The Effect of Stereotype Threat on Group Versus Individual Performance

Nicholas P. Aramovich

Small Group Research 2014 45: 176
DOI: 10.1177/1046496414523508

The Effect of Stereotype Threat on Group Versus Individual Performance

Small Group Research
2014, Vol. 45(2) 176–197

DOI: 10.1177/1046496414523508
sgr.sagepub.com
SAGE

Nicholas P. Aramovich[1]

Abstract

Stereotype threat has been one of the most studied topics in social psychology in recent years. This research shows that subtle reminders of stereotypes about one's social category hurt task performance—an effect replicated across several stereotypes and performance domains. Despite extensive research on individual performance, it is unknown how stereotype threat affects group performance. A question of theoretical and practical importance is whether people who face a common stereotype can overcome it by working together. To answer this question, an experiment was conducted comparing the performance of individual women and groups of women on a math/logic problem when faced with a stereotype threat. Results indicated that when facing a stereotype threat, groups outperformed the best individuals and performed just as well as non-threatened groups. This effect was due to threatened groups avoiding problem-solving errors. The implications for understanding group versus individual performance when facing stereotype threats are discussed.

Keywords

stereotype threat, group problem solving, group performance

Stereotype threat refers to the risk of confirming, as self-characteristic, a negative stereotype about one's social category (Steele & Aronson, 1995).

[1]Harvard University, Cambridge, MA, USA

Corresponding Author:
Nicholas P. Aramovich, Peoplefluent, 300 Fifth Ave., Waltham, MA 02451, USA.
Email: nicholas.aramovich@peoplefluent.com

Over the past 20 years, stereotype threat has been one of the most studied topics in social psychology (Inzlicht & Schmader, 2011). This body of research has revealed that when people are reminded of negative stereotypes about themselves, they underperform relative to their true potential (for a review, see Schmader, Johns, & Forbes, 2008). Moreover, stereotype threat is a robust phenomenon—negative performance effects have been found for people who are stigmatized based on their gender, race, and socioeconomic status across several performance domains including academics, athletics, and interpersonal interactions (Schmader, 2010).

Although the effects of stereotype threat for individual performance are well documented, it is unknown how stereotype threat affects group performance. Because groups rather than individuals are commonly asked to accomplish tasks and solve problems across a variety of organizational and educational contexts, it is important to understand how stereotype threat affects their performance. A question of both theoretical and practical importance is whether people who share a common stigma about their task ability can disconfirm the stereotype by working together to perform well.

Consider, for example, the task of solving a math problem and demonstrating the solution to the other students in the class. Research shows that women face both explicit and subtle stereotypes about their math ability compared with men (Eccles, Jacobs, & Harold, 1990; Murphy, Steele, & Gross, 2007). With this in mind, in a coeducational classroom, would a group of female students be more successful at this task than an individual female student? As I describe below, there are several reasons to believe that the group would be more successful than the single student at performing well in this threatening situation.

Group Performance When Facing a Shared Stereotype Threat

When individuals are reminded of negative stereotypes about their ability, they actively monitor the performance environment for cues suggesting that they will confirm the stereotype (Beilock, Rydell, & McConnell, 2007) and attempt to suppress negative thoughts and feelings of anxiety associated with the stereotype (Johns, Inzlicht, & Schmader, 2008). Engaging in these activities requires working memory, a resource necessary for successful performance on complex, cognitive tasks. Therefore, as a result of these cognitive activities, individuals facing stereotype threats underperform relative to nonthreatened individuals (Schmader, 2010).

However, there are reasons to believe that groups facing a stereotype threat will not show these same performance impairments. The first reason is that groups should be able to avoid performance errors. Because a group has

multiple members, it has more total cognitive resources than an individual does (Hinsz, Tindale, & Vollrath, 1997). Therefore, even if members individually experience cognitive deficits, a group may have the total cognitive capacity to meet the demands of the task and perform well. Moreover, the shared nature of group problem solving should allow groups to capitalize on their increased capacity. When members share their ideas with one another, they are able to evaluate one another's contributions and detect and correct any errors that are committed. Indeed, research shows that groups are better than individuals at detecting and correcting errors (Laughlin, VanderStoep, & Hollingshead, 1991; Laughlin, Zander, Knievel, & Tan, 2003; Shaw, 1932). Therefore, although monitoring the performance environment for signs of failure ironically often hurts individuals (Schmader, 2010), this same behavior should benefit groups because they have the increased cognitive capacity to do so.

In addition, groups facing stereotype threat may be even more vigilant for errors than non-threatened groups. For example, research on alcohol and group performance finds that, when intoxicated, group members engage in monitoring processes aimed at detecting and correcting errors in judgment, which results in the same performance as sober groups (Abrams, Hopthrow, Hulbert, & Frings, 2006; Frings, Hopthrow, Abrams, Hulbert, & Gutierrez, 2008). Similarly, groups facing a stereotype threat should be able to perform at the same level as non-threatened groups to the extent they successfully monitor their performance and correct errors. Therefore, one reason why group performance should not be affected by stereotype threat is a group's increased capacity to monitor its performance and avoid errors.

A second reason why group performance should not be affected by stereotype threat is because group members may simply feel less concern about confirming the stereotype. Among individuals, concerns about confirming the stereotype may stem from lower performance expectations, which have been linked to underperformance (e.g., Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003; Rosenthal, Crisp, & Suen, 2007). However, most people have the implicit belief that groups are effective and also report higher expectations for their personal performance when working in a group compared with when working alone (e.g., Allen & Hecht, 2004; Paulus, Dzindolet, Poletes, & Camacho, 1993). Therefore, group members may be less likely to hold low performance expectations for themselves and their group. As a result, they may not experience the same distracting thoughts and concerns about confirming the stereotype.

Group members may also feel less concern about the stereotype threat because they have the opportunity to affiliate with one another. For example, research shows that when facing threats, people affiliate with others in the

same predicament to compare their thoughts and feelings (e.g., Dooley & Catalano, 1984; Harlow & Cantor, 1995; McGuire, 2007; Schachter, 1959). Affiliating with others has several benefits, such as reducing one's feelings of anxiety and uncertainty (e.g., Kulik, Mahler, & Moore, 1996; Taylor, 2007). Consequently, group members may feel less concern about the stereotype than individuals because of the opportunity to affiliate with other members.

Finally, group members may also experience fewer negative thoughts and concerns about the stereotype threat due to the shared nature of the group problem-solving process described above. Sharing ideas and listening to others should direct members' attention on the task and away from their own negative inner states. In support of this idea, research on high-pressure testing situations reveals that directing problem solvers to articulate their problem steps aloud improves performance (DeCaro, Rotar, Kendra, & Beilock, 2010). Taken together, these factors suggest that a second reason group performance should not be affected by stereotype threat is because group members may simply feel less concern about confirming the negative stereotype.

Overview and Predictions

To test these predictions, I conducted an experiment that compared the performance of groups of women with individual women when working on a problem-solving task involving mathematical and logical reasoning when facing a stereotype threat. Consistent with previous research, I predicted that stereotype threat would impair individual performance relative to non-threatened individuals (Hypothesis 1). However, due to the reasons given above (groups' *collective capacity*, and the potential for members to *feel less concern* about the stereotype threat), I predicted that stereotype threat would not affect group performance. Therefore, threatened groups should perform at the same level as non-threatened groups (Hypothesis 2), and outperform threatened individuals (Hypothesis 3).

Method

Participants and Design

Participants were 171 female undergraduates who received credit toward a class requirement. Eighty-one participated individually and 90 participated in 30 three-person groups. Participants were randomly assigned to a stereotype threat condition in a 2 (stereotype threat: threat vs. no threat) × 2 (participation type: group vs. individual) experimental design. After random assignment, the participant breakdown was as follows: individual participants (threat, $n = 43$; no threat, $n = 38$) and groups (threat, $n = 14$; no threat, $n = 16$).

Task

Individuals and groups solved one letters-to-numbers (l-t-n) problem (Laughlin, Bonner, & Miner, 2002). In a l-t-n problem, 10 letters (A-J) are randomly assigned without replacement to one of the numbers 0 to 9. The goal is to determine the coding of letters to numbers in as few trials as possible, using a maximum of 10 trials. A trial consists of proposing an addition or subtraction equation in letters (e.g., A + B = ?), which is answered by the experimenter in letters (e.g., A + B = E). Participants then propose the value of any letter (e.g., A = 3), and the experimenter gives feedback as to whether this hypothesis is true or false. The full code may be proposed at the end of each trial and participants are told whether the entire code is correct. The full correct mapping of letters to numbers solves the problem; otherwise, additional trials must be used.

Procedure

Laboratory sessions were conducted by a male experimenter. During informed consent, participants learned that the purpose of the research was to understand how people solve problems individually or in groups, and that they would be video-recorded while working on the problem. All participants consented to the video recording.

Participants received instructions on the l-t-n problem adapted from previous research (Laughlin et al., 2002) and also received a handout showing seven illustrative problem trials. After receiving instructions, but before starting the problem, individuals and groups were randomly assigned to one of the stereotype threat conditions. Depending on condition, participants listened to one of two different audio-recorded messages from the head researcher. This procedure, which was adapted from Schmader (2002), served as the stereotype threat manipulation. On both recordings, all participants heard the voice of a man who introduced himself as a researcher in the psychology department who was interested in how individuals and groups solve problems. Both recordings described the l-t-n problem as a reliable indicator of mathematical and logical reasoning ability. Individual participants were told that their score on the problem would be used as an indicator of their ability to think logically. Groups were told that their score on the problem would be used as an indicator of their group's ability to think logically. Individuals were told their score would be compared with all other individuals who participated in the experiment, and groups were told their score would be compared with all other groups that participated in the experiment. Both individuals and groups were encouraged to do their personal best. In the no-threat condition, this is

all participants heard. However, in the stereotype threat condition, the head researcher went on to say that he was interested in how women perform on this problem relative to men. Individual participants were told that their score would be used as an indicator of women's logical reasoning in general, and groups were told that their score would be used as an indicator of women's logical reasoning in general. To bolster this manipulation, participants in the stereotype threat condition identified their gender on their problem worksheets.

After hearing the recording, participants received a problem worksheet and paper and pencils for computations. Groups were told to determine for themselves who would record equations and hypotheses on the worksheet and received no explicit instructions for achieving consensus or resolving disagreements. The experimenter informed participants that they would have 30 min to complete the problem, reminded them that their goal was to solve the problem in as few trials as possible, started or ostensibly started the video recording, and then left the room. In reality, video recordings were only actually made of groups; however, to maintain consistency across experimental conditions, individuals were told that a recording was made. The primary purpose of the video recording was to capture group interaction.[1] However, the video camera also served to create the perception that all participants' individual performance could be evaluated. Research suggests that the potential for evaluation plays a key role in producing stereotype threat performance deficits (e.g., Jamieson & Harkins, 2010; Wout, Danso, Jackson, & Spencer, 2008; Zhang, Schmader, & Hall, 2013). However, performing in a group has the potential to make individual task contributions unidentifiable (e.g., Karau & Williams, 1993), which may make group members feel less concern about personally confirming the stereotype (but see Jamieson & Harkins, 2010; Schmader, 2002). The video camera likely decreased the perception among participants that their individual performance was unidentifiable, and eliminated a possible reason why group members *may* not have felt concern about the stereotype threat.

When working on the problem, to receive feedback on equations and hypotheses, participants brought their worksheets to an adjacent room and the experimenter wrote feedback directly on the worksheet. Each group or individual solved a unique, random code that they worked on until they correctly solved the code, used all 10 trials, or time expired. Consistent with previous research, non-solvers received a score of 11 (e.g., Laughlin et al., 2002).

After working on the problem, participants completed a short questionnaire. The questionnaire contained a manipulation check, which asked participants to indicate (*Yes* or *No*) whether the audio recording indicated that women's scores would be compared with men's scores. The questionnaire

also contained two items, adapted from Schmader and Johns (2003), regarding feelings of concern about the stereotype. Participants rated the extent to which they agreed with the following statements on a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*): "I am concerned that the researcher will judge women as a whole based on my score on the problem," and "I am concerned that the researcher will think women are less logical than men if I did not do well on this problem." Responses were averaged to create a single composite measure of concern about the stereotype (Cronbach's alpha = .72). After completing this questionnaire, participants were debriefed and thanked for their participation.

Identification of the Best Individuals

Hypothesis 3 predicted that, when faced with a stereotype threat, groups would outperform individuals. To provide a stringent test of this hypothesis, the best individual performers were identified. Because groups contained three members, the best individual performers were identified by randomly assigning individual participants within each stereotype threat condition to a nominal (pseudo) group of three (cf. Laughlin et al., 2002). The best individual performer in each nominal group was identified as the individual who used the fewest trials to solve the problem. In the case of tied scores, an individual was randomly selected.

Comparing groups with the best individuals tested the assumption that group performance was attributable in some way to group interaction (Larson, 2010). If groups merely adopt the ideas of their best members, groups should perform at the same level as the best individuals. However, if group performance is the result of interaction processes stemming from the increased capacity of having additional members, groups should outperform the best individuals.

Problem-Solving Errors

As suggested above, one reason why groups might outperform individuals when facing a stereotype threat is because although both may attempt to be vigilant for problem-solving errors, groups will have the cognitive capacity to detect and correct them if made. To test this explanation, a coding scheme was developed based on previous research with l-t-n problems (e.g., Laughlin et al., 2002; Laughlin et al., 2003). Groups' and the best individuals' problem worksheets were coded for three types of performance errors. *Uninformative equations* are those that do not yield any new information about l-t-n mappings in the code or have the potential to do so. For example, if it is already

known from feedback that A + B = C, subsequently proposing C − B = ? is uninformative. A *nonplausible-value hypothesis* proposes a l-t-n mapping that is not mathematically or logically possible based on existing problem feedback. For example, if it is known from feedback that A + B = C and C = 5, then both the values of A and B must be less than 5, and proposing values greater than or equal to 5 for either is nonplausible. *Known-value hypotheses* propose as a hypothesis a correct l-t-n mapping that could have been deduced from feedback or from logical inference. For example, if it is already known that A = 1 and B = 2, and A + B = C, then hypothesizing that C = 3 is an inefficient use of a hypothesis because the value of C is known based on existing feedback. Two raters who were blind to experimental condition coded equations for whether they were uninformative (agreement = 94%, Cohen's κ = .89), and hypotheses for whether they proposed known values or were nonplausible (agreement = 88%, Cohen's κ = .81). Disagreements between raters were resolved via discussion.

Theoretically, the tendency to commit one type of problem-solving error should be positively associated with the tendency to commit the other two. Specifically, individuals facing a stereotype threat should be just as likely to commit one type of error as the others, and groups should be just as likely to avoid committing one type of error as the others. Therefore, I expected that the frequency in which groups or individuals committed each type of error should be positively related. However, because no previous research has coded for these three specific problem-solving errors on l-t-n problems, formal hypotheses about the intercorrelations among these errors were not made.

Results

Manipulation Checks

Audio recording. In the no-threat condition, four individuals incorrectly indicated that they would be compared with men, and in the threat condition, four individuals incorrectly indicated that they would not be compared with men. These eight individuals were excluded from subsequent analyses. Two participants from two separate groups in the threat condition also failed this check. To determine whether these participants affected the other members in the group, the average level of concern about the stereotype reported by members from each of these two groups was compared with the average level of concern reported by all groups in the threat condition (M = 3.20, SD = 0.83). Despite having one member who failed the manipulation check, one of these groups reported an equivalent amount of concern as other groups (M = 3.50,

$z = 0.36$) and was therefore retained in subsequent analyses. The second group reported significantly less concern about the stereotype ($M = 1.33$, $z = -2.26$) and was excluded from subsequent analyses. These exclusions left 73 of the original 81 individual participants (threat, $n = 39$; no threat, $n = 34$) and 29 of the original 30 groups (threat, $n = 13$; no threat, $n = 16$).

Concern about the stereotype. One reason why group performance may be unaffected by stereotype threat is because members of threatened groups were simply not concerned about the stereotype. To test for this possibility and to determine whether the stereotype threat manipulation had its intended effect, the amount of concern participants expressed about the stereotype was analyzed. Because some participants were nested within groups, an intraclass correlation was computed to determine the degree of nonindependence in participants' concern ratings due to group membership. The intraclass correlation was .24 (Wald $z = 1.72$, $p = .086$) indicating a moderate degree of nonindependence. To avoid a downwardly biased estimate of standard error and the increased possibility of a Type I error, group members' concern ratings were averaged and groups' concern ratings were compared with individuals' concern ratings. Results of a 2 (stereotype threat: threat vs. no threat) × 2 (participation type: group vs. individual) factorial ANOVA revealed only a main effect of stereotype threat with individuals and groups in the stereotype threat condition ($M = 3.19$, $SD = 1.28$) reporting more concern than individuals and groups in the no-threat condition ($M = 2.33$, $SD = 1.14$), $F(1, 98) = 11.68$, $p = .001$, $d = 0.72$. This result shows that the manipulation had its intended effect. Participating in a group did not necessarily buffer group members from feeling concern about the stereotype. Group members in the stereotype threat condition reported just as much concern as individual participants in the stereotype threat condition and more than individuals and group members in the no-threat condition. This result also fails to support the explanation that groups might be unaffected by stereotype threat simply because their members were unconcerned about confirming the stereotype.

Problem-Solving Performance

Groups versus the average individual. The number of trials that groups and individuals used to solve the problem were compared using a 2 (stereotype threat: threat vs. no threat) × 2 (participation type: group vs. individual) factorial ANOVA. Results indicated that there was a main effect of participation type; on average, groups ($M = 6.17$, $SD = 0.97$) used fewer trials to solve the problem than individuals did ($M = 8.37$, $SD = 1.85$), $F(1, 98) = 37.52$, $p < .001$, $d = 1.35$. There was also a marginally significant stereotype threat by

participation type interaction, $F(1, 98) = 3.23$, $p = .075$, $\eta_p^2 = .03$. To test hypotheses, planned comparisons of simple effects were conducted. In support of Hypothesis 1, stereotype threat impaired individual performance; individuals in the stereotype threat condition used more trials ($M = 8.82$, $SD = 1.75$) than individuals in the no-threat condition ($M = 7.85$, $SD = 1.86$), $F(1, 98) = 6.52$, $p = .012$, $d = 0.55$. In support of Hypothesis 2, stereotype threat did not affect group performance; groups in the stereotype threat condition ($M = 6.00$, $SD = 1.16$) used the same number of trials as groups in the no-threat condition ($M = 6.31$, $SD = 0.79$), $F(1, 98) < 1$, $d = 0.33$. In addition, in support of Hypothesis 3, when facing a stereotype threat, groups outperformed individuals by using fewer trials than individuals did, $F(1, 98) = 29.75$, $p < .001$, $d = 1.77$. As the marginally significant interaction suggested, groups used fewer trials than individuals in the no-threat condition as well, $F(1, 98) = 9.90$, $p = .002$, $d = 0.93$; however, the effect was not as large as in the threat condition. See Figure 1.

Groups versus the best individuals. The previous analysis showed that, when facing stereotype threat, groups outperformed the average individual participant. However, a more stringent test is to compare group performance with the performance of the best individuals. The best individual performers were identified based on the nominal group procedure described above. The number of trials that groups and the best individuals used to solve the problem were compared using a 2 (stereotype threat: threat vs. no threat) × 2 (participation type: group vs. best individual) factorial ANOVA. Similar to the first analysis, overall, groups ($M = 6.17$, $SD = 0.97$) used fewer trials to solve the problem than the best individuals did ($M = 6.92$; $SD = 1.44$), $F(1, 49) = 4.90$, $p = .032$, $d = 0.63$. However, this main effect was qualified by a significant interaction, $F(1, 49) = 4.23$, $p = .045$, $\eta_p^2 = .08$. In further support of Hypothesis 1, stereotype threat impaired the performance of the best individuals; they also used more trials in the threat condition ($M = 7.38$, $SD = 1.61$) than in the no-threat condition ($M = 6.36$, $SD = 1.03$), $F(1, 49) = 4.54$, $p = .038$, $d = 0.77$. In addition, in further support of Hypothesis 3, when facing a stereotype threat, groups outperformed even the best individuals, $F(1, 49) = 9.10$, $p = .004$, $d = 1.03$. Group performance was equivalent to the best individuals in the no-threat condition, $F(1, 49) < 1$, $d = 0.06$. See Figure 1.

Problem-Solving Errors

Having already ruled out the possibility that groups in the stereotype threat condition were simply less concerned about the stereotype, another possible reason why group performance was unaffected by stereotype threat, and

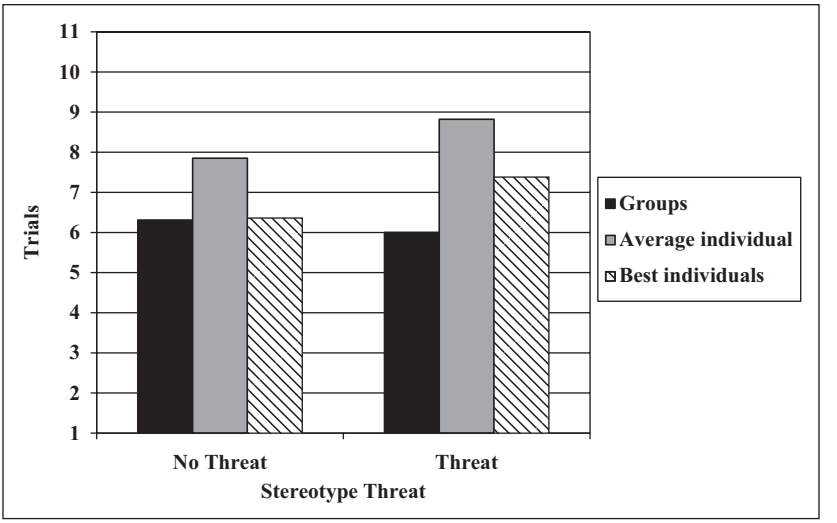

Figure 1. Group performance versus two individual performance baselines under conditions of stereotype threat and no threat.

therefore why threatened groups outperformed threatened individuals, is that threatened groups were able to avoid problem-solving errors. To test this explanation, a series of analyses was performed. Specifically, I tested (a) which problem-solving errors are related to performance, (b) whether there were differences between the experimental conditions in committing those problem-solving errors that predicted performance, and (c) whether committing (avoiding committing) these errors explained the performance difference found between threatened groups and threatened individuals.

Which problem-solving errors predicted performance?. As a preliminary analysis, the number of trials that groups and the best individuals used to solve the problem was regressed simultaneously on the proportion of uninformative equations, nonplausible-value hypotheses, and known-value hypotheses that they committed out of their total numbers of equations and hypotheses.[2] In addition, the average number of letters used per equation was also included because research finds that this variable is associated with using fewer trials to solve the problem (e.g., Laughlin et al., 2002).[3] Results indicated that uninformative equations ($\beta = .30$, $p = .018$), nonplausible-value hypotheses ($\beta = .34$, $p = .007$) were significantly related to the number of trials used in the expected direction. The average number of letters used per equation

Table 1. Intercorrelations, Means, and Standard Deviations for Trials to Solution and Problem-Solving Errors for Groups and the Best Individuals.

Variable	1	2	3	4	5
1. Number of trials used to solve	—				
2. Proportion of uninformative equations	.39**	—			
3. Proportion of known-value hypotheses	.20	−.16	—		
4. Proportion of nonplausible-value hypotheses	.42**	.18	.14	—	
5. Average number of letters used per equation	−.35*	−.19	−.33*	−.05	—
M	6.51	0.05	0.18	0.07	2.18
SD	1.25	0.08	0.16	0.09	0.36

Note. The observed intercorrelations between the problem-solving errors are somewhat lower than expected, which is likely attributable to restricted variance on these variables. As evidenced by the mean levels of these variables, groups and the best individuals were fairly successful at avoiding these errors resulting in floor effects. It is possible that the true relationships among the problem-solving errors are stronger, which may have been observed had the entire sample been included in the problem-solving error analysis.
* p < .05. **p < .01.

was marginally associated with using fewer trials (β = −.24, $p = .067$). The proportion of known-value hypotheses used was unrelated to the number of trials used (β = .12, *ns*) and was therefore not included in subsequent analyses.[4] As a set, these variables explained a significant proportion of variance in the number of trials used, $R^2 = .37$, $F(4, 48) = 6.96$, $p < .001$. Intercorrelations, means, and standard deviations for these variables appear in Table 1.

Group differences in committing errors. I next determined whether groups and the best individuals differed in their use of uninformative equations and nonplausible-value hypotheses using two factorial ANOVAs with these variables as outcomes.[5] For uninformative equations, results revealed only a significant interaction of stereotype threat and participation type $F(1, 49) = 11.99$, $p = .001$, $\eta_p^2 = .20$. Follow-up analyses revealed that the best individuals used a significantly higher proportion of uninformative equations in the stereotype threat condition ($M = 0.10$, $SD = 0.10$) than in the no-threat condition ($M = 0.03$, $SD = 0.06$), $F(1, 49) = 5.23$, $p = .027$, $d = 0.91$. In contrast, groups did not use any uninformative equations in the stereotype threat condition ($M = 0.00$, $SD = 0.00$), which differed significantly from the proportion used by groups in the no threat condition ($M = 0.07$, $SD = 0.09$), $F(1, 49) = 6.93$, $p = .011$, $d = 1.14$. Results also showed that when facing stereotype threat, groups used fewer uninformative equations than the best individuals, $F(1, 49) = 11.19$, $p < .001$, $d = 1.48$. However, groups and the best individuals did not

differ in the no-threat condition, $F(1, 49) = 2.40$, *ns,* $d = 0.52$. For nonplausible-value hypotheses, results revealed only a significant main effect of participation type. The best individuals used a higher proportion of nonplausible-value hypotheses ($M = 0.12$, $SD = 0.10$) than groups did ($M = 0.03$, $SD = 0.07$) across stereotype threat conditions, $F(1, 49) = 12.37$, $p = .001$, $d = 0.99$.

Mediation analysis. Because groups and the best individuals differed in their use of uninformative equations (in the stereotype threat condition) and nonplausible-value hypotheses (in both conditions), it was possible to determine whether these errors mediated the performance difference found between groups and the best individuals when facing stereotype threat. To do so, a moderated mediation analysis was performed using the bootstrapping procedure and corresponding SPSS macro described in Hayes (2013). This procedure allowed for tests of moderated mediation with multiple mediators, and uses a bootstrapping procedure that makes fewer unrealistic assumptions about shape of the sampling distribution of the indirect effect than the Sobel test does, and is more powerful. In this analysis, the number of trials that groups and the best individuals used to solve the problem was regressed on the main effects of stereotype threat (contrast coded: no threat = −½, threat = +½) and participation type (contrast coded: best individual = −½, group = +½), the interaction of stereotype threat and participation type, and the proportions of uninformative equations and nonplausible-value hypotheses used. Five thousand bootstrap samples were performed. The indirect effect is statistically significant if the bootstrap-generated 95% confidence interval (CI) of the effect does not include zero.

Results revealed that in the stereotype threat condition, the performance difference between groups and the best individuals was mediated both by the use of uninformative equations (95% CI = [−1.20, −0.04]) and nonplausible-value hypotheses (95% CI = [−1.28, −0.08]). In the no-threat condition, the effect of participation type on performance was not mediated by the use of uninformative equations (95% CI = [−0.02, 0.65]); however, there was an indirect effect of participation type on performance mediated by nonplausible-value hypotheses (95% CI = [−1.28, −0.03]). This latter effect is at first conceptually difficult to interpret given the lack of an overall performance effect between groups and the best individuals in the no-threat condition. Recall, however, that there was an overall main effect of participation type on performance with groups using fewer trials to solve the problem than the best individuals across threat conditions. Moreover, there was an overall main effect of participation type on the use of nonplausible-value hypotheses, with groups using fewer nonplausible-value hypotheses than the best individuals across threat conditions. Therefore, finding that the effect of participation

type on performance was mediated by the use of nonplausible-value hypotheses across both threat conditions is a reflection of these two main effects.

Discussion

This study found that when faced with stereotype threat, all female groups performed significantly better on a problem related to mathematical and logical reasoning than individual women did, and even outperformed the very best individuals. This latter effect is noteworthy considering very few studies have reported groups surpassing this performance baseline (for reviews, see Kerr & Tindale, 2004; Larson, 2010). In addition, groups facing stereotype threat performed just as well as groups not exposed to a stereotype threat.

An important factor contributing to groups' ability to perform well when facing a stereotype threat was their avoidance of problem-solving errors. Groups facing a stereotype threat used fewer uninformative equations than the best individuals and non-threatened groups, and fewer nonplausible-value hypotheses than the best individuals. These results are consistent with the idea that groups have an increased capacity to monitor their performance compared with individuals. In addition, the finding that threatened groups used fewer uninformative equations than non-threatened groups suggests that they may have been especially vigilant for errors during the problem-solving process.

Importantly, participating in a group did not necessarily buffer participants from feeling concern about confirming the stereotype. When faced with a stereotype threat, participants expressed just as much concern if they participated in a group or as an individual, and more concern than participants not facing a stereotype threat. However, threatened groups were better able to deal with these concerns than threatened individuals and committed fewer errors.

As noted above, the finding that groups performed better than the best individuals when facing a stereotype threat is an important empirical contribution to the small group performance literature because very few studies have documented this "strong synergy" effect (Larson, 2010). Interestingly, research that has consistently documented this effect—including the present study—has used l-t-n problems as the group task (Laughlin et al., 2002; Laughlin, Hatch, Silver, & Boh, 2006; Laughlin et al., 2003). Laughlin (2011) attributes group superiority over the best individuals on this task to the highly intellective nature of the task in which correct (and incorrect) letter-to-number mappings can be demonstrated using arithmetic, algebra, logic, properties of the ordinary number system, and experimenter feedback (see also Laughlin et al., 2002). Larson (2010) proposes that in addition to the ability to detect

correct or incorrect inferences on equations and hypotheses, l-t-n problems may be solved using a variety of strategies. Moreover, he suggests that problem solvers are able to assess the relative efficiency of one strategy versus another for solving the code in the fewest number of trials.

The explanations offered by Laughlin (2011) and Larson (2010) center on the intellective nature of the task itself, and focus more on the problem-solving processes that groups use rather than on the processes that individuals use. However, it should be clear that the task is a constant in the research—both groups and individuals solve the same l-t-n problem. Therefore, the task's properties, including the demonstrability of correct or incorrect inferences and the ability to use multiple, but differentially efficient strategies, are the same for group and individuals.

Therefore, although it may be implicit in these previous explanations, it is worth stating explicitly that what actually differs between groups and individuals and what is ultimately responsible for groups' performance advantage is the way that groups interact with the problem compared with individuals. One of the fundamental differences between groups and individual problem solving is that the former tends to be a socially shared process (cf. Hinsz et al., 1997; Larson & Christensen, 1993). Group members have the opportunity to share their ideas with one another and have the ability to evaluate the quality of these ideas and detect and correct any errors that are committed. In contrast, unless explicitly instructed to think out loud (e.g., Newell & Simon, 1972), the individual problem-solving process is essentially *inside the head*. Individuals working alone simply do not have the benefit of having others present to evaluate their ideas and detect their errors. Instead, individual problem solvers are responsible for the tasks of generating their own ideas, as well as evaluating them. As the present study suggests, the responsibility for idea generation and evaluation may be too cognitively demanding for individuals under certain circumstances (e.g., facing a stereotype threat), but not necessarily for groups. Therefore, although it is certainly true that the intellective nature of l-t-n problems is a critical factor in explaining the performance difference between groups and the best individuals on this task, it is the shared nature of the group problem-solving process that allows groups to take advantage of the intellective features of the task and outperform individuals.

Recall, however, that in the present study groups only outperformed the best individuals when facing a stereotype threat. Groups and the best individuals performed equivalently on the problem when not facing a stereotype threat. These results may have been due to an important methodological difference in this study compared with previous research (e.g., Laughlin et al., 2002; Laughlin et al., 2006; Laughlin et al., 2003). Although they were

adapted from previous research, the instructions used in this study were much more detailed and elaborate than the standard instructions (e.g., Laughlin et al., 2002). Specially, the instructions used here showed problem solvers seven illustrative example trials rather than the standard four. This was done in an attempt to minimize individual differences in problem-solving ability by ensuring all participants had a strong understanding of the problem rules to increase the chances of observing stereotype threat effects in the threat condition. A potential consequence of these enhanced instructions, however, was that a performance ceiling effect was created. This ceiling effect may have been responsible for the best individuals performing at the same level as groups in the no-threat condition. A ceiling effect was not observed in the stereotype threat condition because consistent with predictions, facing a stereotype threat impaired individuals' but not groups' performance.

Limitations and Future Directions

Some limitations of this study are worth mentioning, as they point to directions for future research. Perhaps most importantly, although it was found that threatened groups were able to perform well because they avoided problem-solving errors, the precise psychological mechanisms behind this effect are still somewhat unclear. It was predicted that groups would be able to avoid errors due to their increased cognitive capacity when monitoring the performance environment. However, it was also suggested that higher performance expectations and an increased focus on the task might also play a role in preventing group members from experiencing distracting thoughts and concerns. Although not initially predicted, if group members did experience fewer negative thoughts and concerns, this would also be associated with groups committing fewer errors.

Because group members reported the same amount of concern about confirming the stereotype as individuals did, there is some reason to believe that they experienced similar distracting thoughts and the associated cognitive deficits. Therefore, these results are consistent with the idea that group members individually experienced the negative inner states commonly associated with stereotype threat but compensated for one another's limitations during the problem-solving process. However, without direct measurement of group members' inner states (e.g., working memory), this "compensation" explanation remains speculative and something future research should explore.

In addition, because performance expectations were not measured, this factor may have also played a role in the effects found here. One intriguing possibility is that, when facing a stereotype threat, group members may appraise the performance situation quite differently from individuals. Group

members may perceive that despite the message that the stereotype conveys—that individuals of their social category have low ability on the task—their current, small group does in fact have sufficient resources to perform well. Therefore, compared with individuals, group members may view task performance as a challenge rather than as a threat (e.g., Blascovich, Mendes, Hunter, & Salomon, 1999; Lazarus & Folkman, 1984; Tomaka, Blascovich, Kelsey, & Leitten, 1993), which may explain why threatened groups outperform threatened individuals. Future research in this area should investigate the extent to which performing in a group affects members' expectations and their appraisals of the performance situation.

Finally, the primary goal of this study was to determine whether people who share a common stigma about their task ability can disconfirm the stereotype by working together and performing well. However, there are many other interesting questions related to stereotype threat and group performance that should be investigated. For example, there are many situations in which people work in groups with diverse others. It is important to understand how members who are stigmatized on the basis of their task ability perform in situations in which they work with non-stigmatized group members. It is also important to understand how non-stigmatized group members perceive and treat members who have a stigma about their ability.

Some existing stereotype threat research is related to these issues. For example, research has found that women perform worse in mixed-sex than in same-sex environments (Inzlicht & Ben-Zeev, 2000; Sekaquaptewa & Thompson, 2003). However, this is not always true; a recent meta-analysis found that women do not necessarily benefit more from female-only or female-majority contexts (Picho, Rodriguez, & Finnie, 2013). Moreover, this research has investigated individual performance, so it remains unclear what conclusions can be made about interacting groups. Therefore, future research is necessary to understand how and whether a group's composition of stigmatized and non-stigmatized members affects its members' interaction and the group's collective performance.

Conclusion

The stereotype threat literature abounds with examples of performance failures facing threat. The present research suggests that one way for people to avoid these performance deficits is to collaborate with others from the same social category or with those who face similar stigmas about their ability. In doing so, this research underscores why people work in groups to begin with—they can achieve outcomes working together that individuals working alone cannot.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Due to a technical malfunction with an external microphone, eight recordings (six threat, two no threat) contained no audio. Videos that contained audio were coded for various problem-solving behaviors of interest. No meaningful differences emerged for threatened ($n = 8$) versus non-threatened ($n = 14$) groups in this analysis. However, because of missing data and unequal sample sizes, this data source was considered a less-reliable indicator of groups' interaction process compared with their problem worksheets.
2. The proportions of uninformative equations and nonplausible-value hypotheses were both moderately positively skewed (>1.5). To determine whether this skewness affected results, these variables were arcsine transformed and all analyses were repeated using the transformed variables. Results were equivalent.
3. This variable was extremely positively skewed (skewness = 5.67) due to a single outlier. Across groups and the best individuals, the average number of letters used was 2.25 ($SD = 0.73$). The outlier was a group in the stereotype threat conditions that used an average of seven letters per equation ($z = 6.51$). Because data transformations could not remedy this degree of skew, this group was given the value of its nearest neighbor using a modified winsorizing procedure (Reifman & Keyton, 2010). This procedure did not affect the correlation between the average number of letters used per equation and the number of trials used to solve ($r = -.35$, $p = .010$) before and after winsorizing. The winsorized variable was still strongly skewed, but much less so (skewness = 2.95). In addition, winsorizing the outlier eliminated heterogeneity of variance across experimental conditions. Therefore, all reported results use the winsorized version of this variable.
4. The proportions of known-value hypotheses that problem solvers used were nonsignificantly negatively related to the proportions of uninformative equations and nonsignificantly positively related to the proportions of nonplausible-value hypotheses that they used (see Table 1). Therefore, overall, there is weak empirical evidence from this study that hypothesizing known values is indeed a problem-solving error. It is important to remember, however, that only groups' and the best individuals' problem worksheets were coded for problem-solving errors. Because they performed relatively well on the problem compared with all other participants and likely committed fewer errors than other participants, the variance on these variables is restricted, which may have resulted in attenuated correlation coefficients.

5. The number of letters used per equation was also tested. Results revealed no effects of stereotype threat, participation type, or their interaction on this variable.

References

Abrams, D., Hopthrow, T., Hulbert, L., & Frings, D. (2006). "Groupdrink"? The effect of alcohol on risk attraction among groups versus individuals. *Journal of Studies on Alcohol*, *6*, 628-636.

Allen, N. J., & Hecht, T. D. (2004). The "romance of teams": Toward an understanding of its psychological underpinnings and implications. *Journal of Occupational and Organizational Psychology*, *77*, 439-461. doi:10.1348/0963179042596469

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General*, *136*, 256-276. doi:10.1037/0096-3445.136.2.256

Blascovich, J., Mendes, W. B., Hunter, S. B., & Salomon, K. (1999). Social "facilitation" as challenge and threat. *Journal of Personality and Social Psychology*, *77*, 68-77. doi:10.1037/0022-3514.77.1.68

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*, 267-285. doi:10.1002/ejsp.145

DeCaro, M. S., Rotar, K. E., Kendra, M. S., & Beilock, S. L. (2010). Diagnosing and alleviating the impact of performance pressure on mathematical problem solving. *Quarterly Journal of Experimental Psychology*, *63*, 1619-1630. doi:10.1080/17470210903474286

Dooley, D., & Catalano, R. (1984). The epidemiology of economic stress. *American Journal of Community Psychology*, *12*, 387-409. doi:10.1007/BF00896502

Eccles, J. S., Jacobs, J. E., & Harold, R. D. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues*, *46*, 183-201. doi:10.1111/j.1540-4560.1990.tb01929.x

Frings, D., Hopthrow, T., Abrams, D., Hulbert, L., & Gutierrez, R. (2008). Groupdrink: The effects of alcohol and group process on vigilance errors. *Group Dynamics: Theory, Research, and Practice*, *12*, 179-190. doi:10.1037/1089-2699.12.3.179

Harlow, R. E., & Cantor, N. (1995). To whom do people turn when things go poorly? Task orientation and functional social contacts. *Journal of Personality and Social Psychology*, *69*, 329-340. doi:10.1037/0022-3514.69.2.329

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York, NY: Guilford Press.

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin*, *121*, 43-64. doi:10.1037/0033-2909.121.1.43

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

Inzlicht, M., & Schmader, T. (2011). *Stereotype threat: Theory, process, and application*. New York, NY: Oxford University Press.

Jamieson, J. P., & Harkins, S. G. (2010). Evaluation is necessary to produce stereotype threat performance effects. *Social Influence*, *5*, 75-86. doi:10.1080/15534510903512409

Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *Journal of Experimental Psychology: General*, *137*, 691-705. doi:10.1037/a0013834

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, *65*, 681-706. doi:10.1037/0022-3514.65.4.681

Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annual Review of Psychology*, *55*, 623-655. doi:10.1146/annurev.psych.55.090902.142009

Kulik, J. A., Mahler, H. I., & Moore, P. J. (1996). Social comparison and affiliation under threat: Effects on recovery from major surgery. *Journal of Personality and Social Psychology*, *71*, 967-979. doi:10.1037/0022-3514.71.5.967

Larson, J. R., Jr. (2010). *In search of synergy in small group performance*. New York, NY: Psychology Press.

Larson, J. R., Jr., & Christensen, C. (1993). Groups as problem-solving units: Toward a new meaning of social cognition. *British Journal of Social Psychology*, *32*, 5-30. doi:10.1111/j.2044-8309.1993.tb00983.x

Laughlin, P. R. (2011). *Group problem solving*. Princeton, NJ: Princeton University Press.

Laughlin, P. R., Bonner, B. L., & Miner, A. G. (2002). Groups perform better than the best individuals on letters-to-numbers problems. *Organizational Behavior and Human Decision Processes*, *88*, 605-620. doi:10.1016/S0749-5978(02)00003-1

Laughlin, P. R., Hatch, E. C., Silver, J. S., & Boh, L. (2006). Groups perform better than the best individuals on letters-to-numbers problems: Effects of group size. *Journal of Personality and Social Psychology*, *90*, 644-651. doi:10.1037/0022-3514.90.4.644

Laughlin, P. R., VanderStoep, S. W., & Hollingshead, A. B. (1991). Collective versus individual induction: Recognition of truth, rejection of error, and collective information processing. *Journal of Personality and Social Psychology*, *61*, 50-67. doi:10.1037/0022-3514.61.1.50

Laughlin, P. R., Zander, M. L., Knievel, E. M., & Tan, T. K. (2003). Groups perform better than the best individuals on letters-to-numbers problems: Informative equations and effective strategies. *Journal of Personality and Social Psychology*, *85*, 684-694. doi:10.1037/0022-3514.85.4.684

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal and coping*. New York, NY: Springer.

McGuire, G. M. (2007). Intimate work: A typology of the social support that workers provide to their network members. *Work and Occupations*, *34*, 125-147. doi:10.1177/0730888406297313

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*, 879-885. doi:10.1111/j.1467-9280.2007.01995.x

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Paulus, P. B., Dzindolet, M. T., Poletes, G., & Camacho, L. M. (1993). Perception of performance in group brainstorming: The illusion of group productivity. *Personality and Social Psychology Bulletin*, *19*, 78-89. doi:10.1177/0146167293191009

Picho, K., Rodriguez, A., & Finnie, L. (2013). Exploring the moderating role of context on the mathematics performance of females under stereotype threat: A meta-analysis. *The Journal of Social Psychology*, *153*, 299-333. doi:10.1080/00224545.2012.737380

Reifman, A., & Keyton, K. (2010). Winsorize. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 1636-1638). Thousand Oaks, CA: Sage.

Rosenthal, H. E. S., Crisp, R. J., & Suen, M.-W. (2007). Improving performance expectancies in stereotypic domains: Task relevance and the reduction of stereotype threat. *European Journal of Social Psychology*, *37*, 586-597. doi:10.1002/ejsp.379

Schachter, S. (1959). *The psychology of affiliation*. Stanford, CA: Stanford University Press.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, *38*, 194-201. doi:10.1006/jesp.2001.1500

Schmader, T. (2010). Stereotype threat deconstructed. *Current Directions in Psychological Science*, *19*, 14-18. doi:10.1177/0963721409359292

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, *85*, 440-452. doi:10.1037/0022-3514.85.3.440

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

Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. *Journal of Experimental Social Psychology*, *39*, 68-74. doi:10.1016/S0022-1031(02)00508-5

Shaw, M. E. (1932). Comparison of individuals and small groups in the rational solution of complex problems. *American Journal of Psychology*, *44*, 491-504.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*, 797-811. doi:10.1037/0022-3514.69.5.797

Taylor, S. E. (2007). Social support. In H. S. Friedman & R. C. Silver (Eds.), *Foundations of health psychology* (pp. 145-171). New York, NY: Oxford University Press.

Tomaka, J., Blascovich, J., Kelsey, R. M., & Leitten, C. L. (1993). Subjective, physiological, and behavioral effect of threat and challenge appraisal. *Journal of Personality and Social Psychology*, *18*, 616-624. doi:10.1037/0022-3514.65.2.248

Wout, D., Danso, H., Jackson, J., & Spencer, S. (2008). The many faces of stereotype threat: Group- and self-threat. *Journal of Experimental Social Psychology*, *44*, 792-799. doi:10.1016/j.jesp.2007.07.005

Zhang, S., Schmader, T., & Hall, W. M. (2013). L'eggo my ego: Reducing the gender gap in math by unlinking the self from performance. *Self and Identity*, *12*, 400-412. doi:10.1080/15298868.2012.687012

Author Biography

Nicholas P. Aramovich received his PhD in social psychology from the University of Illinois at Chicago. He was previously a lecturer of psychology at Harvard University and is currently a senior research manager at Peoplefluent in Waltham, Massachusetts. He is interested in small group dynamics and organizational behavior.