**The British
Psychological Society**

Women's stereotype threat-based performance motivation and prepotent inhibitory ability

Russell R. C. Hutter[1]* (iD), Lucy C. Davies[2] (iD),
Constantine Sedikides[3] (iD) and Mark Conner[1] (iD)
[1]School of Psychology, University of Leeds, UK
[2]University College London, UK
[3]University of Southampton, UK

According to the mere effort account of performance, stereotype threat motivates disproval of the negative performance stereotype, which in turn potentiates the overproduction of prepotent responses. In mathematics (maths), prepotent responding facilitates solve type question (e.g., equations) performance, but reduces comparison type question (e.g., estimations) performance. Problematically, the mere effort account indexes performance motivation as task performance. Also, this account posits that performance reduction on non-prepotent tasks derives from the overproduction of prepotent responses, as opposed to failed inhibition of prepotent responses associated with the alternative, namely, the working memory interference perspective. We investigated motivational and prepotent responding as applied to stereotype threat. In Experiment 1, a maths question selection task indexed motivation (independently of performance). Stereotype threat led female test takers to select more solve than comparison maths questions, in accord with the mere effort account. In Experiment 2, higher inhibitory ability protected overall maths performance following stereotype threat, but it did not protect non-prepotentiated comparison question performance (inconsistent with the working memory interference perspective). The results support the mere effort account.

Occupational and educational environments rely on performance-based evaluations. Therefore, knowledge of factors that facilitate or hinder performance is important. The stereotype threat effect has come to the fore as a cause of debilitated performance. Stereotype threat occurs when people become concerned that their performance will be evaluated in the light of a negative stereotype, coupled with fear of confirming it (Steele & Aronson, 1995). One form of stereotype threat refers to gender-mathematics stereotypes (Nosek *et al.*, 2005; Spencer, Steele, & Quinn, 1999), such as 'women are poorer at mathematics' or 'men are better at mathematics,' which can induce performance concerns (Shapiro & Neuberg, 2007) and lower performance in women (Schmader, Johns, & Forbes, 2008).

The prevalent explanation for stereotype threat effects is the working memory interference perspective (Beilock, Rydell, & McConnell, 2007; Bonnot & Croizet, 2007; Croizet *et al.*, 2004; Schmader, 2002; Schmader & Johns, 2003; Schmader *et al.*, 2008;

Correspondence should be addressed to Russell R. C. Hutter, School of Psychology, University of Leeds, Leeds LS2 9JT, UK (email: r.r.c.hutter@leeds.ac.uk).

Turner & Engle, 1989). According to Schmader and Johns (2003), for example, stereotype threat behaves as a stressor in performance situations where an active negative stereotype poses a threat to social identity (Derks, Inzlicht, & Kang, 2008). Both female identity and maths identity are likely threatened. Stereotype threat occupies working memory, competing for cognitive resources with the stereotypically associated task. When undertaking mathematics (maths)-related tasks, worries over the negative performance stereotype compete with the task for phonological loop resources (Beilock *et al.*, 2007). The result of phonological loop overload is performance reduction among persons experiencing the negative stereotype. Steele and Aronson (1995) contend that, although participants are motivated to do well, dividing attention between completing test items and managing frustration affects performance. Thus, it is not motivation that drives stereotype threat effects, but rather frustration at time spent evaluating the source of frustration. An alternative approach, the mere effort account (Jamieson & Harkins, 2007), places motivation at the heart of stereotype threat performance effects. This account is not alone in indexing motivation as a core stereotype threat mechanism. Other approaches include motivation to reduce errors (prevention focus; Seibt & Forster, 2004) or avoid stereotype confirmation (Brodish & Devine, 2009). The mere effort account, however, focusses exclusively on performance motivation to disprove an active stereotype via activation of prepotent responding.

Mere effort account and maths stereotype threat

The mere effort account seeks to explain the effect of evaluation on task performance. Potential similarities exist between the evaluation-performance and stereotype threat literatures, including effort withdrawal and processing interference (Jamieson & Harkins, 2007). Performance evaluation can engender performance concerns, motivating test-takers to do well and thus potentiating a prepotent (i.e., dominant) response. Prepotent responding facilitates answers closely associated, but hampers answers remotely associated, with a task (Harkins, 2001, 2006). When applied to stereotype threat, the mere effort account predicts that an activated negative stereotype linked with performance motivates individuals to perform well and disprove it, activating a prepotent task response (Harkins, 2006; Jamieson & Harkins, 2007, 2009, 2011; McFall, Jamieson & Harkins, 2009). The type of prepotent response activated depends on the task. For example, horizontal maths subtraction problems (e.g., 50 minus 8 =) activate the 'method of adjustment' (adjust the second number to the nearest 10, subtract the second number from the first, add the adjustment) prepotent response (Seitchik & Harkins, 2015). Our interest lies in response to differing maths problem types when experiencing stereotype threat, and particularly in the strength of the relation between experiencing stereotype threat and maths performance as a function of motivation and overproduction of prepotent responses, because motivation and prepotent overproduction are central to the mere effort account.

Two maths problem types are relevant (Jamieson, 2009), solve and comparison (Appendix S1). Solve problems require prepotent solve responding, activating learned material (e.g., formulae), whereas comparison problems do not rely on prepotent solve responding, but instead require logic, estimation, or intuitive responses – material not learned (Gallagher & De Lisi, 1994; Gallagher *et al.*, 2000; Quinn & Spencer, 2001). Stereotype threat undermines women's performance for comparison problems, but facilitates performance for solve problem (Jamieson & Harkins, 2009). Thus, performance differs as a function of the correct application of the prepotent response. Such mere effort

effects have been reported in educational settings (Davies, Conner, Sedikides, & Hutter, 2016).

Mere effort account mechanisms

Two mechanisms form the basis of the mere effort account (Harkins, 2006; Jamieson & Harkins, 2007). Stereotype threat motivates (i.e., test takers must care) that (their) performance could reflect badly on the self (Inzlicht & Ben-Zeev, 2000) or their group (Wout, Danso, Jackson, & Spencer, 2008). Motivation activates well-learned (i.e., prepotent) responses to the task, derived from memory. Such responses may not be correct, depending on task. For example, activated prepotent solve responses on a maths task are useful for solve, but not comparison, questions. We recognize the inextricable link between motivational and prepotent response mechanisms and thus explore consequences of both.

Motivation is a core component in Schmader *et al.*'s (2008) integrated processes model, which explains stereotype threat effects in terms of a physiological stress response, performance monitoring, and suppression of negative thoughts/emotions, overseen by working memory. The model, however, does not explain *how* motivation leads to performance increases on solve questions. In contrast, the mere effort account explains this as heightened performance motivation that strengths automatic response tendencies and the overproduction of the prepotent response. Therefore, in the presence of stereotype threat, performance becomes facilitated or impaired depending on whether the activated prepotent response is correct. In the context of the maths-gender stereotype threat, this explains performance enhancement for solve questions, because motivation activates prepotent solve responses. We consider motivational and inhibitory mechanisms next.

Motivation

Distinctly and independently of other forms of motivation associated with stereotype threat (Smith, Lewis, Hawthorne, & Hodges, 2013; Smith, Sansone, & White, 2007), performance motivation can only be indexed by task performance (Jamieson & Harkins, 2007). Thus, motivation to undermine the negative stereotype relates only to the task linked to the stereotype. For example, the stereotype that women perform poorly relative to men at maths can only be measured on maths or maths-related tasks (i.e., it is task-specific). This renders performance motivation difficult to test directly, and hence potentially unfalsifiable as an explanation, because motivation to disprove the stereotype and task performance is one and the same. Jamieson and Harkins (2007) reported evidence of motivated overproduction of prepotent responses on an antisaccade task for threatened test takers (who believed the task produced gender differences). Although the antisaccade task commendably indexes motivation as task performance (i.e., speed of accurate responses), the process is intertwined with the dependent measures.

How can motivation be measured independently of task, if performance motivation indexes the task itself? A promising measure is maths question selection. Motivation to disprove the active performance stereotype and perform well on all questions for a maths task could influence the choice of maths question. Although question selection constitutes a facet of task performance, it is not measured by task performance. It follows that, if stereotype threat motivates maths performance, threatened test takers, when given a choice, will manifest a preference to answer solve over comparison

questions provided the prepotent solve response is activated. Forbes and Schmader (2010) asked participants whether they would like to work on a maths or verbal remote-associates problem. The more maths questions participants selected, the more time they spent working on them (i.e., the more motivated they were). In our research, question choice, or selection therefore, offers a unique opportunity to test performance motivation independently of maths performance.

Overproduction of prepotent responses

Overproduction of prepotent responses results from stereotype threat motivating participants to perform well. Jamieson and Harkins (2007) informed participants that a 'visospatial capacity' (antisaccade task) task produced gender differences. To perform well, antisaccade tasks require inhibition of the prepotent tendency (i.e., looking at a presented cue), and instead looking at the opposite side of a visual display. Threatened female test takers looked in the wrong direction (prepotent responding) more often than controls, but performed better overall than controls, when given time to implement a correction and launch corrective saccades. Performance motivation seemingly led stereotype threat participants to respond as quickly as possible, even when required first to negate the incorrect prepotent response. Jamieson and Harkins interpret this as evidence for the overproduction of prepotent responses resulting from motivation to perform well following stereotype threat. Schmader *et al.* (2008) contrastingly argued that failure to inhibit prepotent responses offers a more likely explanation, a failing in working memory brought about by stereotype threat. They also pointed out that Jamieson and Harkins' method does not allow for a direct comparison between these two explanations. More recently, Pennington, Litchfield, McLatchie, and Heim (2018) compared the mere effort account and working memory interference perspective, activating female visuospatial and maths stereotype threat, followed by an antisaccade task (Experiment 1), and an antisaccade task and vertical and horizontal arithmetic problems (Experiment 2). Pennington *et al.* reasoned that results supporting the mere effort account would be characterized by negative stereotype threat activating motivation to undermine the stereotype, resulting in more incorrect saccades (i.e., looking towards the presented cue) and accompanied by faster corrective saccades, compared to no-stereotype threat. The authors also argue that a working memory interference perspective might predict more incorrect saccades launched, but crucially would differ from the mere effort account in fewer and slower corrective saccades launched. This argument is based on the premise that stereotype threat loads working memory, interfering with task performance. Stereotype threat did not affect antisaccade performance in either experiment. Similarly, stereotype threat did not undermine maths performance in Experiment 2. In other words, the results support neither the mere effort account nor the working memory perspectives, while challenging the concept of stereotype threat.

Inhibition is an executive function working to keep irrelevant information from entering colour attentional focus and suppress task-inappropriate automatic responses (Friedman *et al.*, 2008; Hasher, Quig, & May, 1997). Inhibition-related processes form a family of functions clustered into distinct categories (Friedman & Miyake, 2004; Hasher, Lustig, & Zacks, 2007). The form of inhibition that Schmader *et al.* (2008) discussed is prepotent response inhibition, the ability to suppress dominant or automatic responses (Friedman & Miyake, 2004). Indeed, Schmader *et al.* (p. 348) make a clear link between ability to inhibit prepotent responses and working memory impairment.

Typically, prepotent response inhibition measurement and testing involve the Stroop task (Friedman & Miyake, 2004): vocalizing the ink colour (in which colour and neutral words names are printed), while attempting to ignore the dominant tendency to read the words (which takes longer and is susceptible to errors when the ink colour and colour word names mismatch; MacLeod, 1991; Stroop, 1935). According to the working memory interference perspective, inhibitory ability associated with prepotent responses (i.e., Stroop performance) should moderate the relation between stereotype threat and maths performance. Answering comparison type questions correctly, after maths stereotype threat, involves inhibiting the prepotent response (solve approach), according to Schmader *et al.* Similarly, Carr and Steele (2009) proposed that experiencing stereotype threat requires the ability to inhibit old strategies in order to develop more successful ones for problem solving. Therefore, individuals with higher inhibitory ability may self-protect (Sedikides, 2012; Sedikides, Green, Saunders, Skowronski, & Zengel, 2016), blocking out the detrimental effects of stereotype threat.

The working memory interference perspective suggests that the ability to deactivate and inhibit cognitions (Friedman & Miyake, 2004; Lustig, Hasher, & Tonev, 2001) protects against stereotype threat effects. Therefore, inhibitory ability may affect stereotype threat on performance by allowing threatened individuals to inhibit incorrect prepotent responses in order to apply other correct approaches. In the current research, we examine whether variation in prepotent response inhibitory ability affects stereotype threat. That is, we expect inhibitory ability to act as a moderator (than a mediator), because this ability might influence the strength of the relation between stereotype threat and maths performance. According to the working memory interference perspective, those with higher prepotent response inhibitory ability will overcome prepotent responding when it is inappropriate to do so. Therefore, individual differences in the ability to inhibit prepotent responding when inappropriate would advantage those with superior inhibitory ability. Lack of such a finding, though, would imply the overproduction of prepotent responses, supporting the mere effort account.

Overview

We compared the overproduction of prepotent responses (mere effort account) with the inhibition of prepotent responses (working memory interference perspective) in an effort to explain stereotype threat effects. Our chosen population and performance stereotype mirrored those of Jamieson and Harkins (2009) and Seitchik and Harkins (2015): women and the stereotype 'women are poorer at maths.' We tested only women, because our chosen stereotype threat manipulation has been shown previously only to affect women (Davies *et al.*, 2016; Jamieson & Harkins, 2009; Seitchik & Harkins, 2015). In Experiment 1, we examined whether activation of a negative performance self-applicable stereotype motivates participants to undermine the stereotype. Although stereotype threat motivates performance on all question types, the activated prepotent response might lead to the selection of more prepotent solve questions to answer than non-prepotent comparison questions. In Experiment 2, we turned to the overproduction of prepotent responding versus the inhibition of prepotent responding, following stereotype threat activation of the motivated preference goal. We tested whether participants with superior inhibitory ability are better placed to inhibit prepotent solve responding when tacking comparison questions. Positive results would lend credence to the working memory interference perspective, and negative results to the mere effort account.

EXPERIMENT 1

In Experiment 1, we were concerned with the first component of the mere effort account in a stereotype threat context, motivation to disprove the active stereotype. We advanced a question selection method as an index of performance motivation. The task-specific nature of motivation in this account implies that such testing can be problematic. We therefore set to develop questions differing in type (i.e., solve vs. comparison), but equal in difficulty, based on Forbes and Schmader's (2010) maths motivation task. Participants' question choice following stereotype threat served as an indicator of performance motivation, because motivation to perform activates prepotent (solve) responding. Thus, we indexed motivated preference with greater likelihood to select solve over comparison questions.

We intended for the maths selection task to comprise both solve and comparison questions typical of a General Certificate of Education (GCSE) exam (as per Jamieson, 2009, footnote 1). General Certificate of Education is a compulsory academic qualification for core knowledge subjects taken by students (14–16 years old) in the United Kingdom. We aimed for the number of solve and comparison questions to be of equal difficulty. We conducted two pilot studies. In Pilot Study 1, we administered a combined maths test comprising 18 questions (nine solve, nine comparison), all set at the GCSE (higher tier) level and taken from non-calculator exams (www.aqa.org.uk). Questions were worth between three and six marks reflecting difficulty and scoring. Thirty women, aged from 18 to 24 years ($M = 22.00$, $SD = 1.20$) completed the test. To control for mathematical ability, we ensured that all participants had a GCSE grade of C or above. All participants identified as British Caucasian and native English speakers. To ensure that question order did not confound question selection choice, we ordered questions alternately as solve and comparison, and counterbalanced (there were five question-order variations). In creating selection task questions of equal difficulty across two question types (solve vs. comparison), we selected five questions (from the original 18), worth three marks, for each question type that elicited similar overall scores. We relied on Jamieson's (2009) method for that. The solve ($M = 50.6$, $SD = 15.08$) versus comparison ($M = 46.6$, $SD = 15.69$) questions did not differ in difficulty, $t(8) = 0.411$, $p = .692$. Our finalized version of the selection task consisted of 10 (five solve, five comparison) equally difficult questions, worth three marks each.

In Pilot Study 2, we set out to test meta-perceptions of question difficulty and self-efficacy related to the 10 questions we derived in Pilot Study 1. We administered these questions to 35 women, aged 18–32 years ($M = 19.03$, $SD = 3.15$). Again, to control for mathematical ability, all participants had a GCSE grade of C or above and identified as British native English speakers. To calculate power for our main dependent variable (question difficulty), based on $d = 0.42$, probability level $= .05$, and a sample size of 35, we used G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). The observed statistical power was .78. We instructed participants to rate 10 maths questions. The questions comprised five solve type and five comparison type. Participants remained unaware of differential question types and were not required to answer the questions. We counterbalanced question order across five question-order variations, and for each maths question, participants completed two meta-perception measure of difficulty: 'How difficult did this question seem?' (1 = *easy*, 11 = *difficult*) and 'For me, answering this question would be' (1 = *easy*, 11 = *difficult*). We adapted the second measure from Norman and Conner (2006). We also administered two self-efficacy measures (adapted from Norman & Conner, 2006): 'How certain are you that you could answer this

question?' (1 = *not at all*, 11 = *extremely so*) and 'How confident are you that you could answer this question?' (1 = *not at all*, 11 = *extremely so*).

To create a *question difficulty* index for solve questions, we added the 'How difficult did this question seem?' and the 'For me, answering this question would be' items and, after dividing by two, calculated correlation: $r = .72$, $p < .01$. We also created a *self-efficacy* index, using the same method 'How certain are you that you could answer this question?' and 'How confident are you that you could answer this question?': $r = .81$, $p < .01$. We repeated the procedure for comparison questions resulting in $r = .81$, $p < .01$ for the difficulty index and $r = .70$, $p < .01$ for the self-efficacy index. Participants perceived solve questions ($M = 7.12$, $SD = 2.62$) as less difficult than comparison questions ($M = 8.18$, $SD = 1.39$), $t(34) = -2.45$, $p = .020$. However, participants were not more certain of their self-efficacy in answering solve questions ($M = 10.65$, $SD = 3.03$) over comparison questions ($M = 10.22$, $SD = 2.24$), $t(34) = 1.31$, $p = .199$.

Taken together, difficulty meta-perceptions and self-efficacy beliefs indicated that participants considered solve questions easier, but did not believe that they possessed the ability to perform better on these questions relative to comparison questions. This occurred despite the finding in Pilot 1 that solve and comparison questions were equally difficult to answer. This reflects a prepotent tendency towards solve questions selection for women (Gallagher *et al.*, 2000; Jamieson & Harkins, 2009).

We hypothesized that, if stereotype threat motivates disproval of the negative stereotype, leading test takers to apply a prepotent solve approach (in accord with the mere effort account), participants would manifest a greater selection of prepotent solve questions than non-prepotent comparison questions to answer. Additionally, we aimed to rule out that increased motivation following stereotype threat is not task-specific, by using as a covariate the inclusion of an (indirect) measure of maths motivation. We did not expect indirect maths motivation to be affected following stereotype threat, if motivation were task-specific, as per the mere effort account. We also examined the roles of female and maths identifications, given that stereotype threat acts as a stressor in performance-based contexts, as per the working memory interference perspective.

Method

Participants and design

We tested an opportunity sample of 103 psychology female undergraduates at a UK university, aged between 18 and 22 years ($M = 19.36$, $SD = 1.07$). The sample size was adequate for achieve power in the region of .80, based on an effect size ($d = 0.64$) from related research (Davies *et al.*, 2016; Jamieson & Harkins, 2009, Experiment 1). All participants (1) identified as British Caucasian, with English as their first language, and (2) had achieved at least a GCSE C grade Maths. We assigned them to a one factor (condition: stereotype threat, no-stereotype threat) between-subjects design, with indirect maths motivation, gender identification, and maths identification as a covariates.

Procedure

A female experimenter escorted participants individually into the laboratory. Participants received the pen-and-paper non-calculator and non-assessed maths tests. After random assignment to experimental conditions, they carefully read the front cover instructions as requested, which included the stereotype threat manipulation, ahead of indicating to the

experimenter their readiness to proceed. We manipulated condition by instructing participants that 'previous research has shown gender differences on this test' (stereotype threat) or that 'previous research has shown no gender differences on this test' (no-stereotype threat). This manipulation (adapted from Steele & Aronson, 1995) successfully induces or removes the maths-gender stereotype threat (Nguyen & Ryan, 2008). Participants completed as many of five questions as possible. The questions were not assessed or performance recorded, serving as a method to ensure the face validity of the question selection task to follow. We derived questions from www.aqa.org.uk; however, the questions were not the same as those outlined in Pilot Study 1 and the maths selection task (below), and were not broken down as solve and comparison ones. On completion of the maths task, participants undertook a question selection task and received instructions to choose a total of any five questions to answer (ostensibly) out of 10 (see pilot studies). The transition between the maths task and the selection task was seamless, and the selection task constituted the second section in the overall procedure. Further, we instructed participants how to complete the selection task (see below). Thus, we designed for the stereotype threat manipulation to impact on the selection task after participants had undertaken the face valid maths test. The selection procedure mimicked that of Pilot Study 2. Following question type selection, participants completed a stereotype threat manipulation check, as well as measures of maths motivation, gender identification, and maths identification.

Dependent measures

The main dependent measure was number of solve questions minus number of comparison questions selected (five from 10) on the maths question selection task. There were five counterbalanced variations of question order, and questions were ordered alternately solve and comparison, to ensure that order did not confound question selection choice. We included an indirect maths motivation measure: 'How motivated were you to perform well on the maths test? ($1 = not at all$, $11 = extremely so$)'. We also administered Schmader's (2002) adapted Self-Esteem Scale to measure gender identification ($1 = strongly disagree$, $5 = strongly agree$): 'Being a woman is an important part of my self-image', 'Being a woman is unimportant to my sense of what kind of person I am [R]', 'Being a woman is an important reflection of who I am', and 'Being a woman has very little to do with how I feel about myself [R]'. A further measure, maths identification, comprised two items ($1 = not at all true$, $7 = completely true$): 'It is important to me that I am good at maths' and 'I am good at maths'. We next administered two stereotype threat manipulation checks (Jamieson & Harkins, 2007, 2009): 'To what extent are there gender differences in performance on this task?' ($1 = no gender differences$, $11 = gender differences$) and 'Who do you believe performs better in this task?' ($1 = males perform better$, $6 = males and females perform the same$, $11 = females perform better$).

Results

Manipulation check

Participants in the stereotype threat condition ($M = 6.11$, $SD = 2.56$) reported that gender differences existed on the test to a greater extent than those in the no-stereotype threat condition ($M = 2.89$, $SD = 2.64$), 95% CI [2.19, 4.24], $t(101) = 6.24$, $p < .001$, $d = 1.24$. Also, participants in the stereotype threat condition ($M = 4.75$, $SD = 2.18$)

reported that males performed better than females than those in the no-stereotype threat condition at a marginal level ($M = 5.26$, $SD = 1.29$), 95% CI [−1.23, 0.22], t (101) = −1.39, $p = .084$, $d = -0.28$. Participants in the stereotype threat condition were therefore aware of, but did not necessarily believe, the negative stereotype.

Question selection[1]

We subjected maths question selection to a single-factor (condition) analysis of covariance (ANCOVA), with indirect maths motivation, gender identification, and maths identification as covariates. Before analysis, we calculated for each participant the number of solve questions minus the number of comparison questions selected, whereby a higher score denoted greater solve responding. Crucially, as hypothesized, stereotype threat participants chose to answer solve questions ($M = 1.40$, $SE = .28$), 95% CI [0.85, 1.96] more often than no-stereotype threat participants ($M = .46$, $SE = .31$), 95% CI [−0.16, 1.08], $F(1, 98) = 4.98$, $p = .028$, $\eta p^2 = .05$.[2] As anticipated, the maths motivation covariate measure did not significantly explain the variance in question type selected, $F(1, 98) = 2.20$, $p = .141$. Similarly, gender identification was not significant, $F(1, 98) = 2.35$, $p = .128$, whereas the effect of maths identification on question type selected was marginal, $F(1, 98) = 3.55$, $p = .063$, $\eta p^2 = .04$ (see Appendix S1 for further moderators analyses: motivation, gender identity, maths identity). We present correlations among variables in Table 1.

Discussion

We found evidence for task-specific motivation to disprove a negative women's maths performance stereotype following activation of stereotype threat. According to the mere

Table 1. Pearson product-moment correlational coefficients among variables in experiment 1 ($N = 103$)

	1	2	3	4	5	6	7
1. Condition	–						
2. Solve minus comparison questions	−.19	–					
3. Maths motivation	.15	−.05	–				
4. Gender identification	.03	−.19	.03	–			
5. Maths identification	.04	.15	.40**	−.16	–		
6. Manipulation check 1 (gender differences)	−.53**	.08	.10	.03	.08	–	
7. Manipulation check 2 (performance beliefs)	.14	.10	.14	.06	.25*	−.16	–

Note. *$p < 0.05$; **$p < 0.01$.

[1] *We conducted a 2 (question type) × 5 (question order on selection task) mixed analysis of variance, with the first factor being between-subjects and the second within-subjects, to examine the potential effects of question order on question selection. The interaction was not significant, $F(1, 98) = 0.19$, $p = .941$, $\eta p^2 = .01$. The counterbalancing question order across the question selection task was successful.*

[2] *When analysed without the indirect maths motivation, gender identification, and maths identification covariates, stereotype threat participants again answered more solve questions ($M = 1.35$, $SE = .29$), 95% CI [0.79, 1.92] than no-stereotype threat participants ($M = .52$, $SE = .32$), 95% CI [−0.12, 1.15], $F(1, 101) = 3.78$, $p = .050$, $\eta p^2 = .04$. Additionally, we did not find a direct effect of condition on indirect motivation, $t(101) = 1.53$, $p = .128$.*

effort account (Jamieson & Harkins, 2007; McFall, Jamieson, & Harkins, 2009), performance motivation activates first, ahead of prepotent responding. Assuming that performance motivation is intrinsically tied to task performance on a given task (Jamieson & Harkins, 2011), it cannot easily be measured independently. Therefore, stereotype threat activated and task-specific motivation is unlike other forms of motivation. Conceptualizing performance motivation in this way creates a problem in testing and falsifying the motivation mechanism. Put otherwise, task performance evinces both stereotype threat and the underlying mechanism. Our adapted Forbes and Schmader's (2010) maths motivation task selection task allowed the investigation of mere effort account's hypothesized motivational mechanisms independently of stereotype threat effects, overcoming the task-specificity problem. We observed evidence of greater solve type maths question selection following stereotype threat. Also, stereotype threat did not affect an indirect measure of maths motivation for solve and comparison questions selected, indicating that motivation to disprove the negative stereotype cannot be indexed via other forms of motivational measure (Jamieson & Harkins, 2009). Additionally, we did not find an effect for gender identification, whereas maths identification reached marginality. This finding is inconsistent with Schmader and Johns' (2003) proposal that stereotype threat as a stressor poses a threat to social identity. Together, the findings support (1) Jamieson and Harkins' (2007) proposition that people become motivated to undermine a negative performance stereotype when threatened by that stereotype, and (2) stereotype threat induces task-specific (as opposed to indirect) motivation.

EXPERIMENT 2

Experiment 2 addressed the ability to overcome prepotent responding in response to the maths-gender stereotype. The mere effort account links the motivational and prepotent response mechanisms. Prepotent responding is triggered by the motivation to disprove a negative performance stereotype and perform well on the task (Jamieson & Harkins, 2007). The overproduction of prepotent response boosts solve question performance, but worsens comparison question performance. This explanation relies on the premise that stereotypically threatened individuals apply the solve response to undermine the stereotype, even where a solve response is not appropriate. The working memory interference perspective, in contrast, states that it is not overproduction of solve responses that reduces comparison question performance, but rather failure to inhibit the prepotent solve response. Therefore, those with superior inhibitory ability would suppress the prepotent solve response when threatened. Inhibiting the prepotent solve approach when it is not required (i.e., for comparison questions) will enable test takers to use the correct comparison approach, improving maths performance. As such, we anticipated inhibitory ability to act as a moderator. As in Experiment 1, we did not expect indirect maths motivation to be affected following stereotype threat and addressed the relevance of female and maths identification.

The two approaches diverge further in their predictions. According to the working memory interference perspective, inhibitory ability moderates stereotype threat effects for question type. If stereotype threat-based performance reduction results from failure to inhibit prepotent responding when working on non-prepotent tasks, then superior ability in suppressing prepotent (solve) responses will protect threatened females' maths

performance against unwanted solve responding when answering comparison questions. Superior inhibitors will be more able to inhibit the solve response and apply the correct comparison response, overcoming stereotype threat effects and performing to their full mathematical ability. Performance on solve questions will be unaffected by inhibitory ability, as the prepotent solve response is inappropriate for these questions. Furthermore, given that it is the experience of stereotype threat that potentiates prepotent responses, the above discussed pattern will not emerge in the control condition. In contrast, according to the mere effort account, performance reduction on non-prepotent tasks derives from the overproduction of prepotent responses. This implies that superior inhibitory ability will not protect comparison type question performance, particularly when experiencing stereotype threat. Finally, we predicted that stereotype threat would lead to performance boost for solve questions (where prepotent responses are correct) and performance decrement for comparison questions (where prepotent responses are incorrect), relative to no-stereotype threat. The working memory interference perspective and mere effort account do not diverge on this last prediction: The two theoretical statements only diverge on the mechanisms likely to be involved (prepotent inhibition vs. overproduction of prepotent responses, respectively).

Method

Participants and design

We recruited an opportunity sample of 165 female undergraduates in a UK university aged from 18 to 25 years ($M = 20.50$, $SD = 1.61$). We conducted a post-hoc power analysis via an online calculator (https://www.danielsoper.com/statcalc/calculator.aspx?id=9) to calculate power for our predictors and moderators, based on $R^2 = .06$, probability level $= .05$, and sample size 165 (Soper, 2018). The analysis confirmed that power was adequate (observed statistical power = .77). A further calculation for our interaction variables, based on $R^2 = .19$, probability level $= .05$, and sample size 165, confirmed that this sample size was sufficient (observed statistical power = .98). All participants identified as British Caucasian, with English as their first language. Also, all participants had achieved at least a GCSE C grade Maths. We assigned them to a 2 (condition: stereotype threat, no-stereotype threat) × 2 (question type: solve, comparison) between-subjects design.[3] We also assessed inhibitory ability as a continuous variable (moderator). The predictor variables were therefore condition and question type, while inhibitory ability was the moderator variable.

Procedure

A female experimenter escorted participants individually into the laboratory and informed them that they would be involved in a series of brief tasks. Participants completed a 5-min Stroop task on-screen involving 10 practice and 48 experimental trials. On each trial, participants were presented with a fixation asterisk for 1,000 ms, followed by a target colour word (i.e., blue, red, green, yellow) or hash key (i.e., ####), in either

---

[3] *We manipulated question type as a between-subjects factor to minimize the effects of exposure to both question types. Jamieson (2009, Experiment 2) offered a similar rationale. Mere effort effects have therefore previously been observed using both between-subjects and within-subjects designs.*

congruent (baseline) or incongruent (interference) coloured font. The hash key represented the neutral condition. The 'Z' response key denoted the correct response for words printed in red or green font, whereas the 'M' key denoted the correct response for words printed in blue or yellow font. Stimuli remained on-screen until participants responded, and the next trial began after an intertrial interval of 1,000 ms. We analysed correct responses. Following a 5-min break, participants were given the maths test and asked to read carefully the front cover. We manipulated stereotype threat as in Experiment 1. All participants subsequently completed as many as possible of five 3-mark comparison questions or five 3-mark solve questions from Davies *et al.* (2016, Experiment 2). All questions appeared as non-calculator mathematics pen-and-paper test(s) closely resembling a GCSE test format. Following random assignment to experimental conditions, participants read carefully the test instructions before signalling to the experimenter that they were ready to begin. We implemented a time constraint of 18 min, resulting in approximately 1 min per mark (analogous to the time allocated in GCSE examinations). We provided a ruler, pencil, and pen to simulate an exam environment. Upon completion, participants completed several measures. Finally, we administered the same two stereotype threat manipulation checks as in Experiment 1.

Dependent measures

The dependent (outcome) measure was correct maths score (out of 15) for solve and comparison questions in the main analyses. We also administered a measure of indirect motivation, Schmader's (2002) adapted measure of gender identification, and a maths identification measure – all the same as in Experiment 1. We treated each as an outcome measure in further analyses.

Results

We used RT difference between congruent and incongruent accurate responses on the Stroop test as a measure of inhibitory ability. We removed 7.9% of trials on which participants made incorrect responses (errors). We then calculated median reaction times (because they are less susceptible to outliers) to incongruent and congruent trials for each participant (Whelan, 2008). Next, we computed a difference score by subtracting the median reaction time on congruent trials from the median reaction time on incongruent trials. Lower scores indicated greater inhibitory ability.

Manipulation check

In regard to the manipulation check item ('To what extent are there gender differences in performance on this task?'), participants in the stereotype threat condition ($M = 5.70, SD = 2.59$) reported that gender differences existed to a greater extent compared to those in the no-stereotype threat condition ($M = 3.04$, $SD = 2.59$), 95% CI [1.80, 3.53], $t(163) = 6.07$, $p < .001$, $d = 0.95$. However, participants in the two conditions did not differ significantly in their belief that males performed better than females at the task, $t(163) = 0.04$, $p = .485$. In all, participants in the stereotype threat condition were aware of the negative group stereotype, but did not necessarily believe it.

Inhibitory ability[4]

We used moderated regression (Baron & Kenny, 1986) to examine inhibitory ability as a potential moderator of maths score. We computed an interaction variable by contrast coding condition as −1 and +1 (stereotype threat vs. no-stereotype threat) and question type as −1 and +1 (solve vs. comparison). We then computed the centred continuous inhibitory ability scores for each participant. Next, we multiplied condition by question type, condition by inhibitory ability, question type by inhibitory ability, and condition by question type by inhibitory ability to create the interaction terms. Finally, we entered the interaction terms into a multiple regression on a second step following the insertion of condition, question type, and inhibitory ability independently at step 1, with the centred continuous maths scores for each participant as the dependent (outcome) variable.

At step 1 (R^2 = .06), the condition main effect was not significant, β = .01, p = .443, t = .14, p = .443. However, the question type main effect was significant, β = −.14, p = .038, t =−1.78.14, p = .038: Participants performed better on the solve than comparison questions. Additionally, the inhibitory ability main effect was significant, β = −.19, p = .007, t = −2.52, p = .007: High (compared to low) inhibitors performed better on the maths test. At step 2 (R^2 = .19), entering the interaction terms explained a significant increase in variance in maths performance over step 1, ΔR^2 = .06, F(4, 157) = 2.63, p = .036 (Table 2). Importantly, the Condition × Question Type interaction was significant, β = .18, p = .016, t = 2.44, p = .016.

We broke down the Condition × Question Type interaction (Figure 1) by conducting simple regressions separately for each level of condition. In no-stereotype threat, we observed no effect of question type on maths performance, R^2 = .003, β = .05, t = .43,

Table 2. Summary of moderated regression analysis for variables predicting maths score in experiment 2 (N = 165)

Variable	Model 1				Model 2			
	B	SE B	β	t	B	SE B	β	t
Condition	.01	.08	.01	0.14	.04	.08	.04	0.53
Question type	−.14	.08	−.14	−1.78*	−.10	.08	−.10	−1.25
Stroop	−.20	.08	−.20	−2.52**	−.03	.11	−.03	−.28
Condition × Question type					.19	.08	.19	2.44*
Condition × Stroop					.26	.11	.25	2.36*
Question type × Stroop					.17	.11	.17	1.53
Condition × Question type × Stroop					.11	.11	.11	0.99
R^2		.06				.19		
F for change in R^2		3.36				2.63		
R^2 change		.06				.06		

Note. *p < .05; **p < .01.

[4] We additionally conducted a 2 (condition) × 2 (question type) between-subjects analysis of covariance, with inhibitory ability as a covariate, to explore the potential confounding role of inhibitory ability on maths performance. We found no significant main effects for condition, F(1, 160) = 0.02, p = .442, or question type, F(1, 160) = 2.48, p = .117. The covariate, inhibitory ability, was significantly related to maths performance, F(1, 160) = 6.10, p = .015, ηp^2 = .04, suggesting that inhibitory ability has a separate effect on maths performance to the interaction. Importantly, when controlling for the covariate effect of inhibitory ability, the interaction remained significant, F(1, 160) = 4.60, p = .033, ηp^2 = .03.
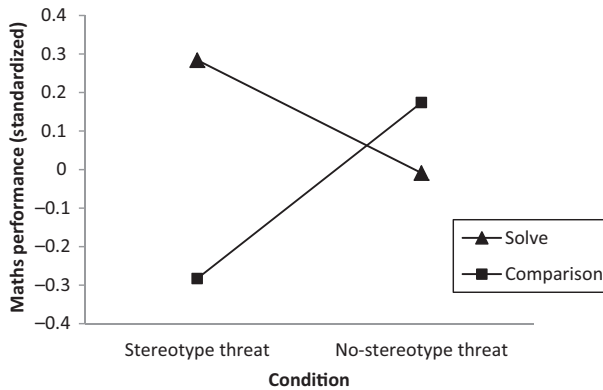
Figure 1. Standardized mean maths performance as a function of condition and question type in Experiment 2.

$p = .669$. However, in stereotype threat, we found a significant effect of question type on maths performance, $R^2 = .08$, $\beta = -.29$, $t = -2.71$, $p = .004$. Participants subject to stereotype threat performed better on the solve than comparison questions. We conducted simple regressions separately for each question type. For comparison questions, participants in stereotype threat underperformed relative to those in no-stereotype threat, $R^2 = .06$, $\beta = -.21$, $t = 2.23$, $p = .014$. For solve questions, no significant differences in maths performance emerged across stereotype threat or no-stereotype threat, $R^2 = .01$, $\beta = -.13$, $t = -1.07$, $p = .144$. These patterns are consistent with the mere effort account.

The Condition $\times$ Question Type $\times$ Inhibitory Ability interaction in the main analysis was not significant, $\beta = .11$, $t = .99$, $p = .324$. The Condition $\times$ Question Type interaction was unmoderated by inhibitory ability, suggesting no advantage for superior inhibitory ability in attempting to inhibit prepotent solve responses in response to comparison questions, following stereotype threat. This finding runs contrary to the working memory interference perspective.

In addition, the Condition $\times$ Inhibitory Ability interaction was significant, $\beta = .26$, $t = 2.36$, $p = .020$ (Figure 2). We used simple regressions to break the interaction down, indicating that in the stereotype threat condition, there was a significant effect of inhibitory ability on maths performance, $R^2 = .30$, $\beta = -.27$, $t = -3.00$, $p = .002$, whereas, in the non-stereotype threat condition, there was no effect of inhibitory ability on maths performance, $R^2 = .005$, $\beta = .11$, $t = .63$, $p = .266$. Stereotype threat resulted in low (vs. high) inhibitors underperforming. Finally, returning to the main analysis, the Question type $\times$ Inhibitory Ability interaction was not significant, $\beta = .17$, $t = 1.53$, $p = .129$ (see Appendix S1 for further moderators analyses: motivation, gender identity, maths identity).

Discussion

Having obtained evidence for increased task motivation to disprove negative performance stereotype activation in Experiment 1, we examined in Experiment 2 competing explanations for maths stereotype threat performance effects: the overproduction of solve responses (mere effort account) and failure to inhibit the prepotent solve response
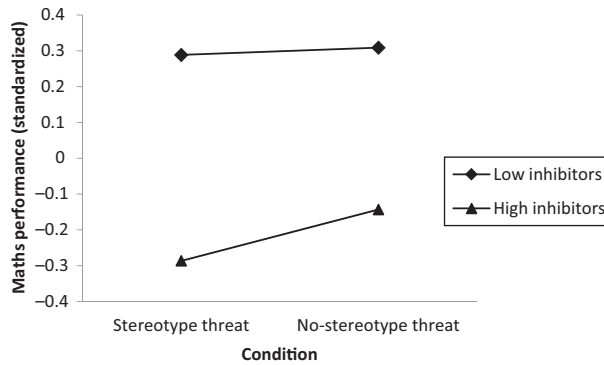
Figure 2. Standardized mean maths performance as a function of condition and inhibitory ability in Experiment 2.

(working memory interference perspective). The results were consistent with the mere effort account. Participants with low (vs. high) inhibitory ability showed greater *overall* test performance decrements following stereotype threat. Therefore, inhibitory ability acted as a buffer to protect high inhibitors against the detrimental consequences of stereotype threat. Superior inhibitory ability did not, however, protect threatened females' maths performance specifically on comparison questions, as would be expected if prepotent responses were inhibited following stereotype threat (working memory interference perspective).

The interaction between condition and question type revealed facilitation for solve over comparison questions in the stereotype threat condition, and, as expected, stereotype threat interfered with performance relative to no-stereotype threat for comparison questions. Yet, facilitation for solve questions was not evident in the stereotype threat (vs. no-stereotype threat) condition. The observed performance effects are broadly consistent with previous findings testing the mere effort account. Stereotype threat did not influence indirect maths motivation, replicating Experiment 1, in support of the idea that motivation to disprove negative stereotype can only be indexed by task-specific motivation (Jamieson & Harkins, 2009). We obtained the same result pattern for gender and maths identification as in Experiment 1: Stereotype threat did not act as a stressor, posing a threat to social identity, a finding inconsistent with the working memory interference perspective (see Appendix S1: motivation, gender identity, maths identity, Experiment 2).

GENERAL DISCUSSION

We investigated motivation to undermine a negative maths performance stereotype in regard to two explanations: overproduction of prepotent solve responses (mere effort account) and inhibition of prepotent solve responses (working memory interference perspective). In Experiment 1, stereotype threat motivated female undergraduates to select more solve than comparison questions to answer. In Experiment 2, female undergraduates undertook a prepotent response inhibition task (Stroop, 1935), followed by a maths test comprising solve and comparison type questions. High prepotent inhibitory ability was unrelated to improved comparison question performance; stated

otherwise, comparison question performance did not require inhibition of the prepotent solve responding.

Contribution

The results suggest that task-based motivation increases following stereotype threat, in support of the mere effort account (Jamieson & Harkins, 2007, 2009, 2012). Stereotype threat motivates test takers to perform well (on solve and comparison questions), undermining the negative performance stereotype through the facilitation of solve responding. Furthermore, the results are inconsistent with the proposition that superior inhibitory ability advantages non-prepotent question performance, a putative consequence of the working memory interference perspective (Schmader *et al.*, 2008). Comparison question performance would be expected to be better for those with superior prepotent inhibitory ability, because, from a working memory inference perspective, comparison question performance relies on inhibition of the prepotent solve response when a threatening performance stereotype becomes active. We did not observe comparison question performance facilitation for high versus low inhibitors. Whereas stereotype threat facilitates solve responses, the ability to inhibit prepotent responses does not improve comparison question performance. This pattern of results fits best with overproduction of prepotent responses (Jamieson & Harkins, 2007), but not with prepotent inhibition (Schmader *et al.*, 2008). To clarify, the argument from the memory interference perspective (Schmader *et al.*, 2008) has been that failure to inhibit prepotent responses, when inappropriate for the task, led to stereotype threat effects in Jamieson and Harkins' (2007) antisaccade task. We reasoned that, if this were so, high inhibitors would be able to suspend prepotent solve responding when answering non-prepotent comparison questions in maths tests. However, we did not observe facilitation on comparison questions for high inhibitors.

Motivation to disprove the negative stereotype

Performance motivation has previously only been measured using task performance (Jamieson & Harkins, 2007, 2011). This is a key requirement of the mere effort account to stereotype threat: Indirect forms of motivation are meaningless, because threat motivates only the drive to undermine the negative stereotype on tasks directly associated with the stereotype itself. We argued for the importance of a task that indexes performance motivation separately from maths performance, but that remained closely related to maths performance, thus allowing testing of the proposed stereotype threat mechanism and maths task independently. Conceptualizing performance and motivation as the same construct creates a problem in testing and falsifying the motivation mechanism.

The adapted Forbes and Schmader's (2010) maths motivation task enabled us to measure performance motivation using maths question type selection (solve vs. comparison). Experiment 1 revealed that stereotype threat led female test takers to select more solve than comparison questions, whereas no differences in question type choice emerged for control participants. Solve question selection indicates that performance motivation led threatened participants to choose questions that allowed the potentiated prepotent solve response to be applied. These findings suggest that experiencing stereotype threat inherently motivates (Jamieson & Harkins, 2011), enhancing test takers' performance resolve to disprove the stereotype and perform well on the task, thus driving selection of questions corresponding with the prepotent solve

response (i.e., solve questions). The indirect measure of maths motivation remained unaffected by stereotype threat, supporting Jamieson and Harkins' (2007) contention that performance motivation is task-specific. Although motivation to perform well following stereotype threat is a core assumption of the integrated process model (Schmader *et al.*, 2008), it is not the main or only driver. The model acknowledges, in line with the mere effort account, that stereotype threat increases *motivation* to perform well and combat the negative stereotype. However, the model predominantly focusses on how threatened test takers are motivated to resolve the cognitive imbalance created by stereotype threat. As threatened test takers struggle against stereotype threat, this burdens working memory resources.

Prepotent solve responses

Experiment 2 addressed the effect on performance of greater ability to inhibit prepotent responding, following stereotype threat. According to the mere effort account, stereotype threat operates via the activation of prepotent responses stemming from motivation to disprove the negative stereotype and perform well on the task. In the context of the maths-gender stereotype, stereotype threat potentiates overproduction of prepotent solve response. However, the working memory interference perspective posits that stereotype threat leads to failure to inhibit prepotent responses on tasks where such an approach is inappropriate. Therefore, as a consequence of potentiated solve responding, threatened female test takers with greater ability to suppress prepotent responses would be well-placed to suppress solve responses.

We used a Stroop task to measure prepotent response inhibition ability (Dao-Castellana *et al.*, 1998; Friedman & Miyake, 2004). Inhibition tasks, such as the Stroop, are reliant on the same mechanisms ostensibly underpinning working memory interference (i.e., working memory disruption affecting the ability to inhibit the dominant habitual response). Hypothetically, the ability to inhibit the solve approach would protect comparison question performance under stereotype threat, enabling the correct comparison approach to be applied. Rather than protecting female maths performance against stereotype threat on comparison questions, inhibitory ability moderated female maths performance *overall* (i.e., for both solve and comparison questions). That is, low inhibitory ability was associated with poorer maths performance in general, following stereotype threat.

In Experiment 2, stereotype threat and maths question type interacted, reflecting Jamieson and Harkins' (2009) results. Stereotype threat facilitated solve relative to comparison question performance. Comparison question performance was reduced in the stereotype threat (vs. no-stereotype threat) condition, although solve question performance did not differ across stereotype threat and no-stereotype threat conditions.

Implications for theory and maths exam performance

Threatened test takers' greater tendency to select solve over comparison questions in Experiment 1 is consistent with the mere effort account. Some researchers from this theoretical tradition have argued that a performance motivation mechanism underlies stereotype threat effects (Jamieson & Harkins, 2007; McFall *et al.*, 2009; Seitchik & Harkins, 2015). The lack of an effect for the indirect maths motivation measure confirms Jamieson and Harkins' (2007) view that only actual performance indexes performance motivation. Maths question selection is a facet of the task itself, allowing measurement of

task motivation without undertaking an actual maths test. The motivational mechanism lies close to the prepotent response mechanism. If, as the working memory interference perspective suggests, failure to inhibit prepotent responses is implicated in stereotype threat effects, then those with high inhibitory ability should be less susceptible to failure to inhibit prepotent responding. Inconsistently with this perspective, we did not observe superior comparison question performance following stereotype threat.

Higher inhibitory ability did, however, protect against overall exam performance following stereotype threat. This finding ostensibly offers some support for the working memory interference perspective. That is, stereotype threat impacted less on high inhibitors' than on low inhibitors' overall maths performance. Beilock *et al.* (2007) explain such effects as the competition between the task at hand and the stereotypically threatening information, resulting in task performance reduction. In the absence of a mere effort account (the literature has not addressed this issue), we speculate that high inhibitors were more motivated to undermine the negative stereotype and familiar with a wider range of question type facilitating overall performance. Alternatively, low inhibitors overproduced prepotent responses regardless of question type.

Higher inhibitory ability would be useful in ignoring performance reducing negative self-applicable information associated with a test. Karbach and Kray (2009) suggested that the performance debilitating effects of a negative stereotype may be alleviated by trained improvements in working memory. These authors showed that task-switching training led to enhancements not only in task-switching performance, but also in inhibitory control (measured using Stroop). If inhibitory ability can be taught, then the skill might curtail unwanted performance deficits prompted by stereotype threat. Thus, the present findings provide a basis for future research efforts on the potential of inhibitory ability training as a stereotype threat intervention. In both experiments, gender identification and maths identification were not influenced by stereotype threat. In other words, stereotype threat was not a stressor leading to social identity threat (see also Appendix S1 for Experiments 1 and 2: motivation, gender identity, maths identity). Experiencing stereotype threat can motivate performance on solve type questions, but interfere with performance on comparison type questions (although participants are motivated to do well on both), suggesting that close attention is required to the contents of maths tests. Crucially, motivation only applies to the task at hand, meaning educators should be aware that female test takers could be affected more than male ones on maths performance.

Limitations

The task choice paradigm in Experiment 1 can be criticized on the grounds that the reason participants chose more solve (than comparison) questions to answer is simply because they considered solve questions easier. Previous research, for example, indicates that participants perform better on solve (than comparison) questions following stereotype threat (Jamieson, 2009). However, in keeping with the mere effort account, stereotype threat facilitates motivation to undermine the stereotype leading to greater activation of the prepotent solve response. Thus, the question type selection task in Experiment 1 acted as a measure of task-based motivation, activating the prepotent solve response. Furthermore, pre-testing of problems in that experiment ensured equality of difficulty.

In both experiments, solve questions were reportedly easier than comparison questions. Also, we observed superior performance on solve questions in Experiment 2. However, the questions in both experiments were piloted for difficulty and were independently assessed for equality of difficulty according to the past exam datable of

Assessment and Qualifications Alliance (http://www.aqa.org.uk/), a body responsible for exam development. In particular, the selected 10 questions were each worth three marks in real-world GCSE exams in England and Wales. Furthermore, Pilot Study 2 of Experiment 1, which assessed meta-perceptions and self-efficacy towards question type difficulty, bolstered our contention that both question types were perceived of as equal in self-efficacy. This pilot also established that the prepotent response was solve, likely to be activated automatically in keeping with previous research (Gallagher *et al.*, 2000; Jamieson & Harkins, 2009).

As expected, in both experiments, the first manipulation check item revealed that stereotype threat led to greater reported gender differences on the maths test relative to no-stereotype threat. However, results on the second manipulation check item (males perform better than females) were marginal following stereotype threat in Experiment 1, and null in Experiment 2. Thus, despite awareness of the negative performance stereotype, stereotype threat participants did not necessarily endorse this view; that is, they did not believe that task performance would reflect the negative stereotype. It is possible that the second manipulation check did not act as a manipulation check *per se*, but rather reflected participants' personal beliefs. A solution could be to implement a stronger threat manipulation. For example, Rosenthal, Quinn, and Seddon (2009), following weak stereotype threat effects, repeated an experiment with a more explicit stereotype manipulation. Specifically, as well as being informed that male and female performance would be compared, participants were also instructed that this comparison owed to women performing worse than men on the task.

Our results offer greater support to the mere effort account over the working memory interference perspective. The results contrast with Pennington *et al.*'s (2018) null stereotype threat performance finding. Although Pennington *et al.* did conduct a power analysis based on Jamieson and Harkins (2007, Experiment 3), the number of participants allocated to the negative group stereotype threat conditions appear low (for Experiment 1, $n = 23$; for Experiment 2, $n = 20$). This raises the prospect of an increased risk of type 2 errors and thus a failure to detect a genuine stereotype threat effect. This prospect is reinforced by the significant results for many of the stereotype threat manipulation checks across both experiments. Pennington *et al.* (p. 7) did obtain some significant higher order effects of stereotype threat, stating 'Although there was a significant main effect of stereotype condition on the proportion of corrective saccades, $F(2, 57) = 3.57, p = .035$, $\eta^2 p = .11$, pairwise comparisons between conditions were non-significant, $ps > .07$'. Again, the failure of the pairwise comparisons could be due to low power. Of further interest is Pennington's *et al.*'s decision to adopt an 'overlap' antisaccade task rather than the 'gap' antisaccade task favoured by Jamieson and Harkins (2007). In the overlap paradigm, the central fixation point remains on-screen while the peripheral target appears in contrast to gap paradigms where this is not so. Pennington *et al.* justify this change based on evidence suggesting that the gap method increases errors and reduces reaction times (Crawford *et al.*, 2005). Nonetheless, their approach represents a departure from Jamieson and Harkins' (2007) original study, thus rendering direct comparisons weaker.

Coda

Consistent with the mere effort account, stereotype threat amplified test takers' performance motivation in Experiment 1, activating the prepotent solve response, increasing selection of solve questions to answer. Also, in Experiment 2, higher inhibitory ability protected overall maths performance, but did not protect comparison question

performance for participants experiencing stereotype threat, as would be expected from the viewpoint of the working memory interference account. Our results suggest instead that the undermining of comparison question performance derives from the overproduction of prepotent solve responses. Taken together, our findings indicate that maths stereotype threat motivates the selection of prepotent (solve type) questions due to the overproduction of prepotent responses rather than interfering with working memory.

References

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

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. https://doi.org/10.1037/0096-3445.136.2.256

Bonnot, V., & Croizet, J. C. (2007). Stereotype internalization and women's maths performance: The role of interference in working memory. *Journal of Experimental Social Psychology*, *43*, 857–866. https://doi.org/10.1016/j.jesp.2006.10.006

Brodish, A. B., & Devine, P. G. (2009). The role of performance-avoidance goals and worry in mediating the relationship between stereotype threat and performance. *Journal of Experimental Social Psychology*, *45*, 180–185. https://doi.org/10.1016/j.jesp.2008.08.005

Carr, P. B., & Steele, C. M. (2009). Stereotype threat and inflexible perseverance in problem solving. *Journal of Experimental Social Psychology*, *45*, 853–859. https://doi.org/10.1016/j.jesp.2009.03.003

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. https://doi.org/10.1016/j.biopsych.2005.01.017

Croizet, J. C., Despres, G., Gauzins, M., Huguet, P., Leyens, J., & Meot, A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin*, *30*, 721–731. https://doi.org/10.1177/0146167204263961

Dao-Castellana, M. H., Samson, Y., Legault, F., Martinot, J. L., Aubin, H. J., Crouzel, C., . . . Syrota, A. (1998). Frontal dysfunction in neurologically normal chronic alcoholic subjects: Metabolic and neuropsychological findings. *Psychological Medicine*, *28*, 1039–1048. https://doi.org/10.1017/s0033291798006849

Davies, L. C., Conner, M., Sedikides, C., & Hutter, R. C. (2016). Maths question type and stereotype threat: Evidence from educational settings. *Social Cognition*, *34*, 196–216. https://doi.org/10.1521/soco.2016.34.3.196

Derks, B., Inzlicht, M., & Kang, S. (2008). The neuroscience of stigma and stereotype threat. *Group Processes and Intergroup Relations*, *11*, 163–181. https://doi.org/10.1177/1368430207088036

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.3758/BF03193146

Forbes, C. E., & Schmader, T. (2010). Retraining attitudes and stereotypes to affect motivation and cognitive capacity under stereotype threat. *Journal of Personality and Social Psychology*, *99*, 740–754. https://doi.org/10.1037/a0020971

Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, *51*, 136–158. https://doi.org/10.1016/j.jml.2004.03.008

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, *137*, 201–225. https://doi.org/10.1037/0096-3445.137.2.201

Gallagher, A. M., & De Lisi, R. (1994). Gender differences in scholastic aptitude test- mathsematics problem solving among high-ability students. *Journal of Educational Psychology*, *86*, 204–211. https://doi.org/10.1037/0022-0663.86.2.204

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathsematical problem solving. *Journal of Experimental Child Psychology*, *75*, 165–190. https://doi.org/10.1006/jecp.1999.2532

Harkins, S. G. (2001). The role of task complexity, and sources and criteria of evaluation in motivating task performance. In S. Harkins (Ed.), *Multiple perspectives on the effects of evaluation on performance: Toward an integration* (pp. 99–131). Norwell, MA: Kluwer Academic. https://doi.org/10.1007/978-1-4615-0801-4_5

Harkins, S. G. (2006). Mere effort as the mediator of the evaluation-performance relationship. *Journal of Personality and Social Psychology*, *91*, 436–455. https://doi.org/10.1037/0022-3514.91.3.436

Hasher, L., Lustig, C., & Zacks, R. T. (2007). Inhibitory mechanisms and the control of attention. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake & J. N. Towse (Eds.), *Variation in working memory* (pp. 227–249). New York, NY: Oxford University Press. https://doi.org/10.1093/ac prof:oso/9780195168648.003.0001

Hasher, L., Quig, M. B., & May, C. P. (1997). Inhibitory control over no-longer-relevant information: Adult age differences. *Memory and Cognition*, *25*, 286–295. https://doi.org/10.3758/BF03 211284

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, *11*, 365–371. https://doi.org/10.1111/1467-9280.00272

Jamieson, J. P. (2009). *The role of motivation in blatant stereotype threat, subtle stereotype threat, and stereotype priming*. Doctoral Dissertation, Northeastern University. Retrieved from http:// iris.lib.neu.edu/cgi/viewcontent.cgi?article=1009&context=psych_diss&sei-redir=1&referer= http%3A%2F%2F

Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology*, *93*, 544–564. https://doi.org/10.1037/0022-3514.93.4.544

Jamieson, J. P., & Harkins, S. G. (2009). The effect of stereotype threat on the solving of quantitative GRE problems: A mere effort interpretation. *Personality and Social Psychology Bulletin*, *35*, 1301–1314. https://doi.org/10.1177/0146167209335165

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

Jamieson, J. P., & Harkins, S. G. (2012). Distinguishing between the effects of stereotype priming and stereotype threat on maths performance. *Group Processes and Intergroup Relations*, *15*, 291–304. https://doi.org/10.1177/1368430211417833

Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science*, *12*, 978–990. https://doi.org/10.1111/j.1467-7687.2009.00846.x

Lustig, C., Hasher, L., & Tonev, S. T. (2001). Inhibitory control over the present and past. *European Journal of Cognitive Psychology*, *13*, 107–122. https://doi.org/10.1080/09541440042000241

MacLeod, C. M. (1991). Half a century of research on the Stroop effect - An intergrative review. *Psychological Bulletin*, *109*, 163–203. https://doi.org/10.1037/0033-2909.109.2.163

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. https://doi.org/10.1037/a0012878

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. https://doi.org/10.1037/a0012702

Norman, P., & Conner, M. (2006). The theory of planned behaviour and binge drinking: Assessing the moderating role of past behaviour within the theory of planned behaviour. *British Journal of Health Psychology*, *11*, 55–70. https://doi.org/10.1016/j.addbeh.2006.12.009

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., & Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex differences in science and maths achievement. *Proceedings of the National Academy of Sciences*, *106*, 10593–10597. https://doi.org/10.1073/pnas.0809921106

Pennington, C. R., Litchfield, D., McLatchie, N. M., & Heim, D. (2018). Stereotype threat may not impact women's inhibitory control or mathematical performance: Providing support for the null hypothesis. *European Journal of Social Psychology*. Advance online publication. https://doi.org/10.17605/osf.io/mdwyv

Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues*, *57*, 55–71. https://doi.org/10.1177/0361684313480045

Rosenthal, H., Quinn, K., & Seddon, E. (2009). *The key processes of stereotype threat: Stereotype activation and self-categorization*. Paper presented at the Social Psychology Section Annual Conference, University of Sheffield, UK.

Schmader, T. (2002). Gender identification moderates stereotype threat effects in women's maths performance. *Journal of Experimental Social Psychology*, *38*, 194–201. https://doi.org/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. https://doi.org/10.1037/0022-3514.85.3.440

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*, 336–356. https://doi.org/10.1037/0033-295X.115.2.336

Sedikides, C. (2012). Self-protection. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (2nd ed., pp. 327–353). New York, NY: Guilford Press.

Sedikides, C., Green, J. D., Saunders, J., Skowronski, J. J., & Zengel, B. (2016). Mnemic neglect: Selective amnesia of one's faults. *European Review of Social Psychology*, *27*, 1–62. https://doi.org/10.1080/10463283.2016.1183913

Seibt, B., & Forster, J. (2004). Stereotype threat and performance: How self-stereotypes influence processing by inducing regulatory foci. *Journal of Personality and Social Psychology*, *87*, 38–56. https://doi.org/10.1037/0022-3514.87.1.38

Seitchik, A. E., & Harkins, S. G. (2015). Stereotype threat, mental arithmetic, and the mere effort account. *Journal of Experimental Social Psychology*, *61*, 19–30. https://doi.org/10.1016/j.jesp.2015.06.006

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. https://doi.org/10.1177/1088868306294790

Smith, J. L., Lewis, K. L., Hawthorne, L., & Hodges, S. D. (2013). When trying hard isn't natural: Women's belonging with and motivation for male-dominated fields as a function of effort expenditure concerns. *Personality and Social Psychology Bulletin*, *39*, 131–143. https://doi.org/10.1177/0146167212468332

Smith, J. L., Sansone, C., & White, P. H. (2007). The stereotyped task engagement process: The role of interest and achievement motivation. *Journal of Educational Psychology*, *99*, 99–114. https://doi.org/10.1037/0022-0663.99.1.99

Soper, D. S. (2018). *Post-hoc statistical power calculator for multiple regression [software]*. Retrieved from http://www.danielsoper.com/statcalc

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's maths performance. *Journal of Experimental Social Psychology*, *35*, 4–28. https://doi.org/10.1006/jesp.1998.1373

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. https://doi.org/10.1037/0022-3514.69.5.797

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *12*, 643–662. https://doi.org/10.1037/h0054651

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*, 127–154. https://doi.org/10.1006/jesp.2001.1500

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*, 475–482. https://doi.org/10.1177/0146167200267001

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. https://doi.org/10.1016/j.jesp.2007.07.005

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Example maths problem types; Additional moderated regression analyses (Experiment 1 and Experiment 2).