

UNDERSTANDING STEREOTYPE LIFT: ON THE ROLE OF THE SOCIAL SELF

David M. Marx
San Diego State University

Diederik A. Stapel
Tilburg University

Stereotype lift occurs when non-targets perform better in a stereotype-relevant testing situation compared to a testing situation that is less stereotype-relevant. The boundary conditions of this effect, however, are not well understood. To this end, the present research focuses on one critical moderator of stereotype lift, namely activation of non-targets' social self in stereotype-relevant (i.e., diagnostic) testing situations. Results showed that simply describing a test as diagnostic of ability did not lead to strong lift effects unless the social self was also activated (either indirectly or directly): non-targets need more "pushing" to activate their social self in diagnostic testing situations because they are *not* threatened by a negative stereotype. In contrast, we found that the test diagnosticity manipulation was enough to cause stereotype threat because targets need less pushing to activate their social self and the associated negative stereotype in diagnostic testing situations.

Considerable research has focused on the adverse effects of negative stereotypes on targets' test performance in stereotype-relevant situations (i.e., *stereotype threat*; Steele, 1997; Steele & Aronson, 1995). In contrast, far less work has focused on the test performance boost of non-targets in those same settings, for example, when male partici-

This research was conducted while the authors were at the University of Groningen and was supported by a "Pionier" Grant "Making sense of Hot Cognition" from the Dutch National Science Foundation and a research grant from the Heymans Institute of the University of Groningen, awarded to the second author. We thank Sei Jin Ko for her helpful comments on earlier versions of this article.

Correspondence concerning this article should be addressed to David M. Marx, Department of Psychology, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182. E-mail: dmarx@sciences.sdsu.edu.

pants' math test performance is higher in a diagnostic testing situation compared to a nondiagnostic one (i.e., *stereotype lift*; Walton & Cohen, 2003). Although both stereotype threat and stereotype lift effects may occur within the same testing situations little is known about *how* and *when* lift occurs relative to what is known about stereotype threat. Moreover, it is interesting to note that lift effects are not always found in experiments examining stereotype-based performance (e.g., Gonzales, Blanton, & Williams, 2002; Steele & Aronson, 1995), and when lift is found, some experiments show stronger lift effects than others. Thus, an important question is why do some experiments find stronger lift effects (e.g., Spencer, Steele, & Quinn, 1999, Experiment 2) than other experiments using similar participant populations (e.g., Gonzales et al., 2002)?¹ In the present research we argue that activation of one's social identity (i.e., the social self; see also Brewer & Gardner, 1996; Bromgard, Trafimow, & Bromgard, 2006; Stapel & Koomen, 2001)—particularly the social self or social identity that is most relevant to the current situation and task—plays a central role in determining how and when lift will occur. We further argue that the way in which one's social self is activated (either indirectly or directly) is an important, yet unexamined, moderating factor of stereotype lift effects.

STEREOTYPE LIFT AND THE SOCIAL SELF

When is stereotype lift more or less likely? We suggest that stereotype lift is more likely when the social self is activated in a diagnostic testing situation, because when the social self is activated the associated stereotypes are also likely to be activated; hence, non-targets should perform better due to the positive stereotype linked to their group. Support for this notion can be found in our past work on the preconditions for stereotype threat (Marx & Stapel, 2006; Marx, Stapel, & Muller, 2005). This research showed that activation of the

1. We are aware that when averaging across experiments, as Walton and Cohen (2003) did in their recent meta-analysis, the lift effect is reliable. Nevertheless, when attempting to further our understanding of stereotype lift and stereotype-based performance more generally, we think it is important to show lift effects within a single experiment, rather than rely on statistical power to achieve such effects. In other words, one reason why strong lift effects were not found in some experiments is because the manipulations used may not have triggered what we consider to be a critical moderating factor, namely activation of non-targets' social self.

social self led to poor test performance for stereotyped targets because their social self is linked to the negative stereotype about their group's ability. Indeed, this reasoning is grounded in one of the core principles of stereotype threat theory: "The more one is identified with the group about whom the negative stereotype exists, or the more one expects to be perceived as a member of that group, the more stereotype threat one should feel in situations where the stereotype applies" (Steele, Spencer, & Aronson, 2002, p. 391). Clearly, the social self plays a critical role in stereotype threat; therefore, it seems logical to assume that the same could be true for stereotype lift. After all, both effects are based upon the stereotypes associated with one's social self. So why do some experiments examining stereotype threat show strong lift effects while others do not?

One compelling reason why stereotype lift does not always occur in stereotype threat experiments is the fact that non-targets need more of a "push" to activate their social self compared to targets, thus some manipulations may be more effective than others in achieving this goal (cf. Walton & Cohen, 2003 for a discussion on the differences in threat manipulations). For stereotyped targets this manipulation issue seems less crucial, because they already have to contend with a negative stereotype about their group in these stereotype-relevant testing situations. That is, for targets, the situation (i.e., taking a diagnostic test) is potent enough to activate their social self and harm their test performance (Major & O'Brien, 2005; Marx, Brown, & Steele, 1999; Marx & Stapel, *in press*), while for non-targets this same situation may not be potent enough to activate their social self and thus cause stereotype lift. Given this, we suggest that activation of the social self in a diagnostic testing situation may lead to stronger lift effects. For instance, having non-targets indicate their gender before taking a diagnostic math test is a more direct reminder of their social self than simply describing a math test as diagnostic of ability: the test diagnosticity manipulation may be too indirect to activate non-targets' social self and the associated stereotype (cf. Walton & Cohen, 2003). In other words, neither one of these manipulations alone may be sufficient to cause strong lift effects. If both manipulations are combined, however, then they may be strong enough to activate non-targets' social self and the associated positive group-based stereotype, thus leading to bigger stereotype lift effects.

EXPERIMENTAL OVERVIEW

Using the meta-analytic work of Walton and Cohen (2003) and our own past theorizing regarding the role of the social self in stereotype-based performance, we argue that non-targets need more “pushing” to show stereotype lift, whereas targets need much less pushing to show stereotype threat. More specifically, we test the hypothesis that stronger lift effects will occur when both a manipulation of test diagnosticity and activation of the social self are used, whereas this should not be the case for stereotype threat because either manipulation alone is sufficient to activate targets’ social self. Moreover, we varied the way in which we activated the social self (either indirectly or directly) to show that lift effects get stronger as the manipulations of the social self become more direct. For some participants we activated the social self by having them circle group-based pronouns (indirect manipulation) while for other participants we activated their social self by having them indicate their gender on the cover of the test booklet (direct manipulation). We also included a condition in which participants’ personal self (“I”-ness) was activated in order to demonstrate that it is not simply self-activation that leads to lift or stereotype threat effects. Other participants just took a diagnostic or nondiagnostic test without any additional manipulations of the self. We examined our reasoning regarding stereotype lift and stereotype threat among male and female participants taking a challenging math test.

We expected that male participants would show stereotype lift and less apprehension when their social self is likewise activated before they take a diagnostic math test. We further expected that the lift effect would get stronger as the manipulation of the social self becomes more direct. For female participants, we predicted that their test performance and apprehension scores should not vary as a function of whether their social self is indirectly or directly activated before taking the test. That is, taking a diagnostic math test is sufficient to increase their apprehension and lower performance—no additional “pushing” is required. For both male and female participants we anticipated that activating the personal self would not have a strong effect on their math test performance or apprehension scores, thus showing that activation of the *social self* is necessary to show both stereotype lift and stereotype threat effects.

METHOD

PARTICIPANTS AND DESIGN

There were 90 female and 94 male Dutch undergraduates who took part in exchange for course credit or pay. For this experiment we used a 2 (Gender of Participant: female, male) \times 5 (Type of Condition: "We" prime + diagnostic test, "I" prime + diagnostic test, indicate gender + diagnostic test, diagnostic test, nondiagnostic test/control) between-participants design.

PROCEDURE

Upon entering the laboratory, participants were informed that they would be involved in a series of short unrelated tasks: all participants then completed a word search task and afterward took a math test. In addition, some participants indicated their gender on the cover of the math test after completing the word search task. The remaining participants merely took the math test without indicating their gender on the cover of the test booklet.

Word Search Task. For this task, participants were told that as part of a proofreading and word search exercise they would read a short paragraph detailing a trip to the city (see Brewer & Gardner, 1996; Marx & Stapel, 2006; Stapel & Koomen, 2001). As a way to manipulate participants' mindsets we asked those in the I Prime condition to circle all the personal pronouns (e.g., I, my, mine, me), those in the We Prime condition to circle all the group-based pronouns (e.g., we, our, ourselves, us), and those in the remaining conditions to circle the letters "x" "y," or "z," while they read the story. Of course for participants in the We Prime condition this "general" manipulation of the social self ("we-ness") may make a variety of social identities accessible, but once the context and situational cues are factored in (I am a person about to take a math test) then the social identity (gender) that is most relevant to the situation (taking a math test) may "win out" in the end (e.g., Davies, Spencer, & Steele, 2005; Major & O'Brien, 2005; Marx & Stapel, 2006, Onorato & Turner, 2004; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). If participants' social identity is linked to a stereotype (be it positive or negative), then it seems reasonable to suggest that their test performance should correspond to that stereotype (e.g., Marx et al., 1999).

Test Description Manipulation. For all conditions the test format resembled a standard Graduate Record Exam (GRE) math section, but varied in a number of important ways. In the Diagnostic conditions the test was described as being diagnostic of math ability as well as one that can identify a person's mathematical strengths and weaknesses. Moreover, written on the cover of the test booklet was the name of a fictitious testing center "Massachusetts Aptitude Assessment Center (MAAC)" followed by the label "Diagnostic Exam." In the Non-Diagnostic/Control condition the same test was described as a reasoning exercise without any mention of gender differences, thus not purposefully activating the negative stereotype about women and math. Test performance could range from 0–20.

Circle Task. After the test description manipulation, participants indicated how close they felt to four social groups (i.e., friends, gender, family, students). For each group, participants were given a set of five diagrams and asked to indicate which of the diagrams best represented the overlap they saw between themselves and the group in question. If, for instance, participants feel close to their gender group, then they should choose a diagram depicting a larger overlap between the self and their gender group (see Aron, Aron, & Smollan, 1992; Marx et al., 2005). We scored this task such that diagrams depicting the most overlap were assigned a (5), and diagrams depicting the least overlap were assigned a (1). Ratings were made separately for each of the social groups.

Apprehension Scores. To assess the amount of apprehension that participants experienced before they took the test we asked them to indicate how much they agreed with the following four statements: "I worry about my ability to perform well on this test"; "I am looking forward to taking this test"; "I worry that I am going to perform poorly on this test"; "I will enjoy taking this test". Responses were recorded on 9-point scales anchored by the terms (1) "strongly disagree" and (9) "strongly agree." After reverse scoring the positive statements we averaged the participant's responses to form a single apprehension score ($\alpha = .69$).²

When participants were finished with the math test they were debriefed and thanked.

2. We found a similar pattern of results when we conducted separate analyses for each of the statements. For reasons of parsimony, we only report effects from the composite item.

RESULTS AND DISCUSSION

MANIPULATION CHECK

Circle Task. To analyze the participants' responses on the Circle Task we conducted a 2 (Participant Gender) \times 5 (Condition) \times 4 (Social Group: gender, student, family, friends) mixed-model ANOVA, with Participant Gender and Condition as the between-participant factors and Social Group as the within-participant factor. Because we had a priori predictions about the Social Group scores, we examined this within-participant variable using a single-degree-of-freedom test. Specifically, we expected that participants would see more overlap between themselves and their gender group relative to the average of the other three social groups (i.e., the gender-overlap effect). As anticipated, there was a main effect of gender overlap, $F(1, 174) = 13.69, p < .05, \eta = .27$. Neither the Participant Gender ($p = .43$) nor the omnibus Condition main effect were reliable ($p = .10$); however, we did find a marginal interaction of Participant Gender and gender overlap, $F(1, 174) = 2.84, p = .09, \eta = .13$, such that the gender-overlap effect was somewhat larger for female participants ($M = 1.40$) compared to male participants ($M = .47$). In addition, we found an omnibus interaction of Condition and gender-overlap, $F(4, 174) = 3.97, p < .05, \eta = .29$ (other $F < 1.00$). To interpret this omnibus interaction we conducted a series of focused tests according to our theoretical framework.

The first comparison demonstrated that the gender-overlap effect was bigger in the Diagnostic ($M = 1.62$) compared to the Control condition ($M = -.13$), $F(1, 174) = 4.41, p < .05, \eta = .16$. Our second analysis showed that the gender-overlap effect was larger for the average of the We Prime and Gender conditions ($M = 1.81$) compared to the Control condition ($M = -.13$), $F(1, 174) = 7.78, p < .05, \eta = .21$. This was similarly the case when we compared the average gender-overlap effect from the We Prime and Gender conditions ($M = 1.81$) to the I Prime condition ($M = -.32$), $F(1, 174) = 9.31, p < .05, \eta = .23$. The final test examined whether our indirect and direct manipulations of the social self led to differences in the gender-overlap effect. Consistent with our reasoning we found that the gender-overlap effect was slightly larger in the Gender ($M = 2.61$)

relative to the We Prime ($M = 1.18$) condition, $F(1, 174) = 2.65, p = .11, \eta = .12$.³

In sum, these findings show that our social self manipulations were effective in activating those aspects of the social self (i.e., gender group) that are most relevant to the particular task and situation at hand. Our results further demonstrated that as the social self manipulation became more direct participants saw greater overlap between themselves and their gender group. The present findings also complement our previous work examining activation of the social self in stereotype threat testing situations (see Marx & Stapel, 2006; Marx et al., 2005).

MAIN ANALYSES

Math Test Performance. To examine the participants' test performance we conducted a 2 (Participant Gender) \times 5 (Condition) between-participants ANOVA (see Table 1). This analysis revealed a main effect of Participant Gender, $F(1, 174) = 44.58, p < .05, \eta = .45$, with male participants performing better than female participants, and an omnibus interaction, $F(4, 174) = 8.95, p < .05, \eta = .41$ (other $F < 1.00$). To interpret this omnibus interaction we conducted a series of focused analyses.

We first wanted to establish that our test description manipulation was effective in creating a stereotype-relevant situation. Accordingly, we compared male and female participants' math test performance within the Diagnostic and Control conditions. Results showed that male participants ($M = 10.55$) performed better than female participants ($M = 9.06$) in the Diagnostic condition, $F(1, 174) = 6.72, p < .05, \eta = .19$, but male ($M = 10.20$) and female participants ($M = 10.80$) performed similarly in the Control condition, $p > .27$, showing that our test description manipulation was successful. Next, we

3. We were quite surprised to find that the gender-overlap effect was larger in the Diagnostic ($M = 1.62$) relative to the We Prime ($M = 1.18$) condition. One reason why this occurred is because female participants in the Diagnostic condition ($M = 2.65$) showed a bigger gender-overlap effect compared to female participants in the We Prime condition ($M = 1.45$). However, consistent with our "push" logic, male participants had slightly higher gender-overlap scores in the We Prime ($M = .89$) compared to the Diagnostic ($M = .75$) condition. But what may be most critical to our reasoning is the finding that when the social self is directly activated (i.e., in the Gender condition) both male ($M = 1.93$) and female ($M = 3.25$) participants show a larger gender-overlap effect relative to the Diagnostic and We Prime conditions.

TABLE 1. Mean (*SD*) Math Test Performance and Apprehension Scores as a Function of Experimental Condition and Participant Gender

	Experimental Condition				Gender
	"I"	Control	Diagnostic	"We"	
Math Test Performance					
Male Participants	10.45 (1.36)	10.20 (2.04)	10.55 (2.01)	11.37 (1.89)	12.33 (2.02)
Female Participants	9.29 (0.85)	10.80 (1.58)	9.06 (2.05)	9.00 (1.59)	8.13 (1.67)
Apprehension Scores					
Male Participants	3.63 (0.71)	4.00 (0.32)	3.99 (0.76)	3.59 (0.76)	2.67 (0.82)
Female Participants	3.51 (0.53)	3.94 (0.91)	4.76 (0.97)	5.21 (0.62)	4.84 (1.06)

turned to the question of whether activation of the social self within a stereotype threat situation has any effect on participants' test performance. We investigated male and female participants' test performance separately because they should show different performance patterns as a function of our manipulations.

Male Participants. Based on previous stereotype lift theorizing (Walton & Cohen, 2003) we first examined the "classic" lift effect, namely whether male participants in the Diagnostic condition ($M = 10.55$) outperformed male participants in the Control condition ($M = 10.20$). This comparison was not reliable, $F < 1.00$, demonstrating that our test diagnosticity manipulation alone was not effective in "lifting" male participants' test performance.⁴ Next, we investigated whether the lift effect got stronger as manipulations of the social self became more direct (i.e., whether the way in which the social self is activated moderates the lift effect). To do this we tested the linear effect of our manipulations, such that we coded the Control condition as (-3), the Diagnostic condition as (-1), the We Prime condition as (+1) and the Gender condition as (+3). As expected this contrast was reliable, $F(1, 174) = 14.87, p < .05, \eta = .28$, such that the lift effect increased as the manipulation of the social self became more direct. Further analyses dem-

4. The Control condition could still have activated the negative stereotype about women and their reasoning ability as well as the positive stereotype about men and their reasoning ability. Hence, it is reasonable to suggest that we did not find a lift effect because our Control condition was not really a "control" condition per se; nevertheless, this alternative explanation seems unlikely considering that we found no math performance or social self activation difference between the I Prime and Control conditions.

onstrated that male participants' test performance was higher in the We Prime ($M = 11.37$) compared to the Control condition ($M = 10.20$), $F(1, 174) = 4.39, p < .05, \eta = .16$. They also performed better in the Gender condition ($M = 12.33$) relative to the Control conditions, $F(1, 174) = 12.81, p < .05, \eta = .26$. However, their test performance was only slightly higher in the Gender compared to the We Prime conditions, $F(1, 174) = 2.54, p = .11, \eta = .12$. Importantly, the results showed that male participants' performance in the I Prime condition ($M = 10.45$) did not differ from male participants' performance in the Diagnostic ($M = 10.55$) or Control conditions ($M = 10.20$), $F_s < 1.00$, indicating that it is *social* rather than *personal* self activation that contributes to lift effects. In conclusion, our data demonstrate that stereotype lift is stronger when male participants' social self is also activated in a diagnostic testing situation. Moreover, the strength of the lift effect increased as manipulations of the social self got more direct.

Female Participants. To examine whether our test description manipulation differentially affected female participants' test performance, we first compared the performance of female participants in the Diagnostic condition ($M = 9.06$) to the performance of female participants in the Control condition ($M = 10.80$). This comparison was reliable, $F(1, 174) = 9.16, p < .05, \eta = .22$, demonstrating a typical threat effect (e.g., Steele, 1997). Moreover, there were no differences between the Diagnostic ($M = 9.06$) condition and the We Prime ($M = 9.00$) condition, $F < 1.00$, but there was a slight difference between the Gender ($M = 8.13$) and Diagnostic ($M = 9.06$) conditions, $F(1, 174) = 2.35, p < .13, \eta = .12$, showing that our direct manipulation of the social self did somewhat enhance the threat effect. Interestingly, and unlike past research on stereotype-based performance and individuation (e.g., Ambady, Paik, Steele, Owen-Smith, & Mitchell, 2004), female participants did *not* show a boost in performance when their personal self was activated relative to the Diagnostic condition, $F < 1.00$. We believe that this "null" effect arose because our manipulation of the personal self (circle personal pronouns) may have been less direct than having participants write down self-descriptions as well as indicate a number of personal qualities (Ambady et al., 2004).

In sum, our performance results demonstrate that activation of the social self can have both deleterious as well as beneficial effects depending on one's group membership. However, for male participants, they needed more of a "push" to activate their social self when taking a stereotype-relevant test compared to female participants.

As a result, we found a stronger lift effect when the social self was activated before taking a stereotype-relevant test and this effect became larger when the social self was indirectly ("we" prime) and directly (indicate gender) activated. Taking a stereotype-relevant test, on the other hand, was sufficient to activate female participant's social self and harm their test performance. Next, we examined male and female participant's apprehension scores as a function of our manipulations.

Apprehension Scores. Is stereotype lift associated with reduced apprehension (Walton & Cohen, 2003)? To address this question we conducted a 2 (Participant Gender) \times 5 (Condition) between-participants ANOVA on the participants' apprehension scores (see Table 1). This analysis revealed a main effect of Gender, $F(1, 174) = 60.17, p < .05, \eta = .51$, with female participants experiencing more apprehension than male participants, and an omnibus main effect of Condition, $F(4, 174) = 8.56, p < .05, \eta = .40$. We also found an omnibus Gender by Condition interaction, $F(4, 174) = 15.46, p < .05, \eta = .51$. To interpret this interaction we conducted simple effects tests within participant gender.

Male Participants. In the Diagnostic condition ($M = 3.99$) male participants felt the same level of apprehension as male participants in the Control condition ($M = 4.00$), $F < 1.00$. Just as we had done with male participants' test performance we tested the linear effect of social self activation on their apprehension scores. This contrast was reliable, $F(1, 174) = 28.53, p < .05, \eta = .38$, showing that as the manipulation of the social self became more direct, male participants felt *less* apprehension prior to taking the test. We also found that male participants felt less apprehension in the Gender condition ($M = 2.67$), $F(1, 174) = 25.87, p < .05, \eta = .36$, relative to the Control condition ($M = 4.00$), male participants in the We Prime condition ($M = 3.59$) experienced slightly less apprehension than male participants in the Control condition ($M = 4.00$), $F(1, 174) = 2.80, p < .10, \eta = .13$, however, male participants felt less apprehension in the Gender condition compared to the We Prime condition, $F(1, 174) = 12.10, p < .05, \eta = .25$. This pattern of effects is consistent with the notion that lift is associated with reduced apprehension or concern (Walton & Cohen, 2003).

Female Participants. Female participants in the Diagnostic condition ($M = 4.76$) had higher apprehension scores than female participants in the Control condition ($M = 3.94$), $F(1, 174) = 10.54, p < .05, \eta =$

.24, as would be consistent with stereotype threat theory. We further found that female participants had higher apprehension scores in the We Prime ($M = 5.21$), $F(1, 174) = 27.52$, $p < .05$, $\eta = .37$, and Gender conditions ($M = 4.84$), $F(1, 174) = 12.29$, $p < .05$, $\eta = .26$, relative to the Control condition ($M = 3.94$). As anticipated, there were no differences in female participants' apprehension scores between the Diagnostic ($M = 4.76$) and Gender conditions ($M = 4.84$), $F < 1.00$. We did find a marginal difference between the We Prime ($M = 5.21$) and Diagnostic conditions ($M = 4.76$), $F(1, 174) = 3.18$, $p < .08$, $\eta = .13$, showing that this particular social self manipulation somewhat enhanced female participants' apprehension scores.

DISCUSSION

The present experiment provides strong support for our notion regarding the role of the social self in stereotype lift. Namely, we found that when the social self is activated (i.e., either indirectly or directly) before taking a diagnostic math test, male participants' performance is enhanced and their apprehension reduced relative to those situations where the social self is not activated (i.e., the Control condition). Our data further showed that stereotype lift was not present for male participants in the Diagnostic condition, demonstrating that male participants may need more of a "push" to show stereotype lift. For female participants, however, the manner in which the social self was activated (be it indirect or direct) had much less of an impact. That is, female participants performed worse in both social self conditions (i.e., We Prime and Gender) and Diagnostic condition relative to the Control condition: "extra" activation of the social self was not necessary because the links between the test description, targets' social self, and the negative group stereotype were already strong enough to harm their test performance (Major & O'Brien, 2005; Marx et al., 1999; Marx & Stapel, 2006).

Considerable research has investigated the adverse effects of negative stereotypes on stereotyped targets' test performance, but far less research has been conducted on the enhanced performance of non-targets in those same testing situations. Although stereotype lift appears now and then in research examining stereotype-based performance effects, the strength of this effect is rather variable. Hence, it seemed important to highlight *when* and *how* lift occurs in order to provide insight into non-targets' performance in stereotype-rele-

vant situations as this may help researchers to understand some of the factors contributing to the gender and racial gaps in high-stakes testing situations (e.g., the Graduate Record Exam, or Scholastic Assessment Test). From our perspective, highlighting how both positive (i.e., lift effects) and negative stereotypes (i.e., threat effects) affect performance may likewise shed light on what contributes to this test score gap, and thus aid researchers and educators in the development of strategies for helping all students perform up their abilities.

Throughout this article, we have argued and demonstrated that additional activation of the social self (i.e., prime “we-ness,” indicate gender) in a diagnostic testing situation contributes to the lift effect, and as manipulations of the social self get more direct the lift effects gets larger. In the current research the test description alone was not enough to cause stereotype lift, presumably because the link between the test description and the positive group-based stereotype was not sufficient to activate non-targets’ social self. But, once that link was made clearer via more direct manipulations of the social self (e.g., indicate gender) the lift effect became more pronounced. Thus, just like for targets in the diagnostic testing situation, activation of the social self can have profound effects for non-targets in similar testing situations. In those situations, however, non-targets benefit whereas targets lose out. Although our findings are only a first experimental step toward an understanding of stereotype lift, the present findings still help explain why lift effects are more “elusive” than stereotype threat effects.

Some may view the present research as contradicting the meta-analytic work of Walton and Cohen (2003) in which they document that both *direct* (i.e., the test indicates gender / racial differences) and *indirect* (i.e., the test is diagnostic of ability) manipulations of threat lead to approximately the same lift effect—non-targets performing better in a stereotype-relevant testing situation compared to a situation that is less stereotype-relevant. On the contrary, we believe that our research highlights one of the core aspects of stereotype threat theory (the social self) and shows how this factor applies to both stereotype threat and stereotype lift effects. Thus, by focusing on “both sides of the coin” we are in a unique position both to extend the important work laid out by Walton and Cohen (2003) as well as to provide additional support for stereotype threat theory—after all, lift and threat are the mirror images of each other, hence pinpointing one of the uni-

fyng factors of lift and threat adds rather than subtracts from our knowledge of stereotype-based performance. Indeed, testing the strength of different manipulations of threat and lift effects within a single experiment has proven to be particularly fruitful in elucidating the underlying processes of stereotype-based performance effects more broadly (see Marx & Stapel, 2006), and thus may provide researchers with the experimental procedures to turn both stereotype threat and stereotype lift “on” and/or “off.”

It is clear that our effects are a part of a larger quest focused on identifying the moderating factors of stereotype-based performance. Indeed, there are a number of important moderating factors that could affect the strength of stereotype-based performance effects more generally and lift effects more specifically (see Walton & Cohen, 2003). For instance, it seems reasonable to suggest that how one views the stereotypes about one’s group (i.e., stereotype endorsement) could contribute to stronger *lift* effects when the stereotype is positive and to stronger *threat* effects when the group stereotype is negative (e.g., Blanton, Christie, & Dye, 2002). Moreover, consistent with Steele et al.’s (2002) reasoning regarding the role of domain identification in stereotype threat effects, it would seem that domain identification could also play an important role in the strength of the lift effect. Finally, the level of one’s group identification (e.g., Schmader, 2002) may reduce the amount of “pushing” needed to create lift. In other words, if men already feel very identified with their gender group, then they may only need a subtle reminder of their social self for lift to occur, whereas they may need much more reminding when they are not strongly identified with their gender group.

CLOSING THOUGHTS

In short, prior research examining stereotype-based performance effects has provided much needed insight into the factors moderating the test performance of both targets and non-targets in high-stakes testing situations. The advantage of the present research is that we provide a global framework for understanding stereotype-based performance more generally. In the end, these findings not only provide the first experimental evidence for how to enhance or moderate the lift effect, but they also provide converging evidence for our earlier reasoning regarding the role of the social self in stereotype threat

effects (Marx & Stapel, 2006; Marx et al., 2005). For that matter, knowing how both targets and non-targets experience a stereotype-relevant testing situation can help provide more insight into the ways in which to reduce the adverse effects of negative stereotypes, as well as shine additional light on the gender and racial test score gap in evaluative testing situations. We are not the first to document or focus on stereotype lift effects, but we are certainly the first to show experimentally *how* and *when* stereotype lift occurs as well as what contributes to strong or weak lift effects.

REFERENCES

- Ambady, N., Paik, S. K., Steele, J., Owen-Smith, A., & Mitchell, J. P. (2004). Deflecting negative self-relevant stereotype activation: The effects of individuation. *Journal of Experimental Social Psychology, 40*, 401–408.
- Aron, A., Aron, E. A., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63*, 596–612.
- Blanton, H., Christie, C., & Dye, M. (2002). Social identity versus reference frame comparisons: The moderating role of stereotype endorsement. *Journal of Experimental Social Psychology, 38*, 253–267.
- Brewer, M. B., & Gardner, W. (1996). Who is this “we”? Levels of collective identity and self-representations. *Journal of Personality and Social Psychology, 71*, 83–93.
- Bromgard, G. D., Trafimow, D., & Bromgard, I. K. (2006). Valence of self-cognitions: The positivity of individual self-statements. *Journal of Social Psychology, 146*, 85–94.
- Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women’s leadership aspirations. *Journal of Personality and Social Psychology, 88*, 276–287.
- 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*, 659–670.
- Major, B., & O’Brien, L. T. (2005). The social psychology of stigma. *Annual Review of Psychology, 56*, 393–421.
- Marx, D. M., Brown, J. L., & Steele, C. M. (1999). Allport’s legacy and the situational press of stereotypes. *Journal of Social Issues, 55*, 491–502.
- Marx, D. M., & Stapel, D. A. (2006). Distinguishing stereotype threat from priming effects: On the role of the social self and threat-based concerns. *Journal of Personality and Social Psychology, 91*, 243–259.
- Marx, D. M., Stapel, D. A., & Muller, D. (2005). We can do it: The interplay of construal orientation and social comparisons under threat. *Journal of Personality and Social Psychology, 88*, 432–446.
- Onorato, R. S., & Turner, J. C. (2004). Fluidity in the self-concept: The shift from

- personal to social identity. *European Journal of Social Psychology*, 34, 257–278.
- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, 38, 194–201.
- Spencer, S. J., Steele, C. M., & Quinn, D. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35, 4–28.
- Stapel, D. A., & Koomen, W. (2001). I, we, and the effects of others on me: How self-construal level moderates social comparison effects. *Journal of Personality and Social Psychology*, 80, 766–781.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613–629.
- Steele, C. M., & Aronson, J. (1995). Stereotype vulnerability and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 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. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 34, pp. 379–440). San Diego, CA: Academic Press.
- Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, UK: Blackwell.
- Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, 39, 456–467.

