

MECHANISMS OF STEREOTYPE THREAT

Title: Stereotype Threat May Not Impact Women's Inhibitory Control or Mathematical Performance: Providing Support for the Null Hypothesis

Date of re-submission: 15/08/2018

Abstract: 150 words

Main text: 10612 words

Tables: 7

Figures: 0

Open Data Statement:

All data are available at: <https://osf.io/mdwyy/> DOI: 10.17605/OSF.IO/MDWYV

MECHANISMS OF STEREOTYPE THREAT

Abstract

Underpinned by the findings of Jamieson and Harkins (2007; Experiment 3, *Journal of Personality & Social Psychology*), the current study pits the mere effort motivational account of stereotype threat against a working memory interference account. In Experiment 1, females were primed with a negative self- or group stereotype pertaining to their visuospatial ability and completed an anti-saccade eye-tracking task. In Experiment 2 they were primed with a negative or positive group stereotype and completed an anti-saccade and mental arithmetic task. Findings indicate that stereotype threat did not significantly impair women's inhibitory control (Experiments 1 & 2) or mathematical performance (Experiment 2), with Bayesian analyses providing support for the null hypothesis. These findings are discussed in relation to potential moderating factors of stereotype threat, such as task difficulty and stereotype endorsement, as well as the possibility that effect sizes reported in the stereotype threat literature are inflated due to publication bias.

Key words: stereotype threat; mathematical performance; working memory; mere effort; Null Hypothesis Significance Testing; Bayesian analysis

MECHANISMS OF STEREOTYPE THREAT

Stereotype Threat May Not Impact Women's Inhibitory Control or Mathematical Performance: Providing Support for the Null Hypothesis

A breadth of research indicates that pejorative societal stereotypes can reduce performance on a range of diverse tasks across populations (Doyle & Voyer, 2016; Lamont, Swift, & Abrams, 2015; Nguyen & Ryan, 2008; Shapiro, 2011). In a seminal series of 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 they performed comparably to their Caucasian peers when the same test was presented as non-diagnostic of ability. Extending these findings, Spencer and colleagues (1999) found that women underperformed when they perceived a quantitative test to be confirmative of gender differences in mathematical aptitude. Conversely, women performed equivalently to men when this negative gender-maths stereotype was dismissed prior to the test. Such findings led to the suggestion that the race and gender-achievement gap might be explained partly by situational cues that heighten the salience of a discredited social identity and shape expectations for success. These initial studies have been criticised, however, for statistically controlling for prior achievement; an approach that exacerbates performance decrements in the stereotype threat condition and reduces them in the control condition (see Brown & Day, 2006; Sacket, Hardison, & Cullen, 2004). Despite this, hundreds of studies have since provided empirical support for the situational phenomenon coined *stereotype threat* (see Pennington, Heim, Levy, & Larkin, 2016; Spencer, Logel, & Davies,

MECHANISMS OF STEREOTYPE THREAT

2016 for theoretical reviews; however see Flore & Wicherts, 2015 for a critical review).

Research has revealed many factors that heighten individuals' susceptibility to stereotype threat. From a methodological viewpoint, performance decrements are more likely to occur under stereotype threat when the task is difficult (Hess, Hinson, & Hodges, 2009; Keller, 2007; Nguyen & Ryan, 2008). However, a recent meta-analysis casts doubt on task difficulty as a significant moderator of stereotype threat (Flore & Wicherts, 2015). It is also proposed that stereotype threat effects are more likely to emerge when individuals attribute worth to their social group membership (i.e., group identification; Brown & Pinel, 2003; Hess et al., 2009; Wout, Danso, Jackson, & Spencer, 2008), endorse the stereotype to be accurate (i.e., stereotype endorsement; Bonnot & Croizet, 2011; Elizaga & Markman, 2008; Schmader, Johns, & Barquissau, 2004), and identify strongly with the stereotyped domain (Davies, Aronson, & Salinas, 2004; Schmader, 2002). Nguyen and Ryan (2008) report in their meta-analytic review, however, that women with moderate relative to high domain identification are more affected by gender-maths stereotypes, with research further suggesting that individuals do not need to identify with the stereotyped domain or group to experience stereotype threat (Keller & Dauenheimer, 2003; Martiny et al., 2011). The complexity of these observed findings has led some scholars to theorise that individuals may experience unique forms of stereotype threat, which are moderated by different factors and underpinned by diverse mechanisms (Shapiro & Neuberg, 2007; Shapiro, Williams, & Hambarchyan, 2013; Wout et al., 2008).

MECHANISMS OF STEREOTYPE THREAT

Traditionally, theories have considered stereotype threat as a singular construct, however more recent research posits that women may be vulnerable to distinct experiences of stereotype threat that impair performance through concerns about their *personal* (i.e., self-as-target) or *social* identity (i.e., group-as-target; Barber, 2017; Shapiro & Neuberg, 2007; Shapiro et al., 2013). The multi-threat framework (Shapiro & Neuberg, 2007) proposes that women may experience 'self-as-target' stereotype threat when they perceive that stereotype-consistent performance will be judged as self-characteristic of personal aptitude. Conversely, they may experience 'group-as-target' stereotype threat when they apprehend that their performance will confirm and reinforce a negative stereotype as accurately representing their social group. Supporting this premise, research suggests that different groups are more susceptible to certain forms of stereotype threat (Shapiro, 2011) and that negative self- and group-based stereotypes may differentially affect performance (Wout et al., 2008). It is therefore important to acknowledge this distinction when evaluating prior stereotype threat studies because different priming techniques may result in somewhat contrasting findings (see Nguyen & Ryan, 2008).

Positively stereotyped social identities have also been shown to diminish performance; 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 suggesting that high expectations for personal success may facilitate performance (Baumeister, Hamilton, & Tice, 1985; Rosenthal & Jacobson, 1968), whereas high group-based expectations can

MECHANISMS OF STEREOTYPE THREAT

diminish performance (Brown & Josephs, 1999). Research also indicates that highly identified male mathematics students underperform when they are primed with both a positive gender and student identity compared to one of these positive social identities alone (Rosenthal & Crisp, 2007). Such research suggests that, just as the fear of confirming a negative stereotype can diminish performance, the pressure to confirm a positive stereotype may lead to difficulties in concentration which translates into poorer performance (Beilock & Carr, 2005; Cheryan & Bodenhausen, 2000; Rosenthal & Crisp, 2007; Tagler, 2012). The opposite effects have been found on tests of spatial ability, however, with women performing better on mental rotation tasks when they are primed with a positive gender-related stereotype compared to either a negative stereotype (Moè, 2009; Moè & Pazzaglia, 2006) or stereotype-nullifying information (Wraga, Duncan, Jacobs, Helt, & Church, 2006). Additional research is therefore required to determine the potential differential impact that stereotype incongruent information (i.e., women are better than men) exerts on women's visuospatial and mathematical performance.

In an effort to understand *how* stereotype priming may diminish or bolster performance, researchers have turned their attention to elucidate psychological mediators of these effects (c.f., Pennington, Heim, Levy, & Larkin, 2016 for a review). One theory that has gathered empirical support is the working memory interference account of stereotype threat (Beilock, Rydell, & McConnell, 2007; Schmader & Johns, 2003; Rydell, Van-Loo, & Boucher, 2014). Through the lens of Baddeley's (1986, 2000) multi-component model, researchers have proposed that the verbal ruminations garnered from negative stereotypes may co-opt the phonological working memory resources required

MECHANISMS OF STEREOTYPE THREAT

to solve complex mathematical problems (Beilock et al., 2007; Schmader & Johns, 2003; Schmader et al., 2008). Consistent with this notion, Schmader and Johns (2003) reported that women under stereotype threat recalled fewer words on an operation span task, with impairments in verbal working memory underpinning the relationship between stereotype threat and mathematical underperformance. Women have also been found to solve fewer difficult problems and report more task-related concerns under stereotype threat (Beilock et al., 2007; Cadinu, Maass, Rosabianca, & Kiesner, 2005; Johns, Inzlicht, & Schmader, 2008). Similarly, researchers have proposed that working memory resources may be depleted when individuals are primed with affirmative group stereotypes because the apprehension of positively representing their social group may lead them to “choke under pressure” (Beilock & Carr, 2005; Cheryan & Bodenhausen, 2000; Tagler, 2012). To this end, stereotype-relevant worries may operate like a resource-demanding secondary task, taxing the phonological component of working memory to diminish performance (Ashcraft & Kirk, 2001; Ashcraft & Krause, 2007).

A competing theory posits that motivation underpins the stereotype threat-performance relationship (Jamieson & Harkins, 2007; Seitchik & Harkins, 2015). The mere effort account (c.f., Jamieson & Harkins, 2007) proposes that the potential for evaluation facilitates the dominant response on a stereotype-relevant task. Within this framework, task performance remains unharmed when the dominant response is correct, but is debilitated when the dominant response is incorrect. Importantly, this theory also suggests that individuals experiencing stereotype threat will take steps to compensate for their

MECHANISMS OF STEREOTYPE THREAT

performance if they recognise that they have made an incorrect response and are provided with the opportunity to correct it.

Providing support for their theory, Jamieson and Harkins (2007; Experiment 3) utilised an anti-saccade eye-tracking task and found that stereotype threatened participants generated more reflexive eye-movements (incorrect saccades) towards a peripherally placed cue before identifying a target. Furthermore, they launched correct and corrective saccades more quickly than those in the control condition (non-threat). They also report a 'tendency' for participants under stereotype threat to launch reflexive saccades more quickly on both pro-saccade ($p = .06$) and anti-saccade trials ($p = .11$). Although some of these findings failed to reach conventional levels of statistical significance, Jamieson and Harkins (2007; Experiment 3) interpret them as providing converging support for the mere effort account relative to the working memory interference account of stereotype threat. They suggest that heightened motivation potentiates a dominant response, resulting in more reflexive saccades to the cue and quicker responses to correct for these errors. Jamieson and Harkins therefore theorise that participants under stereotype threat launch more reflexive saccades because they are vigilant to the negative stereotype and aim to disprove it, rather than lacking the working memory capacity required to inhibit such response.

The distribution of saccadic reaction times, errors and corrective saccades in the anti-saccade task can provide remarkable insights into cognitive functioning and indeed be used as sensitive clinical indicators (c.f., Antoniadou et al., 2013). Nevertheless, one possible issue with Jamieson and Harkins' (2007) interpretation that stereotype threatened participants respond

MECHANISMS OF STEREOTYPE THREAT

faster to a peripherally placed cue is that this behaviour does not necessarily imply intention to disprove the negative stereotype. Schmader, Johns and Forbes (2008) challenge the assertion that increased motivation to correct errors is incompatible with a working memory interference explanation of stereotype threat. Instead they argue that despite appearing motivated to correct for their mistakes, stereotype threatened participants continue to produce incorrect responses; a behaviour indicative of impaired working memory (Kane & Engle, 2003; Mitchell, Macrae, & Gilchrist, 2002). As such, a potential limitation of the task used by Jamieson and Harkins (2007; Experiment 3) is that it is not able to tease apart reliably whether stereotype threat-performance decrements occur due to heightened motivation or impaired working memory.

An additional problem with using the anti-saccade task to arbitrate between these competing theories is that there are known methodological issues that can influence performance on this particular task (for reviews see Hutton & Ettinger, 2006; Munoz & Everling, 2004). For example, Jamieson and Harkins (2007) presented the critical target after the peripheral cue was extinguished. This “gap effect” (Saslow, 1967) is known to increase errors and reduce saccadic reaction times in comparison to overlap conditions where the central fixation point remains onscreen whilst the target appears in the periphery (Crawford et al., 2013; Fischer & Weber, 1993). Employing this gap procedure, Jamieson and Harkins (2007; Experiment 3) report that, on average, participants in the stereotype threat condition launched correct saccades on only 40% of anti-saccade trials compared to control participants who were correct on 73% of trials (Cohen’s $d = 1.68$). The sheer magnitude of the

MECHANISMS OF STEREOTYPE THREAT

accuracy impairment reported by Jamieson and Harkins raises the question as to whether stereotype threat would still impair accuracy on the anti-saccade task in more controlled settings that do not exploit the gap effect (i.e., the overlap paradigm), thus proffering support for the generalisability of these initial findings.

Moreover, in the anti-saccade task employed by Jamieson and Harkins (2007) participants did not simply have to look towards or away from a peripheral target; they first had to launch a pro- or anti-saccade and then identify the orientation of a subsequently presented target. The addition of having a final target makes corrective saccades, in this instance, somewhat misleading because participants will almost certainly make a 'corrective saccade' towards this new target in order to make a response (i.e., a key press indicating target orientation). Indeed, both Jamieson and Harkins (2007) and the formative study by Roberts, Hager and Heron (1994) report that all participants made corrective saccades towards this second target regardless of experimental condition. However, in simple anti-saccade tasks, including those used to discriminate clinical pathology (Antoniades et al., 2013; Crawford et al., 2013; Hutton & Ettinger, 2006), only one target is presented in the periphery so that participants must either look towards or away from it. The critical point is that participants have to make the latter anti-saccade into a blank region of space that approximately mirrors the direction of the peripherally presented target. In this context, a corrective saccade is more useful because participants will only correct if they actually remember that they should not look at the peripheral target but instead direct their gaze towards the opposite blank

MECHANISMS OF STEREOTYPE THREAT

region. Without any cues reminding participants where to look (such as a second target), many errors on the anti-saccade task go uncorrected.

Building upon prior research, the current study therefore employs a simple anti-saccade task and adopts an overlap paradigm in an attempt to evaluate support for the mere effort or working memory explanations of stereotype threat. In doing so, we also draw upon the multi-threat framework (Shapiro & Neuberg, 2007) to distinguish between self- and group-relevant stereotypes to ascertain their influence on women's inhibitory control performance. In line with the priming techniques employed by Jamieson and Harkins (2007), Experiment 1 examined whether a negative self or group stereotype pertaining to women's visuospatial and mathematical ability diminished inhibitory control performance. Building on this, Experiment 2 investigated the influence of negative and positive group stereotype on both inhibitory and mathematical performance.

Experiment 1

Utilising an anti-saccade eye-tracking paradigm, Experiment 1 examined the influence of a negative self- and group-relevant stereotype on women's inhibitory control (termed "visuospatial performance" in Jamieson & Harkins, 2007). Experimental predictions were two-tailed, allowing us to pit the mere effort account against the working memory interference account of stereotype threat. The mere effort account predicts that participants primed with a negative stereotype should make more reflexive eye movements towards the target (incorrect saccades) relative to the control condition because increased motivation facilitates the dominant response. Additionally, it predicts that this heightened motivation will influence stereotype threatened participants to

MECHANISMS OF STEREOTYPE THREAT

launch quicker correct saccades (i.e., eye movements directed correctly away from the target) and quicker corrective saccades (i.e., eye movements directed to the correct location after an incorrect response) compared to participants who are not subject to evaluation (Jamieson & Harkins, 2007). The working memory interference account also predicts that participants who are primed with a negative group stereotype will launch more incorrect saccades towards the target relative to control participants. However, in contrast to the mere effort explanation, this theory predicts that stereotype threatened participants should launch slower correct saccades and be less likely to correct for incorrect responses owing to diminished working memory capacity (Rydell et al., 2014). Table 1 presents an overview of the experimental predictions.

[TABLE 1 HERE]

Method

Participants

A power analysis was conducted based on the response time difference for corrective saccades between the stereotype threat and control condition in Jamieson and Harkins' study (2007; Experiment 3; $f = .397$), accounting for the requirement of three experimental conditions in the present experiments. This was the smallest significant effect size of the focal analyses reported by Jamieson and Harkins (2007; Experiment 3), informing our rationale for inclusion. The power analysis indicated that a sample size of 66 participants was required to detect this lowest reported effect size with 80% power

MECHANISMS OF STEREOTYPE THREAT

(G*Power, Faul et al., 2007). Bayesian analyses were also utilised in addition to NHST to overcome limitations posed by inferences of statistical power (Dienes, 2014).

Sixty-four females were successfully recruited ($M_{age} = 22$ years, $SD = 5.53$; 87.5% White British) from a university in the United Kingdom and received course credits or monetary remuneration for their time. Of this sample, 95.3% were university students, with the majority studying Psychology (40.6%) or Health and Social Sciences (54.7%), and all spoke English as a first language. They were assigned randomly to one of three stereotype conditions: 1) self-as-target stereotype threat ($n = 21$); 2) group-as-target stereotype threat ($n = 23$); and 3) a non-threat control ($n = 20$). In order to control for similar levels of perceived mathematical ability and domain identification in the sample, participants were asked to report their perceived competence in mathematics ("I am good at maths") and the degree to which they valued the domain ("It is important to me that I am good at maths; 1 = Strongly Disagree, 9 = Strongly Agree). There were no significant differences in participants' self-reported maths skills (overall $M = 5.16$, $SD = 1.94$) and domain identification (overall $M = 5.95$, $SD = 1.96$) as a function of experimental condition ($p_s > .05$), and these two factors were not found to moderate stereotype threat effects in any of the forthcoming analyses.

Stereotype Threat Manipulation

Group-as-target stereotype threat. Participants in the group-as-target stereotype threat condition were primed with the identical manipulation employed by Jamieson and Harkins (2007, p. 548). They received the following

MECHANISMS OF STEREOTYPE THREAT

written manipulation, which informed them explicitly that their task performance would be a diagnostic indicator of gender-related ability (c.f., Aronson et al., 1999; Shapiro & Neuberg, 2007):

“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. In line with the multi-threat framework (Shapiro & Neuberg, 2007), we also examined the impact of a self-relevant stereotype on inhibitory control performance. Equivalent to the group-as-target prime, participants were informed of the negative gender-related stereotype based on research suggesting that participants should be knowledgeable of a negative stereotype in order to be susceptible to stereotype threat effects (Shapiro & Neuberg, 2007). They were then informed that the task would be diagnostic of their personal ability (c.f., Shapiro & Neuberg, 2007):

“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

MECHANISMS OF STEREOTYPE THREAT

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”.

Non-threat control. To nullify stereotype threat, participants in the control condition were informed that their anti-saccade performance would not be evaluated (c.f., Steele & Davies, 2003) and that the experiment was investigating the role of working memory (Schmader & Johns, 2003):

“This experiment investigates the role of working memory in problem solving. The task that you are about to undertake is **non-diagnostic** of ability.”

Measures

Anti-saccade eye tracking task.

The anti-saccade task was developed using Experiment Builder (SR Research Ltd) and participants' eye movements were recorded using an EyeLink 1000 desktop eye-tracker, with a sampling rate of 1,000 Hz. Participants completed 4 blocks of 84 anti-saccade trials (including 4 practice trials) followed by 4 blocks of 84 pro-saccade trials, with block order counterbalanced between participants. Each trial started with a fixation cross that was presented on the screen randomly for 800-1000ms. A target then appeared 8° to the left or right of the fixation point and remained onscreen for 1000ms. The targets consisted of a square (neutral stimuli) or a number (numerical stimuli) that were presented randomly and equally across trials.

MECHANISMS OF STEREOTYPE THREAT

These two different target-types were selected because previous research investigating inhibition from a mere effort account has used neutral stimuli (Jamieson & Harkins, 2007; Experiment 3), whereas numerical stimuli have been utilised to elucidate the working memory interference account (Rydell et al., 2014; Experiment 3). Each target was the same size (1.4°) to ensure that this did not influence inhibitory control (see Roberts, Hager, & Heron, 1994).

Procedure

Participants were recruited for a study that examined ostensibly factors relating to problem solving. After reporting normal or corrected-to-normal vision, they were seated in front of the eye-tracker with their heads stabilised by a chin rest 57cm from the computer monitor. Eye movements were calibrated using a 9-point calibration system and calibrations were only accepted if the average error was $< 0.5^\circ$. Before commencing with the anti-saccade eye-tracking task, participants were provided with additional written task information that corresponded to their experimental condition. Given the similarities between the two stereotype threat manipulations, the researcher also reiterated to participants that the proceeding task was diagnostic of personal (self-as-target) or gender-related ability (group-as-target) before they commenced with the task. 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 accurately as possible. During pro-saccade trials, participants were asked to look directly towards the target. Upon completion of the task, participants responded to two questions to evaluate the perceived effectiveness of the stereotype threat

MECHANISMS OF STEREOTYPE THREAT

manipulations. Specifically, participants were asked: “To what extent are there gender differences in visuospatial performance?” (1 = No differences, 10 = gender differences) and “Who do you believe performs better on this task?” (1 = males, 5 = both males and females perform equally, 10 = females; Jamieson & Harkins, 2007)¹. At the end of the experiment, participants were provided with both a verbal and written debrief, which emphasised that the stereotypes they had heard were not a true reflection of their ability and were used as an experimental manipulation to explore the phenomenon of stereotype threat.

Data Preparation

In accordance with trimming and exclusion criteria reported by Jamieson and Harkins (2007; Experiment 3, pp. 553), filters were used prior to data analysis to ensure that eye movements recorded by the eye tracker represented responses to the stimuli presented. Specifically, the initial four practice trials were removed from 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 invalid and were removed from the analysis. A total of 3% of pro-saccade and 3% of anti-saccade trials across all participants were excluded using this criterion. Eye movements were classed as anticipatory if participants initiated saccades in less than 80ms and saccades beginning at 1,000ms or greater were excluded because they could not have been initiated in response to the target (Crevit & Vandierendonck, 2005; Jamieson & Harkins, 2007). This criterion resulted in

¹ Where necessary, measures were scaled and reverse coded to bring them in line with those reported by Jamieson and Harkins (2007; Experiment 3).

MECHANISMS OF STEREOTYPE THREAT

the exclusion of an additional 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 errors on the anti-saccade task (resulting in $n = 60$ participants).

Results

Analysis Strategy

The results were analysed using both Null Hypothesis Significance Testing (NHST) and Bayesian analyses. First, a series of Analysis of Variance tests (ANOVAs) were conducted, with main effects and interactions elucidated with Bonferroni-corrected pairwise comparisons. We report the p -values and associated 95% confidence intervals for mean differences, as well as effect size measures (partial-eta squared and Cohen's d). In line with recommendations by Dienes and McLatchie (2018), we then calculated Bayes factors (B) for all 1 degree of freedom effects (see Martin, Sackur, Anllo, Naish, & Dienes, 2016 for similar analyses). Unlike NHST, Bayes factors do not depend on statistical power and do not attempt to control long-run error rate; rather, they quantify the degree of evidence and use the data themselves to determine the relative probability of different theories (Dienes, 2014; Jeffreys, 1939/1961). Bayes factors were computed using Dienes' (2008) calculator². Here we specify the alternative model using a half-normal distribution scaled by the raw effects reported by Jamieson and Harkins (2007; Experiment 3), unless otherwise stated. Conventionally, $Bs < .10$ are interpreted as strong evidence and Bs

²R Script created by Baguely and Kaye (2010).

MECHANISMS OF STEREOTYPE THREAT

<0.33 are interpreted as moderate evidence for the null hypothesis., (cf. Etz & Vandekerckhove, 2016). Values between 0.33 and 3 are often considered weak or “anecdotal” evidence (Dienes, 2014). Conversely, $B_s > 3$ are interpreted as moderate and $B_s > 10$ indicate strong evidence for the alternative hypothesis.

All raw data are available at: <https://osf.io/mdwyv/>

Stereotype Threat Manipulation Check

There was no significant main effect of stereotype condition on participants' responses to the first manipulation check, $F(2, 59) = 1.81, p = .17, \eta_p^2 = .06$. Bayes factors indicated weak support for the null hypothesis when evaluating whether participants in the self-as-target ($M_{diff} = 1.01, 95\% \text{ CI } [-.53, 2.54], d = .56, B_{H(0, 3.38)} = 1.29$) and group-as-target conditions ($M_{diff} = 1.01, 95\% \text{ CI } [-.46, 2.57], d = .54, B_{H(0, 3.38)} = 1.11$) endorsed that there were gender differences in visuospatial performance relative to the control condition.

There was, however, a significant main effect of stereotype condition on participants' responses to the second manipulation check, $F(2, 59) = 4.95, p = .01, \eta_p^2 = .14$. Participants in the group-as-target stereotype threat condition ($M = 4.68, SD = 1.86$) were more likely to report that men outperformed women relative to the control condition ($M = 6.47, SD = 1.64$), $M_{diff} = 1.80, 95\% \text{ CI } [.34, 3.25], p = .01, d = 1.02, B_{H(0, 1.90)} = 39.75$. There was also moderate evidence for the difference between the self-as-target condition ($M = 5.10, SD = 2.10$) compared to the control condition on this measure, $M_{diff} = 1.38, 95\% \text{ CI } [-.09, 2.85], p = .07, d = .73, B_{H(0, 1.90)} = 6.78$. See Table 2 for descriptive statistics.

[TABLE 2 HERE]

MECHANISMS OF STEREOTYPE THREAT

Anti-saccade task

Two separate analyses were conducted on correct saccades and corresponding saccadic reaction time (SRT) as a function of trial-type (pro and anti-saccade). There was a significant main effect of accuracy, with participants responding more accurately to pro-saccade ($M = .99$, $SD = .02$) relative to anti-saccade trials ($M = .86$, $SD = .16$), $F(1, 59) = 43.95$, $M_{diff} = .13$, 95% CI [.09, .17], $p < .001$, $\eta_p^2 = .43$, $B_{H(0, 0.03)} = 1.45 \times 10^6$. There was also a significant main effect of response time, with participants expectedly faster at responding to pro-saccade ($M = 182.12$, $SD = 24.38$) relative to anti-saccade trials ($M = 243.34$, $SD = 33.40$), $F(1, 59) = 204.97$, $M_{diff} = -.61.22$, 95% CI [-.69.78, -.52.67], $p < .001$, $\eta_p^2 = .78$, $B_{H(0, 37.73)} = 1.98 \times 10^{43}$.

Anti-saccade trials

A series of analyses were conducted for percentage accuracy and SRT of reflexive, corrective and correct saccades as a function of stereotype condition. There were no significant differences on any of the dependent variables as a function of the stimuli used (i.e., number vs. shape) and therefore this variable was collapsed within all analyses.

Correct Saccades. There was no significant main effect of stereotype condition on the percentage of correct anti-saccades, $F(2, 57) = 0.03$, $p = .97$, $\eta_p^2 = .001$. Bayes factors indicated moderate evidence for the null hypothesis when comparing the self-as-target ($M_{diff} = .01$, 95% CI [-.12, .13], $d = -.12$, $B_{H(0, .33)} = 0.13$) and group-as-target conditions ($M_{diff} = .01$, 95% CI [-.12, .14], $d = -$

MECHANISMS OF STEREOTYPE THREAT

.06, $B_{H(0, .33)} = 0.13$) to the control³. There was no significant main effect of stereotype condition on SRT for correct saccades, $F(2, 58) = 0.30$, $p = .75$, $\eta_p^2 = .01$. Bayes factors indicated noteworthy evidence for the null hypothesis when comparing SRT for the self-as-target ($M_{diff} = -4.35$, 95% CI [- 31.49, 22.78], $d = .12$, $B_{H(0, 80.44)} = 0.19$) and group-as-target conditions ($M_{diff} = -8.44$, 95% CI [-35.58, 18.69], $d = -.27$, $B_{H(0, 80.44)} = 0.28$) to the control.

Reflexive Saccades (Incorrect responses). There was no significant main effect of stereotype condition on the percentage of reflexive saccades, $F(2, 57) = 0.03$, $p = .97$, $\eta_p^2 = .001$. Bayes factors indicated strong evidence for the null hypothesis when comparing data for the self-as-target ($M_{diff} = -.01$, 95% CI [-.14, .12], $d = -.06$, $B_{H(0, 0.33)} = 0.12$) and group-as-target conditions ($M_{diff} = .01$, 95% CI [-.14, .12], $d = -.06$, $B_{H(0, 0.33)} = 0.13$) to the control. There was no significant main effect of stereotype condition on SRT of reflexive saccades, $F(2, 56) = 0.25$, $p = .78$, $\eta_p^2 = .009$. Bayes factors indicated weak evidence for the null hypothesis when comparing data for the self-as-target ($M_{diff} = -2.66$, 95% CI [-26.47, 21.15], $d = .08$, $B_{H(0, 25.2)} = 0.44$) and group-as-target conditions ($M_{diff} = -6.65$, 95% CI [-30.16, 16.86], $d = .30$, $B_{H(0, 25.2)} = 0.65$) to the control condition.

Corrective Saccades. Although there was a significant main effect of stereotype condition on the proportion of corrective saccades, $F(2, 57) = 3.57$, $p = .035$, $\eta_p^2 = .11$, pairwise comparisons between conditions were non-significant, $ps > .07$. Bayes factors indicated strong evidence for the null

³All comparisons between the self-as-target and group-as-target conditions provided either weak or strong support for the null hypothesis. These findings are not reported as they were not central to the experimental aims or hypotheses.

MECHANISMS OF STEREOTYPE THREAT

hypothesis when comparing the self-as-target ($M_{diff} = -.21$, 95% CI [-.44, .01], $d = -.74$, $B_{H(0, 0.33)} = 0.08$) and group-as-target conditions ($M_{diff} = -.21$, 95% CI [-.43, .02], $d = -.77$, $B_{H(0, 0.33)} = 0.08$) to the control. There was no significant main effect of stereotype condition on SRT for corrective saccades, $F(2, 53) = 0.30$, $p = .75$, $\eta_p^2 = .01$. Bayes factors indicated weak evidence for the null hypothesis when comparing the self-as-target ($M_{diff} = 21.16$, 95% CI [-47.18, 89.50], $d = -.24$, $B_{H(0, 47.95)} = 0.37$) and group-as-target conditions ($M_{diff} = 12.10$, 95% CI [-55.30, 79.51], $d = .16$, $B_{H(0, 47.95)} = 0.37$) to the control. Table 3 provides a summary of descriptive statistics. Analyses of pro-saccade trials are reported in Supporting Information File 1, as well as within-sample correlations for all dependent measures.

[TABLE 3 HERE]

Discussion

Experiment 1 utilised the anti-saccade eye-tracking paradigm to discern the mere effort and working memory interference accounts of stereotype threat. Findings indicate that priming a negative self- or group-relevant stereotype did not hamper participants' correct, corrective or reflexive saccadic accuracy or associated SRT compared to the control condition. Bayesian analyses corroborated these findings, offering substantial support for the null compared to the alternative hypotheses. This contrasts with the findings reported by Jamieson and Harkins (2007; Experiment 3), who found that participants under stereotype threat launched quicker correct and corrective saccades relative the control condition; a finding they interpret as support for the mere effort motivational account of stereotype threat. They also report that participants

MECHANISMS OF STEREOTYPE THREAT

launched reflexive saccades (incorrect eye-movements towards a peripherally placed cue) on a greater proportion of anti-saccade trials. As such, findings from Experiment 1 offer little support for the mere effort account of stereotype threat when using an anti-saccade task with an overlap paradigm. In addition, our findings do not lend support to a working memory interference account, which suggests that participants will launch slower correct and corrective saccades because of diminished working memory capacity.

In accordance with Jamieson and Harkins' (2007), participants in the current study 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 employed anti-saccade task is a relatively simple task that is used predominantly as a measure of inhibitory control (Munoz & Everling, 2004). Resultantly, participants may not have perceived this particular task to be a valid indicator of their mathematical ability, which may explain why both the self-as-target and group-as-target primes did not influence anti-saccade performance. Furthermore, the simplicity of this task may have obscured stereotype threat effects by not evoking sufficient working memory demand. In order to corroborate the findings of Experiment 1, we therefore conducted a second experiment using the same anti-saccade task, but also included a measure of mathematical performance. Participants were informed that they would complete both of these tasks to strengthen the veracity of the stereotype threat manipulation. In this experiment, as well as re-examining the influence of negative stereotype priming, we also explored the impact that a positive group stereotype exerts on inhibitory control and mathematical performance.

MECHANISMS OF STEREOTYPE THREAT

Experiment 2

Experiment 2 contrasted the impact of a negative gender-related stereotype on inhibitory control and mathematical performance with a positive stereotype that carried opposite performance implications (e.g., women are better at mathematics compared to men). Previous research has reported contrasting findings with regards to the effect that a positive stereotype exerts on performance. Some studies have found that stereotype incongruent information hampers mathematical performance, possibly because heightened expectations for success lead individuals to 'choke under pressure' (Cheryan & Bodenhausen, 2000; Rosenthal & Crisp, 2007). Conversely, other studies demonstrate that the salience of a positive stereotype bolsters spatial performance (e.g., mental rotation) by encouraging the expectation to succeed (Moé, 2009; Wraga et al., 2008).

From a mere effort perspective (Jamieson & Harkins, 2007), it is therefore plausible that the salience of a positive group-relevant stereotype might motivate participants to perform well in a bid to confirm the stereotype. Underpinned to this theory, participants primed with a positive stereotype would therefore be expected to launch more incorrect eye movements (reflexive saccades) towards a peripheral target on the anti-saccade task relative to those in a 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 erroneous responses on a greater proportion of anti-saccade trials. Indeed, these predictions are supported by prior

MECHANISMS OF STEREOTYPE THREAT

research, which suggests that positive group stereotypes bolster visuospatial performance (Moé, 2009; Wraga et al., 2009). Through the lens of a working memory interference account, however, the salience of a positive stereotype is theorised to heighten situational performance pressure, and resultantly lead to underperformance (c.f., Beilock & Carr, 2005; Cheryan & Bodenhausen, 2000; Rosenthal & Crisp, 2007). As a consequence of “choking under pressure”, participants are therefore expected to launch more incorrect saccades, and correct for these incorrect responses more slowly and less often compared to participants in the control condition. Table 4 presents these contrasting experimental predictions derived from the mere effort and working memory interference accounts of stereotype threat.

In line with a wealth of previous research (c.f., Beilock et al., 2007; Rydell et al., 2014; Spencer et al., 1999), it was predicted that women primed with a negative group stereotype would solve fewer mathematical problems compared to the control condition. This is particularly the case for difficult mathematical problems presented horizontally relative to vertically because such problems have been shown to place greater demands on verbal working memory (Beilock et al., 2007; Trbovich & LeFevre, 2003). 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 (Jamieson & Harkins, 2011; O’Brien & Crandall, 2003), 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).

MECHANISMS OF STEREOTYPE THREAT

[TABLE 4 HERE]

Method**Participants**

Decisions regarding sample size followed the same rationale as Experiment 1. Sixty female participants ($M_{age} = 21$ years, $SD = 5.87$; 98.3% White British) were successfully recruited from the same U.K university (66.7% Psychology students) and received course credits or monetary remuneration for their time. They were assigned equally to one of three conditions ($n = 20$ in each): 1) negative group-as-target stereotype; 2) positive group-as-target stereotype; and 3) a non-threat control condition. Participants' self-reported mathematical ability (overall $M = 4.91$, $SD = 1.54$) and domain identification (overall $M = 5.58$, $SD = 1.66$) did not significantly differ as a function of experimental condition, and did not moderate stereotype threat effects in any of the forthcoming analyses (all $ps > .05$).

Stereotype Threat Manipulations

The negative group-as-target prime and the control prime were identical as those used in Experiment 1.

Positive group-as-target stereotype. Participants assigned to the positive stereotype condition were primed with the following information which suggested that women typically outperform men on tests of visuospatial and 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

MECHANISMS OF STEREOTYPE THREAT

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.”

Additional Measures

Modular Arithmetic Task. In accordance with previous research (see Beilock et al., 2017; Beilock & Carr, 2005; Seitchik & Harkins, 2015), we used a modular arithmetic (MA) task to examine mathematical performance. This novel task is advantageous in the study of stereotype threat effects because working memory demand can be manipulated easily, and task familiarity is controlled for to a greater extent compared to using standardised national tests (e.g., SAT/GRE; Beilock et al., 2007; Beilock & Carr, 2005). The task was administered with E-Prime experimental software, and participants were instructed to judge the validity of 64 mathematical problems. Problems such as ‘ $43 = 16 \pmod{3}$ ’ were presented on the screen and participants were instructed to subtract the middle number from the first number (e.g., $43 - 16$) and then divide their answer by the number in brackets (e.g., $27/3$). Participants responded ‘true’ if the division resulted in a whole number and ‘false’ if the division resulted in a decimal number.

Working memory demand was manipulated through problem difficulty and orientation. Specifically, participants completed 32 simple and 32 difficult problems presented either horizontally or vertically. Simple problems required

MECHANISMS OF STEREOTYPE THREAT

a single-digit no borrow subtraction operation (e.g., $7 = 2 \pmod{5}$), whereas difficult problems required a double-digit borrow subtraction (e.g., $43 = 16 \pmod{3}$). Horizontally oriented problems are theorised to tax working memory significantly more than vertically presented problems because they appear in a different format to how individuals typically solve problems in Western cultures (c.f., Beilock et al., 2007; Trbovich & LeFevre, 2003). Accuracy scores were calculated by dividing the number of problems answered correctly by the total number of problems. Given the dichotomous nature of the task (i.e., true/false response), accuracy scores below chance were removed from the final analyses ($< 50\%$, $n = 4$; see Data Preparation).

Procedure

The procedure was equivalent to Experiment 1, with exception to the addition of the MA task, which was presented in a counterbalanced order with the anti-saccade task. Participants completed two practice questions of the MA task, one presented horizontally and the other vertically and stated explicitly that they understood the task instructions before moving onto the test block. Participants were not provided with scratch paper to show their calculations (as in Jamieson & Harkins, 2009; Seitchik & Harkins, 2015) because this lessens the demands placed on working memory resources and limits the extent to which a working memory interference account of stereotype threat can be elucidated (Raghubar, Barnes, & Hecht, 2010; Trbovich & LeFevre, 2003). After completing the anti-saccade task, participants completed the same two manipulation checks as Experiment 1. They then completed the same two questions, this time pertaining to the MA task (“To what extent are there gender

MECHANISMS OF STEREOTYPE THREAT

differences in mathematical performance?” and “Who do you believe performs better on this task?”), which were scored in the same manner.

Data preparation

As in Experiment 1, trimming and exclusion criteria followed that reported by Jamieson and Harkins (2007). A total of 4% of pro-saccade and 5% of anti-saccade trials were excluded because initial saccades exceeded 2.82°. 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. Eye-tracking data from three participants were removed due to excessive invalid center starts and calibration error. In accordance with prior research (Beilock & DeCaro, 2007; DeCaro, Rotar, Kendra, & Beilock, 2010), mathematical accuracy data from four participants were excluded from analyses because they responded with below chance performance (range after exclusion = .61-.98).

Results

Stereotype Threat Manipulation Check

Anti-saccade task. There was a marginally significant main effect of stereotype condition on the first manipulation check, $F(2, 50) = 2.93$, $p = .06$, $\eta_p^2 = .11$, with Bayesian analyses providing support for the alternative hypothesis. Participants in the negative group-as-target stereotype condition ($M = 5.72$, $SD = 2.54$) appeared to endorse gender differences in visuospatial performance to a greater extent than the control condition ($M = 3.71$, $SD = 2.54$), $M_{diff} = 2.02$, 95% CI [-.42, 4.46], $p = .14$, $d = .79$, $B_{H(0, 3.38)} = 3.78$. Participants in the positive stereotype condition ($M = 5.83$, $SD = 3.52$) also

MECHANISMS OF STEREOTYPE THREAT

seemingly endorsed gender differences in visuospatial performance to a greater extent than the control condition, $M_{diff} = .2.13$, 95% CI [-.31, 4.57], $p = .11$, $d = .69$, $B_{H(0, 3.38)} = 4.73$.

There was a significant main effect of stereotype condition on the second manipulation check, $F(2, 50) = 5.08$, $p = .01$, $\eta_p^2 = .17$. Participants in the negative group-as-target stereotype condition ($M = 4.72$, $SD = 1.45$) perceived that men would outperform women on the anti-saccade task relative to the control ($M = 6.59$, $SD = 1.50$), $M_{diff} = -1.87$, 95% CI [-3.45, -.19], $p = .02$, $d = 1.27$, $B_{H(0, 1.90)} = 19.43$. However, there was substantial evidence for the null hypothesis when comparing judgments in the positive stereotype ($M = 6.56$, $SD = 2.75$) to the control condition, $M_{diff} = -.03$, 95% CI [-1.71, 1.64], $d = -.01$, $p = 1.00$, $B_{H(0, 1.90)} = 0.32$.

Modular arithmetic task. There was a significant main effect of stereotype condition on the third manipulation check, $F(2, 50) = 3.53$, $p = .037$, $\eta_p^2 = .12$. Participants in the negative group-as-target stereotype condition were more likely to endorse gender differences in mathematical performance ($M = 6.88$, $SD = 2.00$) relative to the control condition ($M = 4.37$, $SD = 2.29$), $M_{diff} = 2.51$, 95% CI [.14, 4.88], $p = .035$, $d = 1.17$, $B_{H(0, 3.38)} = 12.98$. Conversely, there was strong evidence for the null hypothesis when comparing the positive stereotype ($M = 5.89$, $SD = 3.79$) to the control condition, $M_{diff} = 1.52$, 95% CI [-.78, 3.82], $p = .32$, $d = .49$, $B_{H(0, 3.38)} = 0.10$.

There was a significant main effect of stereotype condition on the fourth manipulation check, $F(2, 50) = 4.24$, $p = .02$, $\eta_p^2 = .15$. Participants in the negative group-as-target stereotype condition were more likely to report that men would outperform women on the MA task ($M = 3.75$, $SD = 1.18$) relative to

MECHANISMS OF STEREOTYPE THREAT

the control condition ($M = 5.74$, $SD = 1.66$), $M_{diff} = -1.99$, 95% CI [-3.74, -.24], $p = .02$, $d = 1.38$, $B_{H(0, 1.92)} = 22.51$. However, there was moderate evidence for the null hypothesis when comparing the positive stereotype ($M = 5.28$, $SD = 2.93$) to the control condition, $M_{diff} = -.46$, 95% CI [-2.15, 1.24], $d = -.19$, $p = 1.00$, $B_{H(0, 1.92)} = 0.22$. See Table 5 for descriptive statistics.

[TABLE 5 HERE]

Anti-saccade task

There was a significant main effect of anti-saccade accuracy, with participants responding more accurately on pro-saccade ($M = .99$, $SD = .05$) relative to anti-saccade trials ($M = .82$, $SD = .20$), $F(1, 56) = 41.90$, $M_{diff} = .17$, 95% CI [.12, .22], $p < .001$, $\eta_p^2 = .43$, $B_{H(0, 0.026)} = 1.46 \times 10^6$. There was also a significant main effect of SRT, with participants responding faster on pro-saccade ($M = 177.35$, $SD = 26.83$) relative to anti-saccade trials ($M = 248.87$, $SD = 47.13$), $F(1, 56) = 93.25$, $M_{diff} = -71.51$, 95% CI [-86.35, -56.68], $p < .001$, $\eta_p^2 = .62$, $B_{H(0, 0.026)} = 1.21 \times 10^{19}$.

Anti-saccade trials.

Correct Saccades. There was no significant main effect of stereotype condition on correct saccades, $F(2, 54) = 0.47$, $p = .63$, $\eta_p^2 = .02$. Bayes factors indicated weak evidence for the null hypothesis when comparing the negative group-as-target ($M_{diff} = -.04$, 95% CI [-.20, .12], $d = .22$, $B_{H(0, .33)} = .36$), and moderate evidence for the null hypothesis when comparing the positive stereotype condition to the control condition ($M_{diff} = .02$, 95% CI [-.14, .18], $d = -.07$, $B_{H(0, .33)} = .15$). There was no significant main effect of SRT for correct

MECHANISMS OF STEREOTYPE THREAT

saccades, $F(2, 54) = 0.50$, $p = .61$, $\eta_p^2 = .02$. Bayes factors indicated moderate evidence for the null when comparing data between the negative stereotype ($M_{diff} = .70$, 95% CI [-37.42, 38.83], $d = .02$, $B_{H(0, 80.44)} = 0.18$), and weak evidence for the null when comparing the positive stereotype ($M_{diff} = 13.72$, 95% CI [-24.41, 51.84], $d = -.27$, $B_{H(0, 80.44)} = 0.44$) to the control condition.

Reflexive saccades. There was no significant main effect of stereotype condition on incorrect saccades, $F(2, 54) = 0.47$, $p = .63$, $\eta_p^2 = .02$. Bayes factors indicated weak support for the null when comparing data between the negative stereotype condition ($M_{diff} = .04$, 95% CI [-.12, 20], $d = .22$, $B_{H(0, 0.33)} = 0.36$), and moderate support for the null when comparing the positive stereotype to the control condition ($M_{diff} = -.02$, 95% CI [-.18, .14], $d = -.07$, $B_{H(0, 0.33)} = 0.15$). There was no significant main effect of stereotype condition on reflexive saccade SRT, $F(2, 53) = 1.43$, $p = .25$, $\eta_p^2 = .05$. Bayes factors indicated strong evidence for the null hypothesis when comparing the negative stereotype ($M_{diff} = .13.63$, 95% CI [-10.75, 38.00], $d = -.48$, $B_{H(0, 25.2)} = .17$), and weak support for the alternative hypothesis when comparing the positive stereotype to the control ($M_{diff} = 15.28$, 95% CI [-9.44, 39.99], $d = -.52$, $B_{H(0, 25.2)} = 1.87$).

Corrective Saccades. There was no significant main effect of stereotype condition on the percentage of corrective saccades, $F(2, 54) = 0.33$, $p = .72$, $\eta_p^2 = .01$. Bayes factors indicated moderate support for the null when comparing the negative ($M_{diff} = .03$, 95% CI [-.20, .24], $d = .11$, $B_{H(0, 0.33)} = 0.33$) and positive stereotype conditions ($M_{diff} = -.05$, 95% CI [-.27, .17], $d = -.14$, $B_{H(0, 0.33)} = 0.18$) to the control condition. There was no significant main effect of stereotype

MECHANISMS OF STEREOTYPE THREAT

condition on corrective saccade SRT, $F(2, 54) = .001$, $p = .999$, $\eta_p^2 < .001$.

Bayes factors indicated weak evidence for the null hypothesis when comparing the negative ($M_{diff} = -1.73$, 95% CI [-87.10, 84.16], $d = .02$, $B_{H(0, 47.95)} = 0.60$) and positive stereotype condition ($M_{diff} = -.52$, 95% CI [-85.89, 84.86], $d = -.005$, $B_{H(0, 47.95)} = 0.59$) to the control condition. See Table 6 for descriptive statistics.

[TABLE 6 HERE]

Mathematical Performance

The Bayesian prior models for the analyses of mathematical performance were specified using the effect sizes reported by Pennington and Heim (2016), who used the same task. A 3 (Condition: Positive ST, Negative ST, Control) x 2 (Difficulty: High, Low) x 2 (Orientation: Horizontal, Vertical) mixed-design ANOVA was conducted on MA accuracy scores. There was a significant main effect of problem difficulty on accuracy scores, with participants solving fewer difficult ($M = .78$, $SD = .11$) compared to simple problems ($M = .96$, $SD = .09$), $F(1, 53) = 141.56$, $M_{diff} = -.18$, 95% CI [-.21, -.15], $p < .001$, $\eta_p^2 = .73$, $B_{H(0, 0.15)} = 2.70 \times 10^{29}$. Expectedly, there was a significant main effect of problem orientation, with participants solving fewer horizontally ($M = .85$, $SD = .10$) relative to vertically oriented problems ($M = .89$, $SD = .10$), $F(1, 53) = 5.36$, $M_{diff} = -.03$, 95% CI [-.06, -.004], $p = .025$, $\eta_p^2 = .09$, $B_{H(0, 0.03)} = 7.92$. There was also a significant two-way interaction between problem difficulty and orientation, $F(1, 53) = 5.49$, $p = .02$, $\eta_p^2 = .09$, $B_{H(0, 0.12)} = 5.70$, with participants solving fewer difficult horizontal ($M = .75$, $SD = .13$) compared to vertical problems ($M = .81$,

MECHANISMS OF STEREOTYPE THREAT

$SD = .14$), $M_{diff} = -.06$, 95% CI $[-.10, -.01]$, $d = -.44$, $p = .01$, $B_{H(0,.07)} = 7.12$.

However, there was weak evidence in favour of the null hypothesis when comparing simple problems as a function of horizontal ($M = .96$ $SD = .11$) and vertical orientation ($M = .96$, $SD = .09$), $M_{diff} = -.006$, 95% CI $[-.03, .02]$, $d = .0$, $p = .59$, $B_{H(0,.05)} = 0.36$.

There was no significant main effect of stereotype condition on mathematical performance, $F(2, 53) = 2.73$, $p = .07$, $\eta_p^2 = .09$. The current study obtained evidence for *non-significant* effect sizes consistent with those reported by Pennington and Heim (2016) when comparing the negative stereotype ($M = .85$, $SD = .10$) to the control condition ($M = .90$, $SD = .05$), $M_{diff} = -.05$, 95% CI $[-.12, .01]$, $p = .18$, $d = -.63$, $B_{H(0, 0.04)} = 3.78$. Similarly, there was a non-significant effect when comparing the positive stereotype ($M = .85$, $SD = .08$) to the control condition, $M_{diff} = -.05$, 95% CI $[-.12, .01]$, $p = .13$, $d = -.75$, $B_{H(0, 0.04)} = 5.14$. There were no significant interactions between experimental condition, problem demand, and orientation (all $p > .08$).

Bayesian Meta-Analysis

A fixed-effects meta-analysis was conducted using Dienes' (2008) calculator to test the main experimental hypotheses that priming a negative group stereotype has a detrimental impact on women's inhibitory control performance. Internal meta-analyses provide a measure of the total weight of evidence across studies (Goh, Hall, & Rosenthal, 2016). Only direct comparisons between the conditions matching those in Jamieson and Harkins (2007; Experiment 3) were included in the meta-analysis (Current Study 1: negative group threat, control; Current Study 2: negative group threat, control). The raw

MECHANISMS OF STEREOTYPE THREAT

effects and single study Bayes factors are shown in Table 7, along with the meta-analytic posterior mean (and SD), and 95% credible intervals. Individually, the level of evidence in support for the null hypothesis in both Study 1 and Study 2 varies from weak to strong. The meta Bayes factors, calculated by combining the two datasets and using this data to test the expected effect sizes specified using the results reported by Jamieson and Harkins (2007; Experiment 3), revealed that the overall body of evidence indicated substantial support for the null relative to the experimental hypothesis⁴.

[TABLE 7 HERE]

General Discussion

Across two experiments, the current research aimed to conceptually replicate and extend Jamieson and Harkins' (2007; Experiment 3) to elucidate whether heightened motivation or deficits in working memory account for the stereotype threat-performance relationship. The mere effort account (Jamieson & Harkins, 2007) theorises that women are motivated to disprove a negative gender-related stereotype pertaining to their visuospatial and mathematical ability. Consequently, this theory predicts that the potential for evaluation motivates participants to disprove the negative stereotype, which triggers prepotent responding (Jamieson & Harkins, 2007; McFall et al., 2009). On the anti-saccade task, this prepotent response influences stereotype threatened women to launch quicker correct saccades and correct for incorrect responses more often and quicker compared to those not subject to evaluation. In contrast, the

⁴Here we report Bayes Factors calculated with a half-normal model, but we also checked these for calculations using a normal distribution. Both models of the alternative hypothesis yielded the same conclusions.

MECHANISMS OF STEREOTYPE THREAT

working memory interference account predicts that women under stereotype threat will make more errors and generate saccades slower on this task because negative verbal ruminations disrupt working memory capacity (c.f., Jamieson & Harkins, 2007; McFall et al., 2009).

Despite having these two-tailed predictions, the current studies were unable to provide support for either the mere effort or working memory explanations of stereotype threat. Specifically, priming a negative self- or group-relevant stereotype (Experiment 1) did not appear to influence reflexive saccades launched incorrectly towards a peripherally placed target, nor the time it took to generate correct and corrective saccades. Moreover, the saliency of a negative or positive group stereotype (Experiment 2) did not influence significantly women's inhibitory control or mathematical performance. These findings garnered from Null Hypothesis Significance Testing (NHST) were augmented by a Bayesian-meta analysis, which proffered substantial evidence in favor of the null over the alternative hypothesis specified using the results of Jamieson and Harkins (2007; Experiment 3).

There are considerable differences between the current study and the original by Jamieson and Harkins (2007; Experiment 3), which may proffer some explanations for these discrepant findings. First, Jamieson and Harkins utilised a gap procedure whereby participants were instructed to look away from a peripherally placed flashing cue and, once this target had been extinguished, identify the orientation of a target that appeared on the opposite side of the screen. The "gap effect" (Saslow, 1967) has been shown to increase errors and reduce saccadic reaction times relative to conditions in which the central

MECHANISMS OF STEREOTYPE THREAT

fixation point remains onscreen whilst a peripheral target is presented (Crawford et al., 2013; Fischer & Weber, 1993). Overcoming the issues posed by this design, the current study utilised an overlap anti-saccade paradigm, whereby participants were instructed to look directly at a fixation cross and then directly away from a peripherally placed target. However, it could be argued that the current design was simpler and did not co-opt working memory to the same extent as Jamieson and Harkins' task. A similar argument could be made for task difficulty on the modular arithmetic task, particularly given that the mean accuracy was 78%. Indeed, such explanation is supported by prior research, which indicates that stereotype threat effects may only be elicited in difficult tasks (Allison, Redhead, & Chan, 2017; Keller, 2007). Other research underpinned by the mere effort hypothesis, however, suggests that stereotype threat facilitates simple task performance (Jamieson & Harkins, 2011; O'Brien & Crandall, 2003), with stereotype threatened participants performing better on a simple Stroop interference task (Jamieson & Harkins, 2011; McFall et al., 2009). Such theoretical rationale does not explain why participants' saccadic response accuracy and latencies (Experiment 1 & 2) or their mathematical performance on simple and difficult problems (Experiment 2) did not differ as a function of stereotype threat in the current study. If women are motivated to disprove a negative gender-related stereotype, then they should launch quicker eye movements away from a peripherally placed target and correct for any incorrect responses on a greater proportion of trials on the anti-saccade task. Further, their mathematical performance would be expected to differ between simple and difficult problems.

The current study employed equivalent stereotype threat primes as

MECHANISMS OF STEREOTYPE THREAT

Jamieson and Harkins (2007). However, it is possible that participants in the original study endorsed these primed stereotypes to a greater extent than participants in the current study. For example, in the current study, the manipulation checks indicated mixed support as to whether participants in the negative group stereotype condition endorsed gender differences in visuospatial performance and the effect sizes for the first manipulation check were substantially smaller than those reported by Jamieson and Harkins (2007; Experiment 3, Cohen's $d = 1.28$ compared to .54 in Exp. 1 and .79 in Exp. 2). Indeed, research has demonstrated that stereotype threat effects typically emerge when participants endorse the stereotype to be an accurate representation of their group membership (Croizet, 2011; Elizaga & Markman, 2008; Schmader, Johns, & Barquissau, 2004). Nevertheless, other research suggests that stereotype endorsement is not necessary to evoke stereotype threat effects, with performance decrements observed in newly created stigmatised groups (Martiny et al., 2011).

Experiment 2 revealed further, that although participants primed with a group-as-target stereotype seemingly endorsed this negative stereotype, it did not appear to significantly influence their inhibitory control or mathematical performance. Closer inspection of the means indicates that, despite there being a significant difference between the stereotype threat and control condition on this measure, those under group-as-target threat reported a neutral response, suggesting that they may have believed males and females performed equivalently on these tasks. This may have masked any potential differences in visuospatial performance because participants in the stereotype threat conditions may have doubted the accuracy of the stereotype. The current

MECHANISMS OF STEREOTYPE THREAT

findings therefore throw into question the replicability of stereotype threat primes across independent studies.

Previous research has also demonstrated that stereotype threatened females underperform to a greater extent when the experimenter is male compared to female (Stone & McWhinnie, 2008), possibly because they hold monitory status in the performance context (see Inzlicht & Ben-Zeev, 2000). In the current study, participants were tested by a female experimenter who was situated outside of the room during testing, whereas in Jamieson and Harkins (2007) study, a male experimenter was present throughout the task. In a bid to control for the potential impact of this, Jamieson and Harkins (2007) instructed participants that the male experimenter was seated so that he could not see their computer screen. However, explicitly informing participants of this may have had the unintended consequence of making them vigilant to the experimenter's gender, particularly given the stereotyped context of the task. It is therefore plausible that differences between the current study and that of Jamieson and Harkins (2007) could arise from subtle environmental cues, such as experimenter gender, that have the potential to modulate the experience of stereotype threat.

Overall, the current findings run contrary to a wealth of studies demonstrating that priming negative gender related stereotypes impairs women's mathematical performance (Beilock et al., 2007; Rydell et al., 2014; Spencer et al., 1999). They also contrast with prior studies indicating that women underperform on mathematical tests (Beilock & Carr, 2005; Cheryan & Bodenhausen, 2000; Rosenthal & Crisp, 2007; Tagler, 2012), but perform

MECHANISMS OF STEREOTYPE THREAT

better on spatial tasks (e.g., Moé, 2009; Wraga et al., 2007) when they are primed with a positive stereotype. It is worth noting, however, that recent research suggests that the stereotype threat literature may be subject to publication bias; a phenomenon whereby significant findings are published and disseminated at a substantially greater rate than non-significant findings (Flores & Wicherts, 2015). Whilst this could have stemmed from the desirable implication that stereotype threat might partly explain real-world achievement outcomes (see seminal papers by Spencer et al., 1999; Steele & Aronson, 1995), the sheer amount of positive findings published in the literature is problematic because it disproportionately inflates effect size estimates and biases meta-analyses. The results reported here suggest that the null hypothesis is a substantially better predictor of the data than the alternative hypothesis specified by previous findings (Jamieson & Harkins, 2007; Experiment 3), with none of the 95% credible intervals of the replicated effects excluded values around zero. As such, the magnitude of the effects that negative gender-related stereotypes exert on women's inhibitory control performance (and other task performance) may be smaller than that reported in original studies and may be inflated by small sample sizes and publication bias (see Flores & Wicherts, 2015).

Limitations

The multi-threat framework contends that individuals may experience 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). Nevertheless, the primes used to evoke negative self- and group-as-target stereotype threat primes in Experiment 1 were very similar, and

MECHANISMS OF STEREOTYPE THREAT

it could be questioned whether participants were able to accurately identify whether their personal or social identity was being targeted. Future research would therefore benefit from assessing whether performance decrements are moderated by the importance that people ascribe to their personal and social identities under different stereotype threat conditions (c.f., Nario-Redmond, Biernat, Eldelman, & Palenske, 2004). Such work might reveal whether stereotype threat is a multi-faceted 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.

Working memory capacity has been implicated in successful anti-saccade task performance by facilitating the top-down inhibition of reflexive, automatic saccades (Meier, Smeekens, Silvia, Kwapil, & Kane, 2018; Munoz & Everling, 2004; Unsworth et al., 2003). As such, this task provides a valid measure to assess psychological phenomenon that is theorised to impact on working memory resources, and has been used extensively as a clinical tool to assess neurological and clinical conditions impacting upon executive functioning (Antoniades et al., 2013; Crawford et al., 2013; Hutton & Ettinger, 2006). Nevertheless, one could question the extent to which temporary manipulations, such as evoking stereotype threat through explicit priming, can exert upon automatic, reflexive eye movements. Future research would therefore benefit from assessing the extent to which explicit priming techniques can impact upon automatic responding and behaviours, as well as examining the duration of these effects in experimental tasks.

MECHANISMS OF STEREOTYPE THREAT

The current experiments were powered based on effect sizes reported by Jamieson and Harkins (2007; Experiment 3) and were only able to detect significantly large effects. This may not be the best approach, however, because it is likely that some effect sizes reported in previous research are inflated due to small sample sizes or publication bias (Ioannidis, 2008; Nuijten, van Assen, Veldkamp, & Wicherts, 2015; Szucs & Ioannidis, 2017). We believe we overcome this limitation by employing Bayesian analyses with a half-normal distribution, which considers smaller effect sizes more plausible than larger effect sizes and is useful when basing predictions on published literature. Moreover, we conducted a Bayesian meta-analysis, pooling the data from Experiment 1 and 2 to provide substantial support for the null hypothesis. Unlike NHST, Bayesian analyses do not rely on inferences based on statistical power and can show that a study unable to detect interesting effect sizes due to low statistical power provides evidence for the null relative to the alternative hypothesis, or that a high-powered non-significant finding proffers no evidence for the null compared to the alternative hypothesis (Dienes & McLatchie, 2018). We therefore recommend future research to power experiments based on the smallest effect size of interest that is deemed theoretically or practically meaningful (see Lakens, McLatchie, Isager, Scheel, & Dienes, 2018; Lakens, Scheel, & Isager, 2018), and to utilise Bayesian analyses to make accurate statistical inferences (see Dienes, 2014; Lakens et al., 2018a).

Conclusion

Looking to the published literature, the effects of stereotype threat on performance appear to be widespread (Nguyen & Ryan, 2008; Spencer et al., 2016; c.f., however, Flores & Wicherts for a critical review), yet the underlying

MECHANISMS OF STEREOTYPE THREAT

mechanisms for the stereotype-threat performance relationship remain debated (Pennington et al., 2016). The current research examined the impact of distinct self- and group-relevant stereotypes on women's mathematical and visuospatial performance. It also set out to elucidate whether diminished working memory or enhanced motivation mediate the stereotype threat-performance relationship. Findings from Experiment 1 indicate that a negative self- or group-relevant stereotype did not appear to influence women's visuospatial performance on the anti-saccade eye-tracking task. Experiment 2 further corroborated these findings, indicating that a negative and positive group-relevant stereotype did not significantly influence visuospatial or mathematical performance. Bayes factors corroborated the inferential analyses, indicating substantial support for the null hypothesis relative to the alternative hypothesis specified by Jamieson and Harkins (2007; Experiment 3). We proffer explanations for the differences between the current study and the original by Jamieson and Harkins, including the difficulty of the anti-saccade task, the degree to which participants endorsed the stereotype, and the potential for subtle environmental cues, such as experimenter gender, to influence the measurement of stereotype threat effects. We also consider the possibility that low-powered published studies exaggerate effect sizes and bias the stereotype threat literature. Future research that employs additional tasks and larger sample sizes would therefore be welcomed to examine the robustness of stereotype threat priming and to investigate further the underlying mechanisms of stereotype threat effects. Multi-lab collaborative studies present one of many ways of rising to these important challenges. Moreover, we advocate the use of Bayesian statistics, which serve to quantify whether the

MECHANISMS OF STEREOTYPE THREAT

data supports experimental predictions to elucidate the evidential value of reported stereotype threat studies within the extant literature (c.f., Dienes & McLatchie, 2018).

MECHANISMS OF STEREOTYPE THREAT

References

- Allison, C., Redhead, E. S., & Chan, W. (2017). Interaction of task difficulty and gender stereotype threat with a spatial orientation task in a virtual nested environment. *Learning and Motivation, 57*, 22-35.
doi:10.1016/j.lmot.2017.01.005
- Antoniades, C., Ettinger, U., Gaymard, B., Gilchrist, I., Kristjansson, A., Kennard, C., . . . Carpenter, R. H. (2013). An internationally standardised antisaccade protocol. *Vision Research, 84*, 1-5.
doi:10.1016/j.visres.2013.02.007
- 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/s10648-01-9200-4
- 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 & Review, 14*, 243-248.
doi:10.3758/BF03194059

MECHANISMS OF STEREOTYPE THREAT

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

Baguely, T., & Kaye, W. S. (2010). Review of: Understanding psychology as a science: An introduction to scientific and statistical inference, by Z. Dienes. *British Journal of Mathematical and Statistical Psychology*, 63(3), 695-698. doi:10.1348/000711009X481027

Barber, S. J. (2017). An examination of age-based stereotype threat about cognitive decline: Implications for stereotype threat theory and development. *Perspectives in Psychological Science*, 12, 62-90. doi:10.1177/1745691616656345

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

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. (2005). When high-powered people fail: Working memory and “choking under pressure” in math. *Psychological Science*, 16, 101-115. doi:10.1111/j.0956-7976.2005.00789.x

MECHANISMS OF STEREOTYPE THREAT

- 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., & 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., 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
- Bonnot, V., & Croizet, J-C. (2011). Stereotype threat and stereotype endorsement: Their joint influence on women's math performance. *International Review of Social Psychology*, *24*, 105-120.
- Brown, R. P., & Day, E. A. (2006). The difference isn't black and white: Stereotype threat and the race gap on raven's advanced progressive matrices. *Journal of Applied Psychology*, *91*(4), 979. doi: 10.1037/0021-9010.91.4.979
- 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

MECHANISMS OF STEREOTYPE THREAT

- Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of experimental social psychology, 39*(6), 626-633. doi:10.1016/S0022-1031(03)00039-8
- 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
- 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
- Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science, 313*, 1307-1310. doi:10.1126/science.1128317
- Crawford, T. J., Higham, S., Renvoize, T., Patel, J., Dale, M., Suriya, A., & Tetley, S. (2005). Inhibitory control of saccadic eye movements and cognitive impairment in Alzheimer’s disease. *Biological Psychiatry, 57*, 1052-1060. doi:10.1016/j.biopsych.2005.01.017
- Crevit, I., & Vandierendonck, A. (2005). Gap effect in reflexive and intentional prosaccades. *Neuropsychobiology, 51*, 39-44 doi:10.1159/000082854
- Davis C., Aronson, J., & Salinas, M. (2006). Shades of threat: Racial identity as a moderator of stereotype threat. *Journal of Black Psychology, 32*(4), 399-417. doi:10.1177/0095798406292464

MECHANISMS OF STEREOTYPE THREAT

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. *Quarterly Journal of Experimental Psychology*, 63, 1619-1630. doi:10.1080/17470210903474286

Dienes, Z. (2008). *Bayesian analysis calculator*. Retrieved from:

http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/bayes_norm_alposterior.swf

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results.

Frontiers in Psychology, 5, 781. doi:10.3389/fpsyg.2014.00781

Dienes, Z. & McLatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review*, 1-12.

doi:10.3758/s13423-017-1266-z

Doyle, R. A., & Voyer, D. (2016). Stereotype manipulation effects on math and spatial test performance: A meta-analysis. *Learning & Individual Differences*, 47, 103-116. doi:10.1016/j.lindif.2015.12.018

doi:10.1016/j.lindif.2015.12.018

Elizaga, R. A., & Markman, K. D. (2008). Peers and performance: How in-group and out-group comparisons moderate stereotype threat

effects. *Current Psychology*, 27(4), 290-300. doi:10.1007/s12144-008-9041-y

Etz, A., & Vandekerckhove, J. (2016). A Bayesian Perspective on the Reproducibility Project: Psychology. *PLoS One*, 11(2), e0149794.

doi:10.1371/journal.pone.0149794

MECHANISMS OF STEREOTYPE THREAT

- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social sciences. *Behavior Research Methods*, *39*, 175-191. doi:10.3758/BF03193146
- Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behavioral Brain Sciences* *16*, 553-610 doi:10.1017/S0140525X00031575.
- Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology*, *53*(1), 25-44. doi:10.1016/j.jsp.2014.10.002
- Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass*, *10*(10), 535-549.
doi:10.1111/spc3.12267
- 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:10.1080/03610730902716413
- Jamieson, J. P., & Harkins, S. G. (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 & Social Psychology Bulletin*, *35*, 1301-1314.
doi:1177/0146167209335165

MECHANISMS OF STEREOTYPE THREAT

- Jamieson, J. P., & Harkins, S. G. (2011). The intervening task method: Implications for measuring mediation. *Personality & Social Psychology Bulletin, 37*, 652-661. doi:10.1177/0146167211399776
- Jeffreys, H. (1939/1961). *The theory of probability*. Oxford, UK: Oxford University Press.
- 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
- Hutton, S. B., & Ettinger, U. (2006). The antisaccade task a research tool in psychopathology: A critical review. *Psychophysiology, 43*, 302-313. doi:10.1111/j.1469-8986.2006.00403.x
- Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology, 19*, 640-648. doi: 10.1097/EDE.0b013e31818131e7
- 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.

MECHANISMS OF STEREOTYPE THREAT

- 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, 323-338. doi:10.1348/000709906X113662
- Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. *Personality and Social Psychology Bulletin*, 29, 371-381. doi:10.1177/0146167202250218
- Lakens, D., McLatchie, N., Isager, P. M., Scheel, A. M., & Dienes, Z. (2018a). Improving inferences about null effects with Bayes factors and equivalence tests. *The Journals of Gerontology, Series B: Psychological Sciences*, Pre-print: <http://doi.org/10.17605/OSF/IO/QTZWR>
- Lakens, D., Scheel, A. M., & Isager, P. M. (2018b). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, 1, 259-269. doi: 10.1177/2515245918770963, pre-print: <http://doi.org/10.17605/OSF.IO/V3ZKT>
- 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. doi:10.1037/a0038586
- Martiny, S. E., Roth, J., Jelenec, P., Steffans, M. C., & Croizet, J-C. (2012). When a new group identity does harm on the spot: Stereotype threat in newly created groups. *European Journal of Social Psychology*, 42, 65-71. doi:10.1002/ejsp.840

MECHANISMS OF STEREOTYPE THREAT

- McFall, S. R., Jamieson, J. P., & Harkins, S. G. (2009). Testing the mere effort account of the evaluation-performance relationship. *Journal of Personality and Social Psychology, 96*, 135-154. doi:10.1037/a0012878
- Meier, M. E., Smeekens, B. A., Silvia, P. J., Kwapil, T. R., & Kane, M. J. (2018). Working memory capacity and the antisaccade task: A microanalytic-macroanalytic investigation of individual differences in goal activation and maintenance. *Journal of Experimental Psychology: Learning, Memory & Cognition, 44*, 68-84. doi: 10/1037/xlm0000431
- 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
- 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 & Individual Differences, 16*, 369-377. doi:10.1016/j.lindif.2007.01.002
- 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. doi:10.1038/nrn1345

MECHANISMS OF STEREOTYPE THREAT

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

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

Nuijten, M. B., van Assen, M. A., Veldkamp, C. L., & Wicherts, J. M. (2015). The replication paradox: Combining studies can decrease accuracy of effect size estimates. *Review of General Psychology*, 19, 172-182. doi:10.1037/gpr0000034

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin*, 29, 782-789. doi:10.1177/0146167203029006010

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*, 83, 353-368. doi:10.1111/bjep.12110

Pennington, C. R., Heim, D., Levy, A., & Larkin, D. (2016). Twenty years of stereotype threat research: A review of psychological mediators. *PLoS One*, 11, e0146487. doi:10.1371/journal.pone.0146487

MECHANISMS OF STEREOTYPE THREAT

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning & Individual Differences, 20*, 110-122.

doi:10.1016/j.lindif.2009.10.005

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

Rosen, V. M., & Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory & Language, 39*, 418-436.

doi:101006/jmla.1998.2590

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart & Winston.

Rydell, R. J., Rydell, M. T., & Boucher, K. L. (2010). The effect of negative performance stereotypes on learning. *Journal of Personality and Social Psychology, 99*, 883-896. doi:10.1037/a0021139

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

MECHANISMS OF STEREOTYPE THREAT

- Sackett, P. R., Hardison, C. M., & Cullen, M. J. (2004). On interpreting stereotype threat as accounting for African American-White differences on cognitive tests. *American Psychologist*, *59*(1), 7. doi:10.1037/0003-066X.59.1.7
- Saslow, M. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America*, *57*(8), 1030–1033. doi:10.1364/JOSA.57.001030
- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, *38*(2), 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., Johns, M., & Barquissau, M. (2004). The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. *Sex Roles*, *50*(11-12), 835-850. 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
- 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

MECHANISMS OF STEREOTYPE THREAT

- 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., & 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:10.1177/1088868306294790
- 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
- 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.
- 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., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology*, *67*, 415-437. doi:10.1146/annurev-psych-073115-103235
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identify and performance. *American Psychologist*, *52*, 613-629. doi:10.1037/0003-066X.52.6.613

MECHANISMS OF STEREOTYPE THREAT

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

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

Stone, J., & McWhinnie, C. (2008). Evidence that blatant versus subtle stereotype threat cues impact performance through dual processes. *Journal of Experimental Social Psychology*, *44*(2), 445-452.
doi:10.1016/j.jesp.2007.02.006

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

Szucs, D., & Ioannidis, J. P. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biology*, *15*, e2000797.

MECHANISMS OF STEREOTYPE THREAT

Tagler, M. J. (2012). Choking under pressure of a positive stereotype: Gender identification and self-consciousness moderate men's math test performance. *Journal of Social Psychology, 152*, 401-416.

doi:10.1080/00224545.2011.615353

Trbovich, P. L., & LeFevre, J. A. (2003). Phonological and visual working memory in mental addition. *Memory & Cognition, 31*, 739-745.

doi:10.3758/BF03196112

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 & Review, 13*, 813-819.

doi:10.3758/BF03194048

MECHANISMS OF STEREOTYPE THREAT

Tables

Table 1.

Experimental predictions for performance on anti-saccade trials based on contrasting theories.

	Negative stereotype (self/group)	
	Mere effort account	Working memory account
Correct %	ST fewer than control	ST fewer than control
Correct RT	ST quicker than control	ST slower than control
Corrective %	ST more than control	ST fewer than control
Corrective RT	ST faster than control	ST slower than control

Note: ST = Stereotype threat.

MECHANISMS OF STEREOTYPE THREAT

Table 2.

Descriptive statistics) for stereotype threat manipulation checks in Experiment 1.

	Self-as- target	Group- as-target	Control	Total
	<i>M (SD)</i>			
1. To what extent are there gender differences in visuospatial performance?	5.95 (1.99)	6.00 (2.25)	4.95 (1.54)	5.66 (1.99)
2. Who do you think performs better on this task?	5.09 (2.10)	4.68 (1.86) _a	6.47 (1.65) _a	5.37 (2.00)

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

MECHANISMS OF STEREOTYPE THREAT

Table 3.

Descriptive statistics for anti-saccade trials as a function of stereotype condition in Experiment 1.

	Self-as- target	Group-as- target	Control	Total
Correct %	.87 (.14)	.86 (.15)	.85 (.19)	.86 (.16)
Correct SRT	242.22 (40.63)	238.13 (32.84)	246.58 (29.36)	242.17 (34.35)
Reflexive %	.14 (.14)	.14 (.15)	.15 (.19)	.14 (.16)
Reflexive SRT	189.84 (40.83)	185.85 (20.88)	192.49 (23.90)	189.33 (29.32)
Corrective %	.54 (.32)	.55 (.28)	.75 (.24)	.61 (.30)
Corrective SRT	401.14 (86.35)	392.08 (78.06)	379.98 (87.51)	390.88 (82.95)

Note: % = percentage correct, SRT = saccadic reaction time, measured in milliseconds. Corrective saccades are a proportion of reflexive saccades.

MECHANISMS OF STEREOTYPE THREAT

Table 4.

Experimental predictions for performance on anti-saccade trials and mathematical performance based on contrasting theories.

	Positive stereotype	
	Mere effort	Working memory
Correct %	PST fewer than control	PST fewer than control
Correct RT	PST quicker than control	PST slower than control
Corrective %	PST more than control	PST fewer than control
Corrective RT	PST faster than control	PST slower than control

Note: PST: Positive stereotype threat

MECHANISMS OF STEREOTYPE THREAT

Table 5.

Descriptive statistics for manipulation checks as a function of experimental condition in Experiment 2.

	Negative	Positive	Control	Total
1. To what extent are there gender differences in visuospatial performance?	5.72 (2.54)	5.83 (3.52)	3.71 (2.54)	5.11 (3.01)
2. Who do you think performs better on this task?	4.72 (1.45) _{bc}	6.56 (2.75) _b	6.59 (1.50) _c	5.94 (2.15)
3. To what extent are there gender differences in mathematical performance?	6.88 (2.00) _d	5.89 (3.79)	4.37 (2.29) _d	5.64 (2.96)
4. Who do you think performs better on this task?	3.75 (1.18) _e	5.28 (2.93)	5.74 (1.66) _e	4.98 (2.21)

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

MECHANISMS OF STEREOTYPE THREAT

Table 6.

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

	Negative	Positive	Control	Total
Correct %	.78 (.28)	.84 (.13)	.83 (.15)	.82 (.20)
Correct SRT	244.76 (32.39)	257.78 (65.00)	244.06 (38.87)	248.87 (47.13)
Reflexive %	.22 (.28)	.16 (.13)	.17 (.15)	.18 (.20)
Reflexive SRT	185.92 (32.38)	187.57 (34.39)	172.29 (23.53)	181.82
Corrective %	.62 (.23)	.55 (.32)	.59 (.27)	.59 (.27)
Corrective SRT	352.50 (89.75)	353.71 (130.47)	354.23 (94.58)	353.48 (104.58)

MECHANISMS OF STEREOTYPE THREAT

Table 7.

Meta-analytic summary of Experiment 1 and 2 compared to Jamieson and Harkins (2007; Experiment 3).

Effect	Study	Mean diff	Study $B_{H(0, J\&H}$ Effect Size)	Meta $B_{H(0, J\&H}$ Effect Size)	Posterior Mean (SD)	Meta 95% Credible Interval
Correct Responses %	J&H	0.33*				
	Study 1	-0.01	0.13			
	Study 2	0.04	0.36	0.15	0.01 (0.04)	-0.07, 0.09
Correct Responses SRT	J&H	80.44*				
	Study 1	8.45	0.28			
	Study 2	-0.70	0.18	0.19	5.36 (8.96)	-12.20, 22.92
Reflexive Saccade %	J&H	0.33*				
	Study 1	-0.01	0.13			
	Study 2	0.04	0.36	0.15	0.01 (0.04)	-0.07, 0.09
Reflexive Saccade SRT	J&H	25.20				
	Study 1	6.65	0.65			
	Study 2	-13.63	0.17	0.19	-3.13 (6.85)	-16.56, 10.29
Corrective Responses %	J&H	0.33*				
	Study 1	-0.21	0.08			
	Study 2	0.03	0.33	0.08	-0.09 (0.06)	-0.21, 0.03
Corrective SRT	J&H	47.95*				
	Study 1	-12.10	0.37			
	Study 2	-1.73	0.60	0.31	-8.12 (21.40)	-50.07, 33.83

Note: The meta-analytical results show the posterior means, SDs and 95% Credible Intervals for replicated effects. The meta Bs quantify the degree that the meta-analytic data support the results obtained in the original study by Jamieson and Harkins (2007). Negative scores indicate different direction of effect from original study. * $p < .05$.