell et al. (2014) utilised a directional prime, whereby female participants were informed that the research was investigating "why women are generally worse at math compared to men" (p. 381). However, within the current study, women in the group-as-target stereotype threat condition were informed that the task was examining the performance norms of women and men. As such, it is plausible that the prime utilised by Rydell et al. (2014) may have had a greater effect on updating ability relative to the prime utilised in the current research because participants were made explicitly aware of the negative societal stereotype pertaining to their lower mathematical ability.

The findings from Experiment 5 indicate that reduced updating ability mediated stereotype threat effects when participants were primed concurrently that both their performance would be diagnostic of personal and gender-related ability. This may suggest that women might be more susceptible to the performance-impinging effects of stereotype threat when both their personal and social identities are the target of a negative societal stereotype. In such situations, targeted individuals have to simultaneously contend with the

implications that a negative stereotype has for both their own ability and that of their valued social group (Wout et al., 2008). This “double jeopardy” may have an additive effect on performance, with negative thoughts diminishing verbal working memory resources and influencing underperformance.

The current research represents one of a few empirical studies that have investigated the impact of distinct stereotype threat primes on women’s mathematical performance (c.f. also, Shapiro et al., 2013; Wout et al., 2008). It is also the first to examine whether reduced updating ability mediates these effects. However, it could be argued that updating ability is implicated more heavily in the modular arithmetic problem solving relative to other mathematical tests because this task requires participants to hold and update calculations in short-term memory (Beilock et al., 2007; Rydell et al., 2014). As such, it could be questioned whether these findings would generalise to other standardised tests, such as the General Record Examination (GRE) and Scholastic Aptitude Test (SAT) which are employed in real world testing environments. Acknowledging this potential limitation in their work, Rydell et al. (2014; Experiment 2) found that reduced updating ability mediated the effects of stereotype threat on women’s performance on the GRE. They also utilised different measures of updating ability to support their findings. Concern for this issue should also be lessened in the current studies because participants in the control condition did not show decrements in updating ability. As such, results from Experiment 5 of this thesis appear to suggest that reduced updating ability accounts for the effects of stereotype threat on women’s mathematical performance, but only when an individual apprehends that poor performance will reflect badly on their own ability and that of their social group.

6.9.1. Chapter Conclusion

Underpinned by the multi-threat framework (Shapiro & Neuberg, 2007), the current study examined the effects that self-as-target and group-as-target stereotype threat had on women's mathematical performance. It also elucidated whether deficits in verbal working memory underpinned these effects. Findings from Experiment 4 indicate that women solved fewer mathematical problems when they perceived their performance to be diagnostic of gender-related ability, but their performance appeared not to suffer when they perceived it to be merely diagnostic of personal ability. Against predictions, however, reduced updating ability did not mediate the relationship between group-as-target stereotype threat and mathematical performance. Taking these findings into consideration, Experiment 5 examined whether concurrently priming a negative personal and social identity had an additive impact on updating and mathematical performance. Findings indicate that women solved fewer mathematical problems and showed poorer updating ability under "combined" stereotype threat relative to the control condition. These findings may suggest that women are more likely to suffer from performance deficits in situations where they apprehend that their personal and social identity may be under evaluative threat. Furthermore, they lend some support to the assertion that stereotype threat may co-opt the verbal working memory resources required to solve difficult mathematical problems.

The empirical studies presented in the current thesis therefore appear to demonstrate the deleterious effects that negative societal stereotypes can have on women's mathematical performance (Experiments 1, 4, & 5). However, the extent to which these effects emerge may be contingent on factors such as

stereotype endorsement and task difficulty (Experiments 2 & 3). The final study of this thesis therefore explores strategies to mitigate stereotype threat. Specifically, it examines whether heightened ingroup representation bolsters mathematical women's mathematical performance.

7. CHAPTER 7: Experiment Six

Creating a Critical Mass May Eliminate the Effects of Stereotype Threat on Women's Mathematical Performance

Abstract

Background: Women in mathematical domains may become attuned to situational cues that signal a discredited social identity, contributing to their lower achievement and underrepresentation. **Aim:** Underpinned by social identity theory, the current study examined whether heightened in-group representation alleviates the effects of self- and group-relevant stereotypes on women's mathematical performance. It investigated further whether single-sex testing environments and stereotype threat influenced participants to believe that their ability was fixed (fixed mindset) rather than a trait that could be developed (growth mindset). **Method:** One hundred and forty-four female participants were assigned randomly and equally to a self-as-target or group-as-target stereotype threat condition or to a control condition. They completed a modular arithmetic maths test and a mindset questionnaire either alone or in same-sex groups of 3-5 individuals. **Results:** Participants solved fewer mathematical problems under self-as-target and group-as-target stereotype threat when they were tested alone but these performance deficits were eliminated when they were tested in single-sex groups. Participants reported a weaker growth mindset when they were tested under stereotype threat and in single-sex groups. Moreover, evidence of inconsistent mediation indicates that single-sex testing environments negatively predicted mindset but positively predicted mathematical performance. **Conclusion:** These findings suggest that single-sex testing environments may represent a practical intervention to alleviate stereotype threat effects but may have a paradoxical effect on mindset.

7.1. Chapter Overview

The empirical studies presented in the current thesis indicate that negative self- and group-relevant stereotypes may have a detrimental impact on women's mathematical performance. Findings also indicate that stereotype threat may diminish phonological working memory resources to bring about decrements in performance. This suggests that stereotype threat may influence negative ruminations or worries about performance, which in turn taxes the verbal component of working which is required to solve mathematical problems. Up until this point in the thesis, the empirical research presented has therefore focused on the adverse effects that stereotype threat can exert on performance, rather than exploring ways in which to remedy it. Capitalising on social identity theory, the final empirical chapter examines whether reinforcing women's sense of belonging in the domain of mathematics, by creating a critical mass, alleviates gender-based concerns to augment performance. However, informed by research into the potential disadvantages of gender-segregated learning environments, it also investigates whether testing females in single-sex groups has a paradoxical effect of influencing a fixed-ability mindset.

7.2. Introduction

Over the past 30 years there has been exponential growth in women's achievement and representation in science-based disciplines. Despite this progress, women continue to underperform and remain underrepresented in maths-intensive fields (Ceci et al., 2009; Ceci & Williams, 2010). Being a minority member can be particularly problematic for women in mathematics classrooms because they may have to contend with negative societal

stereotypes regarding their prescribed inferiority in comparison to men (Inzlicht & Ben-Zeev, 2003). Consequently, being outnumbered may attract a disproportionate amount of attention to a women's social identity (i.e., being female) and increase feelings of responsibility for representing one's group (Saenz, 1994).

Research on stereotype threat (Steele & Aronson, 1995) indicates that women underperform relative to men when they apprehend that their mathematical performance will be interpreted in line with pejorative gender stereotypes (c.f., Spencer et al., 1999; Steele, 1997). These effects appear to be robust (Nguyen & Ryan, 2008) and extend beyond the laboratory (Good, Aronson, & Harder, 2008; Keller, 2007; Hollis-Sawyer, & Sawyer, 2008). As such, researchers have turned their attention to examining the moderating factors that might heighten women's susceptibility to stereotype threat. It has been proposed that seemingly benign and subtle factors, such as the gender composition of a classroom, may undermine women's mathematical performance and contribute further to their underrepresentation in this domain (Bigler & Liben, 2006; 2007; Inzlicht & Ben-Zeev, 2000; 2003; Sekaquaptewa & Thompson, 2003).

In a direct test of this notion, Inzlicht and Ben-Zeev (2000) found that women underperformed on a mathematical test when men outnumbered them. However, these performance deficits were not observed when women completed the test in same-sex groups. Moreover, women's mathematical performance was found to decrease in proportion to the number of men in the testing environment. Extending this, Sekaquaptewa and Thompson (2003) examined the dual influence of solo status and stereotype threat on women's

mathematical performance. Findings indicated that women underperformed to a greater extent when they completed the test in opposite sex groups (solo status) relative to completing the test in same-sex groups. An interaction between solo status and stereotype threat revealed further that experiencing both of these factors simultaneously was more detrimental to performance than experiencing one of these factors alone. These findings support a wealth of research which suggests that the numerical representation of minority group members may interact with their stereotyped status to impact whether an environment will promote or attenuate learning, engagement and performance (Inzlicht & Ben-Zeev, 2000; 2003; Murphy, Steele, & Gross, 2007; Sekaquaptewa & Thompson, 2003).

Research has also examined the extent to which stereotype threat effects are mitigated when women work collaboratively to solve mathematical problems. For example, Aramovich (2014) found that women were buffered from the performance-impinging effects of stereotype threat when they were tested in same-sex groups, relative to alone, because they were able work together to detect errors. Nevertheless, the practical implications of this study may be limited because in real-life testing environments women are typically required to undertake quantitative tests independently as a measure of their personal ability. Overcoming this issue, Huguet and Régner (2007; Experiment 2) conducted a field study and revealed that stereotype threatened females underperformed when they worked alone or in mixed-sex classrooms on a task that measured ostensibly mathematical skills. However, these performance deficits were eliminated when females worked in single-sex groups. These findings suggest that the mere presence of other in-group members (i.e.,

females) may promote women's mathematical performance when they are assessed individually. Other research appears to support these findings, indicating that females may be more susceptible to stereotype threat in co-educational relative to single-sex schools, because they are more exposed to stereotypical beliefs and perceptions pertaining to gender-subject competence (Picho & Stephens, 2012). However, this previous research has not distinguished between distinct stereotype threats and it therefore remains to be ascertained whether single-sex classrooms may alleviate both self-as-target and group-as-target stereotype threat.

Presenting as a further issue, previous work has focused largely on the potential efficacy of single-sex testing environments as a practical means to bolster women's performance against stereotype threat (Huguet & Régner, 2007; Inzlicht & Ben-Zeev, 2000; Picho & Stephens, 2012). Less work has examined the impact that gender-segregated classrooms may exert on attitudinal outcomes, specifically with regard to the potential negative impacts of highlighting gender (c.f., Halpern et al., 2011). Based on a rationale garnered from same-sex schooling (c.f., Halpern et al., 2011; Pahlke, Shibley-Hyde, & Allison, 2014), the current research investigates the notion that gender-segregated environments may influence a fixed-ability mindset (Dweck, 2006; 2008). When placed in same-sex classrooms, females may question why they have been separated from their male peers and attribute this to inherent sex differences (Halpern et al., 2011; Pahlke et al., 2014). Such environmental cues may signal to women, either explicitly or implicitly, that their ability to succeed in mathematics is limited by group membership (Dweck, 2006; 2008; Good et al., 2008). Indeed, this is an important consideration in view of research

indicating that a fixed-ability mindset may have a damaging, and long-term effect on educational outcomes (Verniers & Martinot, 2015; c.f., also Martin, 2015). As such, it is plausible that single-sex classrooms may reduce situational performance pressure by alleviating women's apprehensions about confirming gender-related stereotypes in the eyes of out-group members (Picho & Stephens, 2012; Titze, Jansen, & Heil, 2011). However, the gender-composition of the classroom may signal to women that their mathematical performance is determined by external factors, namely their gender, which may contribute to a fixed-ability mindset.

The first aim of the current study was therefore to examine whether the mere presence of other females could ameliorate the effects of self- and group-relevant stereotypes on women's mathematical performance. Here it was predicted that female participants would solve fewer mathematical problems under self-as-target and group-as-target stereotype threat when they were tested alone relative to those in a control condition. This prediction was derived from research suggesting that, when tested alone, women may apprehend that they are single representatives of their social group, which may exacerbate situational performance pressure (c.f., Aramovich, 2014; Huguet & Régner, 2007; Steele, 1997). It was also predicted that these performance decrements would be alleviated when females were tested in single-sex groups. At first blush, it may seem that women should be susceptible to group-as-target stereotype threat in single-sex groups because this concerns their devalued group membership in the stereotyped domain (c.f., Shapiro et al., 2013). However, in line with previous research (Inzlicht & Ben, Zeev, 2000; Murphy et al., 2007; Sekaquaptewa & Thompson, 2003), the numerical representation of

other females within the mathematics classroom should lessen concerns about representing positively the in-group to bolster performance. Furthermore, when tested in single-sex groups, women may be less susceptible to self-as-target stereotype threat because they strive to disconfirm the negative group stereotype as being a true representation of their personal ability.

The second aim of the current study was to examine the effects of stereotype threat and group composition on mindset. Underpinned by research on single-sex schooling (c.f., Halpern et al., 2011; Pahlke et al., 2014), it was predicted that female participants would become more cognisant of the differences between women and men when they were tested in single-sex groups relative to alone. Under such conditions, it was predicted that they would attribute their mathematical ability to internal, fixed factors (i.e., fixed mindset) rather than a trait that could be shaped and developed (i.e., growth mindset). Given that stereotypes are essentially fixed mindset labels (Blackwell, Trzesniewski, & Dweck, 2007; Dweck, 2006; 2008), it was also predicted that females who were primed explicitly with information regarding gender differences in mathematical performance would report a weaker growth mindset compared to those in the control condition.

7.3. Method

7.3.1. Participants

One hundred and forty-four females ($M_{age} = 21.60$, $SD = 4.67$, 88.9% White British, 83.3% university students) signed up via an online participation website and arranged a time to come into the lab. They received £3 remuneration for their participation. In a between-participants design, they were allocated

randomly and equally to one of three experimental conditions: 1), self-as-target stereotype threat, 2), group-as-target stereotype threat, and 3), a control condition. To examine the effects of in-group representation on performance, half of the participants in each experimental condition completed the study alone, whereas the other half were tested in groups of 3-5. The study consisted of a 3 (condition: self-as-target, group-as-target, control) x 2 (group composition: alone, group) between-participants design.

7.3.2. Stereotype Threat Manipulations

The same self-as-target and group-as-target stereotype threat primes were utilised as in Experiment 1 (Chapter 4).

7.3.3 Measures

Mathematical performance. The same modular arithmetic test was employed as in Experiment 1.

Mindset questionnaire. Participants' mindset was measured using a 20-item self-report questionnaire (McKenzie, 2013; adapted from Dweck, 2006). This questionnaire was modified to ensure that all questions were related to mathematical ability, rather than general intelligence (See Appendix C). Participants responded on a 4-point Likert scale anchored between 'Strongly Agree' and 'Strongly Disagree'. Questions related to a growth mindset included "Mathematical talent can be learned by anyone" and questions related to a fixed mindset included "Maths is much easier to learn if you are male". Scores were totalled out of 60, with higher scores indicative of a growth-ability

mindset. The questionnaire resulted in high internal consistency in the current study, Cronbach's $\alpha = .81$.

7.3.3. Procedure

After being assigned randomly to one of three experimental conditions, participants completed two self-report questions; "I am good at maths" and "It is important to me that I am good at maths". Responses were recorded on a 9-point Likert scale anchored between 1 (Strongly Disagree) and 9 (Strongly Agree). These questions were included in order to control for any differences in perceived mathematical ability and domain identification as a function of experimental condition (c.f., Keller, 2007; Steele, 1997). Upon implementing the stereotype threat prime, participants completed the mindset questionnaire and the maths test, with the order of these measures counterbalanced. Participants were then introduced to the maths test with written instructions presented on a computer. They were instructed to judge the validity of each maths problem, indicating whether the answer was true (i.e., a whole number) or false (i.e., a decimal number) using the 'Z' and 'M' buttons on a standard keyboard, respectively. Participants completed the maths test on individual computers, which had screens on either side to ensure that participants could not observe others' answers. Upon completion of the study, participants received a verbal and written debrief.

7.4. Results

Perceived Mathematical Ability and Domain Identification

A MANOVA indicated that participants in the self-as-target stereotype threat condition reported lower perceived mathematical ability ($M = 5.08$, $SD = 1.75$) compared to the control condition ($M = 6.04$, $SD = 1.49$), $F(2, 141) = 4.03$, $p < .05$, $\eta_p^2 = .05$. Moreover, participants in the self-as-target condition attributed less importance to the domain of mathematics ($M = 5.38$, $SD = 1.91$) compared to the control condition ($M = 6.27$, $SD = 1.50$), $F(2, 138) = 3.53$, $p = .03$, $\eta_p^2 = .05$. Although these responses were above average, participants' self-reported mathematical ability and domain identification were entered as covariates in all analyses to ensure that they did not influence performance. Participants in the group-as-target and control condition did not differ in their reports of mathematical ability and domain identification (both $p > .05$). Moreover, participants' responses to these two questions did not differ as a function of group composition (group vs. alone), $p > .05$.

Mathematical Performance

Modular arithmetic accuracy was examined in a 3 (Condition: self-as-target, group-as-target, control) x 2 (Group composition; alone, group) x 3 (Problem difficulty; simple, moderate, difficult) x 2 (Problem presentation: horizontal, vertical) mixed factorial ANCOVA. Experimental condition and group composition were analysed as between-participants factors and problem difficulty and presentation were input as within-participants factors.

Problem difficulty and presentation. There was a significant main effect of problem difficulty, $F(2, 272) = 19.84, p < .001, \eta_p^2 = .13$. Bonferroni-corrected pairwise comparisons reveal that participants solved fewer difficult ($M = .78, SD = .17$) relative to simple problems ($M = .93, SD = .10$), $p < .001, d = -1.08$. They also solved fewer moderate ($M = .76, SD = .17$) relative to simple problems ($M = .93, SD = .10$), $p < .001, d = -1.22$. There was a significant two-way interaction between problem difficulty and presentation, $F(2, 272) = 3.22, p = .04, \eta_p^2 = .02$. Participants solved fewer horizontally presented difficult ($M = .75, SD = .19$) and moderate problems ($M = .72, SD = .18$) compared to simple problems ($M = .95, SD = .07$), $p < .001, d = -1.40$ and -1.68 , respectively. Participants also solved fewer vertically presented difficult ($M = .82, SD = .19$) and moderate problems ($M = .80, SD = .20$) compared to simple problems ($M = .90, SD = .13$), $p < .001, d = -.49$ and $-.59$, respectively.

Stereotype threat. There was a significant main effect of experimental condition on maths performance, $F(2, 136) = 4.67, p = .01, \eta_p^2 = .06$. Pairwise comparisons indicate that participants assigned to the self-as-target condition solved significantly fewer problems ($M = .79, SD = .12$) compared to the control condition ($M = .86, SD = .12$), $p < .01, d = -.58$. There was no significant difference in performance between the group-as-target relative to the self-as-target stereotype threat ($p > .05$) and the control condition ($p > .05$). A three-way interaction was obtained between experimental condition, problem difficulty and presentation, $F(4, 135) = 3.78, p < .01, \eta_p^2 = .05$. Participants in the self-as-target condition solved more difficult and moderate problems when they were presented horizontally ($M = .72, SD = .19, M = .64, SD = .18$) relative

to vertically ($M = .82$, $SD = .19$, $M = .78$, $SD = .21$), $p < .001$, $d = -.53$ and $-.72$, respectively. They solved fewer simple problems when they were presented vertically ($M = .85$, $SD = .13$) relative to horizontally ($M = .94$, $SD = .08$), $p < .001$, $d = -.83$. Participants in the group-as-target stereotype threat condition solved fewer difficult questions when they were presented horizontally ($M = .74$, $SD = .19$) relative to vertically ($M = .80$, $SD = .19$) $p < .05$, $d = -.32$. Participants in the control condition solved fewer difficult and moderate problems when they were presented horizontally ($M = .79$, $SD = .19$, $M = .78$, $SD = .18$) relative to vertically ($M = .84$, $SD = .19$, $M = .85$, $SD = .21$), $p < .05$, $d = -.26$ and $-.36$. They solved fewer simple problems when they were presented vertically ($M = .93$, $SD = .13$) relative to horizontally ($M = 1.0$, $SD = .08$), $p < .001$, $d = -.74$.

Participants under self-as-target condition solved fewer horizontally presented moderate problems ($M = .64$, $SD = .18$) compared to females in the group-as-target ($M = .73$, $SD = .18$), $p < .05$, $d = -.50$, and control conditions ($M = .78$, $SD = .18$), $p = .001$, $d = -.78$. Participants in the self-as-target condition solved fewer horizontally presented simple problems ($M = .94$, $SD = .08$) compared to females in the control condition ($M = 1.0$, $SD = .08$), $p < .001$, $d = -.75$. They also solved fewer vertically presented simple problems ($M = .85$, $SD = .13$) compared to females in the group-as-target ($M = .94$, $SD = .13$), $p < .01$, $d = -.69$, and control conditions ($M = .93$, $SD = .13$), $p = .01$, $d = -.62$. Participants under group-as-target stereotype threat solved fewer horizontally presented simple problems ($M = .93$, $SD = .08$) compared to the control condition ($M = 1.0$, $SD = .08$), $p < .001$, $d = -.87$. Accuracy scores for participants in the group-as-target condition did not significantly differ from the

control condition on all other problems, $p > .05$. All other pairwise comparisons were non-significant, $p > .05$. See Figure 8 for three-way interaction between experimental condition, problem demand and presentation.

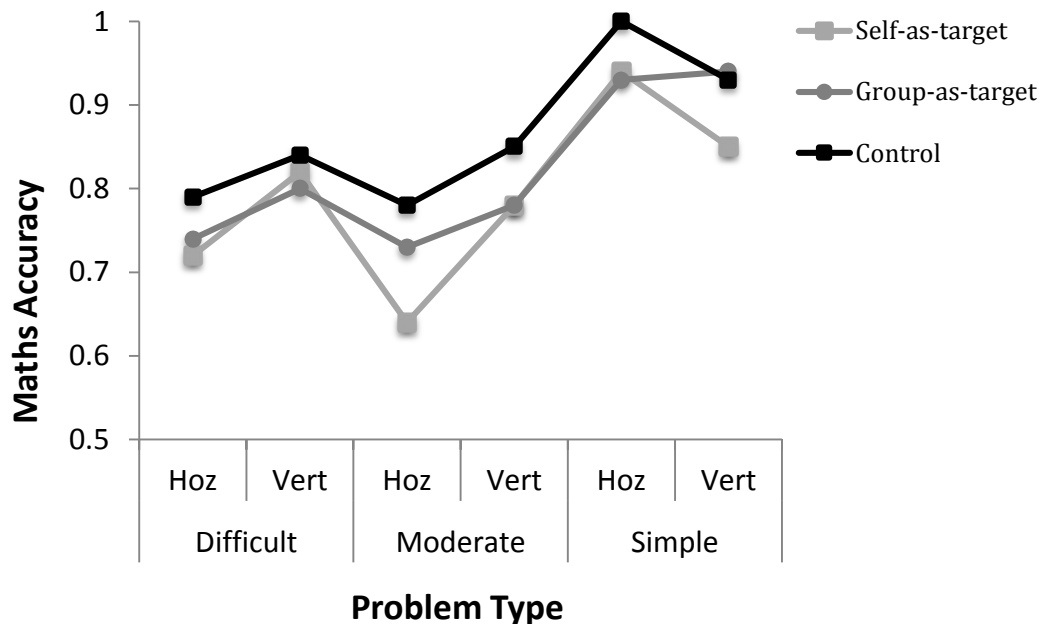


Figure 8. Mean modular arithmetic accuracy scores (%) as a function of experimental condition, problem demand and presentation.

Group Composition. Of central importance to the aim of the current study, there was a significant main effect of group composition, $F(1, 136) = 3.96, p < .05, \eta_p^2 = .03$. Pairwise comparisons indicated that participants solved fewer maths problems when they were tested alone ($M = .81, SD = .12$) relative to in single-sex groups ($M = .84, SD = .12$), $p < .05, d = -.25$. This was qualified by a significant three-way interaction between experimental condition, group composition and problem presentation, $F(2, 136) = 3.58, p < .05, \eta_p^2 = .05$. When tested alone, participants who were primed with a self-as-target stereotype threat solved fewer horizontally presented problems ($M = .73, SD =$

.11) relative to participants in the control condition ($M = .86$, $SD = .11$), $p < .001$, $d = -1.18$. Participants who were tested alone under group-as-target stereotype solved fewer horizontally presented problems ($M = .77$, $SD = .11$) compared to the control condition ($M = .86$, $SD = .11$), $p < .05$, $d = -.82$. Accuracy did not significantly differ for vertically oriented problems, $p > .05$. Importantly, there were no significant performance decrements as a function of experimental condition when females were tested in groups, $p > .05$. These results suggest that the mere presence of other females bolstered participants' mathematical performance from the effects of self-as-target and group-as-target stereotype threat.

Further confirming this, females primed with a self-as-target stereotype solved fewer horizontally presented problems when they were tested alone ($M = .73$, $SD = .11$) compared to when they were tested in groups ($M = .80$, $SD = .11$), $p = .01$, $d = -.64$. They also solved fewer vertically presented problems when tested alone ($M = .77$, $SD = .14$) relative to in a group ($M = .85$, $SD = .14$), $p < .05$, $d = -.57$. There was also a trend for participants primed with a group-as-target stereotype threat to underperform on horizontally presented problems when tested alone ($M = .77$, $SD = .11$) compared to when they were tested in a group ($M = .83$, $SD = .11$), $p = .058$, $d = -.55$. Females' performance in the control condition did not differ significantly as a function of group composition, $p > .05$. When tested alone, females assigned to the self-as-target condition solved fewer horizontally presented problems ($M = .73$, $SD = .11$) relative to vertically presented problems ($M = .77$, $SD = .14$), $p < .05$, $d = -.32$. They also solved fewer horizontally presented problems ($M = .80$, $SD = .11$) compared to vertically presented problems when they were tested in a group ($M = .85$, $SD =$

.14), $p = .01$, $d = - .40$. Females under group-as-target threat solved fewer horizontally presented problems ($M = .77$, $SD = .11$) compared to vertically presented problems ($M = .84$, $SD = .14$) when tested alone ($p < .001$, $d = - .56$) but not when they were tested in groups ($p > .05$). Females in the control condition solved horizontally and vertically presented problems with equivalent accuracy when tested alone and in a group, $p > .05$. Overall, these results suggest that women were susceptible to stereotype threat when they are tested individually, however, single-sex testing environments alleviated these performance deficits. See Figure 9 for interaction between experimental condition, group composition and problem presentation. See Table 15 for descriptive statistics.

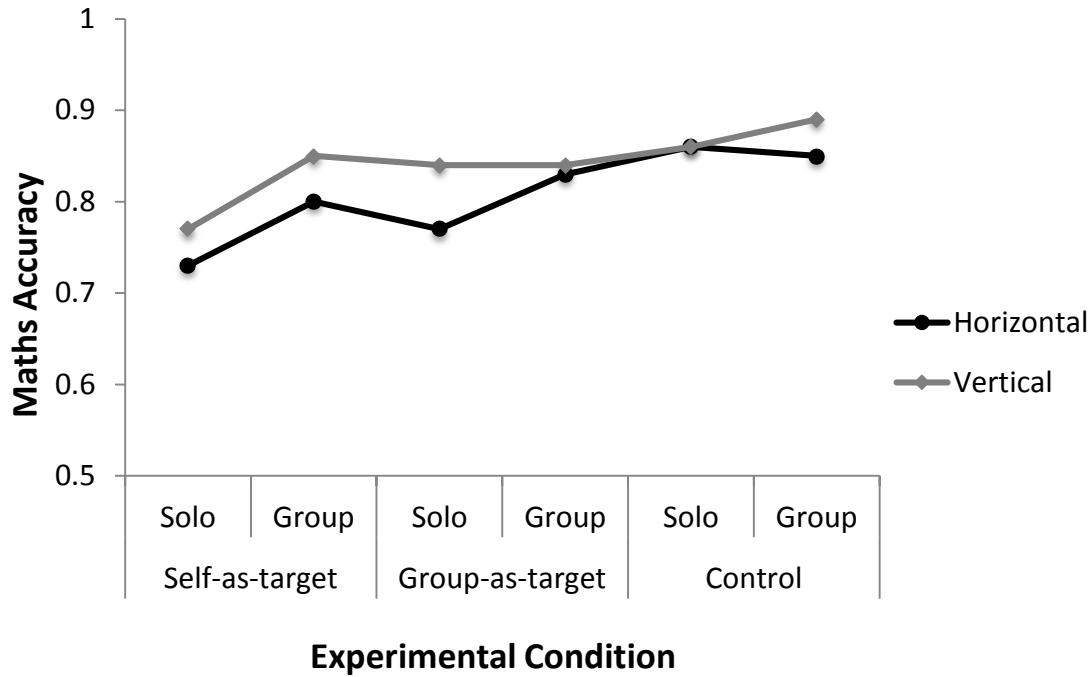


Figure 9. Mean modular arithmetic accuracy scores (%) as a function of experimental condition, group composition and problem presentation.

Table 15.

Mean arithmetic accuracy scores and corresponding deviations as a function of experimental condition, group composition, and problem presentation.

	Self-as-target		Group-as-target		Control	
	Solo	Group	Solo	Group	Solo	Group
	Mean (SD)					
Horizontal % Correct	.73 (.11)	.80 (.11)	.77 (.11)	.83 (.11)	.86 (.11)	.85 (.11)
Vertical % Correct	.77 (.14)	.85 (.14)	.84 (.14)	.84 (.14)	.86 (.14)	.89 (.14)

Note: Estimated marginal means with perceived mathematical ability and domain identification as covariates.

Mindset

Females' self-reported mindset did not significantly differ dependent on whether they completed the questionnaire before or after the maths test ($p > .05$), indicating an absence of order effects. There was a significant main effect of mind-set as a function of experimental condition, $F(2, 138) = 4.45$, $p = .01$, $\eta_p^2 = .06$. Participants assigned to the self-as-target stereotype threat condition ($M = 38.58$, $SD = 6.07$) reported a weaker growth mind-set compared to the control condition ($M = 41.35$, $SD = 6.03$), $p < .05$, $d = -.46$. Participants in the group-as-target stereotype threat condition ($M = 38.46$, $SD = 4.45$) also reported a weaker growth mind-set compared to the control condition, $p < .05$, $d = -.55$. There was a significant main effect of group composition, $F(1, 138) = 13.04$, $p < .001$, $\eta_p^2 = .09$. Participants who completed the test in groups reported a weaker growth mind-set ($M = 37.85$, $SD = 5.32$) compared to those who completed the test alone ($M = 41.08$, $SD = 5.62$), $p < .001$, $d = -.59$. There was no significant interaction between stereotype threat and group composition, $p > .05$. These results suggest that negative gender-maths stereotypes pertaining to women's personal or social identity may hamper a growth-ability mindset. Furthermore, testing females in same-sex groups did not appear to have a positive effect on mindset.

Mediation Analysis

Mediation analysis was conducted using ordinary least squares path analysis (Hayes, 2013). This analysis examined the influence that the single-sex testing environment exerted on mindset and mathematical performance. Results

indicate that group composition indirectly influenced mathematical performance through its effect on mindset. Specifically, group composition negatively influenced mindset ($a = -3.24$) but positively predicted mathematical performance ($b = .19$). A bias-corrected bootstrap confidence interval for the indirect effect ($ab = -.63$) did not include zero (LLCI, = -1.48 , ULCI = $-.10$). However, there was still evidence that being tested in a group influenced mathematical performance independent of its effect on mindset ($c' = 2.36$), $p < .05$. This provides evidence of partial inconsistent mediation (MacKinnon, Fairchild, & Fritz, 2007; MacKinnon, Krull, & Lockwood, 2000), with mindset acting as a suppressor variable⁸ (MacKinnon et al., 2000; Tzelgov & Henik, 1991). See Figure 10 for mediator model.

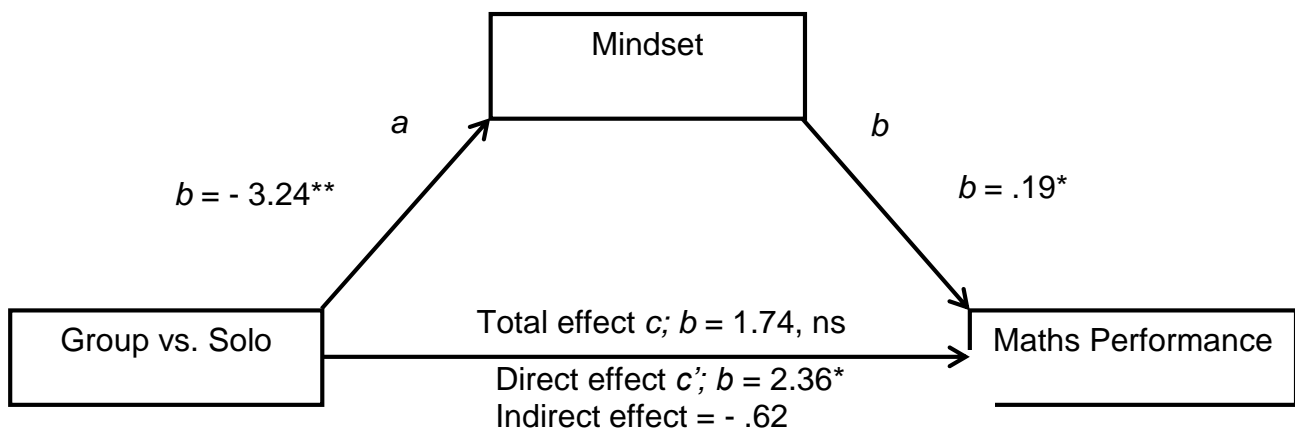


Figure 10. Mediator model: Relationship between group composition on mathematical performance through mindset. Note: * $p < .05$, ** $p < .001$.

⁸ Suppression can occur when the indirect effect has a sign that is opposite to that of the total effect, and therefore the omission of a suppressor variable might lead the total effect to appear small or non-significant (c.f., Rucker et al., 2011).

7.5. Discussion

The current study evaluated the efficacy of single-sex testing environments as a practical means to eliminate stereotype threat effects. Moreover, it examined whether testing women in single-sex groups or under stereotype threat influenced them to perceive that their ability was a fixed trait. Results indicate that female participants underperformed when they were tested alone and were primed with either a self- or group-relevant stereotype. However, these performance decrements did not emerge when they were tested in single-sex groups. These findings suggest that in-group members may function as “social vaccines” who increase social belonging and inoculate fellow group members’ performance against the experience of stereotype threat (Dasgupta, 2011; c.f., also Jetten, Haslam, Haslam, & Branscombe, 2009). Nevertheless, participants reported a weaker growth mindset when they were tested in groups relative to alone and under stereotype threat. As such, single-sex testing environments may reduce group members’ concerns about confirming a negative stereotype to bolster women’s mathematical performance but may have a paradoxical effect on mindset.

Female participants were susceptible to group-as-target stereotype threat when they were tested alone in comparison to those in a control condition. In this situation, women find themselves as single representatives of their social group, which may heighten the salience of negative stereotypes that accompany their group status. Being a minority member may result in added pressure because women apprehend that performance will confirm, and thereby reinforce, pejorative stereotypes as a true representation of their in-

group (Huguet & Régner, 2007; Inzlicht & Ben-Zeev, 2000). Findings also appear to reveal that self-as-target stereotype threat had a greater negative effect than group-as-target stereotype threat. Participants may have been more vulnerable to self-as-target stereotype threat when they were tested alone because they perceived that performance would be evaluated in line with their personal ability. As such, the salience of a negative self-relevant stereotype may have interacted with the experience of being alone in the testing environment to attract a disproportionate amount of attention to one's personal identity.

Findings indicate further that women's mathematical performance was protected from the effects of self-as-target and group-as-target stereotype threat when they were tested in same-sex groups. This finding may be explained by distinctiveness theory (Abrams, Thomas, & Hogg, 1990; Cota & Dion, 1986), which posits that group saliency increases relative to the number of out-group members in a particular setting (McGuire, McGuire, & Winton, 1979). Resultantly, the mere presence of other in-group members may have decreased women's apprehensions about representing the group positively to bolster performance (Inzlicht & Ben-Zeev, 2000; 2003; Sekaquaptewa & Thompson, 2003). This finding may have practical implications in relation to gender-segregated learning environments. For example, research suggests that women may feel marginalised in mathematics classrooms when men outnumber them, which may influence deficits in mathematical performance (Inzlicht & Ben-Zeev, 2000; Huguet & Régner, 2007; Sekaquaptewa & Thompson, 2003). As such, increasing the number of women in counter-stereotypical domains, to create a critical mass, may present as a strategy to

alleviate experiences of stereotype threat and encourage more women into maths-intensive fields.

Despite the positive impact that in-group representation had on performance, women who were tested in single-sex groups reported a weaker growth mindset compared to those who were tested alone. Evidence of inconsistent mediation reveals that being in a group negatively predicted mindset but positively predicted mathematical performance. When tested in single-sex groups, females may have become aware that they had been segregated from their male peers, and attributed this to alleged gender differences in mathematical ability (c.f., Halpern et al., 2011; Pahlke et al., 2014). This may have led females to believe that gender is a fundamental characteristic of ability, weakening a growth-ability mindset. Participants also reported a weaker growth mindset when they were primed with a self-as-target and group-as-target stereotype relative to participants in the control condition. This is consistent with research suggesting that negative gender-maths stereotypes may influence women to believe that their mathematical ability is limited because of their group membership (Dweck, 2008; Dar-Nimrod & Heine, 2006; Good, Aronson, & Inzlicht, 2003).

A number of limitations in the present research should be noted. First, the study did not employ a fully matched design in that females' mathematical performance was not compared to that of males within single and mixed-sex testing environments. The rationale to only recruit female participants was underpinned by research which has demonstrated consistently that women's mathematical ability is hampered in the presence of men (e.g., Inzlicht & Ben-Zeev, 2000; 2003; Murphy et al., 2007; Sekaquaptewa & Thompson, 2003),

and by findings indicating that men are less susceptible to stereotype threat in the maths domain (Rydell et al., 2014). However, it may be asserted that females experiences of distinct stereotype threats may be different when they are tested in the presence of outgroup members. That is, women may be more susceptible to the effects of group-as-target stereotype threat, relative to self-as-target stereotype threat, when they find themselves as a single representative of their social group. This may be particularly the case for women in a vanguard position, such as those in STEM-related disciplines, because their social identity is highly salient. It is therefore recommended that future research investigates the presence of moderating factors that may accentuate the effects of self-as-target and group-as-target stereotype threat.

The findings of the current study indicate that participants in the group-as-target and self-as-target stereotype conditions solved fewer simple and moderate problems relative to participants in the control condition. This contrasts with research indicating that stereotype threat effects are more pronounced on difficult questions (Keller, 2007). Within-participant analyses indicated that across all experimental conditions participants solved fewer difficult problems relative to moderate and simple problems. Resultantly, participants in the control condition may have also found these problems difficult, with this diminishing any potential differences between experimental conditions. This may particularly be the case given that a novel laboratory task was employed to ensure that participants were not familiar with the format of the test (c.f., Beilock & Carr, 2005). Future research that examines stereotype threat effects as a function of problem type and difficulty, and utilises more

ecologically valid tests, such as the General Certificate in Secondary Education (GCSE), is therefore recommended.

Results also reveal that mindset partially mediated the effects of in-group representation on women's mathematical performance. However, given that partial mediation was found, this suggests that additional (unmeasured) variables may explain the relationship between single-sex testing environments and performance further. For example, previous research suggests that single-sex testing environments may mitigate stereotype threat by decreasing anxiety (Ben-Zeev, Fein, & Inzlicht, 2005). Additional research would therefore benefit from exploring explanations for the potential efficacy of single-sex testing environments in the elimination of stereotype threat, with researchers acknowledging both the advantages and limitations of implementing such strategies.

7.5.1. Chapter Conclusion

The current research indicates that the salience of a negative self or group-relevant stereotype may have a detrimental impact on women's mathematical performance, with these effects emerging after controlling for participants' perceived mathematical ability and domain identification. However, these performance deficits were reduced when women were tested in same-sex groups. This may suggest that heightened in-group representation may bolster women's performance in the domain of mathematics. This finding may be particularly noteworthy when considering research which suggests that poor numerical representation may be a key determinant in women's decisions to avoid or leave math-intensive fields, even for those who are highly skilled and

identify with the domain (Murphy et al., 2007). However, findings also reveal that females reported a weaker growth mindset when they were tested in single-sex groups. This underscores the importance of examining the potential efficacy of gender-segregated learning environments on both attitudinal and behavioural outcomes, suggesting that whilst testing females in single-sex classrooms may alleviate experiences of stereotype threat, the saliency of gender may influence a fixed-ability mindset.

8. CHAPTER 8 – Thesis Discussion and Conclusions

8.1. General Discussion

The past twenty years of research indicates that the salience of a pejorative societal stereotype impacts negatively on females' mathematical performance (c.f., Doyle & Voyer, 2016; Nguyen & Ryan, 2008; Picho et al., 2013 for meta-analyses). However, the majority of previous research has utilised group-based primes to elicit stereotype threat, and has seemingly overlooked the role of the self in stereotype-salient environments. More recently, researchers have proposed that stereotype threat may operate through multiple, distinct pathways, which target an individual's personal or social identity (Shapiro & Neuberg, 2007; Shapiro et al., 2013; Wout et al., 2013). Although the multi-threat framework (Shapiro & Neuberg, 2007; Shapiro et al., 2013) advances the theoretical understanding of this situational phenomenon, relatively less research has examined empirically the influence that distinct stereotype threats exert on performance outcomes. Accordingly, the first aim of this thesis was to examine the extent to which self-as-target and group-as-target stereotype threat reduce females' mathematical performance.

In addition, there is still debate with regards to the mechanisms which are proposed to underpin the stereotype threat-performance relationship. The working memory interference account (Beilock et al., 2007; Rydell et al., 2014; Schmader & Johns, 2003) and the mere effort account (Jamieson & Harkins, 2007; 2009; Seitchik & Harkins, 2015) appear to be two of the most robust theories in the current stereotype threat literature. However, they both represent opposing theories of *how* stereotype threat diminishes performance (c.f., Schmader et al., 2008 for an overview). The working memory interference account (Beilock et al., 2007; Rydell et al., 2014; Schmader & Johns, 2003)

suggests that stereotype threat impedes the cognitive resources required to solve mathematical problems. Conversely, the mere effort account (Jamieson & Harkins, 2007; 2009; Seitchik & Harkins, 2015) proposes that stereotype threat motivates females to perform optimally, which is seen to facilitate the most dominant response on a stereotype-relevant task. Whilst progress has been made in recognising the possible existence of distinct stereotype threats, and elucidating the mechanisms that may underpin the stereotype threat-performance relationship more generally, these two research traditions have not been studied conjointly to date. That is, research has yet to elucidate the underlying mechanisms of distinct stereotype threats. Bridging this gap in the literature, the second aim of the thesis was to investigate whether deficits in verbal working memory or heightened motivation account for the relationship between self- and group-relevant stereotypes on females' lowered mathematical performance.

8.1.1. Overview of Empirical Studies

Experiment 1 examined whether females' mathematical performance is reduced, to a greater or lesser extent, when they are primed with self-as-target or group-as-target stereotype threat. It also assessed whether deficits in verbal working memory may account for these effects. Results indicate that participants primed with a self-relevant stereotype displayed lower accuracy scores on a modular arithmetic task compared to participants in the control condition. However, there was no main effect of group-as-target stereotype threat on females' mathematical performance. This finding suggests that self-relevant gender stereotypes may have a greater detrimental impact on females'

mathematical performance, possibly because heightened self-awareness may exacerbate situational performance pressure (c.f., Baumeister, 1984; Beilock & Carr, 2001; 2005; Schmader et al., 2013; Van-Loo et al., 2013). Furthermore, participants in both the self-as-target and group-as-target stereotype threat conditions solved fewer horizontally oriented problems in comparison to those in the control condition. They also solved fewer problems when they were presented horizontally relative to vertically. In accordance with previous research (Beilock et al., 2007; Cadinu et al., 2005; Steele et al., 2002; Schmader et al., 2008), this suggests that the salience of a negative self- or group-relevant stereotype may influence negative thoughts and worries, which resultantly diminishes verbal working memory resources to impair mathematical performance.

Findings from Experiment 1 therefore lend some support to the working memory interference theory of stereotype threat. However, they do not rule out a motivational explanation. The mere effort account (Jamieson & Harkins, 2007; 2009; Seitchik & Harkins, 2015) proposes that stereotype threat arouses individuals' concerns about their ability to perform well on a given task, which facilitates the most dominant response. This is seen to influence stereotype threatened participants to respond more rapidly in comparison to their non-threatened counterparts, and can have a paradoxical effect of hampering performance (i.e., a speed-accuracy trade-off; Jamieson & Harkins, 2007). In Experiment 1, there was no significant difference in the time it took participants to answer the modular arithmetic problems as a function of experimental condition. This may suggest that participants engaged in similar amounts of effort irrespective of the stereotype primes. However, the mere effort account

also proposes that stereotype threatened participants should compensate for any erroneous responses when they recognise that their answer is incorrect and have the opportunity to correct it (Jamieson & Harkins, 2007). In Experiment 1, participants were not given the opportunity to amend their responses, meaning that the mere effort account of stereotype threat could not be explored.

Addressing this issue, Experiment 2 utilised the anti-saccade eye-tracking task, which is able to isolate the motivational processes that may underpin stereotype threat effects (i.e., saccade launch latencies and percentage of corrections). It examined whether priming a negative self- or group-relevant stereotype heightened participants' motivation or reduced their verbal working memory capacity. Findings indicate that participants subject to self-as-target and group-as-target stereotype threats launched marginally fewer corrective saccades relative to control participants. In support of the working memory interference account (e.g., Beilock et al., 2007; Schmader & Johns, 2003), this may suggest that stereotype threat interferes with females' ability to recognise that they have made an incorrect response due to the additional demands that a negative stereotype places on working memory.

Contrary to predictions, there were no significant differences in the time it took participants in both of the stereotype threat conditions to launch correct and corrective saccades compared to the control condition. These results are inconsistent with the findings reported by Jamieson and Harkins (2007; Experiment 3) who found that participants launched quicker correct and corrective saccades when they were primed with a negative gender-maths stereotype, perhaps owing to their motivation to disprove it. However, a key

limitation of Experiment 2 was that females' mathematical performance was not assessed, and whilst this decision was made to ensure that the study was equivalent to that of Jamieson and Harkins (2007; Experiment 3), it is apparent that studies of this nature are unable to determine whether heightened motivation mediates the effects of stereotype threat on females' mathematical performance.

Experiment 3 therefore employed both the anti-saccade and modular arithmetic tasks to elucidate whether the mere effort motivational account, or the working memory interference account, may explain the stereotype threat-performance relationship. Expanding on this, it also examined the effects of a positive gender-related stereotype on females' visuospatial and mathematical performance. Findings indicate that participants primed with a positive stereotype solved marginally fewer difficult problems compared to participants in the control condition ($p = .058$, $d = - .86$). This finding supports previous research, suggesting that the expectation to perform in line with a positive stereotype may lead individuals to choke under pressure (Beilock et al., 2001; 2004; Cheryan & Bodenhausen, 2000; Rosenthal & Crisp, 2007). Against predictions, however, a negative group-relevant stereotype did not appear to have an adverse impact on mathematical performance. Furthermore, there were no significant differences in participants' visuospatial performance (correct, corrective saccades RT and % correct) as a function of experimental condition. Taken together, Experiments 2 and 3 are unable to provide support for the mere effort or working memory interference account of stereotype threat.

A number of plausible explanations and study limitations were noted for Experiments 2 and 3, which may explain the null findings obtained. First, the

sample size in these two studies was relatively small (although in line with Jamieson & Harkins, 2007), and this may have resulted in insufficient statistical power. Second, participants who were assigned to the stereotype threat conditions did not appear to strongly endorse the stereotype presented, whereas in Jamieson and Harkins (2007; Experiment 3) study they did. Finally, findings from the current experiments reveal that all participants responded to anti-saccade trials with accuracy rates of 80%, and to pro-saccade trials with 98% accuracy. These high accuracy scores may indicate that the anti-saccade eye-tracking task was too simple and resulted in ceiling effects. Consistent with this explanation, research suggests that performance deficits under stereotype threat may be more likely to emerge on difficult tasks (Keller, 2007; Neuvillle & Croizet, 2007; Spencer et al., 1999).

Experiments 4 and 5 utilised separate and more difficult cognitive tasks to examine whether the executive functions of updating, inhibition and shifting underpin the effects of distinct stereotype threats on females' mathematical performance. Whilst the mere effort account of stereotype threat can be elucidated using inhibitory control tasks (Jamieson & Harkins, 2007; 2009), the working memory interference account can be explored using tasks of updating ability (Beilock et al., 2007; Rydell et al., 2014; Schmader & Johns, 2003). Furthermore, a measure of threat-based concern (Marx, 2012) was employed in these studies to assess whether participants were apprehensive about the implications that a negative self- or group-relevant stereotype may have for their personal or social identity, or whether both of these identities can be simultaneously active under such conditions.

At this point, it was acknowledged that the self-as-target prime utilised in Experiments 1-3 also included reference to the social group (i.e., females' mathematical ability). Such manipulations were designed in line with conditions set out by Shapiro and Neuberg (2007), which posit that females need to be knowledgeable about the negative gender-related stereotype and the implications that this may have for their personal performance to be susceptible to self-as-target stereotype threat. However, in their empirical study, Shapiro et al. (2013) found that females underperformed when they were merely primed that their performance would be diagnostic of personal ability, possibly suggesting that the saliency of group membership is not a prerequisite condition to elicit stereotype threat effects. Accordingly, Experiment 4 utilised a self-relevant prime from the extant literature (Shapiro et al., 2013) to investigate whether performance decrements can occur when females perceive a mathematical test to be indicative of their personal ability, without reference to the negative gender-maths stereotype that governs their social group. It also examined the impact of group-as-target stereotype threat on females' mathematical performance.

Findings from Experiment 4 indicate that participants who were primed with group-as-target threat solved fewer mathematical problems compared to participants in the control condition. However, those in the self-as-target stereotype threat condition did not appear to underperform relative to those in the control condition. Manipulation checks appeared to show that whilst participants in the self-as-target stereotype threat condition reported concerns that the task was diagnostic of personal ability, they were seemingly unaware that their performance may be used as a general indicator of females' ability.

Conversely, participants in the group-as-target stereotype threat condition reported concerns that their performance may be evaluated in line with both their personal and gender-related aptitude. These findings may suggest that experiences of stereotype threat operate through the interplay of personal and social identity to diminish females' mathematical performance. Consistent with this assertion, other research indicates that an individual's personal and social identity can become fused in so far that they value the outcomes of the group as their own, and may come to see these two identities as functionally equivalent (Bilewicz, & Kofta, 2011; Swann et al., 2009).

In contrast to the findings reported by Rydell et al. (2014), reduced updating ability was not found to mediate the effects of a negative group stereotype on females' mathematical performance. However, it should be noted that the differences between the group-based primes utilised in this study and that of Rydell et al.'s (2014) may explain these discrepant findings. Specifically, Rydell et al. (2014) utilised a directional prime, in which females were informed explicitly that males had superior mathematical ability compared to females. Conversely, the current study merely informed participants that their performance would be used to assess the norms for women and men. The prime utilised in Experiment 4 of the current thesis was therefore subtler and may have placed fewer demands on females' verbal working memory resources.

Although the multi-threat framework (Shapiro & Neuberg, 2007; Shapiro et al., 2013) makes a distinction between self- and group-relevant stereotypes, the findings from Experiment 4 therefore suggest that thoughts relating to both

the self and the social group are likely to operate in stereotype-salient environments. From this perspective, it could be argued that individuals may experience multiple stereotype threats concomitantly. For example, females completing a mathematical test may worry that their performance will be self-characteristic of personal ability (self-as-target) and consequently lend credence to the negative stereotype pertaining to their social group (group-as-target). With this in mind, Experiment 5 examined whether stereotype threat may be more likely to deplete working memory resources, and consequently diminish mathematical performance, when females apprehend that both their personal and social identities will be the target of negative gender-related expectations.

Results from Experiment 5 indicate that female participants who were primed concurrently with a negative self-as-target and group-as-target stereotype threat solved fewer mathematical problems and displayed reduced updating ability relative to those in a control condition. Moreover, updating ability appeared to mediate the stereotype threat-performance relationship. These findings suggest that stereotype threat may operate through both concerns to the self and the social group. In such situations, females may apprehend that their own mathematical ability may confirm the negative stereotype pertaining to their social group, with this 'double jeopardy' taxing the verbal working memory resources needed to solve difficult mathematical problems.

In both Experiments 4 and 5, however, there was no significant difference in participants' inhibitory control as a function of experimental

condition. These findings are inconsistent with a mere effort explanation (Jamieson & Harkins, 2007; 2009; McFall et al. 2009), which predicts that participants under stereotype threat would show better inhibitory control because the potential for evaluation motivates them to perform well. As such, findings from these studies appear to provide some support for the working memory interference account of stereotype threat-performance effects, rather than a mere effort motivational account.

Experiments 1-5 examined systematically the influence that distinct stereotype threats exert on females' mathematical performance, and investigated the mechanisms that may account for these effects. Moving beyond this, the final empirical study of this thesis investigated whether testing females in single-sex groups, relative to alone, alleviated self-as-target and group-as-target stereotype threat. In line with previous research which has demonstrated that promoting an incremental mindset alleviates stereotype threat (Good et al., 2003), it also examined whether stereotype threat and gender-segregated classroom environments may contribute to a fixed-ability mindset.

Findings indicate that female participants solved more problems under self-as-target and group-as-target stereotype threat when they completed a mathematical test in the presence of other females. This suggests that heightened in-group representation may decrease females' apprehensions about positively representing their in-group to augment mathematical performance (Inzlicht & Ben-Zeev, 2000; Sekaquaptewa & Thompson, 2003). However, females also reported a weaker growth-ability mindset when they were primed with a negative self- or group-relevant stereotype, and when they

were tested in single-sex groups relative to alone. Although the presence of other in-group members may alleviate situational performance pressure, gender-segregated learning environments may influence females to question *why* they have been separated from their male peers, which may relay a message that gender is a fundamental characteristic of ability.

Experiment 6 therefore highlights the possible positive and unintended negative consequences of interventions designed to mitigate stereotype threat. Such findings may lead researchers and educationalists to question what the best strategy is to reduce the influence of stereotypes on performance and related attitudes. One approach could be to teach females in single-sex classrooms and to encourage them to view mathematical ability as a malleable rather than as a fixed attribute (Aronson, Fried, & Good, 2002; Dweck, 2015; Good et al., 2003; 2008; Spitzer & Aronson, 2015). Explaining to students that they have not been separated from males based upon their ability, but rather to promote their learning may have a positive effect on attainment, motivation and engagement (c.f., Dweck, 2008; Elliot & Dweck, 2005).

Other research, however, suggests that same-sex educational environments may come at a longer-term cost to successful gender-role socialisation, particularly when females are re-integrated with males within further education and the workplace (Halpern et al., 2011). An alternative strategy may therefore lie within tackling negative gender-related stereotypes within co-educational classrooms. Teaching students about the pervasive effects of stereotype threat and the direct influence it can exert on performance could be one way to achieve this (c.f., Johns et al., 2005). Given that stereotypes about ability are likely to represent fixed mindset beliefs (Blackwell

et al., 2007; Dweck, 2006; 2008), this strategy, in itself, may encourage students to adopt a growth mindset and increase females' participation and performance in mathematics.

8.1.2. Thesis Limitations and Research Implications

Specific theoretical and methodological issues have been discussed within each chapter's empirical findings. Before considering the overall conclusions of this thesis, a number of general limitations should be borne in mind when interpreting findings as a whole.

8.1.2.1. Theoretical/Conceptual issues. Underpinned by the multi-threat framework (Shapiro & Neuberg, 2007; Shapiro et al., 2013), the current thesis examined whether performance is debilitated by self- or group-relevant stereotypes, and focused accordingly on the 'target' of stereotype threat. However, another component of this theory acknowledges the 'source' of stereotype threat. Whilst the target of stereotype threat refers to the implications that performance may have for either the self or the social group (i.e., who will one's stereotypical actions reflect upon: the self, or the group?), the source of stereotype threat refers to apprehensions regarding who may judge this behaviour (i.e., the self, in-group others, or out-group others). For instance, females may be at risk of experiencing in-group threat when they perceive that members of their own social group (i.e., other females) will evaluate their performance and treat them in line with the negative stereotype. Conversely, females may be at risk of experiencing out-group stereotype threat when they perceive that members of a different social group (i.e., males) may evaluate

their performance and discriminate against them on the basis of this negative stereotype (Shapiro & Neuberg, 2007).

In the present series of studies, the 'source' of the stereotype was not explicitly primed, however it is possible that participants may have been concerned that their performance would be evaluated by ingroup or outgroup others. From this perspective, it is conceivable that females may experience the 'target' and the 'source' of stereotype threat in concert. Consistent with this reasoning, previous research indicates that blurring intergroup boundaries can reduce stereotype threat effects (Rosenthal & Crisp, 2006; 2007), suggesting that concerns regarding both the ingroup and the outgroup are likely activated by negative stereotypes.

Additional research is therefore recommended to explore the impact that ingroup and outgroup stereotype threats exert on performance. This could be accomplished by influencing females to believe that their mathematical performance will be evaluated by other females or males, or by providing them with feedback regarding how other ingroup and outgroup members have performed. It would also be beneficial to measure the importance that people ascribe to their personal and social identities under different stereotype threat conditions (c.f., questionnaire employed by Nario-Redmond, Biernat, Eldelman, & Palenske, 2004). Such work might reveal whether stereotype threat is a multifaceted situational phenomenon, which operates separately through concerns for an individual's personal and social identity, or whether it represents a singular construct, in which both the concepts of the self and the social group are interlinked.

Empirical support has been accrued for many of the moderators of stereotype threat proposed in the literature, such as task difficulty (Hess et al., 2003; Keller, 2007) and domain identification (Appel et al., 2011). Resultantly, the current thesis focused on the mediating mechanisms that may explain the stereotype threat-performance relationship. Nevertheless, limited research exists which has explored the potential moderating factors that heighten females' susceptibility to distinct forms of stereotype threat (c.f., Wout et al., 2008 for an exception). For example, it is possible that individuals with lower personal self-esteem may be vulnerable to self-as-target stereotype threats, whereas those with lower collective self-esteem may be susceptible to group-as-target stereotype threat (Shapiro & Neuberg, 2007). Group-as-target stereotype threat may also have a greater detrimental impact on performance for individuals who are more likely to define their self-concept as interconnected with others (i.e., interdependent self-construal), relative to those who separate their sense of self from others (i.e., independent self-construal) (c.f., Voci, Hewstone, Crisp, & Rubin, 2008).

Additionally, females who encounter numerical asymmetry in terms of their gender within their daily environment (e.g., educational discipline or workplace) may be more susceptible to group-as-target stereotype threat relative to self-as-target stereotype threat because they are conscious of their minority status in the activities they pursue (c.f., Inzlicht & Ben-Zeev, 2000; Sekaquaptewa & Thompson, 2003). However, it is important to note that other research has found that female engineering students, who have successfully entered a gender counter-stereotypic domain, show better performance when primed with a group-relevant stereotype compared to Psychology majors

(Crisp, Bache, & Maitner, 2009). It is therefore apparent that future research is required to examine the factors that may exacerbate, or protect against, self-as-target and group-as-target stereotype threat.

In a similar vein, a recent line of enquiry suggests that interventions should be tailored to the specific type of stereotype threat in order to be effective. Research by Shapiro et al. (2013) indicates that whilst a self-affirmation intervention mitigated self-as-target stereotype threat, it did not alleviate group-as-target stereotype threat. Similarly, a positive role model intervention was found to bolster females' mathematical performance under group-as-target stereotype threat but not under self-as-target stereotype threat. The authors suggest that group-based interventions may safeguard only against group-as-target stereotype threat, whereas self-based interventions mitigate self-as-target stereotype threat. However, it could be argued that findings from Experiment 6 of the current thesis contradict Shapiro et al.'s (2013) work. Specifically, this study indicates that heightening the numerical representation of women in mathematics domains – a strategy that could be regarded as a group-based intervention – may alleviate both self-as-target and group-as-target stereotype threats. In contrast to the multi-threat framework (Shapiro & Neuberg, 2007), these findings suggest that group-based interventions may be effective in reducing different forms of stereotype threat because both the self and the social group are likely implicated in experiences of stereotype threat.

8.1.2.2. Methodological issues. In each study participants reported their perceived mathematical ability and the importance they attributed to this domain. These measures were employed in accordance with previous research

suggesting that participants may only be vulnerable to stereotype threat when they believe that they are skilled in mathematics and value the ability domain (Beilock et al., 2007; Nguyen & Ryan, 2008; Spencer et al., 1999; Steele, 1997). Sensitivity analyses were also conducted in this regard to ensure that participants did not significantly differ on these measures as a function of experimental condition. Despite this, it could be argued that participants may have differed in their actual (rather than perceived) mathematical ability, particularly because no baseline measures of mathematical performance were obtained. Whilst this is a documented limitation of stereotype threat research (c.f., Boucher, Rydell, & Murphy, 2015; Kaye & Pennington, 2016), studies tend not to adopt baseline measures to ensure that participants do not become too comfortable with the task, which may consequently weaken the influence of stereotype threat manipulations. It is also plausible that providing participants with a stereotype threat manipulation after they have completed part of the task may influence demand characteristics (i.e., when people infer the true aims of the study and act accordingly).

In line with previous research (Beilock et al., 2007; Rydell et al., 2014; Seitchik & Harkins, 2015; Ståhl, Van Laar, & Ellemers, 2012), a modular arithmetic task was utilised to examine the effects of stereotype threat on females' mathematical performance. This novel laboratory task was deemed most appropriate because participants are not familiar with the task requirements, which may eliminate confounding factors such as practice effects and prior expertise. Nonetheless, recent research suggests that, in addition to reducing performance, stereotype threat may interfere with learning (Rydell, Shiffrin, Boucher, Van-Loo, & Rydell, 2010b; Jones-Taylor & Walton, 2011).

For example, Rydell et al. (2010a) found that female participants who received a stereotype threat manipulation before being taught how to solve modular arithmetic solved fewer easy problems compared to participants who were presented with the manipulation after the instructions. They were also less able to explain the mathematical operations required to solve such problems. In the current experiments, participants received the stereotype threat manipulation prior to reading through task instructions, and consistent with Rydell et al. (2010a) stereotype threatened participants in Experiment 6 solved fewer easy problems relative to the control condition. It could therefore be argued that stereotype threat may have interfered with females' ability to learn the operations required to solve modular arithmetic problems, which consequently impeded their performance. A future line for research would be to examine whether stereotype threat exerts its negative effects by influencing performance or undermining knowledge acquisition. Given the limited literature on stereotype threat-based learning effects, and research investigating the different effects that self- and group-relevant stereotypes may exert on performance, exploring these two areas of research in conjunction would present as a fruitful avenue for future research.

The empirical work presented in the current thesis provides some support for a working memory interference account of stereotype threat (Beilock et al., 2007; Rydell et al., 2014; Schmader & Johns, 2003), and extends previous research by assessing whether deficits in verbal working memory may account for the effects of self-as-target and group-as-target stereotype threat on females' mathematical performance. However, due to the measures employed, it is not known whether participants experience different thoughts,

which relate to either their personal identity (i.e., personal performance) or their social group (i.e., females' performance), when they are primed with a self- or group-relevant stereotype threat.

In order to investigate the hypothesis that stereotype threat depletes phonological aspects of verbal working memory, other research has employed thought-listing techniques such as asking participants to spontaneously report on their thoughts and feelings under stereotype threat (Beilock et al., 2007; Cadinu et al., 2005; Mrazek et al., 2011). Nevertheless, such research has examined stereotype threat as a general construct, and has not distinguished explicitly between self- and group-relevant primes. Future research may therefore benefit from employing thought probing techniques to explore individuals' different experiences of self-as-target and group-as-target stereotype threat. For example, negative thoughts relating to the self may influence deficits in working memory under self-as-target stereotype threat (i.e., ruining one's opportunities, letting oneself down, feeling the stereotype is applicable to oneself). Conversely, group-based intrusive thoughts may tax verbal working resources under group-as-target threat because individuals view their performance in line with their social group (i.e., letting the group down, confirming the societal stereotype) (Shapiro & Neuberg, 2007).

8.1.3. Thesis Conclusions

As a whole, the findings of the current thesis indicate that females' mathematical performance may be stifled when both their personal and social identities are concurrently made salient in a stereotype-relevant domain (Experiments 1, 4, 5 & 6). However, females' mathematical performance appears not to be reduced in situations in which only their personal identity is made salient (Experiment 5). In contrast to the multi-threat framework (Shapiro & Neuberg, 2007), this may suggest that both the concepts of the self and the social group are implicated in experiences of stereotype threat.

Findings also appear to provide some support for the working memory interference account of stereotype threat, suggesting that negative thoughts and ruminations about being viewed as stereotypic may consume the verbal working memory resources required to solve mathematical problems. These findings are consistent with previous work suggesting that the cognitive imbalance between the concepts of the self, the social group and the stereotyped domain may influence verbal ruminations or worries about confirming the stereotype, with a consequent reduction in mathematical performance (Beilock et al., 2007; Rydell et al., 2014; Schmader & Johns, 2003).

Whilst future work is advocated to examine stereotype threat in more ecologically valid environments, the findings from the current thesis may have practical implications for females in the domain of mathematics. For example, in school settings students are required to sit standardised exams to evaluate their personal ability. However, it is possible that female students may be at risk

of underperforming on mathematics tests when they are knowledgeable of negative gender-related stereotypes. It is therefore important that educationalists are aware of the deleterious effects that negative gender-related stereotypes can exert on females' mathematical performance, and take steps to minimise the salience of such stereotypes in classroom settings.

Experiment 6 of the current thesis indicates that one potential way of achieving this may be to increase the presence of other ingroup members by teaching students in single-sex groups. This finding may be of considerable importance given research which suggests that the poor numerical representation of females in mathematics may influence even the most highly confident and domain-identified students to avoid or leave math-intensive fields (Murphy et al., 2009). From this perspective, enhancing females' sense of social identity may present as a practical means for enhancing their performance and participation in mathematics. Nevertheless, Experiment 6 also indicates that both stereotype threat and single-sex environments appear to influence a weaker growth mindset, which negatively predicts performance. Interventions that aim to teach females to view their mathematical ability as a malleable trait that is not determined by their gender may therefore be beneficial in fostering a supportive educational environment that deters negative stereotypes and promotes mathematical performance.

References

- 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 of Aging, 23*, 934-939. doi: 10.1037/a0014293
- Abrams, D., Thomas, J., & Hogg, M. A. (1990). Numerical distinctiveness, social identity and gender salience. *British Journal of Social Psychology, 29*, 87-92. doi: 10.1111/j.2044-8309.1990.tb00889.x
- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed). Washington, DC: American Psychological Association.
- Appel, M., & Kronberger, N. (2012). Stereotypes and the achievement gap: Stereotype threat prior to test taking. *Educational Psychology Review, 24*, 609-635. doi: 10.1007/s10648-012-9200-4
- Appel, M., Kronberger, N., & Aronson, J. (2011). Stereotype threat impairs ability building: Effects on test preparation among women in science and technology. *European Journal of Social Psychology, 41*, 904-913. doi: 10.1002/ejsp.835
- Aramovich, N. P. (2014). The effect of stereotype threat on group versus individual performance. *Small Group Research, 45*, 176-197. doi: 10.1177/1046496414523508

- Aronson, J., & Inzlicht, M. (2004). The ups and downs of attributional ambiguity: Stereotype vulnerability and the academic self-knowledge of African-American students. *Psychological Science, 15*, 829-836. doi: 10.1111/j.0956-7976.2004.00763.x
- 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. doi: 10.1006/jesp.2001.1491
- Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*, 29-46. doi: 10.1006/jesp.1998.1371
- Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General, 130*, 224-237. doi: 10.1037/0096-3445.130.2.224
- Ashcraft, M. H., & Krause, J. A. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin and Review, 14*, 243-248. doi: 10.3758/BF03194059
- Ashmore, R. D., & Del Boca, F. K. (1981). Conceptual approaches to stereotypes and stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 1-35). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, 417-423. doi: 10.1016/S1364-6613(00)01538-2
- Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). New York: Cambridge University Press.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science*, 1, 164–180. doi: 10.1111/j.1745-6916.2006.00011.x
- Barber, S. J. (in press). An examination of age-based stereotype threat about cognitive decline: Implications for stereotype threat research and theory development. *Perspectives on Psychological Science*.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173–1182. doi: 10.1037/0022-3514.51.6.1173

- Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology*, *46*, 610–620. doi: 10.1037/0022-3514.46.3.610
- Baumeister, R. F., & Showers, C. J. (2006). A review of paradoxical performance effects: Choking under pressure in sports and mental tests. *European Journal of Social Psychology*, *16*, 361-383. doi: 10.1002/ejsp.2420160405
- Baumeister, R. F., Hamilton, J. C., & Tice, D. M. (1985). Public versus private expectancy of success: Confidence booster or performance pressure? *Journal of Personality and Social Psychology*, *48*, 1447-1457. doi: 10.1037/0022-3514.48.6.1447
- Beasley, M. A., & Fischer, M. J. (2012). Why they leave: The impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors. *Social Psychology of Education*, *15*, 427-448. doi: 10.1007/s11218-012-9185-3
- Beaton, A., Tougas, F., Rinfret, N., Huard, N., & Delisle, M-N. (2009). Strength in numbers? Women and mathematics. *European Journal of Psychology of Education*, *22*, 291-306. doi: 10.1007/BF03173427
- Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psychological Science*, *17*, 339-343. doi:10.1111/j.1467-8721.2008.00602.x

- Beilock, S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology: General*, *130*, 701-725. doi: 10.1037//0096-3445.130.4.701
- Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail: Working memory and “choking under pressure” in math. *Psychological Science*, *16*, 101-105. doi: 10.1111/j.0956-7976.2005.00789.x
- Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 983-998. doi: 10.1037/0278-7393.33.6.983
- 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*, 1059–1071. doi: 10.1177/0146167206288489
- Beilock, S. L., Kulp, C. A., Holt, L. E., & Carr, T. H. (2004). More on the fragility of performance: Choking under pressure in mathematical problem solving. *Journal of Experimental Psychology: General*, *133*, 584-600. doi: 10.1037/0096-3445.133.4.584

- 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*, 256-276. doi: 10.1037/0096-3445.136.2.256
- Bellinger, D. B., DeCaro, M. S., & Ralston, P. A. S. (2015). Mindfulness, anxiety, and high-stakes mathematics performance in the laboratory and classroom. *Consciousness and Cognition*, *37*, 123-132. doi: 10.1016/j.concog.2015.09.001
- Bem, S. L., & Lewis, S. A. (1975). Sex role adaptability: One consequence of psychological androgyny. *Journal of Personality and Social Psychology*, *31*, 634-643. doi: 10.1037/h0077098
- Ben-Zeev, T., Carrasquillo, C. M., Ching, A., Kliengklom, T. J., McDonald, K. L., Newhall, D. C., et al. (2005). "Math is hard!" (Barbie, 1994): Responses of threat vs. challenge- mediated arousal to stereotypes alleging intellectual inferiority. In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender differences in mathematics: An integrative psychological approach* (pp. 189-206). Cambridge, MA: Cambridge University Press.
- Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology*, *41*, 174-181. doi: 10.1016/j.jesp.2003.11.007

- Benbow, C. P. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects and possible causes. *Behavioural and Brain Sciences*, *11*, 169-183. doi: 10.1017/S0140525X00049244
- Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science*, *210*, 1262–1264. doi: 10.1126/science.7434028
- Benbow, C. P., & Stanley, J. C. (1983). Sex differences in mathematical reasoning ability: More facts. *Science*, *222*, 1029-1031. doi: 10.1126/science.6648516
- Berjot, S., Roland-Levy, C., & Girault-Lidvan, N. (2011). Cognitive appraisals of stereotype threat. *Psychological Reports*, *108*, 585-598. doi: 10.2466/04.07.21.PR0.108.2.585-598
- Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudice. *Advances in Child Development and Behaviour*, *34*, 39-89. doi: 10.1016/S0065-2407(06)80004-2
- Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science*, *16*, 162-166. doi: 10.1111/j.1467-8721.2007.00496.x

- Bilewicz, M., & Miroslaw, K. (2011). Less biased under threat? Self-
verificatory reactions to social identity threat among groups with negative
self-stereotypes. *Journal of Applied Social Psychology, 41*, 2249-2267. doi:
10.1111/j.1559-1816.2011.00802.x
- Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories
of intelligence predict achievement across an adolescent transition: A
longitudinal study and an intervention. *Child Development, 78*, 246-263.
doi: 10.1111/j.1467-8624.2007.00995.x
- Bleeker, M. M., & Jacobs, J. E. (2004). Achievement in math and science: Do
mothers' beliefs matter 12 years later? *Journal of Educational Psychology,*
96, 97-109. doi: 10.1037/0022-0663.96.1.97
- Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and doing
diverge: The effect of stereotype threat on self-reported versus non-verbal
anxiety. *Journal of Experimental Social Psychology, 40*, 247-255. doi:
10.1016/S0022-1031(03)00099-4
- Boucher, K. L., Rydell, R. J., & Murphy, M. C. (2015). Forecasting the
experience of stereotype threat for others. *Journal of Experimental Social
Psychology, 58*, 56-62. doi: 10.1016/j.jesp.2015.01.002
- 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*, 180-185.
doi: 10.1016/j.jesp.2008.08.005

- Brown, R. P., & Josephs, R. A. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. *Journal of Personality and Social Psychology, 76*, 246-257. doi: 10.1037/0022-3514.76.2.246
- 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*, 626-633. doi: 10.1016/S0022-1031(03)00039-8
- Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (Don't expect an easy answer). *Journal of Personality and Social Psychology, 98*, 550-558. doi: 10.1037/a0018933
- Buss, D. M. (1995). Evolutionary psychology: A new paradigm for psychological science. *Psychological Inquiry, 6*, 1-30. doi: 10.1207/s15327965pli0601_1
- 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*, 267-285. doi: 10.1002/ejsp.145
- 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*, 183-197. doi: 10.1002/ejsp.303
- Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science, 16*, 572, 578. doi: 10.1111/j.0956-7976.2005.01577.x

- Caplan, P. J., & Caplan, J. B. (1994). *Thinking critically about research on sex and gender*. New York: Harper Collins College Publishers.
- Carlson, R. A. (1997). *Experienced cognition*. Mahwah, NJ: Erlbaum.
- Carr, P. B., & Steele, C. M. (2009). Stereotype threat and inflexible perseverance in problem solving. *Journal of Experimental Social Psychology, 45*, 853-859. doi: 10.1016/j.jesp.2009.03.003
- Ceci, S. J., & Williams, W. M. (2010). Sex differences in math-intensive fields. *Current Directions in Psychological Science, 19*, 275-279, doi: 10.1177/0963721410383241
- Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin, 135*, 218-261. doi: 10.1037/a0014412
- 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, 20*, 143-158. doi: 10.1016/j.psychsport.2012.10.005
- Cheryan, S., & Bodenhausen, G. V. (2000). When positive stereotypes threaten intellectual performance: The psychological hazards of "model minority" status. *Psychological Science, 11*, 399-402. doi: 10.1111/1467-9280.00277

- 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 and Organization Management, 35*, 77-107. doi: 10.1177/1059601109354839
- Clark-Blickenstaff, J. (2005). Women and science careers: Leaky pipeline or gender filter? *Gender and Education, 17*, 369-386. doi: 10.1080/09540250500145072
- Cohen, G. L., & Garcia, J. (2005). "I am us": Negative stereotypes as collective threats. *Journal of Personality and Social Psychology, 89*, 566-582. doi: 10.1037/0022-3514.89.4.566
- Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science, 1*, 1307-1310. doi: 10.1126/science.1128317
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159. doi: 10.1037/0033-2909.112.1.155
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.
- 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. doi: 10.1177/0269881105058572

- Collaer, M. L., Reimers, S., & Manning, J. T. (2007). Visuospatial performance on an internet line judgment task and potential hormonal markers: Sex, sexual orientation, and 2D:4D. *Archives of Sexual Behavior, 36*, 177-192. doi: 10.1007/s10508-006-9152-1
- Collins, D. W., & Kimura, D. (1997). A large sex difference on a two-dimensional mental rotation task. *Behavioral Neuroscience, 111*, 845-849. doi: 10.1037/0735-7044.111.4.845
- Cota, A. A., & Dion, K. L. (1986). Salience of gender and sex composition of ad hoc groups: An experimental test of distinctiveness theory. *Journal of Personality and Social Psychology, 50*, 770-776. doi: 10.1037/0022-3514.50.4.770
- Cragg, L., & Gillmore, C. (2014). Skills underlying mathematics: The role of executive function in the developmental of mathematics proficiency. *Trends in Neuroscience and Education, 3*, 63-68. doi: 10.1016/j.tine.2013.12.001
- Crevit, I., & Vandierendonck, A. (2005). Gap effect in reflexive and intentional prosaccades. *Neuropsychobiology, 51*, 39-44 doi: 10.1159/000082854
- Crisp, R. J., & Abrams, D. (2009). Improving intergroup attitudes and reducing stereotype threat: An integrated contact model. *European Review of Social Psychology, 19*, 242-284. doi: 10.1080/10463280802547171
- 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*, 171-184. doi: 10.1080/15534510802607953

- Crocker, J., & Luhtanen, R. (1990). Collective self-esteem and ingroup bias. *Journal of Personality and Social Psychology*, 1, 60-67. doi: 10.1037/0022-3514.58.1.60
- Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review*, 96, 608-630. doi: 10.1037/0033-295x.96.4.608
- Croizet, J-C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin*, 6, 588-594. doi: 10.1177/0146167298246003
- Croizet, J-C., Després, G., Gauzins, M-E., Huguet, P., Leyens, J-P., & Méot, A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin*, 30, 721-731. doi: 10.1177/0146167204263961
- Cronin, P., Ryan, F., & Coughlan, M. (2008). Undertaking a literature review: A step-by-step approach. *British Journal of Nursing*, 17, 38-43. doi: 10.1968/bjon.2008.17.1.28059
- Cullen, M. J., Waters, S. D., & Sackett, P. R. (2006). Testing stereotype threat theory predictions for math-identified non-math-identified students by gender. *Human Performance*, 19, 421-440. doi: 10.1207/s15327043hup1904_6

- Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math-gender stereotypes in elementary school children. *Child Development, 82*, 766–779. doi: 10.1111/j.1467-8624.2010.01529.x
- Danaher, K., & Crandall, C. S. (2008). Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology, 38*, 1639-1655. doi: 10.1111/j.1559-1816.2008.00362.x
- Dar-Nimrod, I., & Heine, S. J. (2006). Exposure to scientific theories affects women's math performance. *Science, 314*, 435. doi: 10.1126/science.1131100
- Dasgupta, N. (2011). Ingroup experts and peers as social vaccines who inoculate the self-concept: The stereotype inoculation model. *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory, 22*, 231-246. doi: 10.1080/1047840X.2011.607313
- Davies, C., Aronson, J., & Salinas, M. (2006). Shades of threat: Racial identity as a moderator of stereotype threat. *Journal of Black Psychology, 32*, 399-417. doi: 10.1177/0095798406292464
- DeCaro, M. S., Rotar, K. E., Kendra, M. S., & Beilock, S. L. (2010). Diagnosing and alleviating the impact of performance pressure on mathematical problem solving. *The Quarterly Journal of Experimental Psychology, 63*, 1619-1630. doi: 10.1080/1747021090347486
- Derks, B., Inzlicht, M., & Kang, S. (2008). The neuroscience of stigma and stereotype threat. *Group Processes and Intergroup Relations, 11*, 163-181. doi: 10.1177/136843020708803

- DeStefano, D., & LeFevre, J. A. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology, 16*, 353–386. doi: 10.1080/09541440244000328
- Devine, P. G., & Brodish, A. B. (2003). Modern classics in social psychology. *Psychological Inquiry, 14*, 196–202. doi: 10.1080/1047840X.2003.9682879
- Dickersin, K. (2005). Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 11–33). Chichester, UK: John Wiley and Sons.
- Dickersin, K., Min, Y-I., & Meinert, C. L. (1992). Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA, 267*, 374-378. doi: 10.1001/jama.1992.03480030052036
- Doyle, R. A., & Voyer, D. (2016). Stereotype threat manipulation effects on math and spatial test performance: A meta-analysis. *Learning and Individual Differences, 47*, 103-116. doi: 10.1016/j.lindif.2015.12.018
- Drach-Zahavy, A., & Erez, M. (2002). Challenge versus threat effects on the goal-performance relationship. *Organizational Behavior and Human Decision Processes, 88*, 667-682. doi: 10.1016/S0749-5978(02)00004-3
- Dweck, C. (2006). *Mindset: The new psychology of success*. New York, NY: Random House.

Dweck, C. (2008). *Mindset: How you can fulfil your potential*. New York, NY: Random House.

Dweck, C. S. (2015). Growth, Discussant. *British Journal of Educational Psychology, 85*, 242-245. doi: 10.1111/bjep.12072

Eccles, J. S., Jacobs, J. E., & Harold, R. D. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues, 46*, 183-201. doi: 10.1111/j.1540-4560.1990.tb01929.x

Elizaga, D. A., & Markman, K. D. (2008). Peers and performance: How in-group and out-group comparisons moderate stereotype threat effects. *Current Psychology, 27*, 290-300. doi: 10.1007/s12144-008-9041-y

Ellemers, N., Gilder, D. D., & Haslam, S. A. (2004). Motivating individuals and groups at work: A social identity perspective on leadership and group performance. *Academy of Management Review, 29*, 459-478. doi: 10.1177/1368430212462497

Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology, 72*, 218-232. doi: 10/1037/0022-3514.72.1.218

Elliot, A. J., & Dweck, C. S. (Eds.) (2005). *Handbook of competence and motivation*. New York, NY Guilford.

- Else-Quest, N. M., Shibley Hyde, J., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*, 103-127. doi: 10.1037/a0018053
- Fiske, S. T. (2003). The discomfort index: How to spot a really good idea whose time has come. *Psychological Inquiry*, *14*, 203-208. doi: 10.1080/1047840X.2003.9682880
- Franks, D. D., & Marolla, J. M. (1976). Efficacious action and social approval as interacting dimensions of self-esteem: A tentative formulation through construct validation. *Sociometry*, *39*, 324-241. doi: 10.2307/3033498
- Frey, F. E., & Tropp, L. R. (2006). Being seen as individuals versus as group members: Extending research on metaperception to intergroup contexts. *Personality and Social Psychology Review*, *10*, 265-280. doi: 10.1207/s15327957pspr1003_5
- Friedman, N. P., & Miyake, A. (2000). Differential roles for visuospatial and verbal working memory in situation model construction. *Journal of Experimental Psychology: General*, *129*, 61–83. doi: 10.1037//0096-3445.129.1.61
- Fritz, M. S., & Mackinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, *18*, 233-239. doi: 10.1111/j.1467-9280.2007.01882.x
- Galdi, S., Cadinu, M., & Tomasetto, C. (2014). The roots of stereotype threat: When automatic associations disrupt girls' math performance. *Child Development*, *85*, 250-263. doi: 10.1111/cdev.12128

- Ganley, C. M., & Vasilyeva, M. (2014). The role of anxiety and working memory in gender differences in mathematics. *Journal of Educational Psychology, 106*, 105-120. doi: 10.1037/a0034099
- Geary, D. C. (1995). Sexual selection and differences in spatial cognition. *Learning and Individual Differences, 7*, 289-301. doi: 10.1016/1041-6080(95)90003-9
- Geary, D. C. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences, 19*, 229–284. doi: 10.1017/S0140525X00042400
- Geary, D. C. (1998). *Male, female: The evolution of human sex differences*. American Psychological Association: Washington DC.
- Geary, D. C. (1999). Sex differences in mathematical abilities: Commentary on the math-fact retrieval hypothesis. *Contemporary Educational Psychology, 24*, 267-274. doi: 10.1006/ceps.1999.1007
- Geary, D. C. (2010). *Male, female: The evolution of human sex differences* (2nd Ed.). Washington, DC: American Psychological Association.
- Geary, D. C., & DeSoto, C. (2001). Sex differences in spatial abilities among adults from the United States and China: Implications for evolutionary theory. *Evolution and Cognition, 7*, 172-177.
- Gerstenberg, F. X. R., Imhoff, R., & Schmitt, M. (2012). 'Women are bad at math, but I'm not, am I?' Fragile mathematical self-concept predicts vulnerability to a stereotype threat effect on mathematical performance. *European Journal of Personality, 26*, 588-599. doi: 10.1002/per.1836

- 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*, 17-28. doi: 10.1016/j.appdev.2007.10.004
- Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Applied Developmental Psychology, 24*, 645-662. doi: 10.1016/j.appdev.2003.09.002
- Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science, 320*, 1164-1165. doi: 10.1126/science.1154094
- Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles, 66*, 153-166. doi: 10.1007/s11199-011-9996-2
- Halari, R., Hines, M., Kumari, V., Mehrotra, R., Wheeler, M., Ng, V., & Sharma, T. (2005). Sex differences and individual differences in cognitive performance and their relationship to endogenous gonadal hormones and gonadotropins. *Behavioral Neuroscience, 119*, 104–117. doi: 10.1037/0735-7044.119.1.104
- Hall, N. R., Crisp, R. J., & Suen, M-W. (2009). Reducing implicit prejudice by blurring intergroup boundaries. *Basic and Applied Social Psychology, 31*, 244-254. doi: 10.1080/01973530903058474

- Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Shibley Hyde, J., & Gernsbacher, M-A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, 8, 1-51. doi: 10.1111/j.1529-1006.2007.00032.x
- Halpern, D. F., Eliot, L., Bigler, R. S., Fabes, R. A., Hanish, L. D., . . . Martin, C. L. (2011). The pseudoscience of single-sex schooling. *Science*, 333, 1706-1707. doi: 10.1126/science.1205031
- Hausman, M. (2014). Art versus science – Academic background implicitly activates gender stereotypes on cognitive abilities with threat raising men’s (but lowering women’s) performance. *Intelligence*, 46, 235-245. doi: 10.1016/j.intell.2014.007.004
- Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, 76, 408–420. doi: 10.1080/03637750903310360
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford Press.
- Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, 67, 451-470. doi: 10.1111/bmsp.12028
- Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, 270, 365. doi: 10.1126/science.7604277

- Hess, T. M., Auman, C., Colcombe, S. J., & Rahhal, T. A. (2003). The impact of stereotype threat on age differences in memory performance. *Journal of Gerontology: Psychological Sciences*, *58*, 3-11. doi: 10.1093/geronb/58.1.P3
- Hess, T. M., Hinson, J. T., & Hodges, E. A. (2009). Moderators of and mechanisms underlying stereotype threat effects on older adults' memory performance. *Experimental Aging Research*, *35*, 153-177. doi: 10.1080/03610730802716413
- Hill, C., Corbett, C., & St. Rose, A. (2010). Why so few? Women in science, technology, engineering, and mathematics. Washington DC: AAUW.
- Hirsh, J. B., & Kang, S. K. (2015). Mechanisms of identity conflict: Uncertainty, anxiety and the behavioral inhibition system. *Personality and Social Psychology Review*, Advanced online article. doi: 10.1177/1088868315589475
- Hoelter, J. W. (1986). The relationship between specific and global evaluations of the self: A comparison of several models. *Social Psychology Quarterly*, *49*, 129-141.
- Hogg, M. A., & Abrams, D. (1988). *Social identifications: A social psychology of intergroup relations and group processes*. London: Routledge.
- Hollis-Sawyer, L. A., & Sawyer, T. P. (2008). Potential stereotype threat and face validity effects on cognitive-based test performance in the classroom. *Educational Psychology: An International Journal of Experimental Educational Psychology*, *28*, 291-304. doi: 10.1080/01443410701532313

- Hornsey, M. J. (2008). Social identity theory and self-categorization theory: A historical review. *Social and Personality Psychology Compass*, 2, 204-222. doi: 10.1111/j.1751-9004.2007.00066.x
- Huguet, P., & Régner, I. (2007). Stereotype threat among schoolgirls in quasi-ordinary classroom circumstances. *Journal of Educational Psychology*, 99, 545-560. doi: 10.1037/0022-0063.99.3.545
- Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, 11, 365-371. doi: 10.1111/1467-9280.00272
- Inzlicht, M., & Ben-Zeev, T. (2003). Do high-achieving female students underperform in private? The implications of threatening environments on intellectual processing. *Journal of Educational Psychology*, 95, 796–805. doi: 10.1037/0022-0663.95.4.796
- Inzlicht, M., & Schmader, T. (2010). *Stereotype Threat: Theory, Process, and Application*. New York: Oxford University Press.
- Jamieson, J. P., & Harkins, S. D. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology*, 93, 544-564. doi: 10.1037/0022-3514.93.4.544
- 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, 1301-1314. doi: 10.1177/0146167209335165

Jamieson, J. P., & Harkins, S. G. (2011a). Distinguishing between the effects of stereotype priming and stereotype threat on math performance. *Group Processes and Intergroup Relations*, *15*, 291-304. doi: 10.1177/1368430211417833

Jamieson, J. P., & Harkins, S. G. (2011b). The intervening task method: Implications for measuring mediation. *Personality and Social Psychology Bulletin*, *37*, 652-661. doi: 10.1177/0146167211399776

Jetten, J., Branscombe, N. R., Haslam, H. A., Haslam, C., Cruwys, T., Jones, J. M., ... Zhang, A. (2015). Having a lot of a good thing: Multiple important group memberships as a source of self-esteem. *PLoS One*, *10*. doi: 10.1371/journal.pone.0124609

Jetten, J., Haslam, C., Haslam, S. A., & Branscombe, N. R. (2009). The social cure. *Scientific American Mind*, *20*, 26-33.

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*, 691-705. doi: 10.1037/a0013834

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*, 175-179. doi: 10.1111/j.0956-7976.2005.00799.x

- Jones-Taylor, V., & Walton, G. M. (2011). Stereotype threat undermines academic learning. *Personality and Social Psychology Bulletin*, *37*, 1055-1067. doi: 10.1177/0146167211406506
- Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*, 47-70. doi: 10.1037/0096-3445.132.1.47
- Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory as variation in executive attention and control. In: A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (21-48). Oxford, England: Oxford University Press.
- Kaye, L. K., & Pennington, C. R. (2016). "Girls can't play": The effects of stereotype threat on females' gaming performance. *Computers in Human Behavior*, *59*, 202-209. doi: 10.1016/j.chb.2016.02.020
- 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. doi: 10.1023/A:1021003307511
- 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*, 323-328. doi: 10.1348/000709906X113662

- Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting effect of women's math performance. *Personality and Social Psychology Bulletin*, 29, 371-381. doi: 10.1177/0146167202250218
- Keller, J., & Sekaquaptewa, D. (2008). Solo status and women's spatial test performance: The role of individuation tendencies. *European Journal of Social Psychology*, 38, 1044-1053. doi: 10.1002/ejsp.490
- Khan, K. S., Riet, G. T., Popay, J., Nixon, J., & Kleijnen, J. (2009). Study quality assessment (Stage 2 Conducting the review, Phase 5). In Centre for Reviews and Dissemination (Ed.), *Systematic Reviews* (pp. 1–20). Available from: <http://www.medicine.mcgill.ca/rtamblyn/Readings%5CUK%20Guide%20-%20Study%20Quality%20Assessment.pdf>
- Kiefer, A. K., & Sekaquaptewa, D. (2007a). 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, 825-832. doi: 10.1016/j.jesp.2006.08.004
- Kiefer, A. K., & Sekaquaptewa, D. (2007b). Implicit stereotypes, gender identification, and math-related outcomes: A prospective study of female college students. *Psychological Science*, 18, 13-18. doi: 10.1111/j.1467-9280.2007.01841.x
- Kimura, D. (1999). *Sex and cognition*. Cambridge, MA: MIT Press.

- Kimura, D., & Hampson, E. (1994). Cognitive pattern in men and women is influenced by fluctuations in sex hormones. *Current Directions in Psychological Science*, 3, 57-61. doi: 10.1111/1467-8721.ep10769964
- 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, 942-958. doi: 10.1037/0022-3514.80.6.942
- Lamont, R. A., Swift, H. J., & Abrams, D. (2015). A review and meta-analysis of age-based stereotype threat: Negative stereotypes, not facts, do the damage. *Psychology and Aging*, 30, 180-193.
- Laurin, R. (2013). Stereotype threat and lift effects in motor task performance: The mediating role of somatic and cognitive anxiety. *The Journal of Social Psychology*, 153, 687-699. doi: 10.1080/00224545.2013.821098
- Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.
- Leary, M. R., Terry, M. L., Allen, A., & Tate, E. B. (2009). The concept of ego threat in social and personality psychology: Is ego threat a viable scientific construct? *Personality and Social Psychology Review*, 13, 151-164. doi: 10.1177/1088868309342595
- Lee, K. M., & Kang, S. Y. (2002). Arithmetic operation and working memory: Differential suppression in dual tasks. *Cognition*, 83, 63-68. doi: 10.1016/S0010-0277(02)00010-0

- LeFevre, J-A., Sadesky, G. S., & Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem-size effect in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 216-230. doi: 10.1037/0278-7393.22.1.216
- Levy, B. R. (1996). Improving memory in old age through implicit self-stereotyping. *Journal of Personality and Social Psychology*, *71*, 1092–1107. doi: 10.1037/0022-3514.71.6.1092
- Leyens, J-P., Désert, 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*, 1189-1199. doi: 10.1177/0146167200262002
- Logel, C., Iserman, E. C., Davies, P. G., & Spencer, S. J. (2009). The perils of double consciousness: the role of thought suppression in stereotype threat. *Journal of Experimental Social Psychology*, *45*, 299-312. doi: 10.1016/j.jesp.2008.07.016
- Logie, R. H. (1995). *Visuo-spatial working memory*. Hillsdale, NJ: Erlbaum.
- London Mathematics Society [LMS]. (2013). Advancing women in mathematics: Good practice in UK university departments. The London Mathematical Society: London. Available from: <http://www.lms.ac.uk/sites/lms.ac.uk/files/files/reports/LMS-BTL-17Report.pdf>

- MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58*, 593-614. doi: 10.1146/annurev.psych.58.110405.085542
- MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science, 1*, 173-181. doi: 10.1023/A:1026595011371
- Maddux, J. E. (1993). Social cognitive models of health and exercise behavior: An introduction and review of conceptual issues. *Journal of Applied Sport Psychology, 5*, 116- 140. doi: 10.1080/10413209308411310
- Mandler, G. (2004). Consciousness: Respectable, useful and probably necessary. In: Solso, R. (Eds) *Information processing and cognition: The Loyola symposium*. New Jersey: Lawrence Erlbaum.
- Martin, A. J. (2015). Implicit theories about intelligence and growth (personal best) goals: Exploring reciprocal relationships. *British Journal of Educational Psychology, 85*, 207-223. doi: 10.1111/bjep.12038
- Marx, D. M. (2012). Differentiating theories: A comparison of stereotype threat and stereotype priming effects. In M. Inzlicht & T. Schmader (Eds.), *Stereotype threat: Theory, process, and application* (pp. 124-140). New York, NY: Oxford University Press.
- Marx, D. M., Stapel, D. A., & Muller, D. (2005). We can do it: The interplay of construal orientation and social comparison under threat. *Journal of Personality and Social Psychology, 88*, 432-446. doi: 10.1037/0022-3514.88.3.432

- 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*, 207-230. doi: 10.1207/S15327043HUP1603_3
- McCauley, C., Stitt, C. L., & Segal, M. (1980). Stereotyping: From prejudice to prediction. *Psychological Bulletin*, *87*, 195-208. doi: 10.1037/0033-2909.87.1.195
- McFall, S., Jamieson, J. P., & Harkins, S. (2009). Testing the mere effort account of the evaluation–performance relationship. *Journal of Personality and Social Psychology*, *96*, 135-154. doi: 10.1037/a0012878
- McGarty, C., Yzerbyt, V. Y., & Spears, R. (2002). *Stereotypes as explanations: The formation of meaningful beliefs about social groups*. Cambridge: Cambridge University Press.
- McGuire, W. J., McGuire, C. V., & Winton, W. (1979). Effects of household sex composition on the salience of one’s gender in the spontaneous self-concept. *Journal of Experimental Social Psychology*, *15*, 77-90. doi: 10.1016/0022-1031(79)90020-9
- McKenzie, K. (2013). Developing a growth mindset: The secret to improving your grades. Academic Success Summit Program. East Stroudsburg University, East Stroudsburg, PA. Retrieved from http://www4.esu.edu/academics/enrichment_learning/documents/pdf/developing_growth_mindset.pdf

- McKown, C., & Weinstein, R. S. (2003). The development and consequences of stereotype consciousness in middle childhood. *Child Development, 74*, 498-515. doi: 10.1111/1467-8624.7402012
- Miller, G. A. (1962). *Psychology: The science of mental life*. Oxford: Harper & Row.
- Mitchell, J. P., Macrae, N. C., & Gilchrist, I. D. (2002). Working memory and the suppression of reflexive saccades. *Journal of Cognitive Neuroscience, 14*, 95-103. doi: 10.1162/089892902317205357
- Miyake, A., & Shah, P. (1999). Toward unified theories of working memory: Emerging general consensus, unresolved theoretical issues, and future research directions. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 442–481). New York: Cambridge University Press.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100. doi: 10.1006/cogp.1999.0734
- Moè, A. (2009). Are males always better than females in mental rotation? Exploring a gender belief explanation. *Learning and Individual Differences, 19*, 21-27. doi: 10.1016/j.lindif.2008.02.002
- Moè, A., & Pazzaglia, F. (2006). Following the instructions!: Effects of gender beliefs in mental rotation. *Learning and Individual Differences, 16*, 369-377. doi: 10.1016/j.lindif.2007.01.002

- Moher, D., Liberato, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and analyses: The PRISMA statement. *PLoS Medicine*, *6*, e1000097. doi: 10.1371/journal.pmed1000097
- Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology*, *81*, 111-121. doi: 10.1111/j.2044-8295.1990.tb02349.x
- Mrazek, M. D., Chin, J. M., Schmader, T., Hartson, K. A., Smallwood, J., & Schooler, J. W. (2011). Threatened to distraction: Mind-wandering as a consequence of stereotype threat. *Journal of Experimental Social Psychology*, *47*, 1243-1248. doi: 10.1016/j.jesp.2011.05.011
- Mullholland, J., Hansen, P., & Kaminski, E. (2004). Do single-gender classrooms in coeducational settings address boys' underachievement? An Australian study. *Educational Studies*, *30*, 19-32. doi: 10.1080/0305569032000159714
- Munoz, D. P., & Everling, S. (2004). Look away: The anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience*, *5*, 218-228.
- 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*, 879-885. doi: 10.1111/j.1467-9280.2007.01995.x

- Nadler, J. T., & Clark, M. H. (2011). Stereotype threat: A meta-analysis comparing African Americans to Hispanic Americans. *Journal of Applied Social Psychology, 41*, 872-890. doi: 10.1111/j.1559-1816.2011.00739.x
- Nario-Redmond, M. R., Biernat, M., Eldelman, S., & Palenske, D. J. (2004). The social and personal identities scale: A measure of differential importance ascribed to social and personal self-categorizations. *Self and Identity, 3*, 143-175. doi: 10.1080/13576500342000103
- National Science Foundation [NSF]. Division of Science Resources Statistics. (2008). *Science and engineering degrees: 1966–2006* (Detailed Statistical Tables) (NSF 08-321). Arlington, VA: Available at <http://www.nsf.gov/statistics/nsf08321/pdf/nsf08321.pdf>.
- National Science Foundation [NSF], National Center for Science and Engineering Statistics. (2013). *Women, minorities, and persons with disabilities in science and engineering*. Special Report NSF 13-204. Arlington: VA. Available at <http://www.nsf.gov/statistics/wmpd/>.
- Neuvilte, E., & Croizet, J-C. (2007). Can salience of gender identity impair math performance among 7-8 year old girls? The moderating role of task difficulty. *European Journal of Psychology of Education, 22*, 307-316. doi: 10.1007/BF03173428
- Nguyen, H-H, D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology, 93*, 1314-1334. doi: 10.1037/a0012702

- Nisbett, R. E., & Wilson, T. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review*, *84*, 231-260. doi: 10.1037/0033-295X.84.3.231
- Nosek B. A., Banaji, M. R., & Greenwald, A. G. (2002). Maths = male, me = female, therefore maths \neq me. *Journal of Personality and Social Psychology*, *83*, 44-59. doi: 10.1037/0022-3514.83.1.44
- Nosek, B. A., Smith, F. L., Sriram, N., Lindner, N. M., Devos, T., . . . Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 10593-10597. doi: 10.1073/pnas.0809921106
- O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's maths performance. *Personality Social Psychology Bulletin*, *29*, 782-789. doi: 10.1177/0146167203029006010
- OECD (2015). *The ABC of gender equality in education: Aptitude, behaviour, confidence*. PISA: OECD Publishing.
- Oyserman, D., Harrison, K., & Bybee, D. (2001). Can racial identity be promotive of academic efficacy? *International Journal of Behavioural Development*, *25*, 379-385. doi: 10.1080/01650250042000401
- Oyserman, D., Kemmelmeier, M., Fryberg, S., Brosh, H., & Hart-Johnson, T. (2003). Racial-ethnic self-schemas. *Social Psychology Quarterly*, *66*, 333-347. doi: 10.1016/j.jrp.2008.03.003

- Pahlke, E., Shibley-Hyde, J., & Allison, C. M. (2014). The effects of single-sex compared with coeducational schooling on students' performance and attitudes: A meta-analysis. *Psychological Bulletin*, *140*, 1042-1072. doi: 10.1037/a0035740
- Pavlova, M. A., Weber, S., Simoes, E., & Sokolov, A. N. (2014). Gender stereotype susceptibility. *PLoS One*, *9*(12): e114802. doi: 10.1371/journal.pone.0114802
- Pennington, C. R., Qureshi, A., Monk, R. L., & Heim, D. (2016). The effects of stereotype threat and contextual cues on alcohol users' inhibitory control. *Addictive Behaviors*, *54*, 12-17. doi: 10.1016/j.addbeh.2015.11.014
- Picho, K., & Stephens, J. M. (2012). Culture, context and stereotype threat: A comparative analysis of young Ugandan women in coed and single-sex schools. *The Journal of Educational Research*, *105*, 52-63. doi: 10/1080/00220671.2010.517576
- Picho, K., Rodriguez, A., & Finnie, L. (2013). Exploring the moderating role of context on the mathematics performance of females under stereotype threat: A meta-analysis. *The Journal of Social Psychology*, *15*, 299-333. doi: 10.1080/00224545.2012.737380
- Prati, F., Crisp, R. J., Meleady, R., & Rubini, M. (2016). Humanizing outgroups through multiple categorization: The roles of individuation and threat. *Personality and Social Psychology Bulletin*, *42*, 526-539. doi: 10.1177/0146167216636624

- Preacher, K. J., & Kelley, K. (2011). *Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. Psychological Methods, 16*, 93-115. doi: 10.1037/a0022658
- Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research, 42*, 185-227. doi: 10.1080/00273170701341316
- Purdie-Vaughns, V., Steele, C. M., Davies, P. G., Dittmann, R., & Randall-Crosby, J. (2008). Social identity contingencies: How diversity cues signal threat or safety for African Americans in mainstream institutions. *Journal of Personality and Social Psychology, 94*, 615-630. doi: 10.1037/0022-3514.94.4.615
- Puts, D. A., Cárdenas, R. A., Bailey, D. H., Burriss, R. P., Jordan, C. L., & Breedlove, S. M. (2010). Salivary testosterone does not predict mental rotation performance in men or women. *Hormones and Behavior, 58*, 282–289. doi: 10.1016/j.yhbeh.2010.03.005
- 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*, 803-815. doi: 10.1177/0146167204264088
- Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences, 20*, 110-122. doi: 10.1016/j.lindif.2009.10.005

- Régner, I., Smeding, A., Gimmig, D., Thinus-Blanc, C., Monteil, J-M., & Huguet, P. (2010). Individual differences in working memory moderate stereotype-threat effects. *Psychological Science, 21*, 1646-1648. doi: 10.1177/0956797610386619
- Reilly, D., Neumann, D. L., & Andrews, G. (2014). Sex differences in mathematics and science achievement: A meta-analysis of national assessment of educational progress assessments. *Journal of Educational Psychology, 107*, 645-662. doi: 10.1037/edu0000012
- Roberts, R. J., Hager, L. D., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General, 123*, 374-393. doi: 10.1037/0096-3445/123/4/374
- Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*, 207-231. doi: 10.1037/0096-3445.124.2.207
- Rosen, V. M., & Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory and Language, 39*, 418-436. doi: 101006/jmla.1998.2590
- 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*, 586-597. doi: 10.1002/ejsp.379

- Rosenthal, H. E. S., & Crisp, R. J. (2006). Reducing stereotype threat by blurring intergroup boundaries. *Personality and Social Psychology Bulletin*, 32, 501-511. doi: 10.1177/0146167205281009
- Rosenthal, H. E. S., & Crisp, R. J. (2007). Choking under pressure: When an additional positive stereotype affects performance for domain identified male mathematics students. *European Journal of Psychology of Education*, 22, 317-326. doi: 10.1007/BF03173429
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86, 638-641. doi: 10.1037/0033-2909.86.3.638
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart & Winston.
- Rothgerber, H., & Wolsiefer, K. (2014). A naturalistic study of stereotype threat in young female chess players. *Group Processes and Intergroup Relations*, 17, 79-90. doi: 10.1177/1368430213490212
- Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass*, 5/6, 369-371. doi: 10.1111/j.1751-9004.2011.00355.x
- 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, 239-250. doi: 10.1177/0146167209355062

- 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*, 949-966. doi: 10.1037/a0014846
- Rydell, R. J., Rydell, M. T., & Boucher, K. L. (2010a). The effect of negative performance stereotypes on learning. *Journal of Personality and Social Psychology, 99*, 883-896. doi: 10.1037/a0021139
- Rydell, R. J., Shiffrin, R. M., Boucher, K. L., Van-Loo, K., & Rydell, M. T. (2010b). Stereotype threat prevents perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 14042-14047. doi: 10.1073/pnas.1002815107
- Rydell, R. J., Van Loo, K. J., & Boucher, K. L. (2014). Stereotype threat and executive functions: Which functions mediate different threat-related outcomes? *Personality and Social Psychology Bulletin, 40*, 377-390. doi: 10.1177/0146167213513475
- Sacket, 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*, 7-13. doi: 10.1037/0003-066X.60.3.271
- Saenz, D. S. (1994). Token status and problem-solving deficits: Detrimental effects of distinctiveness and performance monitoring. *Social Cognition, 12*, 61-74. doi: 10.1521/soco.1994.12.1.61

- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38*, 194-201. doi: 10.1006/jesp.2001.1500
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*, 440-452. doi: 10.1037/0022-3514.85.3.440
- Schmader, T., Croft, A., & Whitehead, J. (2014). Why can't I just be myself? A social cognitive analysis of the working self-concept under stereotype threat. *Social Psychological and Personality Science, 5*, 4-11. doi: 10.1177/1948550613482988
- 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*, 835-850. doi: 10.1023/B:SERS.0000029101.74557.a0
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*, 336-356. doi: 10.1037/0033-295X.115.2.336
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide*. Pittsburgh, PA: Psychology Software Tools.
- Schunk, D. H. (1989). Self-efficacy and achievement behaviors. *Educational Psychology Review, 1*, 173-208. doi: 10.1007/BF01320134

- Schuster, C, Martiny, S. E., & Schmader, T. (2015). Distracted by the unthought: Suppression and reappraisal of mind wandering under stereotype threat. *PLoS One*, *10*(3): e0122207. doi: 10.1371/journal.pone.0122207
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*, 93–105. doi: 10.1037//0003-066X.54.2.93
- 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*, 38-56. doi: 10.1037/0022-3514.87.1.38
- Seitchik, A. E., & Harkins, S. G. (2015). Stereotype threat, mental arithmetic, and the mere effort account. *Journal of Experimental Social Psychology*, *61*, 19-30. doi: 10.1016/j.jesp.2015.06.006
- 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. doi: 10.1016/S0022-1031(02)00508-5
- Shackman, A. J., Sarinopoulos, I., Maxwell, J. S., Pizzagalli, D., Lavrie, A., & Davidson, R. J. (2006). Anxiety selectively disrupts visuospatial working memory. *Emotion*, *6*, 40-61. doi: 10.1037/1528-3542.6.1.40

- Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125*, 4–27. doi: 10.1037/0096-3445.125.1.4
- Shapiro, J. R. (2011). Different groups, different threats: A multi-threat approach to the experience of stereotype threats. *Personality and Social Psychology Bulletin*, *37*, 464-480. doi: 10.1177/0146167211398140
- Shapiro, J. R. (2012). Types of threats: From stereotype threat to stereotype threats. In T. Schmader & M. Inzlicht (Eds.), *Stereotype Threat: Theory, Process, and Application* (pp. 71-88). New York: Oxford University Press.
- Shapiro, J. R., & Neuberg, S. L. (2007). From stereotype threat to stereotype threats: Implications of a multi-threat framework for causes, moderators, mediators, consequences, and interventions. *Personality and Social Psychology Review*, *11*, 107-130. doi: 10.1037/0033-295X.115.2.336
- Shapiro, J. R., & Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields. *Sex Roles*, *66*, 175-183. doi: 10.1007/s11199-011-0051-0
- Shapiro, J. R., Williams, A. M., & Hambarchyan, M. (2013). Are all interventions created equal? A multi-threat approach to tailoring stereotype threat interventions. *Journal of Personality and Social Psychology*, *104*, 277-288. doi: 10.1037/a0030461

- Shibley-Hyde, J. (2014). Gender similarities and differences. *Annual Review of Psychology, 65*, 373-398. doi: 10.1146/annurev-psych-010213-115057
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*, 422-445. doi: 10.1037//1082-989X.7.4.422
- Siegler, R. S., & Shrager, J. (1984). A model of strategy choice. In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229-293). Hillsdale, NJ: Erlbaum.
- Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2005). Parents' socializing behavior and children's participation in math, science, and computer out-of-school activities. *Applied Developmental Science, 9*, 14–30. doi: 10.1207/s1532480xads0901_3
- Skorich, D. P., Webb, H., Stewart, L., Kostyanaya, M., Cruwyz, T., McNeill, K. ... O'Brien, K. J. (2013). Stereotype threat and hazard perception among provisional license drivers. *Accident Analysis and Prevention, 54*, 39-45. doi: 10.1016/j.aap.2013.02.002
- Smith, J. L. (2004). Understanding the process of stereotype threat: A review of mediational variables and new performance goal directions. *Educational Psychology Review, 16*, 177-206. doi: 10.1023/B:EDPR.0000034020.20317.89
- Smith, P. K., Jostmann, N. B., Galinsky, A. D., & van Dijk, W. W. (2008). Lacking power impairs executive functions. *Psychological Science, 19*, 441-447. doi: 10.1111/j.1467-9280.2008.02107.x.

- Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist*, *60*, 950-958. doi: 10.1037/0003-066X.60.9.950
- Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology*, *67*, 415-437. doi: 10.1146/annurev-psych-073115-103235
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*, 4–28. doi: 10.1006/jesp.1998.1373
- Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, *89*, 845-851. doi: 10.1037/0022-3514.89.6.845
- Spitzer, B., & Aronson, J. (2015). Minding and mending the gap: Social psychological interventions to reduce educational disparities. *British Journal of Educational Psychology*, *85*, 1-18. doi: 10.1111/bjep.12067
- Ståhl, T., Van Laar, C., & Ellemers, N. (2012). The role of prevention focus under stereotype threat: Initial cognitive mobilization is followed by depletion. *Journal of Personality and Social Psychology*, *102*, 1239-1251. doi: 10.1037/a0027678
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613–629. doi: 10.1037/0003-066X.52.6.613

- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*, 797-811. doi: 10.1037/0022-3514.69.5.797
- Steele, C. M., & Davies, P. G. (2003). Stereotype threat and employment testing: A commentary. *Human Performance*, *16*, 311-326. doi: 10.1207/S15327043HUP1603_7
- Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pp. 379–440). New York, NY: Academic Press.
- Sternberg, R. J., Grigorenko, E.L., & Kidd, K. K. (2006). Intelligence, race, and genetics. *American Psychologist*, *60*, 46-59. doi: 10.1037/0003-066X.60.1.46
- Stoet, G., & Geary, D. C. (2012). Can stereotype threat explain the gender gap in mathematics performance and achievement? *Review of General Psychology*, *6*, 93-102. doi: 10.1037/a0026617
- Stoet, G., & Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within- and across-nation assessment of 10 years of PISA data. *PLoS One*, *8*, e57988. doi: 10.1371/journal.pone.0057988
- 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*, 1667-1678. doi: 10.1177/014616702237648

- Stone, J., & McWhinnie, C. (2008). Evidence that blatant versus subtle stereotype threat cues impact performance through dual processes. *Journal of Experimental Social Psychology, 44*, 445-452. doi: 10.1016/j.jesp.2007.02.006
- 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*, 1213-1227. doi: 10.1037/0022-3514.77.6.1213
- Streets, V. N., & Major, D. A. (2014). The limited utility of stereotype threat research in organizational settings. *Industrial and Organizational Psychology, 7*, 447-449. doi: 10.1111/iops.12177
- 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*, 665-693. doi: 10.1111/j.1559-1816.2004.tb02564.x
- Stricker, L. J., & Ward, W. C. (2008). Stereotype threat in applied settings re-examined: A reply. *Journal of Applied Social Psychology, 38*, 1656-1663. doi: 10.1111/j.1559-1816.2008.00363.x
- Stroop, J. R. (1935). Studies of interference in serial variable reactions. *Journal of Experimental Psychology, 18*, 643-662.
- Stryker, S. (1980). *Symbolic interactionism: A social structural version*. Menlo Park: Benjamin Cummings.

- Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin*, 135, 859-884. doi: 10.1037/a0017364
- Swann Jr, W. B., Gómez, Á, Conroy-Seyle, D., Morales, J. F., & Huici, C. (2009). Identity fusion: The interplay of personal and social identities in extreme group behavior. *Journal of Personality and Social Psychology*, 96, 995-1011. doi: 10.1037/a00113668
- Swann, W. B., Jetten, J., Gómez, Á, Whitehouse, H., & Bastian, B. (2012). When group membership gets personal: A theory of identity fusion. *Psychological Review*, 119, 441-456. doi: 10.1037/a0028589
- Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics. Essex, England: Pearson Education Limited.
- Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Social Issues*, 25, 79-97. doi: 10.1111/j.1540-4560.1969.tb00620.x
- Tajfel, H. (1981). Human Groups and Social Categories: Studies in Social Psychology. Cambridge: Cambridge University Press.
- Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-48). Monterey, CA: Brooks/Cole.
- Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behaviour. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations*, (pp. 7–24). Chicago: Nelson-Hall.

- Tempel, T., & Neumann, R. (2014). Stereotype threat, test anxiety, and mathematics performance. *Social Psychology of Education, 17*, 491-501. doi: 10.1007/s11218-014-9263-9
- Titze, C., Jansen, P., & Heil, M. (2011). Single-sex school girls outperform girls attending a co-educative school in mental rotation accuracy. *Sex Roles, 65*, 704-711. doi: 10.1007/s11199-011-9947-y
- Tosto, M. G., Hanscombe, K. B., Haworth, C. M. A., Davis, O. S. P., Petrill, S. A., Dale, P. S., . . . Kovas, Y. (2014). Why do spatial abilities predict mathematical performance? *Developmental Science, 17*, 462-470. doi: 10.1111/desc.12138
- Trbovich, P. L., & LeFevre, J. A. (2003). Phonological and visual working memory in mental addition. *Memory and Cognition, 31*, 739-745. doi:10.3758/BF03196112
- Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford: Blackwell.
- Turner, J. C., Oakes, P. J., Haslam, A., & McGarty, C. (1994). Self and collective: Cognition and social context. *Personality and Social Psychology Bulletin, 20*, 454-463. doi: 10.1177/0146167294205002
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127-154.

- Tzelgov, J., & Henik, A. Suppression situations in psychological research: Definitions, implications, and applications. *Psychological Bulletin*, 109, 524-536. doi: 10.1037/0033-2909.109.3.524
- U.S. Department of Education. (2015). National Center for Education Statistics. *Gender differences in science, technology, engineering, and mathematics (STEM) interest, credits earned, and NAEP performance in the 12th grade*. Retrieved March 28, from <http://nces.ed.gov/pubs2015/2015075.pdf>
- Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: The maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114, 104-132. doi: 10.1037/0033-295X.114.1.104
- Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1302-1321. doi: 10.1037/0278-7393.30.6.1302
- van Veelan, R., Otten, S., Cadinu, M., & Hansen, N. (2016). An integrative model of social identification: Self-stereotyping and self-anchoring as two cognitive pathways. *Personality and Social Psychology Review*, 20, 3-26. doi: 10.1177/1088868315576642

- Van-Loo, K. J., Boucher, K. L., Rydell, R. J., & Rydell, M. T. (2013). Competition in stereotyped domains: Competition, social comparison, and stereotype threat. *European Journal of Social Psychology, 43*, 648-660. doi: 10.1002/ejsp.1977
- Verniers, C., & Martinot, D. (2015). Perception of students' intelligence malleability and potential for future success: Unfavourable beliefs towards girls. *British Journal of Educational Psychology, 85*, 289-299. doi: 10.1111/bjep.12073
- Voci, A., Hewstone, M., Crisp, R. J., & Rubin, M. (2008). Majority, minority, and parity: Effects of gender and group size on perceived group variability. *Social Psychology Quarterly, 71*, 114-142. doi: 10.1177/019027250807100203
- Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*, 1174-1204. doi: 10.1037/a0036620
- Wai, J., Cacchio, M., Putallaz, M., & Makel, M. C. (2010). Sex differences in the right tail of cognitive abilities: A 30 year examination. *Intelligence, 38*, 412-423. doi: 10.1016/j.intell.2010.04.006
- Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate intellectual ability of negatively stereotyped students. *Psychological Science, 20*, 1132-1139. doi: 10.1111/j.1467-9280.2009.02417.x

- Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: Bradford Books.
- Wheeler, S. C., Demarree, K. G., & Petty, R. E. (2007). Understanding the role of the self in prime-to-behavior effects: The active-self account. *Personality and Social Psychology Review*, *11*, 234-261. doi: 10.1177/1088868307302223
- Wheeler, S. C., Jarvis, W. B. G., & Petty, R. E. (2001). Think unto others: The self-destructive impact of negative racial stereotypes. *Journal of Experimental Social Psychology*, *37*, 173–180. doi: 10.1006/jesp.2000.1448
- White J. B. (2008). Fail or flourish? Cognitive appraisal moderates the effect of solo status on performance. *Personality and Social Psychology Bulletin*, *34*, 1171-1184. doi: 10.1177/0146167208318404
- Wickware, P. (1997). Along the leaky pipeline. *Nature*, *390*, 202-203. doi: 10.1038/36639
- Wout, D., Danso, H., Jackson, J. & Spencer, S. (2008). The many faces of stereotype threat: Group- and self-threat. *Journal of Experimental Social Psychology*, *44*, 792-799. doi: 10.1016.j/jesp.2007.07.005
- 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 and Review*, *13*, 813–819. doi: 10.3758/BF03194002

- Yeung, N. C., & von Hippel, C. (2008). Stereotype threat increases the likelihood that female drivers in a simulator run over jaywalkers. *Accident Analysis and Prevention*, *40*, 667-674. doi: 10.1016/j.aap.2007.09.003
- Zhang, S., Schmader, T., & Hall, W. M. (2013). L'eggo my ego: Reducing the gender gap in math by unlinking the self from performance. *Self and Identity*, *12*, 400-412. doi: 10.1080/15298868.2012.687012
- Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, *37*, 197–206. doi: 10.1086/651257

Appendices

Appendix A: Participant Information Sheet and Consent Form.

Participant Information Sheet

Purpose and Background:

The purpose of this study is to explore factors related to problem solving. It is hoped that studying such differences may lead to the development of interventions to increase individuals' performance in the domain of mathematics. The study will take a maximum of 30 minutes to complete, and requires you complete a short questionnaire and 48 mathematical problems. If you wish to participate in this study, please sign the consent form provided.

Potential Benefits of Participation:

The data and information collected will be analysed and written up into a report. This report will contain information to help researchers better understand factors relating to problem solving. You are eligible to receive either **1** Psychology course credit, or a payment of **£3** for your participation in this study.

Possible Risks of Participation:

You may experience a degree of discomfort during the problem-solving task if you come across a question which you do not know the answer to. Before completing the task, you will be given some practice questions. If you are still unsure of how to complete the test, please ask the researcher and they will demonstrate how to complete the problems. During the test, if you do not know an answer please just hazard a guess. Participant's test scores will only be analysed as a group, and your individual score will not be individually identifiable in **any** documentation.

Confidentiality:

All information you provide will be anonymous and will be word coded to ensure full anonymity (your memorable word). All data will be stored in a locked office to which only the investigators will have access. In line with ethical procedures, data will be retained for a period of five years following publication of any articles resulting from this work, after which they will be destroyed. No information reported will ever be directly attributed to you and it will not be possible to identify individual respondents.

Freedom to withdraw:

You have the right to withdraw from the study at any time, and can do so without penalty. You may also withdraw your data within 4 weeks of completing the study. You will need to recall and provide your memorable word when contacting the researcher to withdraw your data. Your data will then be removed and destroyed, and will not be used within any analyses or publications.

Appendix B. Debrief Sheet.

Debrief Sheet

Thank you for your participation in this study. **It is important to state that any negative stereotypes you have heard are not a true reflection of your mathematic ability and were only used to examine the effects that negative stereotypes may have on performance.**

The purpose of this study was to examine the effects that negative stereotypes have on individuals' performance. It was important that this specific information was withheld from you until completion of the study. This is so we can reliably investigate the research question. Participants were assigned to one of three groups: (1) self-as-target, (2) group-as-target, and (3) a non-threat control. Participants in the self-as-target condition were primed with a negative stereotype regarding their own **personal** maths ability. Participants in the group-as-target condition were primed with a negative stereotype that highlighted their **groups** (i.e. female) mathematical ability. The non-threat control did not hear any negative stereotype regarding their mathematical ability. All participants then completed a modular arithmetic maths test and a mindset questionnaire.

It was hypothesised that female participants' who were randomly assigned to either the self-as-target, or group-as-target threat conditions would perform significantly less maths problems correctly compared to the control condition. This is because we predict that negative stereotypes will hinder performance. We also predict that individuals who heard a negative stereotype regarding either their personal or groups mathematical ability will reveal a fixed mindset on the mindset quiz (i.e. intelligence is fixed and cannot be changed), compared to the control condition who will reveal a growth mindset (i.e. intelligence is malleable).

Please feel free to ask the research any questions regarding this research. If you have any more questions once you have participated please contact the lead research:

Charlotte Pennington: penninc@edgehill.ac.uk

Appendix C: Mindset Questionnaire, Experiment 6.

- 1. Mathematical ability is something people are born with that can't be changed.**

Strongly Agree Agree Disagree Strongly Disagree

- 2. No matter how intelligent you are at maths, you can always be more intelligent.**

Strongly Agree Agree Disagree Strongly Disagree

- 3. You can always substantially change how intelligent you are at maths.**

Strongly Agree Agree Disagree Strongly Disagree

- 4. You are a certain kind of person, and there is not much that can be done to really change that.**

Strongly Agree Agree Disagree Strongly Disagree

- 5. You can always change basic things about the kind of person you are.**

Strongly Agree Agree Disagree Strongly Disagree

- 6. Mathematical talent can be learned by anyone**

Strongly Agree Agree Disagree Strongly Disagree

7. Only a few people will be truly good at maths—you have to be “born with it.”

Strongly Agree Agree Disagree Strongly Disagree

8. Math is much easier to learn if you are male.

Strongly Agree Agree Disagree Strongly Disagree

9. The harder you work at maths, the better you will be at it.

Strongly Agree Agree Disagree Strongly Disagree

10. No matter what kind of mathematical person you are, you can always change substantially.

Strongly Agree Agree Disagree Strongly Disagree

11. Trying new maths problems is stressful for me and I avoid it.

Strongly Agree Agree Disagree Strongly Disagree

12. Some people are good at maths, and some are not – it’s not often that people change.

Strongly Agree Agree Disagree Strongly Disagree

13. I appreciate when people give me feedback about my maths performance.

Strongly Agree Agree Disagree Strongly Disagree

14. I often get angry when I get negative feedback about my maths performance.

Strongly Agree Agree Disagree Strongly Disagree

15. All human beings are capable of learning maths.

Strongly Agree Agree Disagree Strongly Disagree

16. You can learn new things, but you can't really change how intelligent you are at maths.

Strongly Agree Agree Disagree Strongly Disagree

17. You can do things differently, but the important parts of who you are can't really be changed.

Strongly Agree Agree Disagree Strongly Disagree

18. Human beings are basically good, but sometimes make terrible decisions.

Strongly Agree Agree Disagree Strongly Disagree

19. An important reason why I do my work is that I like to learn new things.

Strongly Agree Agree Disagree Strongly Disagree

20. Truly smart people do not need to try hard in maths.

Strongly Agree Agree Disagree Strongly Disagree