

Running head: Does Stereotype threat Require Stereotypes?

Does Stereotype Threat Require Stereotypes?

Sarah LeSturgeon & David Phelps

Hanover College

Abstract

When males take a test assessing verbal skills, unlike females, they are at risk of confirming the widely-shared negative stereotype targeted at their group namely, that females outperform males in verbal ability. Consistent with previous research, this risk is termed stereotype threat and has been shown to negatively impact performance. Would performance be as negatively impacted, however, if males were still told that their gender performed worse on the same academic task, but the task was framed as a test of 'analogical-reasoning,' which does not evoke any widely-shared stereotypes? Furthermore, would males do better if they faced a negative expectation that was not tied to their social group (gender) nor fit to a widely-held belief (females outperform males on verbal tasks)? To explore this question, participants were randomly assigned to receive either negative or positive expectations that (a) targeted gender, a meaningful social group, or (b) targeted a meaningful social group fit to a widely-shared stereotype or (c) targeted neither of these two elements. Results were accounted for by comparing the mean scores (the average number of correct answers) on a test of verbal ability. Using a between subjects, one-way ANOVA, a trend emerged revealing that when male and female participants experienced a threat towards their meaningful social group (gender), they tended to perform worse than participants presented with simply a negative expectation not tied to a meaningful social group. Although their scores were not significantly different, it is evident that stereotype threat may have a greater negative impact when threat is tied to one's social identity.

Does Stereotype Threat Require Stereotypes?

Stereotype threat refers to the threat people feel when they are at risk of confirming a negative stereotype targeted at their group. The term was coined by Steele and Aronson (1995) to describe why Black students perform poorly on academic tests compared to White students when their race is emphasized, but perform equivalent to or even better than White students when race is not made salient. Although much debate remains over the exact mechanism at play in reducing performance, Steele and Aronson (1995) speculate that anxiety is produced when negatively-stereotyped participants unlike regular participants, are at risk of being judged by the negative stereotype. This anxiety, in turn, interferes with the participant's ability to perform optimally on the test. In effect, Stereotype threat is a self-fulfilling prophecy: One is apprehensive about potentially confirming a negative-stereotype, the apprehension lowers one's score, and the reduced score confirms the stereotype.

Additional research has demonstrated Stereotype Threat in the presence of a wide variety of academic stereotypes. Spencer, Steele, and Quinn (1999) investigated gender performance on math ability whereby females are negatively stereotyped to perform worse than males. Spencer and colleagues created a telling manipulation by creating an expectation that females would perform just as well as males in one condition, whereas the other condition's expectations were fit to the widely-held stereotype that females perform worse than males. The results demonstrated that females' performance fit the given expectation for each condition.

A similar phenomenon termed expectancy effects has been demonstrated in research on teacher's expectations such that students perform at levels that match said expectations

(Brophy & Good, 1974; Rosenthal & Jacobson, 1968). Expectancy effects are also described as a self-fulfilling prophecy (Merton, 1948) whereby negative expectations generate negative performance. What if Spencer et al. (1999) changed the nature of the expectations given to participants such that some participants received negative expectations that did not fit a widely-shared stereotype (males would be outperformed by females on math)? What if other participants received negative expectations that did not target their gender but some fabricated grouping variable (persons who have never participated in a book club are expected to do poorly on this test)? Would the threat still exist? More provocatively, if the threat still exists could it be explained by simple expectancy effects? If, however, the threat is greater as the expectation is tied to a meaningful social group and fit to a widely-held stereotype it is still worth clarifying how much of an impact each element contributes to stereotype threat. A study assessing the differences in impact on performance between negative-expectations, negative-expectations tied to a meaningful social group, and negative expectations tied to a meaningful social group and fit to a widely-shared belief can help begin to clarify the role of each element in stereotype threat.

To create an experiment that distinguishes between expectancy effects that are tied to meaningful groups (or not) and widely-shared (or not), we needed a measure of performance that could be portrayed as reflecting either a widely-shared stereotype or an arbitrary expectation. We felt that stereotypes concerning math performance were too widely-shared and math items themselves too easily identified as measuring mathematical ability to suit our purposes. We settled instead on analogies, which could be labeled as a measure of “verbal ability” and linked to the expectation that males are outperformed by females on tests

measuring verbal abilities (Maccoby & Jacklin, 1974). Alternatively, the task could be labeled as a measure of “analogical reasoning,” for which there is no widely-shared expectation of gender differences in performance. In the first case, expectations for performance would be both tied to a meaningful group (gender) and widely-shared (reflecting the stereotype that women outperform men on verbal tasks). To create a condition in which the expectation is tied to a meaningful group but is not widely shared, we told some participants that their gender had performed either well or poorly on the upcoming analogical reasoning task. To create a condition in which the expectation is neither widely shared nor tied to a meaningful group, we prompted participants with a dot-estimation task, automatically assigned them to the dot over-estimator group, and then gave told them that their group had performed either well or poorly (compared to dot under-estimators) on the upcoming analogical reasoning task.

Researchers of stereotype threat calculate the effect size of the threat by comparing the performance scores of the targeted participants under threat to the performance scores of the targeted participants not under threat. If males are negatively-stereotyped to be outperformed by females on tests measuring verbal abilities (Maccoby & Jacklin, 1974), the difference between the score of the threatened males compared to the non-threatened males (who are told the task measures “analogical reasoning” and given high performance expectations) constitutes the effect size of threat. If stereotype threat is explainable by expectancy effects alone, the effect size of threat for a group facing a negative-expectation should be similar to the effect size of threat for a group suffering from a negative-expectation tied to their social identity and fit to a widely-held belief.

Lee and Kent (2005) performed a literature review of teacher expectations and found that the self-fulfilling prophecy of performance matching expectation occurs amongst students from stigmatized social groups. This leads the researchers to hypothesize that the effect size of threat will be significantly greater for participants facing a negative expectation tied to their social identity and fit to a widely-shared belief than participants undergoing a simple negative-expectation.

Steele and Aronson (1995) speculate that the mechanism behind stereotype threat is not internalization. They argue that persons do not even need to believe in the stereotype to be affected by it, they just need to be aware of its presence. This leads the researchers to further hypothesize that a stereotype does not need to be widely-shared in order to be effective as long as it is tied to one's social identity. That is, the researchers expect the effect size of a negative expectation tied to social identity to be similar to a negative expectation tied to social identity and fit to a widely-shared belief. Together these hypotheses recognize that although all elements may contribute to stereotype threat, social identity may be the most threatening aspect of stereotypes altogether.

Method

Participants

Participants were obtained through a popular webpage for on-line psychological experiments (Krantz, 2008). The total number of participants was 124, but six were removed due to incomplete data such as failing to indicate their gender. Therefore, data was collected for 118 participants, 32 of which were males. The sample included 96 Caucasian subjects and 22 who were not Caucasian. They ranged in age from 18-59 with a mean age of 38.5.

Materials and Procedure

Upon selecting the study, all participants were randomly assigned to one of five conditions and then presented with an informed consent briefly describing the purpose of the experiment, indicating the minimum age for inclusion, as well as listing contact information. Following the informed consent, participants within all conditions were directed to a dot estimation task, which was used to create a meaningless group designation (Tajfel, Billig, Bundy, & Flament, 1971) for participants in two of the five conditions (discussed below). Four consecutive web-pages contained one task per page. Each page showed a box containing a large number of dots along with five multiple choice possibilities which indicated different numbers (160, 200, 240, 280, and 300); each meant to represent the total amount of dots in the adjacent box. Participants were instructed to estimate the quantity of dots shown by selecting one of the given choices. This process was repeated four times. Completion of the final dot estimation task led participants to a new web-page. At the top of this page was a manufactured manipulation graph which varied by the condition to which the participant was assigned. Through the manipulation, participants were given an expectation for their future performance that was (1) either threatening or non-threatening,(2) either tied to a real group such as gender or to a meaningless group, and (3) either tied to a widely-held stereotype linked to a real group or not. These manipulations produced the 2 x 3 design depicted in Table 1 and explained below.

Table 1: *Overview of Elements of Manipulation across Conditions*

	Expectations Alone	Expectations for Real Group	Expectations for Real Group with Widely Shared Stereotype
Threatened	Dot overestimators told that overestimators do poorly	Told that upcoming task is analogical and their gender does poorly	Men told that upcoming task is verbal and that men do poorly
Not Threatened	Dot overestimators told that overestimators do well	Told that upcoming task is analogical and their gender does well	Women told that upcoming task is verbal and that women do well

In the Expectations Alone condition, participants were given expectations for their performance on an upcoming analogy test. In order to ensure that these expectations were not related to anything the participant may have previously experienced through stereotypes, we gave them predictions based on the previous dot estimation task. The estimation task was a superficial task whose purpose was solely to be used as a perceived predictor of ability to solve analogies even though this correlation was completely manufactured. To create the illusion of inclusion in a false group, participants were assigned to a superficial group, “dot over-estimators,” which, to participants, appeared to be a result of their performance on the dot estimation task. The purpose of assigning them to this group after completing the dot estimation task was to secure identification with a meaningless group, dot over-estimators. [keep with next paragraph] Participants were shown a manipulation graph which depicted the past performance of dot over-estimators and dot under-estimators on the upcoming analogies task. Participants randomly assigned to the threatened condition saw that over-estimators historically performed worse on the task, while those assigned to the non-threatened condition saw that over-estimators historically performed better. Analogical reasoning was a phrase used to describe the upcoming test because calling it a test of verbal ability may have elicited confounding stereotypes associated with verbal tests, such that women are thought to outperform men on tests of verbal ability. Thus, our intention in Expectation Alone condition was to give participants an expectation for their performance that was based only on their membership in a meaningless group. In order to help ensure that participants did, in fact, observe the manipulation graph, they were asked to check a box next to the graph which would give the researchers permission to use their data to update the graph.

Under the graph was a block of 28 analogies obtained from a GRE test preparation book. The directions asked participants to select an answer from five pairs of words that best expressed the relationship expressed in a given pair. For example, “apple is to fruit” might be the given analogy and the participants, in order to answer correctly, would select “as carrot is to vegetable.” The relationship

between an apple and fruit is that an apple is an instance of the category *fruit* and, likewise, a carrot is an instance of the category *vegetable*.

Following the analogies task, participants filled out the *Identification with School Questionnaire*, developed by Voelkl (1996). This questionnaire was used to measure the extent to which a participant identified with school and its sub-domains. Questions on the scale assess sub-domains of school such as belonging (“I feel comfortable when I am in school, like I belong there,” “Teachers don’t care about me”) or the degree to which the participant values school (“School is important in life,” “The things we do in class are useless” (reversed)). After reading each statement, participants rated the statements on a scale of 1 (strongly disagree) to 5 (strongly agree). The *Identification with School Questionnaire* has been found to be reliable with an alpha of .78 (Osborne & Walker, 2006). Finally, participants filled out a short demographic survey which asked for their age, ethnicity, and current location. Once all sections of the study were completed, participants submitted their data and were presented with an online debriefing form.

Participants in the other conditions also proceeded through the informed consent and dot estimation tasks, were given a graph communicating their expected performance, completed the analogies task, Identification with School Questionnaire, and demographics questions, and were debriefed. However, participants in the Expectations for Real Groups condition were given an expectation tied to a real group (their gender) rather than a meaningless group (dot over-estimators) used in the Expectations Alone condition. The graph illustrated past performance of men and women on the analogical reasoning task. Participants were randomly assigned either a graph showing that males tended to outperform females or that females tended to outperform males. A participant’s own gender interacted with the graph to sort participants into either threatened or non-threatened conditions. For example, female participants assigned to a graph showed men outperforming women were in the threatened condition. In order to increase the likelihood that participants read the graph,

they were asked to check a box next to the graph that would give the researchers permission to use their data to update the graph. Furthermore, directly under the graph, participants were required to type out their gender, making their own gender more salient and, therefore, increasing the effects of the graph on their performance. The Expectations for Real Groups with Widely-Shared Stereotype condition is one that contained the classic stereotype threat. Participants randomly assigned to this condition received a manipulation graph that always depicted females outperforming men on the upcoming “verbal test” in order to elicit the widely-held stereotype that females are better at verbal tests compared to men. As a result, men in this condition were always threatened and women were never threatened. This particular scenario targeted not only a real group (gender), but it was also linked to a widely-held stereotype, the belief that women outperform men on verbal tasks. Similar to the previous condition, participants checked a box next to the graph indicating that the researchers could use their data to update the graph and they typed their gender below the graph to increase the salience of their gender and its relevance to the graph. In this condition, the analogy task was referred to as a “verbal test” for the sole purpose of relating the term “verbal” to the widely-held stereotype of gender and verbal ability.

Results

Participants’ scores on the verbal test represented the number of questions answered correctly. Using their scores, a one-way between-subjects ANOVA was used to compare the performance of threatened males within each condition (expectations alone, expectations for real group, and expectations for real group that are widely shared). Findings revealed that when men were threatened with a negative expectancy alone, they obtained a mean score of 17.43. This means that on average, males answered approximately 17.43 analogy questions correctly when threatened with a negative expectancy. On the other hand, when males were targeted with a negative expectancy tied to their social identity, they performed worse with a mean score of 13.75. Although the difference in

performance between these two groups was not significant ($p = .24$), the trend fits the researchers' hypothesis that targeting one's social identity with a negative expectation would lower a subject's performance on the test. This trend is further made evident when looking at the effect of threat paired with a wide-held stereotype towards one's social identity. Threatened males in this condition, on average, obtained a mean score of 12.22; however, according to Post-Hoc Tukey pairwise comparisons, this score was not significantly different from those threatened in the expectation and real group condition ($p = 0.74$) or the expectancy condition ($p = 0.09$). This trend is made clear below in Table 2 such that subjects' scores lowered when the threat was tied to a social group and lowered even further when the social group was targeted by a widely-held stereotype.

Table 2: *Mean Number of Questions Answered Correctly by Threatened Males*

Expectations for Real Group with Widely Shared Stereotype	Expectations for Real Group	Expectations Alone
12.22	13.80	17.43

In order to demonstrate the effect of threat on male performance within the conditions, an independent t-test was used. Within the expectancy alone condition, there was not a significant change in performance when males were threatened ($p = .13$). Furthermore, threatened and non-threatened males in the expectancy tied to ones social identity condition did not perform significantly different on the verbal test, ($p = .07$). While statistical significance is not present, it is still apparent from these results that threat tied to one's social identity had a greater negative effect on men's ability to answer analogy questions.

Scores of female participants were examined with another one-way, between subjects ANOVA in order to assess their performance when threatened in two conditions (expectations alone and

expectations for real group). When presented with negative expectations, females performed better ($M = 16.29$) compared to when they were targeted with negative expectations tied to a real group ($M = 14.13$). A summary of these means is presented in Table 3. However, the two different elements of stereotype threat did not show to have a significant difference, for the p -value when comparing the two groups was 0.74.

Table 3: *Mean Number of Questions Answered Correctly by Threatened Females*

Expectations for Real Group with Widely Shared Stereotype	Expectations for Real Group	Expectations Alone
No Threat	14.13	16.29

A 2 (threatened or non-threatened) \times 3 (expectations alone, expectations for real group, and expectations for real group that are widely shared) completely between-subjects ANOVA was used to examine whether the effect of threat varied across the three conditions. The ANOVA shows that there was no main effect of threat on performance on the verbal test such that participants who were threatened did not perform significantly worse than those who were not-threatened, $F(1, 113) = 0.17$, $p = 0.70$. Also, participants did not perform significantly differently when in the expectancy condition, expectancy and real group condition, or the classic stereotype condition, $F(2, 113) = 1.27$, $p = .29$. The interaction between threat and condition was not significant ($p = .13$). These results suggest that regardless of condition and whether participants experienced a threat or not within a condition, these elements did not significantly affect their performance on the analogy test.

To break these results down into smaller components, the researchers tested for simple main effects of threat within individual conditions. Looking at the expectancy condition, participants who were threatened ($M = 16.91$) did not perform significantly different than those who were not

threatened ($M = 14.05$), $p = .17$. Similarly, in the expectancy and real group condition, participants who were threatened ($M = 13.70$) did not perform significantly different than those who were not threatened ($M = 16.50$), $p = .16$. Finally, when presenting participants with a widely-held stereotype targeted towards a real group (gender), participants who were in the threatened condition did not perform significantly different on the verbal test ($M = 12.22$) than those who did not experience any threat ($M = 13.63$), $p = .43$. Although significance was not reached, there are general trends that are worthy to note. For example, across conditions participants who were not threatened performed relatively the same. Their scores ranged from 13.63-16.50. Also predicted was those participants experiencing a classic stereotype threat, in this case when men were told that females performed better on tests of verbal ability, they performed worse than the expectancy and real group condition. This indicates that the strongest threat to participants encompassed a widely-held stereotype targeted at a real group like gender rather than simply exposing participants to a threat not linked to a widely-held stereotype. Overall, those subjects who perceived a threat to their social identity answered the least amount of correct answers on the analogy test.

Discussion

Although the study generated no significant results between the elements of stereotype and threat, the interaction was marginally significant and suggests that some difference exists between the conditions. This difference, although not significant, fit the pattern of the hypothesis. The hypothesis predicted that the negative expectation unattached from a meaningful social identity and a widely-shared belief would not significantly contribute to stereotype threat. This was demonstrated by the male participants whose non-threatened scores matched their threatened scores. That is, the addition of a negative expectation in of itself did not significantly negatively impact performance. Curiously, the female participants scored higher when threatened than when not threatened, as long as the expectation was not tied to their social identity or a widely-shared belief.

There are two plausible explanations for this increase in performance under threat. First, perhaps the female participants after undergoing the dot task, receiving an identity as dot over-estimator, and viewing the manipulation graph depicting dot over-estimators previous performance as low felt challenged or a need to compensate for their or their group's apparently poor performance. Secondly, the female participants may have found the threat to be productively arousing. The Yerkes-Dodson law reflects how level of arousal affects performance. Fit to a bell-curve, the Yerkes-Dodson law, claims that peak performance occurs under moderate levels of arousal. Too much or too little arousal, however, negatively impacts performance. It is possible to re-conceptualize the various conditions in terms of arousal. Female participants that were in the non-threatened dot-overestimation group may have been the least aroused such that they did not adequately engage the task. Their threatened counterparts, however, were pushed closer to their peak performance by the resultant arousal. This would not be the first documented instance of Stereotype Threat research drawing from the Yerkes-Dodson law to interpret data. Keller (2007) reviewed the Stereotype Threat literature noting that a study by O'Brien and Crandall (2003) provided evidence that supports an arousal-based understanding of stereotype threat. On an easy math test, in which females are stereotyped to be outperformed by males, threatened females were found to outperform their non-threatened counterparts. On a difficult math test, however, the reverse was true – non-threatened females outperformed threatened females. To explain this inconsistency Keller (2007) employs Yerkes-Dodson law arguing that the non-threatened females when answering easy math questions were not aroused enough to approach their peak performance, while threatened females answering difficult math questions were too aroused.

Participants receiving a negative expectation tied to their social identity but not fit to a widely-shared belief performed worse than participants receiving a negative expectation untied to both their social identity and a widely-shared belief. These participants performed worse than their non-

threatened counter parts, demonstrating that the negative expectation tied to social identity did negatively affect participants, but not significantly so.

Participants receiving a negative expectation tied to their social identity and fit to a widely-shared belief performed only slightly worse than participants faced with a negative expectation tied to social identity alone. That the addition of a widely-shared belief did not significantly negatively impact the participants' scores fit the researchers' hypothesis.

Only males could be threatened in this condition for no widely-shared belief exists that females are outperformed by males on tests measuring verbal skills. This restriction allowed the researchers to more closely examine which type of negative-expectation is the most debilitating to performance from the point of view of the males in the study. The pattern here, clearly but again not significantly fits the hypothesis. Males threatened by a negative expectation alone perform much better than males threatened by a negative expectation tied to social identity, whom in turn, perform slightly better than males threatened by a negative expectation tied to social identity fit to a widely-shared belief. These results are not without their limits. The manipulation constructs may have been ineffective for certain participants. That is, participants may have overlooked the negative expectation provided for them altogether. The manipulation graphs may have been too abstract to take effect for some participants because the Y-axis displayed no numbers. The manipulation, thus, displayed the message that participants were doing worse or better, but worse and better were abstract, not tied to any concrete estimates of how many GRE analogy questions a participant was expected to answer correctly.

One discrepancy uncontrolled for by the conditions is also worth noting. The manipulation for participants in conditions where expectations were tied to social identity, and then again to a widely-shared belief, participants were asked to type in their gender under the manipulation graph. In the condition where expectation was not tied to social identity or fit to a widely-shared belief participants were assigned the role of dot over-estimator. They were not asked to type in their group under the

manipulation graph. Controlling for this discrepancy by asking participants to type in the group that the study had just assigned them, however, may have appeared strange to participants and perhaps even acted as a demand characteristic giving away that the researchers were manipulating the expectation of the participant's assigned group. Thus, the discrepancy was allowed to persist, even though it may act have acted as a confounding variable and skew results in performance between the groups.

Although demographic information was recorded participants were never asked to self-report whether English was their first language or not. As such, some participants' scores may have been affected not by their level of threat but by their inability to adequately understand English. Prior level of education and other assessments of previous ability on verbal tests were also not gathered from participants. Future research may benefit from collecting such data in order to tease out the confounding variable of absence of ability from threatened ability.

Future directions for stereotype threat research include a closer analysis of the role of Yerkes-Dodson law in explaining stereotype threat. Perhaps participants perform best in the presence of threat, and perform worst when said threat is too underwhelming or overwhelming. To extend the relationship between Yerkes-Dodson and Stereotype threat researchers can compare performance on difficult and easy academic tests as well as compare participants who do not identify with academic performance to participants who do identify with academic performance. Perhaps those who identify highly with academic performance are too aroused by threat, while those who identify lowly find stereotype threat to be productively arousing.

Works Cited

- Brophy, J. E., & Good, T. L. (1974). *Teacher-student relationships*. Oxford: Holt, Rinehart & Winston.
- Jussim, L., & Herber, K. (2005) Teacher expectations and self-fulfilling prophecies. *Personality and Social Psychology Review*, 9(2), 131-155.

- Keller, J. (2007). Stereotype threat in classroom settings. *British Journal of Educational Psychology*, 77(2), 323-338.
- Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. California: Stanford University Press.
- Merton, R. K. (1948). The bearing of empirical research upon the development of social theory. *American Sociological Review*, 13, 505-515
- O'Brien, L., & Crandall, C. (2003). Stereotype threat and arousal. *Personality and Social Psychology Bulletin*, 29(6), 782-789.
- Osborne, J. & Walker, C. (2006). Stereotype Threat, identification with Academics, and Withdrawal from School: Why the most successful students of colour might be most likely to withdraw. *Educational Psychology*, 26(4), 563-577.
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart & Winston, 1968.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1), 225-229.
- 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.
- Tajfel, H., Billig, M.G., Bundy, R.P., & Flament, C. (1971). Social categorization and intergroup behavior. *European Journal of Social Psychology*, 1(2), 149-178.
- Voelkl, K. E. (1996). Measuring Students' Identification with School. *Educational and Psychological Measurement*, 56(5), 760-770.

