



Reports

Stereotype threat and inflexible perseverance in problem solving

Priyanka B. Carr *, Claude M. Steele

Department of Psychology, Stanford University, Jordan Hall, Bldg 420, Stanford, CA 94305, United States

ARTICLE INFO

Article history:

Received 24 July 2008

Revised 19 February 2009

Available online 17 March 2009

Keywords:

Stereotype threat

Identity threat

Stereotype suppression

Cognitive flexibility

Inflexibility

Perseverance

Mental set

ABSTRACT

The present research examines whether women burdened by stereotype threat, a threat of confirming negative ingroup stereotypes (Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797–811), are less able to abandon old strategies and employ newer, more efficient ones when conditions change. In two studies, stereotype threat was found to increase inflexible perseverance: women made to believe they were taking a diagnostic math/spatial ability test, compared to those not threatened by stereotypes, were more likely to use previously successful but presently inefficient or incorrect strategies. In Study 1, participants under stereotype threat also suppressed relevant stereotypes to the greatest degree, and their inflexible perseverance was predicted by the degree to which they suppressed these stereotypes. Implications for test performance and potential decision-making effects of stereotype threat are discussed.

© 2009 Elsevier Inc. All rights reserved.

While developing and following routines in thinking and persisting in previously successful strategies often serve us well, such routines and persistence implicitly rely on an unchanging world. However, change is inevitable and constant, and cognitively adapting to new contexts becomes necessary: test-takers find that approaches that succeeded on easy problems fail on difficult ones. Those who change jobs discover that their previous ways of thinking no longer suffice. People given novel problems learn that their old approaches do not lend themselves to success. And, academics who venture into an interdisciplinary world discover that they must adapt a new, more flexible way of thinking.

Flexibly changing with changing situations is important for adaptive decision-making (Payne, Bettman, & Johnson, 1993), and persevering in old strategies that are no longer optimal can interfere with success. Are there factors that interfere with our ability to do this? The central hypothesis of the present research is that the burden of negative stereotypes about one's group interferes with one's capacity to adapt to new situations. Specifically, we propose that stereotype threat, the threat of confirming a negative stereotype about one's group (Steele & Aronson, 1995), may induce a perseverant way of thinking in those who experience it, interfering with their ability to replace old strategies with more successful ones when the situation changes.

Stereotype threat

Stereotype threat theory proposes that the performance situation, infused with cues about the devaluation of one's social identity, creates an extra burden for individuals from stereotyped groups, undermining their performance in the stereotyped domain (Steele & Aronson, 1995). Research has demonstrated this impaired performance for many stigmatized groups in many domains: women in math and science domains (Ben-Zeev, Fein, & Inzlicht, 2005; Davies, Spencer, Quinn, & Gerhardtstein, 2002; Johns, Schmader, & Martens, 2005; O'Brien & Crandall, 2003; Pronin, Steele, & Ross, 2004; Spencer, Steele, & Quinn, 1999), African-Americans in intellectual domains (Aronson, Fried, & Good, 2002; Blascovich, Spencer, Quinn, & Steele, 2001), Latinos in intellectual domains (Gonzales, Blanton, & Williams, 2002), White males in athletics (Stone, 2002; Stone, Lynch, Sjomeling, & Darley, 1999), and children from lower socioeconomic status backgrounds in intellectual domains (Croizet & Claire, 1998).

However, while there is overwhelming evidence that stereotype threat interferes with performance and achievement, and while many emotional and cognitive mediators—anxiety (Bosson, Haymovitz, & Pinel, 2004; Spencer et al., 1999), arousal (Ben-Zeev et al., 2005; Blascovich et al., 2001), changing performance expectancies (Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003), working-memory interference (Schmader, 2002), and cognitive load (Croizet et al., 2004)—have been examined, little research has explored what changes in cognitive strategies and styles of thinking—independent of how it might affect performance—this threat induces. *Inflexible perseverance*, perseveration in strategies that were successful once but that are no longer

* Corresponding author.

E-mail address: pbangard@stanford.edu (P.B. Carr).

efficient, may be one such change that follows from stereotype threat.

Stereotype threat and inflexible perseverance

Flexibly abandoning old strategies and generating new ones requires many processes. If the context has changed and requires new strategies, the problem solver must notice the shift in context. In addition, the problem solver must have the cognitive resources to inhibit the previous response and develop a new response. Much research has shown that, as imperfect information processors, people are generally prone to inflexibility: for example, we apply non-relevant past experiences to present situations (Gilovich, 1981), we try to confirm pre-existing hypotheses (Nickerson, 1998), and we persist in mental sets (Luchins & Luchins, 1994). The research we review below suggests that stereotype threat may exacerbate these tendencies and contribute to inflexible perseverance.

Stress and arousal

Stereotype threat increases anxiety and arousal (Ben-Zeev et al., 2005; O'Brien & Crandall, 2003; Spencer et al., 1999) and triggers a physiological stress response (Blascovich et al., 2001; Murphy, Steele, & Gross, 2007). This high stress and arousal, which tends to narrow attention to more focal, stress-related cues (Easterbrook, 1959; Murphy et al., 2007; Wells & Matthews, 1994), may render a stereotype-threatened individual more likely to miss cues in the environment signaling that the testing situation has changed and a new strategy is more adaptive: as attention is drawn to cues signaling an identity threat, attention may be drawn away from the environment. Thus, those stressed and aroused by stereotype threat may be less likely to adapt a new strategy in a changed situation.

Ego depletion and working memory

Stereotype threat, associated with resource-demanding tasks such as stereotype suppression (Logel, Iserman, Davies, Quinn, & Spencer, 2008; Wenzlaff & Wegner, 2000), leads to ego depletion (Inzlicht, McKay, & Aronson, 2006) and working-memory depletion (Schmader & Johns, 2003). Ego-depletion and working-memory interference are states in which cognitive resources necessary for successful performance and generating new solutions (Crinella & Yu, 1999) are drained. Furthermore, working-memory interference leads to difficulty inhibiting prepotent responses (Kane, Bleckley, Conway, & Engle, 2001; Kane & Engle, 2003). Therefore, individuals experiencing stereotype threat may not have the resources to generate new responses and inhibit previous ones.

Prevention focus

Being primed with stigmatized identities (Oyserman, Uskul, Yoder, Nesse, & Williams, 2007) and experiencing stereotype threat (Seibt & Forster, 2004) have been found to increase prevention focus, a state associated with concerns about responsibility and safety and cued when safety needs are made salient (Higgins, 1997). Prevention focus creates a risk-averse style of behavior that is not explorative and is committed to avoiding errors (Crowe & Higgins, 1997; Higgins, 1996). Stereotype threat cuing prevention focus, therefore, may lead to reluctance to deviate from what is known to be correct and to difficulty generating novel approaches. In support of this hypothesis, Seibt and Forster (2004) found that inducing prevention focus through negative stereotypes leads to decreased creativity.

The present research

The above research leads us to hypothesize that threat may produce inflexible perseverance. In two studies, using different measures of perseverance, we test whether stereotype threat leads its target to persevere in previously successful strategies that are no longer optimal or correct. In Study 1, we also test whether the unique experience of stereotype threat—contending with negative stereotypes—fosters inflexible perseverance. In Study 2, we test whether stereotype threat leading to concern about making mistakes (one aspect of prevention focus) fosters inflexible perseverance.

Study 1

Method

Participants and design

Men ($N = 39$) and women ($N = 28$) were randomly assigned to a diagnostic math test (20 men, 14 women) or a non-diagnostic puzzle (19 men, 14 women) condition.¹ All participants were undergraduates at Stanford University and participated in exchange for \$10 or partial course credit.

Participants completed a modified Luchins' Water-Jar task (Luchins, 1942) to measure perseverance followed by a lexical decision task to measure stereotype suppression. Mood, performance expectancy, liking, and motivation measures followed. Participants completed domain-identification measures at pre-testing, several weeks before the experiment. Only those moderate to high (score at 5 or greater on a 7-point scale for each of our two domain-identification measures) in identification were invited for participation because stereotype threat effects are weak or non-existent among those who do not care about the domain (Aronson et al., 1999).

Threat manipulation

All participants took the same test (the Water-Jar task). Participants in the diagnostic condition read that the test was developed by a prestigious organization and was a highly validated measure of mathematical ability. Participants indicated their gender before they started the test.

Participants in the non-diagnostic condition read that the task was a puzzle-solving exercise, developed by students and used to gain insight into how people solve puzzles. They were told it produced no gender differences. These participants indicated their gender after the task.

Measures

Perseverance: Water-Jar task. The Water-Jar task has been widely used to study perseveration in mental set and thus lends itself well to our study of inflexible perseverance (Luchins & Luchins, 1994; Schultz & Searleman, 2002). In our computerized version of the Water-Jar task, participants were told they had a barrel full of water and a pot which they needed to fill with a desired quantity of water. To get the desired quantity of water, they could use three jars (A, B, C) each of which could hold a certain quantity of water. Participants were instructed to generate the simplest strategy as quickly as possible for getting the desired quantity and to not use any jar more than three times in a problem. A strategy of $2A + 2C$ would mean that the pot should be filled using Jar A twice and using Jar C twice to get the desired quantity.

After an example problem, participants were given three practice problems that had novel solutions that did not overlap with

¹ Originally, 44 men and 44 women participated. While we invited only those who were math-identified to participate, through sharing of passwords, those low in math identification were able to participate. We excluded participants with low math identification.

test trial solutions. During the test trials, participants were presented with six “set-establishing” trials in which the fairly complex strategy of B–A–2C was the only successful solution. Five “Critical” trials in which the old, complex strategy still reached the solution, but there was also a simpler, more efficient way to do so (e.g., A–C) followed. Because perseverance in mental set occurs at high rates in non-threatening conditions (see Luchins & Luchins, 1994; Schultz & Searleman, 2002), we modified the test to tap extreme perseverance. The last three critical trials were very obvious about the simple solution: an obvious trial may ask for five cups when Jar A holds 20 and Jar C holds 15; a non-obvious trial may ask for 14 cups when Jar A holds 32 and Jar C holds 18.

Stereotype suppression: lexical decision task. Research has found that stereotype threat leads to activation of the relevant stereotype before a threatening test begins (Davies et al., 2002). However, once the threatening task begins, those targeted by the stereotype engage in stereotype suppression, exerting effort to push negative stereotypes out of conscious awareness. Research using a lexical decision task found that those who were stereotype-threatened, compared to those not threatened, responded *slower* to stereotypic words during the threatening task, indicating they were exerting effort to suppress the relevant stereotype (Logel et al., 2008). We presented the lexical decision task in our study at the very end of our measure and suggest it is a continuation of the task. Thus, we expected evidence for suppression under stereotype threat.

The lexical decision task asks participants to judge whether a string of letters is a word or a non-word. A list of words associated with negative stereotypes about women in math and words not associated with the stereotype but matched for length and frequency of usage (using norms established by Kucera & Francis, 1967) were rated by 50 undergraduates for their relevance to stereotypes about women in math. The 10 words that were rated as most associated were chosen to be our “stereotypic” words (inferior, weak, slow, dumb, stupid, woman, female, lesser, deficient, inadequate). The 10 “non-stereotypic” words (disagree, curt, nose, debt, firing, voice, wooden, denial, animosity, reflection) were rated as low in association. We also generated 10 non-words matched for length.

Participants were presented with the stimuli in randomized order on a computer. Participants saw the following order of events: a fixation cross for 300 ms, a blank screen for 500 ms, the target array of letters until a word or non-word decision was made, and a blank screen for 1000 ms.

Performance expectancy. Participants were given our stereotype suppression measure after the perseverance measure to ensure we did not prime perseverance. We asked participants to rate on a scale of 1 (not well at all) to 7 (extremely well) how well they thought they performed on the Water-Jar task to make certain that perceived performance did not predict stereotype suppression.

Math identification. Participants rated on a scale of 1 (not at all true of me) to 7 (very true of me) two statements assessing math identification: “I am good at math” and “It is important to me that I do well at math”.

Other measures. To rule out alternative explanations, we included a mood measure (PANAS; Watson, Clark, & Tellegen, 1988) and measured participants’ liking for and motivation to do well on the Water-Jar task.

Results and summary

One female participant (in the non-diagnostic condition) was excluded for not following instructions.

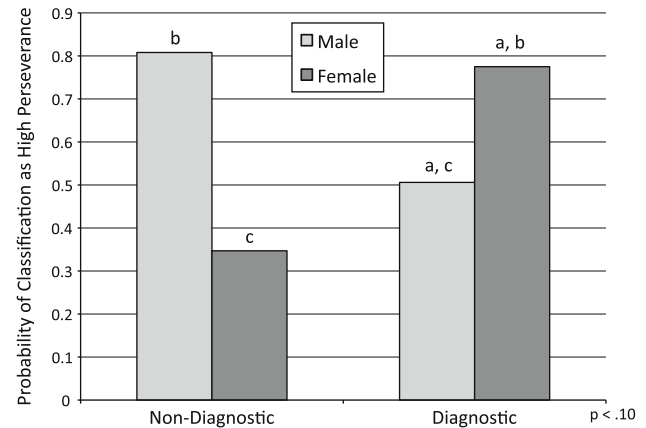


Fig. 1. Probability of being classified as high in perseverance by condition, derived from logistic regression.

Set establishment

All participants established set, using B–A–2C on at least four of the six set-establishing trials.

Does identity threat lead to inflexible perseverance?

We hypothesized that women in the diagnostic condition would be most perseverant. Diagnosticity should not have the same effect on men, for whom the math test is not identity threatening. The Water-Jar task was designed so that the last three critical trials made the simple solution very obvious. We did so to make this measure more sensitive to condition effects as people in general have a strong tendency to maintain set.²

To tap perseverance, we created the variable, “perseverance level.” Participants who used the set solution on at least one of these last three Critical trials—trials in which the easier solution is obvious—were categorized as *high* in perseverance. Participants who used the set solution on none of the last three critical trials were classified as *low* in perseverance.

A logistic regression, controlling for math identification, revealed that women ($\beta = -2.07$, odds-ratio = .13, $p < .05$) and those in the diagnostic condition ($\beta = -1.4$, odds-ratio = .243, $p = .078$), in general, were less likely to show high perseverance. These main effects were qualified by a gender by condition interaction, ($\beta = 3.28$, odds-ratio = 26.44, $p < .05$), with those in the stereotype threat cell (women in the diagnostic condition) more likely to be high in perseverance.³ Entering the interaction term in the model made it a significantly better predictor of level of perseverance ($\Delta\chi^2 = 7.38$, $p < .01$).

The estimated probabilities from the logistic regression (see Fig. 1) revealed that women in the diagnostic condition were more likely (probability = .774) to be high in perseverance than women in the non-diagnostic condition (probability = .347, $\Delta\chi^2 = 4.05$, $p < .05$; $\beta = .93$, odds-ratio = 2.54, $p < .05$). The means also revealed a trend in the opposite direction for men: men in the diagnostic condition

² Including the last three trials with obvious solutions allowed us to analyze the data in a way that minimizes variance in solutions created by differences in individual problems. Our concern about variance created by differences in individual problems was borne out. Our logistic regression analysis using a dichotomous variable yielded a significant interaction. We also analyzed the data by examining the effects of gender and condition on the total number of critical trials solved using the set solution. A 2 (male or female) \times 2 (diagnostic or non-diagnostic) ANCOVA, controlling for math identification, revealed the same pattern of results as the logistic regression. However, the gender by condition interaction was only marginally significant, $F(1,61) = 3.28$, $p = .075$. Women in the diagnostic condition tended to solve more problems with the set solution ($M_{adj} = 2.59$) than women who were not threatened ($M_{adj} = 1.89$, $t(61) = 1.24$, $p > .2$). Men in the diagnostic condition ($M_{adj} = 2.15$) tended to solve fewer problems with the set solution than men in the non-diagnostic condition ($M_{adj} = 2.80$, $t(61) = 1.74$, $p > .15$).

³ The gender by condition interaction term essentially contrasts the stereotype threat cell (coded as 1) with all other cells (coded as 0).

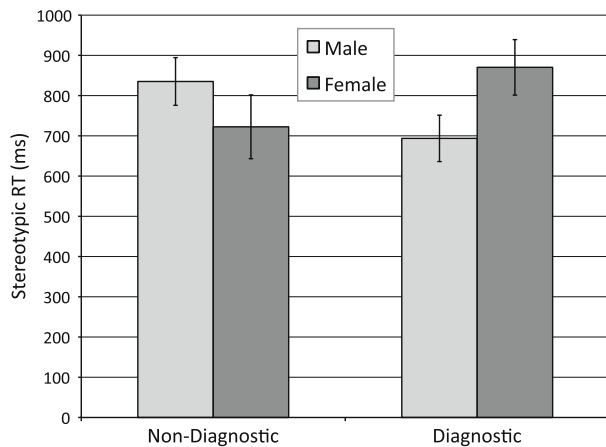


Fig. 2. Reaction time to lexically classify words related to negative stereotypes about women in math (controlling for reaction time to non-stereotypic words) by condition.

were less likely to be high in perseverance (*probability* = .506) than men in the non-diagnostic condition (*probability* = .808, $\Delta\chi^2 = 3.15$, $p = .076$; $\beta = -.65$, *odds-ratio* = .524, $p = .084$).

Does identity threat lead to stereotype suppression?

Drawing from Logel et al. (2008), when participants in one condition take a significantly longer amount of time to decide if a stereotypic word is a word, we can infer that they are trying to suppress the stereotype to a greater degree.

All trials in which participants made an incorrect response or had reaction times less than 200 ms or more than 2000 ms (1.2% of trials) were excluded. A 2 (male or female) \times 2 (diagnostic or non-diagnostic) ANCOVA examining reaction time to classify stereotypic words, with reaction time for the classification of non-stereotypic words as a covariate, revealed a gender by condition interaction, $F(1,61) = 4.59$, $p < .05$.⁴ There were no main effects for condition or gender, $F_s < 1$.

As Fig. 2 reveals, women in the diagnostic condition took significantly longer to classify stereotypic words ($M_{adj} = 870.3$ ms) than men in the diagnostic condition ($M_{adj} = 693.7$ ms, $t(61) = 2.14$, $p < .05$), evidencing effort to suppress negative stereotypes under threat. They also tended to take longer to classify stereotypic words than women in the non-diagnostic condition ($M_{adj} = 722.4$ ms, $t(61) = 1.79$, $p = .07$) but not men in the non-diagnostic condition ($M_{adj} = 835.1$ ms, $t < 1$).

Does stereotype suppression predict perseverance?

We examined whether the reaction time to stereotypic words significantly predicted classification as high or low in perseverance. A logistic regression revealed that time taken to lexically classify stereotypic words did not significantly predict classification as high or low in perseverance ($\Delta\chi^2 = 1.17$, $p = .28$; $\beta = .001$, *odds-ratio* = 1.0, $p > .25$).

However, we did not expect the effect of stereotype suppression to evidence itself across all groups: those in the non-diagnostic condition and men, in general, may not be much affected by suppressing words related to incompetence. Thus, we conducted the previous analysis separately for each of the four cells in our study. This analysis revealed that, for women in the diagnostic condition, time taken to lexically classify a stereotypic word was a significant predictor of classification as high or low in perseverance ($\Delta\chi^2 = 6.0$, $p < .05$; $\beta = .01$, *odds-ratio* = 1.01, $p = .15$). Women who took longer to react to the stereotypic words (women who were suppressing

the stereotype the most) were more likely to be classified as high in perseverance. However, men in the diagnostic condition showed the opposite pattern: men who were faster to react to the stereotypic words (presumably men who were activating the concept of incompetence) were marginally more likely to be classified as high in perseverance ($\Delta\chi^2 = 3.522$, $p = .06$; $\beta = -.01$, *odds-ratio* = .99, $p = .13$). In the non-diagnostic condition perseverance was not correlated with time to lexically classify stereotypic words ($p_s > .25$).

Other analyses

Participants' mood, performance estimates, liking, and motivation did not predict perseverance level or reaction time on the lexical decision task, $F_s < 1$.

Summary

In Study 1, we find that stereotype threat leads to greater probability of engaging in inflexible perseverance and increased effort to suppress stereotypes. Illuminating that the process is indeed driven by concerns about negative stereotypes, we find that perseverance by stereotype-threatened participants is correlated with their stereotype suppression.

Study 2

In Study 2, we conceptually replicate the first study using a different measure of perseverance, the Wisconsin Card Sorting Test (WCST), an even more widely used measure of perseverance.

A weakness of Study 1 was that the condition effect, though larger for women, was almost as large for men, albeit in the opposite direction. We posit that this finding may be akin to the effects of stereotype lift, whereby the non-threatened group performs better in the diagnostic condition compared to the non-diagnostic condition, and that flexibility may be a mechanism through which stereotype lift operates. While, in general, stereotype threat effects are much larger than those of stereotype lift, individual studies have occasionally found stereotype lift effects as large as those of stereotype threat (see Walton & Cohen, 2003). We expect a second study to show effects of stereotype threat—gender differences in the diagnostic condition and condition differences for women. However, we do not expect to replicate the relatively rare magnitude of stereotype lift effects from Study 1. These findings would clarify that the inflexible perseverance effect is driven by stereotype threat and individuals who are threatened.

In the second study, we also do not measure stereotype suppression, which did not mediate our effect in all cells in the first study, but instead examine another potential mediator—activation of thoughts related to making mistakes. Being concerned with making mistakes might lead participants to avoid giving up a strategy that has had a history of success. Stereotype threat has been found to increase prevention focus and one aspect of prevention focus is concern with mistakes and errors. Thus, in Study 2, we measure activation of words related to making mistakes as a potential mediator.

Method

Participants and design

As in Study 1, moderately to highly math-identified men ($N = 32$) and women ($N = 31$) were randomly assigned to a diagnostic math/spatial ability test (17 men, 13 women) or non-diagnostic puzzle (15 men, 18 women) condition.⁵ All participants were

⁴ Excluding stereotypic words "female" or "woman" from the analysis did not change the significance or pattern of results.

⁵ Twenty-one participants (11 men, 10 women) who participated from an introductory psychology class after hearing a lecture on stereotype threat were excluded.

undergraduates at Stanford University and participated in exchange for \$10 or partial course credit. Participants completed the WCST to measure perseverance, followed by a lexical decision task to measure activation of mistake-related words. Mood, performance expectancy, liking, and motivation measures followed.

Threat manipulation

Threat was manipulated as in Study 1, with the only difference being that the diagnostic test (the WCST) was described as a test of spatial and analytical ability predictive of success in mathematical and spatial fields.

Measures

Perseverance: Wisconsin card sorting test. The WCST (Berg, 1948) is a widely used test of the ability to flexibly shift in the face of changing situations and contingencies. In our modified and computerized version of the WCST, the participant is presented with a target card with a picture on it (e.g., two yellow stars) and the participant is asked to match this card to one of four cards that vary in the number, color, and shape of objects depicted (specifically, one red circle, two green stars, three blue squares, and four yellow pluses). The sorting rule (match by number, color or shape) is not revealed to the participant, but is to be deciphered through accuracy feedback. The rule changes after 10 correct sorts, and the participant must deduce and sort by the new rule. Participants receive between 84 and 128 trials and between 5 and 9 shifts in the sorting rule. The number of errors and perseverative errors (errors that match on the basis of the previous sorting rule after a rule shift) are recorded.

"Mistake" activation: lexical decision task. The lexical decision task to measure activation of thoughts related to mistakes was structured as in Study 1, except the target words were about "mistakes" (mistake, error, inaccurate, blunder, incorrect, wrong, imperfect, miscalculation, flaw, careless) and the matched control words were unrelated to that construct (dispute, dirty, hemorrhage, hostage, poisonous, piece, repulsive, culture, uncontrollable, wail, withdraw). The relatedness of these words to the construct was confirmed through pre-testing with 50 undergraduates.

Other measures. As in Study 1, we measured mood, liking, motivation, and performance expectancy.

Results and summary

Does stereotype threat lead to inflexible perseverance?

We expected women in the diagnostic condition to be most perseverant on the WCST. We calculated a mean perseverance score for each participant by dividing the number of perseverative errors the participant made by the number of sorting rule shifts experienced by the participant. This score tells us on average how many trials a participant perseverated in the previous sorting rule after the sorting rule had changed.

An ANOVA revealed the expected gender by condition interaction, $F(1,59) = 6.05, p < .05$. There were no significant main effects. Women in the diagnostic condition had the highest perseverance scores ($M = 3.25$), significantly higher than those of women in the non-diagnostic condition ($M = 1.75, t(59) = 2.42, p < .05$) and those of men in the diagnostic condition ($M = 1.78, t(59) = 2.36, p < .05$). Men did not significantly differ by condition ($M = 2.40$ for men in the non-diagnostic cell) and there was no significant gender difference in the non-diagnostic condition, $ps > .2$ (see Fig. 3).

Does stereotype threat lead to thinking about mistakes? Does thinking about mistakes predict perseverance?

If concern with making mistakes was driving our effect, we might expect women in the diagnostic cell to show the greatest

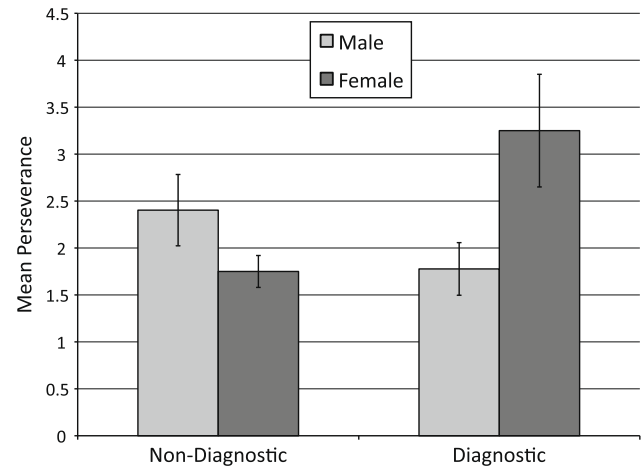


Fig. 3. Mean perseverance score (perseverative errors/rule changes) by condition.

activation of "mistake" words. We would also expect such activation to correlate with perseverance. However, we failed to find support for this account. After excluding all incorrect trials and trials with reaction times below 200 ms and above 2000 ms (1% of trials), an ANCOVA, with the mean response time for classification of mistake-unrelated words as a covariate, revealed no significant gender by condition interaction or main effects for classification of mistake-related words, $F_s < 1$. Correlations of reaction time with perseverance scores were not significant across participants or within-cell, $ps > .2$.

Other analyses

Participants' mood, performance estimates, liking, and motivation did not predict perseverance or reaction time, $F_s < 1$.

Summary

Further supporting our hypothesis, we found in a different paradigm that stereotype threat causes its targets to be perseverant. However, we failed to find support for the prediction that thinking about mistakes drives the perseverance effect.

Discussion

How do cognitive strategies and behaviors change pursuant to threat? In two studies, using two different measures of perseverance, we show that stereotype threat pressures people toward inflexible perseverance in a problem-solving strategy that is no longer optimal or correct. Moreover, in the first study, we found that the degree to which stereotype-threatened individuals suppressed stereotypes was associated with their likelihood of being high in perseverance.

To test performance

Being stereotype-threatened leads to impaired test performance (Steele, 1997). The current research suggests that one reason test performance suffers may be that stereotype threat leads to inflexibility. Flexibility, which we define as a cognitive state of mind that allows for adaptation and change, is important for successful decision-making (Payne et al., 1993). Flexibility, we imagine, is also important for doing well on standardized tests which are often designed to prevent mastery by previous learning: from question to question, new strategies and approaches must be adapted. It is our hypothesis that inflexible perseverance would impair performance on such tests and might mediate the effects of threat on test performance. Flexibility may also be a potential

behavioral mediator of stereotype lift effects on test performance. In Study 1, we found that men tended towards greater flexibility in the diagnostic condition compared to the non-diagnostic condition. Study 2 did not replicate the magnitude of that effect (as is to be expected since stereotype lift effects are smaller than those of stereotype threat) but the means followed the same pattern.

However, further research is needed to demonstrate that inflexibility mediates the effect of threat and lift on test performance. Discovering behavioral mediators like inflexible perseverance, though, may be particularly useful in designing interventions to alleviate the burden faced by those who are under stereotype threat: interventions could be designed to target specific, tangible behavior.

Why does threat cause inflexible perseverance?

In our first study, we found significantly increased stereotype suppression under stereotype threat. We also found that suppression by stereotype-threatened participants was correlated with degree of inflexibility. However, since we did not find mediation, we examined another potential mediator in Study 2—concern with mistakes. Our second study failed to find an effect of threat on thinking about mistakes or any evidence for such thinking predicting inflexibility. Because we did not find complete mediation in our first study, this research is inconclusive on what drives the effects of threat on perseverance. However, it is suggestive that stereotype suppression rather than being concerned with mistakes may be the most fruitful mediator for future research to consider. Such research could use different measures of stereotype suppression (for example, post-suppressional rebound) or examine possible effects of suppression as mediators. One such effect may be working-memory depletion: exerting effort to push stereotypes out of awareness, as women under stereotype threat seem to be doing, depletes the limited ego and working-memory resources necessary for generating new solutions and noticing context changes. Future research should examine depletion of resources as a more proximal mediator of stereotype threat effects on inflexibility to uncover the mechanism driving the inflexible perseverance effect.

Beyond test performance

While much research on stereotype threat is focused on test performance, the current finding that threat results in inflexible perseverance suggests that threat may affect much more than test performance. Inflexible mindsets are implicated in many decision-making errors such as confirmatory hypothesis testing (Nickerson, 1998), groupthink (Janis, 1972), status-quo bias (Kahneman, Knetsch, & Thaler, 1991), and escalation of commitment (Staw, 1976). And, identity threats exist beyond the classroom and the test-taking situation. Women in a technology firm and African-American CEOs are also faced with threats based on negatively stereotyped identities. Chronically stigmatized minorities, facing an ongoing and almost ever-present *social identity threat* spanning many situations (Steele, Spencer, & Aronson, 2002), may be constantly burdened by identity threat. Thus, stereotype threat, spanning many situations and resulting in a perseverant mindset, may negatively impact decision-making and job performance.

Our research is limited in that it examines inflexibility in a single domain (a testing domain). It is important for future research to explore how threat affects mindsets and strategies in other domains. In the domain of decision-making, for example, the current research would suggest that threat may result in greater susceptibility to the sunk-cost fallacy or escalation of commitment effects: individuals with a perseverant mindset may be more likely to persist in an initially chosen way, even when that strategy is failing or costly. Our findings highlight the need to examine the effects of

threat beyond test performance and to alleviate the burden of threat in domains that extend beyond the classroom door.

Acknowledgments

We thank Tito Balsamo and Nathan Grossman for their assistance with data collection.

References

- Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, 38(2), 113–125.
- Aronson, J., Justina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When White men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, 35(1), 29–46.
- Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology*, 41(2), 174–181.
- Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology*, 39(1), 15–22.
- Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science*, 12(3), 225–229.
- Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of Experimental Social Psychology*, 40(2), 247–255.
- Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology*, 33(2), 267–285.
- Crinella, F. M., & Yu, J. (1999). Brain mechanisms and intelligence: Psychometric “g” and executive function. *Intelligence*, 27(4), 299–327.
- Croizet, J., & Claire, T. (1998). Extending the concept of stereotype and threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin*, 24(6), 588–594.
- Croizet, J., Despres, G., Gauzins, M., Huguet, P., Leyens, J., & Moet, A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin*, 30(6), 721–731.
- Crowe, E., & Higgins, E. T. (1997). Regulatory focus and strategic inclinations: Promotion and prevention in decision making. *Organizational Behavior and Human Decision Processes*, 69, 117–132.
- Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin*, 28(12), 1615–1628.
- Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, 66(3), 183–201.
- Gilovich, T. (1981). Seeing the past in the present: The effect of associations to familiar events on judgments and decisions. *Journal of Personality and Social Psychology*, 40, 797–808.
- Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and Social Psychology Bulletin*, 28(5), 659–670.
- Higgins, E. T. (1996). *Emotional experiences: The pains and pleasures of distinct regulatory systems*. Hillsdale, NJ: Lawrence Erlbaum.
- Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist*, 52(12), 1280–1300.
- Inzlicht, M., McKay, L., & Aronson, J. (2006). Stigma as ego depletion: How being the target of prejudice affects self-control. *Psychological Science*, 17(3), 262–269.
- Janis, I. (1972). *Victims of groupthink*. Boston, MA: Houghton Mifflin Company.
- Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. *Psychological Science*, 16(3), 175–179.
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The Journal of Economic Perspectives*, 5(1), 193–206.
- Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, 130(2), 169–183.
- Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, 132(1), 47–70.
- Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Logel, C., Iserman, E. C., Davies, P. G., Quinn, D. M., & Spencer, S. J. (2008). The perils of double consciousness: The role of thought suppression in stereotype threat. *Journal of Experimental Social Psychology*. doi:10.1016/j.jesp.2008.07.016.
- Luchins, A. S. (1942). Mechanization in problem-solving: The effect of Einstellung. *Psychological Monographs*, 54 (6, Serial No. 248).

- Luchins, A. S., & Luchins, E. H. (1994). The water jar experiments and Einstellung effects: I. Early history and surveys of textbook citations. *Gestalt Theory*, 16(2), 101–121.
- Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science*, 18(10), 879–885.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220.
- O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin*, 29(6), 782–789.
- Oyserman, D., Uskul, A. K., Yoder, N., Nesse, R. M., & Williams, D. R. (2007). Unfair treatment and self-regulatory focus. *Journal of Experimental Social Psychology*, 43(3), 505–512.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision-maker*. New York, NY: Cambridge University Press.
- Pronin, E., Steele, C. M., & Ross, L. (2004). Identity bifurcation in response to stereotype threat: Women and mathematics. *Journal of Experimental Social Psychology*, 40(2), 152–168.
- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, 38(2), 194–201.
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85(3), 440–452.
- Schultz, W. P., & Searleman, A. (2002). Rigidity of thought and behavior: 100 years of research. *Genetic, Social, and General Psychology Monographs*, 128(2), 165–207.
- 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(1), 38–56.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1), 4–28.
- Staw, B. M. (1976). Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action. *Organizational Behavior and Human Decision Processes*, 16(1), 27–44.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52(6), 613–629.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797–811.
- Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pp. 379–440). San Diego, CA, US: Academic Press.
- Stone, J. (2002). Battling doubt by avoiding practice: The effects of stereotype threat on self-handicapping in white athletes. *Personality and Social Psychology Bulletin*, 28(12), 1667–1678.
- Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. *Journal of Personality and Social Psychology*, 77(6), 1213–1227.
- Walton, G., & Cohen, G. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, 39(5), 456–467.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070.
- Wells, A., & Matthews, G. (1994). *Attention and emotion: A clinical perspective*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.
- Wenzlaff, R. M., & Wegner, D. M. (2000). Thought suppression. *Annual Review of Psychology*, 51, 59–91.