Article

Group Processes & Intergroup Relations

GPIR

Group Processes & Intergroup Relations
1–14

DOI: 10.1177/1368430211417833
gpir.sagepub.com
SAGE

Distinguishing between the effects of stereotype priming and stereotype threat on math performance

Jeremy P. Jamieson[1] and Stephen G. Harkins[2]

Abstract

Stereotype threat and stereotype priming have both been shown to impair test performance. Although research suggests threat-based concerns distinguish the experience of threat from priming (Marx & Stapel, 2006), it is not clear whether these psychological phenomena impact performance via similar or distinct mechanisms. The current work demonstrates that priming and threat produce distinctive patterns of performance via different mechanisms. Motivation was found to play a proximal role in the effect of stereotype threat on females' math performance. Threatened females were motivated to disconfirm the negative stereotype, but performed more poorly because they were more likely than controls to use the incorrect, but prepotent conventional solution approach. Gender-math stereotypes do not incorporate the notion that females are motivated to disconfirm stereotypes. Instead the results are consistent with the argument that participants primed with female gender constructs performed poorly because they withdrew effort.

Keywords

mere effort, motivation, priming, stereotype threat

Paper received xxx; paper accepted xxx

Stereotype threat—one of the most heavily studied topics in social psychology over the past decade—refers to the concern that is experienced when one feels at risk of confirming, as self-characteristic, a negative stereotype about one's group (Steele & Aronson, 1995). A wide range of stereotypes have been tested (e.g., women's lack of ability in math and science: Spencer, Steele, & Quinn, 1999; African-Americans' underperformance on standardized tests: Steele & Aronson, 1995; White males' athletic inferiority: Stone, 2002). In each case, the threat of confirming the stereotype undermines the performance of stigmatized individuals.

[1]Harvard University
[2]Northeastern University

Corresponding author:
Jeremy P. Jamieson, 1206 William James Hall, Cambridge, MA 02138, USA.
Email: jamieson.jp@gmail.com

Stereotypes can also impact the performance of non-stigmatized individuals. For example, Wheeler, Jarvis, and Petty (2001) primed racial stereotypes by requiring White participants to write a story from the first person perspective about an individual named either "Erik" or "Tyrone." Participants who wrote the story from the perspective of Tyrone (a stereotypically African-American name) performed more poorly on a subsequent test of intellectual ability, which is consistent with racial stereotypes regarding intellectual ability (Wheeler et al., 2001).

Although stereotype threat and stereotype priming are widely studied topics, little has been done to determine whether priming and threat share underlying mechanisms. Some research has suggested this possibility: Stereotype activation leads to stereotype-consistent behaviors (e.g., Ambady, Paik, Steele, Owen-Smith, & Mitchell, 2004; Dijksterhuis & Bargh, 2001; Wheeler et al., 2001; Wheeler & Petty, 2000). More recently, however, Marx and Stapel (2006) have argued that despite the fact that stereotype threat and priming have the same effect on performance, the cognitive processes that underlie the effects are different. Specifically, they proposed that stereotype threat effects stem from the fact that the target of the stereotype knows about the stereotype and is a member of the stigmatized group, whereas priming effects simply require knowledge of the stereotype. As they write: "in stereotype threat situations, targets (but not nontargets) are affected because they know the group stereotype ("women are bad at math") and because they are members of the group that is targeted by the stereotype ("I am a woman")" (p. 244). It is this combination of "knowing and being" that gives rise to threat-based concerns, which distinguishes threat from priming.

The goal of the current work was to extend the theoretical framework provided by Marx and Stapel (2006) by testing the possibility that stereotype threat and priming differ not only in the threat-based concerns produced by the former, but not by the latter, but also in the pattern of performance that results. To do so, we relied on a motivation-based account of stereotype threat performance effects (Jamieson & Harkins, 2007, 2009, 2011).

Motivation and stereotype threat

Given the reliability of stereotype threat performance effects, much of the recent work in this area has focused on identifying mechanisms. Converging evidence suggests that threat effects are multiply mediated by cognitive, affective, and motivational processes (Steele, Spencer, & Aronson, 2002). Though most of the initial work on mechanisms focused on cognitive (e.g., Schmader & Johns, 2003) and affective processes (e.g., Bosson, Haymovitz, & Pinel, 2004), recent research has considered motivational processes (e.g., Carr & Steele, 2009; Forbes, Schmader, & Allen, 2008; Jamieson & Harkins, 2007, 2009, 2011; Rydell, Rydell, & Boucher, 2010; Rydell, Shiffrin, Boucher, Van Loo, & Rydell, 2010). The crux of these motivational explanations is that stigmatized individuals' efforts to avoid confirming the stereotype can have the paradoxical effect of harming performance (e.g., Jamieson & Harkins, 2009; Rydell, Shiffrin et al., 2010; Seibt & Forster, 2004).

In the current work, we focused on motivational processes because efforts to disconfirm stereotypes are implicated in stereotype threat, but not in ideomotor priming processes. More specifically, we examined the processes underlying threat and priming using Jamieson and Harkins's (2007, 2009, 2011) motivation-based mere effort model. This model (and others: O'Brien & Crandall, 2003) argues that one consequence of stigmatized participants' motivation to disconfirm negative stereotypes is that it potentiates whatever response is prepotent, or most likely to be produced (e.g., reading the color-word instead of naming the color in the Stroop Color-Word Test). If the prepotent response is correct, stereotype threat can improve performance—this can be seen in research showing threat facilitates simple task performance (Jamieson & Harkins, 2011; O'Brien & Crandall, 2003). If the prepotent response is incorrect, and participants do not know, as is quite often the case, or lack the knowledge or time for correction, potentiating that response will debilitate performance. However, in those occasional cases where participants are able to recognize that their prepotent tendencies are incorrect and they are given the opportunity to implement correction, performance can be facilitated.

Recently Jamieson and Harkins (2009) tested this model in research examining participants' performance on a test comprised of two types of quantitative GRE problems—"solve" and "comparison"—which differ in the solution approach that tends to be most efficient. Solve type problems tend to be most efficiently solved by using the traditional solution approach taught in US schools: Directly applying an equation or algorithm and computing an answer. On the other hand, comparison problems require the test-taker to take an unconventional approach, such as simplifying terms or using logic, estimation, and/or intuition to get the correct answer. In fact, comparison problems often do not require any computations at all.

Research in the educational psychology literature indicates that when given math problems, American students predominantly take a conventional approach using known rules and equations. For example, Gallagher, De Lisi, Holst, McGillicuddy-De Lisi, Morely, and Cahalan (2000, Exp. 2) found that participants used a conventional approach 55.5% of the time, whereas unconventional (e.g., logic, estimation, etc.) approaches were used only 10% of the time, with the remainder consisting of guesses (16.5%), omissions (9%), and unknowns (9%). This pattern has been replicated in other studies (Gallagher & De Lisi, 1994; Jamieson & Harkins, 2009; Quinn & Spencer, 2001), and it holds true regardless of problem type. For instance, on solve type problems researchers found that the conventional approach was used 66% of the time versus 9% for the unconventional approach, and on the comparison problems the conventional approach was used 45% of the time versus 11% for the unconventional approach (Gallagher et al., 2000, Exp. 2).[1]

Given the training American students receive in school, this pattern of findings is not particularly surprising. As Stigler and Hiebert (1999) note, in the US, "teachers present definitions of terms and demonstrate procedures for solving specific problems. Students are then asked to memorize the definitions and practice the procedures" (p. 27). In other words, students are taught to memorize equations, then directly apply those equations to the problems, and finally to calculate the answers.

On solve type problems, the conventional approach is typically the most effective approach. As long as the test-taker knows the correct equation to apply, solving it will produce an answer. Thus, if threatened females are motivated to disconfirm the stereotype, this motivation should translate into facilitated performance on these problems. However, the form facilitation can take is constrained by the task. Because the conventional approach is prepotent, everyone tends to use it on solve type problems, and threatened participants should not differ from controls in their knowledge of formulas or operations. As a result, effort cannot aid participants' success in solving the problems that they attempt. Rather, heightened motivation can only be reflected in the number of problems attempted or answered. And, of course, time constraints place an upper bound for this effect.

In contrast, the conventional solving approach is far less efficient on comparison problems and often does not work at all. Because of their math training, it is also unlikely that participants even recognize that they are not using the best approach on these problems. Thus, potentiating the prepotent conventional approach should debilitate performance. Given the fact that the effect of debilitation on comparison problems is much stronger than the facilitation effect on solve problems, threatened participants' overall performance should suffer compared to controls. Jamieson and Harkins (2009) found support for each of these predictions: Females subject to stereotype threat completed more solve problems, solved fewer comparison problems correctly, and performed worse overall compared to controls.

Motivation and priming stereotypes

However, there is no reason to expect that motivational processes would underlie stereotype priming effects. Research on behavioral priming indicates that primes impact behavior by activating information associated with whatever construct is primed (e.g., Bargh, Chen, & Burrows, 1996). Stereotypes about females' math ability suggest that females do not have the same degree of mathematical ability as males. This gender-math stereotype does not incorporate the notion that females are motivated by their concern about confirming math stereotypes, which potentiates their use of conventional solution approaches leading to divergent patterns of performance on solve and comparison type problems. Instead, the activated stereotypic information passively guides behavior (e.g., Bargh et al., 1996). As a result, activating the female gender construct should debilitate performance on both types of problems (e.g., Marx & Stapel, 2006; Wheeler et al., 2001).

To test this hypothesis, participants were randomly assigned either to one of the conditions in the stereotype threat paradigm (threat vs. control) or to one of the conditions in the priming paradigm (priming vs. control), and then completed the same GRE-Q test used by Jamieson and Harkins (2009). We used the same explicit threat manipulation as Jamieson and Harkins (2009). In the threat condition, females were informed that they would be taking a test of quantitative ability and that the test exhibited gender differences. On the other hand, no threat control participants were told the test did not exhibit any gender differences.

Our priming manipulation was adapted from Wheeler et al. (2001). This manipulation required participants to write a short story from the perspective of a female undergraduate student, which has been shown to increase the salience of information associated with female gender stereotypes in the active self-concept (Wheeler, DeMarree, & Petty, 2007). In the control prime condition, participants wrote a similar story from a university-wide perspective that did not implicate any particular group. Because motivational processes should be absent from prime to behavior effects, participants given in the female prime condition were expected to perform more poorly than controls on both solve and comparison problems because of the stereotype that females perform more poorly than males. In contrast, the threat participants should perform differentially as function of problem type: Profound debilitation on comparison problems, a smaller facilitation effect on solve problems, and debilitated performance overall compared to their controls.

Method

Participants

Seventy-six Northeastern University female undergraduates participated in this experiment in exchange for course credit.

Materials

The math test consisted of 30 multiple-choice problems taken from the quantitative section of the GRE. The test included 15 comparison type problems and 15 solve type problems, and was presented as a paper and pencil test. See Appendix A for examples of each type.

Problems were sampled from a GRE preparation book that included performance norms, as indexed by the proportion of test-takers answering each problem correctly out of all those who attempted that problem. First, we randomly picked 12 problems of each type from problems that varied in their solution rates from 35% to 65%. We then picked the final three problems for each type so that mean overall accuracy averaged 50% for each problem set (unconventional range = 38% to 60%; conventional range = 42% to 63%). Problems were randomized throughout the test with the constraint that no one type of problem could appear in more than three consecutive problems. Problem order was identical for all participants.

Participants worked on the test for 20 minutes and were instructed to complete as many problems as they could as accurately as possible. All participants were given two practice problems (one of each type) prior to beginning and were not permitted the use of calculators. Ample scratch-paper was provided for participants to show their work.

Procedure

Participants were randomly assigned to one of the four conditions in a 2 (Paradigm: Stereotype Threat vs. Priming) x 2 (Condition: Experimental vs. Control) design. Stereotype threat was manipulated explicitly by having a male experimenter inform participants that they would be completing a math test that either had (threat) or had not (control) been shown to produce gender differences. No specific mention was made as to whether men outperformed women or vice versa, only that gender differences did or did not exist on the task. Participants were expected to infer that women would perform more poorly than men based on the societal stereotype. This manipulation has been successfully used in previous research (e.g., Brown & Pinel, 2003; Jamieson & Harkins, 2007, 2009, 2011; Keller, 2002; Keller & Dauenheimer, 2003; O'Brien & Crandall, 2003; Spencer et al., 1999).

The priming manipulation was adapted from Wheeler et al. (2001). Participants were informed that they would be completing math problems, but prior to the math test the experimenter explained that he was helping another lab pilot test another experiment "examining the role of hemispheric dominance on creativity," and asked if the participants would also complete this experiment. All but one participant, who was excluded from the analysis, agreed. Participants were then told that they would write a creative story about an assigned topic with either their dominant or non-dominant hand. All participants wrote with their dominant hands.

Participants in the female prime condition were asked to write about a day in life of a female Northeastern University student named "Ashley" for five minutes. Participants assigned to the gender-neutral control condition wrote about a typical day at Northeastern University, as an institution. That is, participants wrote about general events that went on at a university on a daily basis, such as the when classes start or an event. This control ensured that gender-specific information was not activated.

After these manipulations, the experimenter gave the participants the two practice math problems and told them that they would have 20 minutes to work on a 30-item test. Upon completion of the math test, each participant responded to a questionnaire. Two questions allowed us to evaluate the effects of the manipulations: "To what extent are there gender differences in performance on this task?" (1 = "no gender differences" and 11 = "gender differences"); and "Who do you believe performs better on this task?" (1 = "males perform better," 6 = "males and females perform the same," and 11 = "females perform better"). We expected these measures to reflect the stereotype threat manipulation but not to be affected by the priming manipulation.

Results

Manipulation checks

Manipulation check items were analyzed in 2 (Paradigm: Priming vs. stereotype threat) x 2 (Condition: Control vs. experimental) between-subjects ANOVAs. First, we examined the extent to which participants believed gender differences existed on the test. Consistent with past stereotype threat research, a contrast between the means in the stereotype threat and its control condition showed that threatened participants believed gender differences existed to a greater extent ($M = 5.57$, $SD = 2.76$) than those in the threat control condition ($M = 3.45$, $SD = 2.61$), $F(1, 71) = 6.05$, $p = .016$, $d = .58$. On the other hand, participants in the gender prime and control prime conditions did not differ on this measure (overall $M = 4.59$, $SD = 2.82$), $F < 1$. This pattern of means produced a marginal Paradigm x Condition interaction in the overall 2 x 2 analysis, $F(1, 71) = 2.79$, $p = .099$, $d = .40$.

We then analyzed whether participants believed males or females performed better. A contrast showed that females subject to stereotype threat indicated that they believed males performed better on the math test to a greater extent (M = 4.62, SD = 1.46) than no threat control participants (M = 5.85, SD = 1.53), $F(1, 71) = 7.31$, $p = .009$, $d = .64$. Gender prime and control prime participants did not differ on this measure (overall M = 5.31, SD = 1.42), $F < 1$. This pattern of effects produced a Paradigm x Condition interaction in the 2 x 2 analysis, $F(1, 71) = 7.00$, $p = .01$, $d = .63$.

Thus, females subject to stereotype threat believed gender differences existed and that males performed better on the GRE test than no threat controls. Additionally, participants in the priming conditions did not differ in their ratings on either of these self-report measures, consistent with the notion that priming influences behavior via ideomotor processes (e.g., Wheeler et al., 2001).

Performance

To examine performance we analyzed the number of problems participants answered correctly.[2] We first looked at overall performance in a 2 (Paradigm) x 2 (Condition) between-subjects ANOVA. Consistent with past research, we found a main effect for condition. Participants in the experimental conditions (stereotype threat and priming) answered fewer problems correctly (M = 7.82, SD = 3.33) than participants in the control conditions (M = 9.43, SD = 2.44), $F(1, 71) = 6.02$, $p < .02$, $d = .58$. No other effects were significant, ps > .20.

To examine performance as a function of problem type, the results were analyzed in a 2 (Paradigm) x 2 (Condition) x 2 (Problem Type: Comparison vs. solve) ANOVA with paradigm and condition as between subjects factors and problem type as a within subjects factor. This analysis produced main effects for condition, $F(1, 71) = 6.29$, $p = .014$, $d = .60$, and problem type $F(1, 71) = 24.04$, $p < .001$, $d = 1.16$, and a Condition x Problem type interaction $F(1, 71) = 6.13$, $p = .016$, $d = .59$. However, these effects must be interpreted in the context of the three-way Paradigm x Condition x Problem Type interaction, $F(1, 71) = 3.97$, $p = .05$, $d = .47$. To interpret this effect, we examined performance on comparison and solve problems in two two-way ANOVAs (paradigm x condition), using the error term from the three-way interaction.

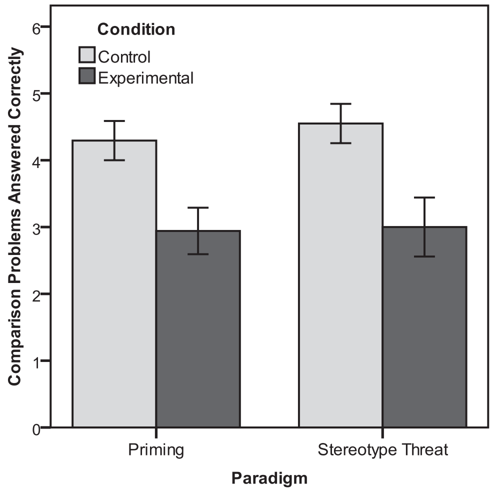

Figure 1. Number of comparison type problems answered correctly as a function of paradigm and condition. Error bars = +/- standard error of the mean.

Comparison problems This analysis produced a main effect for condition, $F(1, 71) = 16.21$, $p < .001$, $d = .96$. As shown in Figure 1, participants in each of the experimental conditions performed more poorly than the participants in the control conditions. That is, both stereotype threat and gender prime participants answered fewer comparison problems correctly than females assigned to no threat and institutional prime control conditions.

Solve problems Analysis of the solve problems produced a Paradigm x Condition interaction, $F(1, 71) = 6.41$, $p = .014$, $d = .60$ (see Figure 2). Participants in the female gender prime condition performed more poorly than the control prime participants, $F(1, 71) = 4.87$, $p = .03$, $d = .52$. On

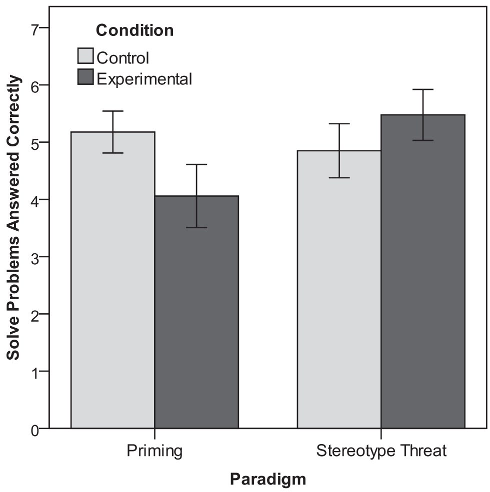

Figure 2. Number of solve type problems answered correctly as a function of paradigm and condition. Error bars = +/- standard error of the mean.

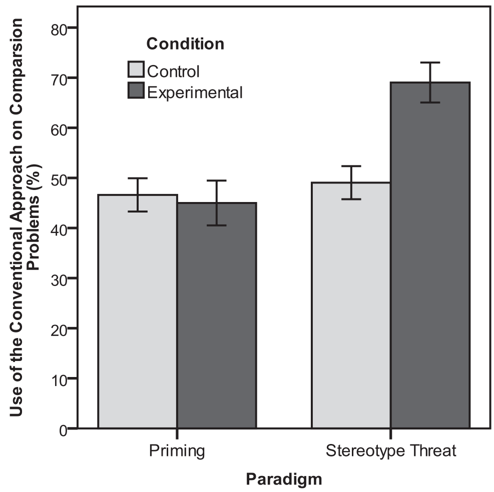

Figure 3. The percentage of comparison type problems participants used the prepotent conventional solution approach as a function of paradigm and condition. Error bars = +/- standard error of the mean.

the other hand, there was a trend for participants subject to stereotype threat to perform *better*, not worse, than their no threat control counterparts, $F(1, 71) = 2.06$, $p = .155$, $d = .34$. Thus, unlike on comparison problems, threat and priming impacted performance on solve problems differently. Whereas the female gender prime impaired performance on solve problems, stereotype threat actually tended to facilitate performance.[3]

Solution approach

As noted before, participants were given scratch-paper to work problems on. An independent rater, blind to condition, coded the scratch paper and the percentage of problems on which the participants' work showed evidence of using the conventional approach was analyzed in the 2 (Paradigm) x 2 (Condition) x 2 (Problem Type) design.[4] This analysis produced significant Paradigm x Condition, $F(1, 71) = 6.68$, $p = .012$, $d = .61$, and Condition x Problem Type, $F(1, 71) = 10.77$, $p = .002$, $d = .78$, interactions. However, we interpreted these interactions in the context of the marginal Paradigm x Condition x Problem Type interaction, $F(1, 71) = 2.82$, $p = .097$, $d = .40$. Like the performance analysis, we decomposed this interaction by examining the different problem types separately.

Comparison problems Analysis of these problems produced a Paradigm x Condition interaction, $F(1, 71) = 18.76$, $p < .001$, $d = 1.03$. As can be seen in Figure 3, participants in the priming conditions did not differ in their use of the conventional approach as a function of condition (overall $M = 45.80\%$, $SD = 16.01$), $F < 1$, but females assigned to the stereotype threat condition used the conventional approach significantly more than the no threat control participants, $F(1, 71) = 35.43$, $p < .001$, $d = 1.41$.

To test whether potentiation of the conventional approach directly debilitated threatened participants' comparison problem performance, we conducted a mediation analysis following the procedures suggested by Kenny, Kashy, and Bolger (1998) on the number of comparison problems solved incorrectly. Replicating previous research (Jamieson & Harkins, 2009), participants' use of the prepotent conventional approach mediated the debilitating effect of stereotype threat on comparison problem performance,

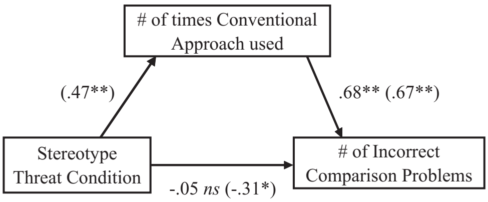

Figure 4. Number of times the prepotent conventional approach was used on comparison problems as a mediator of number of comparison problems answered incorrectly. Coefficients in parentheses indicate zero-order correlations. Coefficients not in parentheses represent parameter estimates for a recursive path model including both predictors. Asterisks (*) indicate parameter estimates or correlations that differ from zero at $p < .05$, double asterisks (**) indicate parameter estimates or correlations that differ from zero at $p < .01$. Stereotype threat condition is dummy coded (stereotype threat = 1, no stereotype threat = 0).

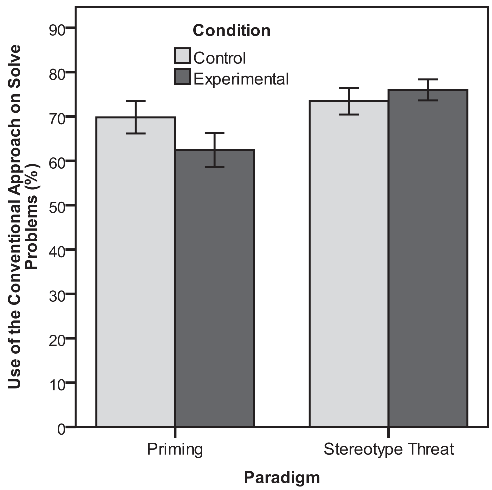

Figure 5. The percentage of solve type problems participants used the prepotent conventional solution approach as a function of paradigm and condition. Error bars = +/- standard error of the mean.

Sobel $Z = 2.63$, $p = .009$ (see Figure 4).[5] Thus, on comparison problems, threatened females used the conventional approach significantly more than controls, and this use of the prepotent, conventional approach mediated the effect of threat on performance.

In contrast, participants in the female prime condition did not differ from control prime participants in their use of the conventional approach on comparison problems, but, nonetheless, performed significantly more poorly than these participants. In an effort to gain some insight into this finding, we looked at performance as a function of whether or not participants showed evidence of using the conventional approach. As would be expected, on those comparison problems on which the conventional approach was used, performance was debilitated to the same extent in the gender prime (M = 30.52%, SD = 28.62%) and control conditions (M = 24.58%, SD = 23.07%), $F < 1$. However, when participants did not show evidence of using the conventional approach, controls outperformed (M = 59.53%, SD = 20.31%) prime participants (M = 38.87%, SD = 22.31%), $F(1, 69) = 6.10$, $p < .05$, $d = .62$.

Solve problems Analysis of performance on the solve problems produced a main effect for paradigm, $F(1, 71) = 11.84$, $p = .001$, $d = .82$, that must be interpreted in the context of the Paradigm x Condition interaction, $F(1, 71) = 3.93$, $p = .05$, $d = .47$ (see Figure 5). Participants in the stereotype threat paradigm did not differ as a function of condition in their use of the conventional approach (overall M = 74.76%, SD = 12.12), $F < 1$, but females assigned to the gender prime condition used the prepotent conventional approach less often than control prime participants, $F(1, 71) = 4.32$, $p = .041$, $d = .49$.

Once again, we looked at math performance as a function of whether or not participants used the conventional approach. On the subset of problems on which both prime and control participants showed evidence of using the conventional approach, the gender prime participants solved the same percentage of solve type problems (M = 54.62%, SD = 28.62%) as controls (M = 60.64%, SD = 20.03%), $F < 1$. However, when participants assigned to the gender prime condition did not show any written evidence of using the conventional approach, their performance dropped (M = 40.78%, SD = 31.49%),

$F(1, 69) = 3.29$, $p = .074$, $d = .44$, whereas the performance of controls was unchanged ($M = 54.95\%$, $SD = 28.02\%$), $F < 1$, interaction $F(1, 69) = 2.21$, $p = .141$, $d = .36$.

Discussion

This research replicates Marx and Stapel's (2006) priming finding: Overall, participants primed with female gender constructs performed more poorly than controls. Of course, numerous stereotype threat studies have also found that, overall, females subject to stereotype threat perform more poorly than no threat controls on math tests, an effect that we replicated in the present research. However, an examination of performance as a function of problem-type reveals that stereotype threat and priming produced these outcomes in different ways.

We found that stereotype threat tended to facilitate females' performance on solve problems, an effect reliable across the three experiments (present research plus two experiments in Jamieson & Harkins, 2009) that have examined performance as a function of problem type, combined $Z = 1.94$, $p = .052$. However, stereotype threat profoundly debilitates performance on comparison problems, combined $Z = -4.62$, $p < .001$. This pattern of results is consistent with the mere effort model (Jamieson & Harkins, 2007; 2009, 2011), which suggests that stereotype threat motivates participants, potentiating prepotent response tendencies. On these math problems, the prepotent response is to use the conventional approach, which is the appropriate approach on solve problems, but not on comparison problems.

In contrast to the pattern of findings under stereotype threat, the female gender prime debilitated performance on *both* types of problems, and for both genders (see note 3). An analysis of solution approaches and solution rates affords some insight into the mechanisms underlying these performance effects. On comparison problems, females under stereotype threat used the conventional approach significantly more than controls, and their use of the prepotent, conventional approach mediated the effect of threat on comparison problem performance (cf., Jamieson & Harkins, 2009). However, participants in the female prime condition did not differ from control prime participants in their use of the conventional approach on these problems.

Our analysis of performance as a function of whether or not participants exhibited evidence of using the conventional solution approach showed that when participants used the conventional approach on comparison type problems, performance was debilitated to the same extent in both prime conditions. However, when participants did not show evidence of using the conventional approach (i.e. when no work was shown), controls outperformed prime participants. On these problems, all participants had an equal opportunity to use the appropriate approach (logic and/or estimation), but female prime participants appear to have done so less than controls as the latter participants outperformed the former. These findings suggest the possibility that prime participants performed poorly because they withdrew effort, which would be consistent with the content of gender-math stereotypes (e.g., Steele & Ambady, 2006).

On solve problems, we found that females under stereotype threat and no threat controls used the conventional approach to the same extent. In contrast, female prime participants used the prepotent conventional approach *less* than their controls. Because use of the conventional approach predicts better performance on solve type problems for all participants ($\beta = .29$, $p = .013$), gender prime participants may have performed poorly because they did use the conventional approach to the same extent as their controls. Consistent with this interpretation, on the subset of solve type problems on which both prime and control participants showed evidence of using the conventional approach, they performed similarly. It was when female prime participants did not show evidence of using the conventional approach that their performance dropped, whereas the performance of controls was unchanged.

When participants did not show any work, they could still be working the problem in their head using the conventional approach. When both prime and control participants showed written evidence, the two groups did not differ in performance. As a result, if we assume prime and control participants were using the conventional approach in their heads to the same extent on the problems on which there was no written work, there is no reason to believe that they would differ in their solution rates. However, controls outperformed those in the prime condition on these problems, suggesting that the female prime participants withdrew effort, not even bothering to use the conventional solving approach. Of course, all our evidence for effort withdrawal on comparison and solve problems is indirect, and thus future research should seek to obtain evidence of this.

This research builds on previous work by Marx and Stapel (2006) by demonstrating that threat and priming are not only accompanied by different thoughts and concerns, but also produce different patterns of performance and operate via different mechanisms. The findings suggest that when females are threatened by the possibility of confirming negative gender-math stereotypes, they are motivated to disconfirm the stereotype, whereas priming the female gender construct does not increase motivation, but instead appears to cause females to withdraw effort. Thus, knowledge of a stereotype alone does not tell us what behavior it will produce in its target.

To disentangle the effects of threat and priming on performance, the research presented here relied on the framework provided by motivational explanations of stereotype threat (Jamieson & Harkins, 2007, 2009, 2011). It may seem surprising that one of the reasons that females underperform on math tests is because they are trying not to underperform (e.g., Seibt & Forster, 2004). However, the research presented here, as well as other work, suggests that explicit efforts to perform well often undermine performance outcomes. For instance, Beilock, Jellison, Rydell, McConnell, and Carr (2006) found that expert male golfers who were told that females are much better putters tried to execute each component of their putting stroke as flawlessly as possible in an effort to disconfirm this negative stereotype. However, motivation to perform well broke down the proceduralization of their putting strokes, debilitating performance. Thus, research shows that the approaches individuals think will help them do well may actually work against them.

Distinguishing between threat and priming performance outcomes and mechanisms has direct relevance for developing interventions to reduce the negative impact of gender-math stereotypes. If ideomotor processes passively guided behavior in situations of stereotype threat, then it would be difficult to implement interventions short of eliminating the stereotype at the societal level. The current research demonstrates that women are, indeed, motivated to disconfirm the negative stereotypes directed at their group. Because females do not want to be typecast as poor at math, researchers can use this motivation when developing interventions. For example, in a recent study (Jamieson & Harkins, 2009), females subject to stereotype threat were told the correct solution approaches to take on problems on a math test. This simple instruction improved the performance because it reduced reliance on conventional approaches. Thus, that intervention redirected the motivation females experienced under threat from inflexible perseverance (see also Carr & Steele, 2009) to channeling efforts into more diverse solution approaches.

Additionally, this and other research (e.g., Carr & Steele, 2009; Gallagher et al., 2000; Jamieson & Harkins, 2009) suggests changes that might be made in the design of standardized tests in an effort to reduce group differences in performance. At its core, stereotype threat is a source of measurement bias (Wicherts, Dolan, & Hessen, 2005). That is, stereotype threat does not impair females' latent quantitative ability, but rather affects the measurement of their ability. Thus, one way to improve performance under threat is to limit measurement bias in our metrics of math ability. For instance, the quantitative section of the GRE general test currently includes both comparison type and solve type problems. The research presented here shows that females subject to stereotype threat use conventional approaches to solving math problems to the exclusion of other approaches. Instead of requiring test-takers to employ non-traditional solution approaches (i.e. logic/estimation) on standardized tests, test developers could design problems that assess latent quantitative ability that rely on solution approaches widely taught in the mathematical curricula. Similarly, changing the way math is taught would also go a long way towards reducing group differences in high-level standardized tests. That is, educators could seek to teach students to use both traditional and non-traditional approaches to solving math problems, rather than emphasizing the conventional approach and memorization.

This research also has relevance beyond the study of group stereotypes by suggesting that we should exercise caution when generalizing from findings produced in priming paradigms to the actual experience of some psychological state. For example, Smith, Jostmann, Galinsky, and van Dijk (2008) recently reported finding that low power impaired performance on executive-function tasks. This finding was subsequently generalized to real world settings, which argues that empowering employees could help reduce costly organizational errors. However, that research used manipulations aimed at activating stereotypes concerning power (e.g., scrambled-sentences priming task). Thus, although that research certainly informs us about the content of stereotypes associated with the powerful and powerless, it may not reflect the actual cognitive ability of people who occupy these positions.

Notes

1 In some cases, it is possible to answer solve type problems by using logic/estimation, and to solve comparison problems by taking a conventional approach. In fact, on some problems, the alternative approach is actually more efficient, but this is not the case for the majority of problems. For example, of the 30 problems used in this research, it would have been possible to use logic and estimation on five of the 15 solve problems and to solve the equations on six of the 15 comparison problems. However, solving the equations was the most efficient approach for 14 of the 15 solve problems and using logic and estimation was the most efficient approach for 14 of the 15 comparison problems.

2. Using percentage correct as the dependent measure did not change the pattern of the results.

3. Jamieson and Harkins (2009) found that males were not affected by the stereotype threat manipulation. Given this, we did not run males in the stereotype threat paradigm in the current research. However, we did run males in the priming paradigm, and found that there were no gender effects. That is, males showed exactly the same pattern of priming effects as was produced by females, $ps < .05$.

4. No identifying information was visible during coding. To code a problem, the rater went through the work shown step-by-step to ascertain exactly what the participants were doing to solve each problem. There were instances in which the participant showed written evidence of using an unconventional approach. Of course, these instances were not included in the conventional approach analyses. We should note that solution approaches could have been measured in other ways. For instance, we could have asked participants to report what they were doing while they solved problems. We chose to use the more implicit method of coding scratch paper in an effort to increase external validity. That is, explicit reporting may potentially introduce demand characteristics—participants could have tried to figure out what we wanted them to do—or talking through a problem out loud may have altered participants' approach. With that said, we doubt that using an explicit measure would have impacted the pattern of results because the current research and previous research using this implicit method (Jamieson & Harkins, 2009) have found the same pattern of results as studies that have used explicit reporting measures (Gallagher & DeLisi, 1994; Gallagher et al., 2000; Quinn & Spencer, 2001).

5. The number of times participants used the conventional approach was used in the mediation analysis, rather than percentage, because the former provides a more valid indicator of the role that the conventional approach plays in comparison problem performance. Percentage assumes that whenever participants did not write anything down, they were not using this method, even though they may have been doing so in their heads. On the other hand, although number could

underestimate the number of times the conventional approach was used, it cannot misclassify the approach as unconventional when, in fact, it was conventional. In addition, we predicted the number of incorrect problems, rather than number correct, because our hypothesis is that using the conventional approach on comparison problems directly leads to poor performance. However, not using the conventional approach is only one requirement for success. To solve the problem correctly, participants must also recognize that an unconventional approach should be used, and then implement it successfully. Thus, the most direct test of the hypothesis requires prediction of incorrect responses.

References

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

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*, 230–244.

Beilock, S. L., Jellison, W. A., Rydell, R., J., McConnell, A. R., & Carr, T. H. (2006). On the causal mechanisms of stereotype threat: Can skills that don't rely heavily on working memory still be threatened? *Personality and Social Psychology Bulletin*, *32*, 1059–1071.

Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of Experimental Social Psychology*, *40*, 247–255.

Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology*, *39*, 626–633.

Carr, P. B., & Steele, C. M. (2009). Stereotype threat and inflexible perseverance in problem solving. *Journal of Experimental Social Psychology*, *45*, 853–859.

Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 33, 1–40). San Diego, CA: Academic Press.

Forbes, C. E., Schmader, T., & Allen, J. J. B. (2009). The role of devaluing and discounting in performance monitoring: A neuropsychological study of minorities under threat. *Social Cognitive Affective Neuroscience*, *3*, 253–261.

Gallagher, A. M., & De Lisi, R. (1994). Gender differences in Scholastic Aptitude Test-Mathematics problem solving among high-ability students. *Journal of Educational Psychology*, *86*, 204–211.

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. *Journal of Experimental Child Psychology*, *75*, 165–190.

Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology*, *93*, 544–564.

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

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

Keller, J. (2002). Blatant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles*, *47*, 193–198.

Keller, J., & 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.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 1, pp. 233–265). New York, NY: McGraw-Hill.

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

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.

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.

Rydell, R. J., Rydell, M. T., & Boucher, K. L. (2010). The effect of negative stereotypes on learning. *Journal of Personality and Social Psychology*, *99*, 883–896.

Rydell, R. J., Shiffrin, R. M., Boucher, K. L., van Loo, K., & Rydell, M. T. (2010). Stereotype threat prevents perceptual learning. *Proceedings of the National Academy of Sciences*, *107*, 14042–14047.

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.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, *85*, 440–452.

Smith, P. K., Jostmann, N. B., Galinsky, A. D., & van Dijk, W. W. (2008). Lack power impairs executive functions. *Psychological Science*, *19*, 441–447.

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.

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.

Steele, J. R., & Ambady, N. (2006). "Math is Hard!" The effect of gender priming on women's attitudes. *Journal of Experimental Psychology*, *42*, 428–436.

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 379–440). New York, NY: Academic Press.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap*. New York, NY: Simon & Schuster, Inc.

Stone, J. (2002). Battling doubt by avoiding practice: The effects of stereotype threat on self-handicapping in White athletes. *Personality and Social Psychology Bulletin*, *28*, 1667–1678.

Wheeler, S. C., DeMarree, K. G., & Petty, R. E. (2007). Understanding the role of the self in prime-to-behavior effects: The active self account. *Personality and Social Psychology Review*, *11*, 234–261.

Wheeler, S. C., Jarvis, W. B., & Petty, R. E. (2001). Think unto others: The self-destructive impact of negative racial stereotypes. *Journal of Experimental Social Psychology*, *37*, 173–180.

Wheeler, S. C., & Petty, R. E. (2000). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin*, *127*, 797–826.

Wicherts, J. M., Dolan, C. V., & Hessen, D. J. (2005). Stereotype threat and group differences in test performance: A question of measurement invariance. *Journal of Personality and Social Psychology*, *89*, 696–716.

Biographical notes

Jeremy Jamieson is a Postdoctoral Fellow in the Psychology Department at Harvard University. His research examines the effects of social stress on performance, decision making, and physiology.

Stephen Harkins is Professor of Psychology at Northeastern University. His research interests focus on the effect of motivation on performance.

Appendix A

Examples of the problems used in this research. These problems were sampled from a GRE preparation book.

Solve type

If the total surface area of a cube is 24, what is the volume of the cube?

a) 8
b) 24
c) 64
d) 48√6
e) 216

For this problem, the individual must apply the formula for the volume of a cube, which is: length x width x height (all of which are the same value for a cube). To get the length of a side, the individual divides 24 by 6 (there are 6 faces on a cube) to obtain the area of one face, 4. The length of one side is 2 (area = length x width). To compute volume, the test-taker then cubes 2 to get the answer, 8. Thus, solve problems involve the application and computation of equations.

Comparison type

$n = (7\ )(19^3)$

Column A	Column B
The number of distinct positive factors of *n*	10

a) The quantity in Column A is greater
b) The quantity in Column B is greater
c) The two quantities are equal
d) The relationship cannot be determined from the information given

This problem can be solved by using intuition. First, the test-taker must realize that each number presented (7, 19, 3) is a prime number. Thus, the test-taker can logically deduce that the factors of the end product can only be multiples of 7 and 19. Thus, the factors of the final product are: 7, 7*19, $7*19^2$, 19, 19^2, 19^3, plus the final product itself ($7*19^3$) and 1. Because the goal of the problem is not to compute the value of *n*, but simply to determine whether the number of positive factors of *n* is greater than, less than, or equal to 10, all the test-taker now needs to do is to add up the number of distinct positive factors (8) to find that Column B is greater than Column A. Thus, the correct answer choice is "b," and no calculations were necessary.