

We Can Do It: The Interplay of Construal Orientation and Social Comparisons Under Threat

David M. Marx and Diederik A. Stapel
University of Groningen

Dominique Muller
University of Colorado and University of Pierre Mendes France

The authors investigated how a collective self-construal orientation in combination with positive social comparisons “turns off” the negative effects of stereotype threat. Specifically, Experiment 1 demonstrated that stereotype threat led to increased accessibility of participants’ collective self (“we”). Experiment 2 showed that this feeling of “we-ness” in the stereotype threat condition centered on the participants’ stereotyped group membership and not on other important social groups (e.g., students). Experiment 3 indicated that in threat situations, when participants’ collective self is accessible, positive social comparison information led to improved math test performance and less concern, whereas in nonthreat situations, when the collective self is less accessible, positive comparison information led to worse test performance and more concern. Our final experiment revealed that under stereotype threat, only those comparison targets who are competent in the relevant domain (math), rather than in domains unrelated to math (athletics), enhanced participants’ math test performance.

Keywords: stereotype threat, self-construal orientation, social comparisons, performance

Over the years an increasing amount of research has focused on how negative stereotypes can lead the targets of those stereotypes to underperform on challenging tests. This phenomenon, known as *stereotype threat* (Steele, 1997; Steele, Spencer, & Aronson, 2002), seeks to explain, from the target’s perspective, why certain groups perform worse than their motivations and prior performances suggest they should. Indeed, Steele et al. (2002) claimed that “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” (p. 391). By now this notion of a group-based threat is well documented in the literature. For instance, research has demonstrated the adverse

effects of stereotype threat on female students’ math test performance (R. P. Brown & Josephs, 1999; Spencer, Steele, & Quinn, 1999) and Black students’ verbal test performance (Marx & Goff, in press; Steele & Aronson, 1995) in addition to a variety of other groups and performance domains (e.g., Croizet & Claire, 1998; Gonzales, Blanton, & Williams, 2002; Leyens, Desert, Croizet, & Darcis, 2000; Stone, Lynch, Sjomeling, & Darley, 1999).

To date, the bulk of stereotype threat research has focused largely on cataloging the types of groups and areas in which the theory applies; thus, the challenge currently facing researchers is identification of the primary underlying mechanisms (cf. Maass & Cadinu, 2003; Marx & Stapel, 2004; Steele et al., 2002; Wheeler & Petty, 2001). Clearly, understanding stereotype threat is critical because it can point researchers toward ways to reduce the threat, particularly if the mechanism is connected to the core principle of stereotype threat—a concern rooted in beliefs about one’s group image (Steele, 1997; Steele et al., 2002). In the current research, we contend that a collective self-construal orientation (i.e., a mindset in which thoughts about one’s group are highly accessible) is directly related to this primary aspect of the theory. In our opinion, when attempting to further understanding of stereotype threat, it seems sensible to focus on processes related to thoughts about one’s group and the associated stereotype because those thoughts are clearly relevant to the main assumption of stereotype threat theory (Steele, 1997; Steele et al., 2002). Hence, if stereotype threat occurs because it leads to heightened accessibility of a group-based stereotype, then the salience of group memberships associated with that stereotype should be greater in those settings in which the negative stereotype applies than it is in ones in which it does not. From this perspective, it seems quite apparent that a collective self-construal orientation (Aron, Aron, & Smollan, 1992; Brewer & Gardner, 1996; Gardner, Gabriel, & Hochschild, 2002) is the mindset for those individuals targeted by stereotype threat, whereas in nonstereotype threat situations those same indi-

David M. Marx and Diederik A. Stapel, Department of Social and Organizational Psychology, University of Groningen, Groningen, The Netherlands; Dominique Muller, Department of Psychology, University of Colorado, and Department of Psychology, University of Pierre Mendes France, Grenoble, France.

Dominique Muller is now at the Department of Psychology, Paris 5 University, Paris, France.

This research was supported in part by Pionier Grant “Making Sense of Hot Cognition,” awarded to Diederik A. Stapel by the Dutch National Science Foundation. Dominique Muller benefited from a Lavoisier scholarship from the Ministère des Affaires Étrangères, France. We thank Karlijn Massar and Lennart Renkema for translation of the experimental materials. We also thank Marcel Zeelenberg for his help with execution of the experiments at the University of Tilburg, The Netherlands. We are grateful to Ernestine Gordijn, Sei Jin Ko, and Sabine Otten for their helpful suggestions and comments on earlier versions of this article.

Correspondence concerning this article should be addressed to David M. Marx or Diederik A. Stapel, Department of Social and Organizational Psychology, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands. E-mail: d.marx@ppsw.rug.nl or d.a.stapel@ppsw.rug.nl

viduals' mindset should be less focused on the collective self. Given this mindset, we further argue that stereotyped targets in stereotype threat situations should be more sensitive to group-based information, which if positive (e.g., learning about a fellow group member who disconfirms the negative stereotype), could reduce their impression-related concerns and ultimately lead to enhanced test performance.

Understanding Stereotype Threat

Research exploring the mechanisms of stereotype threat has made great strides in understanding stereotype threat (e.g., Maass & Cadinu, 2003; Steele et al., 2002; Wheeler & Petty, 2001) as well as in pinpointing its effects on stereotyped targets' physiological (e.g., Blascovich, Spencer, Quinn, & Steele, 2001), cognitive (e.g., Davies, Spencer, Quinn, & Gerardstein, 2002; Schmader & Johns, 2003), and emotional reactions (e.g., Gonzales et al., 2002; O'Brien & Crandall, 2003; Spencer et al., 1999). However, this research has not focused on what it is that defines stereotype threat: contending with one's group image (Steele, 1997; Steele et al., 2002). Thus, we know very little about the *mindset* of those individuals targeted by the stereotype and how those mindsets relate to the targets' test performance in stereotype threat and nonthreat situations (but see Davies et al., 2002; Quinn & Spencer, 2001; Steele & Aronson, 1995, for work on stereotype activation among stereotyped participants under stereotype threat conditions).

In the present experiments, we approached this mindset question by targeting processes and thoughts related to one's group membership because those thoughts are theorized to be the most applicable to stereotype threat (Steele & Aronson, 1995). To this end, we argue and demonstrate that because of the salience of an individual's collective self ("we") in those contexts in which that individual's group image is at stake, learning about another group member who is talented on the stereotyped task may actually "turn off" the negative effects of the threat (Marx & Roman, 2002). This line of reasoning suggests that *positive* social comparisons in stereotype threat situations may lead to different performance results than would such comparisons in nonstereotype threat situations.¹ That is, exposure to positive social comparisons could provide individuals with stereotype-inconsistent information, thereby leading to lowered impression-related concerns in the testing situation and, ultimately, to enhanced test performance. We base this argument (i.e., stereotype threat → accessibility of the collective self → assimilative social comparison effects) on recent experiments exploring construal orientation ("I" vs. "we") and the direction of self-evaluative and performance-related social comparison effects (e.g., Mussweiler & Strack, 2000; Stapel & Suls, 2004). Specifically, this research demonstrates that when the collective self is accessible (as would be the case under stereotype threat conditions) both self-evaluations and performance can result in assimilation effects.

Assimilation Versus Contrast

We use social comparisons to obtain self-knowledge by relating thoughts and behaviors to those around us (Collins, 1996; Festinger, 1954; Wood, 1989). However, when it has involved the

specific direction of social comparison effects, prior research has been somewhat mixed (Blanton, 2001). Namely, what determines whether the social comparison process results in assimilation or contrast effects? To address this question, Stapel and Koomen (2001) conducted a series of experiments that showed that when a collective self-construal orientation ("we-ness") is activated social comparisons lead to assimilative self-evaluations, but when a personal self-construal orientation ("I-ness") is activated social comparisons lead to contrastive self-evaluations. These researchers interpreted their results as follows. When an individual's personal self is more accessible that individual is thought to be in an "I" frame of mind and is likely to value being distinct, therefore focusing on how the self and comparison others are different or distinct (thus *not* deriving much benefit from a fellow group member's success). However, when a person's social self is more accessible that person shifts into a "we" frame of mind and is likely to value being part of a group and, as a result, stress *similarities* with comparison others—thus deriving more benefit from a fellow group member's success (Aron et al., 1992; Brewer & Gardner, 1996; Brewer & Weber, 1994; Gardner et al., 2002; Schubert & Otten, 2002; Tajfel & Turner, 1979).

Similar to the Stapel and Koomen (2001) perspective, other social comparison research found assimilative effects when the comparison target and the perceiver shared the same group membership (e.g., Blanton, Crocker, & Miller, 2000; Spears, Gordijn, Dijksterhuis, & Stapel, 2004). Research by Blanton et al. (2000), for instance, showed that positive social comparisons in a stereotyped domain led to assimilative effects on participants' self-esteem after receiving negative feedback, but only when the comparison target was from the same stereotyped group. Hence, in those settings in which individuals are focused on their group membership, they should demonstrate assimilation when they compare with a talented in-group member (Blanton, Christie, & Dye, 2002; Brewer & Weber, 1994). Given that past research has shown that a collective mindset leads to assimilative effects, we believe that the same pattern of results could apply to *performance* in a stereotype threat situation. That is, because reminding stereotyped targets about a group-based stereotype likewise increases accessibility of their stereotyped identity (see also, Steele, 1997; Steele et al., 2002) and thus their collective self (Blanton et al., 2000, 2002), we argue that the same type of assimilation effects could occur for performance as they do for self-evaluations (Mussweiler & Strack, 2000; Stapel & Suls, 2004).

Interplay of Construal Orientation and Social Comparisons Under Threat

On the basis of this theoretical framework, we claim that because a stereotype threat situation enhances feelings of "we-ness," exposure to positive social comparison information can actually turn off the negative effects of stereotype threat on the performance of stereotyped individuals. This reasoning suggests that when provided with positive social comparison information under stereotype threat conditions, targets of the stereotype could use this

¹ We chose to use the terms *positive* and *negative* social comparison because we think these terms are more comprehensive and clearer than the more commonly used terms, *upward* and *downward*.

information as evidence that the stereotype is not always applicable to their group and that they too may be able to overcome its negative effects. Moreover, this positive group-based information may lessen the impression-related concerns of stereotyped targets (Blanton et al., 2002) because knowing that a fellow group member has done well in the stereotyped domain could likewise reduce targets' unique concerns about the impression they are making in the testing situation. After all, when the collective self is activated, it seems reasonable to argue that group-based success might also be incorporated into the self, thus leading to assimilation effects for both performance (behavior) and self-evaluations (Mussweiler & Strack, 2000; Stapel & Suls, 2004).

When stereotype threat is absent, however, positive social comparison information should cease to be as effective given that the participants' collective self is less salient at that time. To illustrate, when Mary is taking a test under stereotype threat conditions her collective self should be more salient than her personal self. Because of this, assimilative performance effects are more likely in the former case because of Mary's "we-ness" mindset and contrastive performance effects are more likely in the latter case because of her "I-ness" mindset (e.g., Brewer & Weber, 1994; Stapel & Koomen, 2001). So, in a stereotype threat situation, when presented with positive comparison information about Susan's math ability, Mary's impression-related concerns might be reduced and her math test performance enhanced, but when presented with negative comparison information about Susan's math ability, Mary's concern may be increased and her test performance may suffer. As can be seen from this example, Mary could easily overcome stereotype threat after receiving positive social comparison information about Susan or succumb to it after receiving negative social comparison information about Susan. Indeed, when stereotype threat is not present, Mary's impression-related concerns and test performance should show a different pattern of effects because her collective self is less accessible in that situation. In this case, Mary should feel more concern and perform more poorly when exposed to positive social comparisons and less concern accompanied by better performance when exposed to negative social comparisons (e.g., J. Brown, Novick, Lord, & Richards, 1992; Festinger, 1954).

To test this logic, we conducted four experiments. Specifically, we demonstrate that in stereotype threat settings (when a "we-ness" mindset is more salient) targets of a negative stereotype report less concern and thus perform better when provided with positive social comparison information, whereas they report more concern and perform worse when provided with negative social comparison information. However, in a nonstereotype threat situation, when the collective self is less salient, positive social comparisons result in more concern and lower performance and negative social comparisons in less concern and higher performance. We believe that these assertions have critical implications for stereotype threat theory, as they not only highlight one of the core aspects of stereotype threat but also highlight how the negative effects of the stereotype about women and math can be turned off.

Research Overview

In this article we set out to test our reasoning that stereotype threat leads to heightened salience of the collective self and, because of this, positive social comparison information leads to

enhanced performance, whereas in those situations in which the group-based stereotype does not apply, such comparison information does not enhance performance as much. By comparing positive and negative social comparisons under stereotype threat and nonstereotype threat conditions, we are in a better position to identify how the effects of social comparisons made under threat differ from social comparisons made in nonthreatening circumstances.

Specifically, in Experiment 1 we measured the number of group-based pronouns female and male participants listed on a language translation exercise. We did this to assess which construal orientation was most accessible in a stereotype threat situation that subtly reminded the participants about the negative gender stereotype concerning women and math compared with a nonstereotype threat situation that did not remind them about the relevant stereotype. For Experiment 2, we measured which aspect of the female participants' collective self showed the biggest difference in accessibility effects between a stereotype threat and a nonstereotype threat setting by asking them to choose diagrams depicting increasing levels of overlap between the self and a number of different social groups (e.g., gender, students). We further measured aspects of the collective self by having the participants respond to questions about their gender group identification. For Experiment 3, we examined how positive social comparison information can alter typical stereotype threat effects (Marx & Roman, 2002) as well as how such comparison information leads to assimilative performance when the stereotype is relevant and to less of a boost in performance when the stereotype is irrelevant. Inclusion of these two conditions (a stereotype threat and a nonstereotype threat condition) was essential because they permitted us to highlight the powerful buffering nature of positive social comparisons in a stereotype threat context and how the comparison process functions differently when the group-based stereotype does not apply. Furthermore, we investigated a potential mediator of positive social comparisons in stereotype threat situations, namely impression-related concerns. Experiment 4 was undertaken to examine whether stereotyped individuals benefit more from a comparison target's success that maps on to the relevant dimension; that is, under stereotype threat conditions do female participants perceive more similarity and increased math test performance when the comparison target has excelled in math rather than athletics? Accordingly, we provided female participants with two positive comparison targets who demonstrated competence either in math (relevant domain) or in athletics (irrelevant domain) and then assessed their impact on the participants' perceived similarity and math test performance.

Experiment 1: Construal Orientation Under Threat

The purpose of Experiment 1 was to investigate whether self-construal orientation differs in a stereotype threat versus a nonstereotype threat situation, particularly for the targets of the stereotype. To do this, we described a math test as diagnostic of math ability (the diagnostic condition), thus making a group-based stereotype relevant to the targets of the stereotype but not to those individuals who were not targeted by the stereotype (see Steele & Aronson, 1995). In the nondiagnostic condition the same test was described as a reasoning exercise; therefore, it did not purposefully link performance on the test to the negative gender stereotype. Our

main interest was whether the diagnostic rather than the nondiagnostic condition led to increased accessibility of a collective self-construal orientation (“we-ness”), particularly for female participants. After learning about the test they would take, the female and male participants were asked to complete a short language translation exercise to assess the accessibility of their collective self (“we”).

We predicted that female participants would indicate more group-based pronouns in the diagnostic condition relative to female participants in the nondiagnostic condition because the negative stereotype about women and math should not be relevant in the latter case; hence, their collective self should be less salient to them. For male participants, we expected no difference between the diagnostic and nondiagnostic conditions because the negative stereotype is not applicable to their gender group. This pattern of results would provide initial evidence that under stereotype threat conditions the collective self is more accessible for those participants (women) who are the targets of the stereotype compared with conditions in which the negative stereotype is not as relevant (the nondiagnostic condition).

Method

Participants and Design

Participants were 22 female and 20 male Dutch undergraduates who took part in exchange for course credit or pay.² For this experiment we used a 2 (sex of participant: female, male) \times 2 (test description: diagnostic, nondiagnostic) between-participants design.

Procedure

On entering the laboratory, participants were told that the experiment was about different aspects of academic ability and that part of the experiment involved an exercise to measure their current analytical thinking (the math test) as well as a reading comprehension exercise (the language translation exercise). Test description served as the manipulation of stereotype threat.

Test description manipulation. In the diagnostic condition the test was described as being diagnostic of math ability as well as being able to identify a person’s mathematical strengths and weaknesses. This procedure has successfully created a situation of stereotype threat in previous research (Steele & Aronson, 1995). In the nondiagnostic condition the math test was described as a reasoning exercise, thus not purposefully activating the negative stereotype about women and math. Furthermore, in both conditions participants were told that they would receive feedback about their test performance at the conclusion of the experiment (in actuality they never took the test, therefore no feedback was given). By including this bogus feedback information we hoped to equate the participants’ task motivation across conditions.

Language translation exercise. After being told about the test, the participants were asked to complete a short language translation exercise (see Dijksterhuis & van Knippenberg, 2000; Stapel & Tesser, 2001) as a way to assess how reading comprehension affects analytic ability. For this exercise, the experimenter explained to the participants that earlier research had shown that while reading a foreign language people are sometimes able to guess the correct translation of pronouns. After introducing this task the participants were given a paragraph purportedly written in a foreign language (Wezwe; a language spoken only in New Guinea) with 20 pronouns underlined. The participants’ assignment was to translate each of the pronouns. Of primary interest was the number of group-based pronouns (e.g., *we*, *us*) each participant listed. On completion of the language translation exercise the participants were thanked and debriefed.

Results and Discussion

Our main hypothesis was that female participants in the diagnostic condition would list more group-based pronouns compared with female participants in the nondiagnostic condition but that male participants would not show a difference in the number of group-based pronouns listed as a function of test description. To examine this prediction, we conducted a 2 (sex of participant) \times 2 (test description) analysis of variance (ANOVA) on the number of group-based pronouns listed by each participant. This analysis yielded only the expected Sex of Participant \times Test Description interaction, $F(1, 38) = 4.37, p = .04, \eta = .32$ (all other effects, $ps > .20$).

Tests of simple effects revealed that female participants in the diagnostic condition ($M = 6.83$) listed more group-based pronouns than did female participants in the nondiagnostic condition ($M = 4.90$), $F(1, 38) = 5.75, p = .02, \eta = .36$. For male participants, there was no reliable difference between the diagnostic ($M = 4.80$) and nondiagnostic conditions ($M = 5.30$), $F(1, 38) = 0.35, p = .56, \eta = .10$. A final comparison revealed that, in the diagnostic condition, female participants listed more group-based pronouns ($M = 6.83$) than did male participants ($M = 4.80$), $F(1, 38) = 6.36, p < .02, \eta = .38$, demonstrating that when the negative gender stereotype is activated, the collective self (“we”) was more accessible for female participants than it was for male participants.

In short, these results clearly support our hypothesis that when targets of a negative stereotype are in a stereotype threat situation the salience of their collective self is increased compared with when the stereotype is not applicable (i.e., in the nondiagnostic condition). Furthermore, this experiment demonstrates that this difference in construal orientation is unique to stereotyped individuals, therefore providing critical empirical support for one of the core theoretical assumptions of stereotype threat theory: Stereotype threat primes a collective mindset for the targets of the stereotype (Steele, 1997; Steele et al., 2002).

Having established that stereotype threat increases accessibility of the collective self, our next step was to investigate whether this construal orientation is a general “we-ness” effect, indicating feelings of closeness with a number of important social groups (i.e., family, friends, gender, students) or whether it focuses on a specific type of “we-ness,” indicating identification with the group (i.e., women) that is most relevant to the particular situation (i.e., a math test-taking situation). This issue also relates to one of the main features of social identity (Tajfel & Turner, 1979) and self-categorization theory (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987), which posits that the self and other social groups can be defined at varying levels of overlap—that is, ranging from very little overlap with some groups to almost complete overlap with other groups. The goal of Experiment 2, therefore, was to address which aspect of the participants’ collective self would show the biggest difference in accessibility effects as a function of our test description manipulation (diagnostic vs. nondiagnostic). As far as we know, this has not been tested directly within a stereotype threat framework until now.

² All of the experiments were conducted in Dutch. Throughout this article, examples of materials used are the English equivalents of the Dutch materials used in those experiments.

Experiment 2: Type of Collective Self-Construal Under Threat

The purpose of this experiment was to investigate which aspect of the collective self showed the biggest difference in accessibility between a stereotype threat setting and a setting in which the stereotype does not apply. To do this, we gave participants a measure of identity salience after our stereotype threat manipulation to assess whether being in the diagnostic condition led to heightened salience of their stereotyped group membership in contrast to the nondiagnostic condition. As a further test of “witness,” participants responded to two statements regarding how strongly they identified with their gender group.

Our first measure was modeled after the Inclusion of Other in the Self Task (Aron et al., 1992) and was used to determine, as a function of the test description, the difference in overlap participants saw between the self and a number of social groups. The second measure used items adapted from the Collective Self-Esteem Scale (Luhtanen & Crocker, 1992). Specifically, we included two items from the Identity subsection to estimate identification with the participants’ gender group, whereas the remaining items asked general questions about other important social groups but did not measure identification, per se, with those groups. After completing these measures, the participants looked over a difficult math test (that was described as either being diagnostic or nondiagnostic of math ability) and then indicated how they believed they would perform on the test. For the final portion of the experiment, participants responded to four questions that were designed to measure the amount of stereotype threat they experienced. This Stereotype Threat Scale was added to establish that differences in the salience of the collective self and performance expectancies were in fact because of concerns about confirming the negative stereotype about their gender group in the diagnostic compared with the nondiagnostic condition.

For this experiment, we expected that the female participants’ gender group would show the biggest difference in accessibility effects between the diagnostic condition and the nondiagnostic condition, because in the diagnostic condition their gender is linked to the negative stereotype about math. In other words, there should be a bigger difference between the diagnostic and nondiagnostic conditions for the self-gender group ratings compared with the differences between the diagnostic and nondiagnostic conditions for the other three social groups (i.e., friends, family, students). Moreover, we predicted that participants would show more gender group identification in the diagnostic condition and less gender group identification in the nondiagnostic condition. We also anticipated that these participants would show typical stereotype threat effects on their math performance expectancies (Stangor, Carr, & Kiang, 1998; Stone et al., 1999), thus demonstrating that even performance expectancies are sensitive to how a test is described. We further hypothesized that the female participants would have higher stereotype threat scores in the diagnostic compared with the nondiagnostic condition.

Method

Participants and Design

Thirty-two female Dutch undergraduates took part in exchange for course credit or pay. For this experiment we used test description (diagnostic, nondiagnostic) as the single between-participants factor.

Procedure

Participants came into the laboratory where they learned that the experiment was about student life and academic ability. Specifically, the female participants were told that they would be asked to complete a short questionnaire about student life and afterward they would take a test to assess their academic ability. As before, test description served as the stereotype threat manipulation (see Experiment 1 for details).

Circle task. After the stereotype threat manipulation participants indicated how close they felt to four specific social groups (i.e., friends, gender, family, students). For each social 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, a female participant feels close to her gender group, then she should choose a diagram depicting a larger overlap between the self and her gender group (see Aron et al., 1992). The Circle Task was scored such that diagrams depicting the most overlap were assigned a 5, and the diagrams depicting the least overlap were assigned a 1. The closeness ratings were made separately for each social group.

Gender group identification. For this task, participants responded to six statements about the social groups used in the Circle Task (e.g., family; “Being with my family during the holidays is important to me”), of which two focused specifically on gender group identification: “Being a woman is an important part of my self-image”; “Being a woman has very little to do with how I feel about myself” (reverse coded). Responses to the six statements were recorded on a 7-point scale anchored on the endpoints with 1 (*strongly disagree*) and 7 (*strongly agree*). Because the two gender group identification statements were positively correlated ($r = .40, p = .02$), we averaged them to form a single gender identification score.

Math performance expectancies. After completing the gender measures, the participants were asked to look over a challenging math test and then to indicate how they believed they would perform on the test. In the diagnostic condition the name of a fictitious testing organization “Massachusetts Aptitude Assessment Center (MAAC)” followed by the label “Diagnostic Exam” were written on the cover of the test booklet. In the nondiagnostic condition the test booklet had only “Reasoning Exercise” written on the cover. For each condition the test format resembled a standard Graduate Record Exam math section and consisted of 20 difficult problems. Math performance expectancies could range from 0 to 20.

Stereotype threat. To assess whether our stereotype threat manipulation was successful, we asked participants to mark down how much they agreed with the following statements: “I worry that my ability to perform well on math tests is affected by my gender”; “I worry that if I perform poorly on this test, the experimenter will attribute my poor performance to my gender”; “I worry that, because I know the negative stereotype about women and math, my anxiety about confirming that stereotype will negatively influence how I perform on math tests.” Responses were recorded on a 7-point scale anchored on the ends by 1 (*strongly disagree*) and 7 (*strongly agree*). We averaged participants’ responses to form a measure of stereotype threat (Cronbach’s $\alpha = .76$). When the participants were finished they were thanked and debriefed.

Results and Discussion

Manipulation Check

To confirm that we had activated feelings of threat, we submitted the participants’ stereotype threat scores to a one-way ANOVA with test description as the between-participants factor. As can be seen in Table 1, female participants were more concerned about confirming the negative stereotype in the diagnostic ($M = 2.75$) relative to the nondiagnostic condition ($M = 2.09$), $F(1, 30) = 4.17, p = .05, \eta = .35$. This analysis shows that female participants reported experiencing more stereotype threat when the test

Table 1
Mean (Standard Deviation) Variables as a Function of Test Description

Variable	Test description	
	Diagnostic	Nondiagnostic
Stereotype threat score	2.75 (1.10)	2.09 (0.72)
Gender identification score	6.00 (0.50)	5.09 (1.23)
Self-gender overlap	2.47 (1.46)	1.59 (0.62)
Self-friends overlap	1.87 (0.64)	1.94 (0.66)
Self-family overlap	1.87 (0.35)	2.06 (0.43)
Self-students overlap	2.40 (0.51)	2.35 (1.06)
Math performance expectancies	10.53 (2.95)	13.59 (1.97)

was described as diagnostic of ability rather than when it was described as a reasoning exercise.

Main Analyses

Circle Task. To analyze the participants' responses on the Circle Task, we conducted a 4 (social group: student, friend, gender, family) \times 2 (test description: diagnostic, nondiagnostic) ANOVA with social group as a within-subject factor and test description as a between-participants factor (see Table 1). This analysis showed that, on average, participants saw more overlap between themselves and the group students (i.e., a main effect for student group) compared with the other three social groups, $F(1, 30) = 6.87, p = .01, \eta = .43$, which makes sense considering that the participants were taking a test, something that certainly makes their student identity salient. More critical to our reasoning, however, was the single-degree-of-freedom interaction that examined whether the test description effect was larger for participants' gender group compared with the other three social groups.³ This analysis revealed the expected interaction, $F(1, 30) = 5.71, p = .02, \eta = .40$, such that our test description manipulation showed the biggest difference in accessibility effects for the participants' gender group relative to the three other social groups. Simple effects tests further demonstrated that a stereotype threat situation led female participants to see more overlap between the self and their gender group ($M = 2.47$) compared with when they were not reminded about the negative stereotype ($M = 1.59$), $F(1, 30) = 5.15, p = .03, \eta = .38$. Of importance, simple effects tests between the diagnostic and nondiagnostic conditions for the other social groups were not reliable ($ps > .18$).

Gender group identification. To test our predictions concerning identification with one's stereotyped group membership, we submitted the participants' gender identification score to a one-way ANOVA with test description as the between-participants factor (see Table 1). This analysis revealed the hypothesized main effect for test description, $F(1, 30) = 7.20, p = .01, \eta = .44$. Participants had higher identification scores with their stereotyped group membership in the diagnostic condition ($M = 6.00$) in contrast to the nondiagnostic condition ($M = 5.09$). In short, the results from these two gender measures provide converging evidence for our contention that a stereotype threat situation, relative to a nonthreat situation, heightens accessibility of the collective self and that this "we-ness" centers on the participants' stereotyped

group membership (i.e., gender) and not on other important social groups (e.g., students).⁴

Math performance expectancies. Having established that stereotype threat leads to stronger identification with the participants' stereotyped group membership, we turned to the question of whether female participants expected to perform worse on the math test in the diagnostic relative to the nondiagnostic condition. Accordingly, we conducted a one-way ANOVA on the participants' math performance expectancies with test description as the between-participants factor. We found the predicted difference, $F(1, 30) = 12.14, p < .01, \eta = .54$. Table 2 shows that female participants had higher performance expectancies in the nondiagnostic condition ($M = 13.59$) and lower expectancies in the diagnostic condition ($M = 10.53$), demonstrating stereotype threat effects on participants' math performance expectancies. Next, we turned to the question of whether participants' gender identification scores moderate this effect (Schmader, 2002)?

To address this question, we conducted a 2 (test description) \times Gender Identification model on participants' performance expectancies, with gender identification serving as a continuous predictor. Results showed that higher gender identification scores were related to lower performance expectancies, $F(1, 28) = 4.84, p < .04, \eta = .38$ (there was no effect for test description, $F < 1.00$). We also found a Test Description \times Gender Identification interaction, $F(1, 28) = 5.29, p < .03, \eta = .40$, such that there was a bigger difference in test performance expectancies between the diagnostic and nondiagnostic conditions at higher levels of gender identification compared with lower levels of gender identification. This result is consistent with previous work demonstrating the moderating role of gender identification on female participants' math test performance (Schmader, 2002), as well as with manipulations intended to prime the specific stereotyped identity (Cheryan & Bodenhausen, 2000; Shih, Pittinsky, & Ambady, 1999), but also extends this past research by showing that gender identification can be heightened or increased simply by how a test is described (diagnostic or nondiagnostic of ability).

These first two experiments reveal strong support for our hypothesis that in stereotype threat situations the collective self ("we") is more accessible for stereotyped participants (but not for nonstereotyped participants—see Experiment 1) compared with situations in which they are not reminded about the negative group-based stereotype. Moreover, when their group image is at stake (i.e., a diagnostic math test situation), these participants view their stereotyped group membership (e.g., women) as more central to their self-concept in the diagnostic relative to the nondiagnostic condition. Our test description manipulation did not have as strong an effect for the other social groups (e.g., students), most likely because these group memberships are not as strongly related to the negative stereotype about women's math ability. The notion that a

³ We do not report the results of the omnibus interaction involving test description and social group because it is a design involving a within-subject variable with multiple levels (i.e., ratings of the four social groups), hence the pooled error term for the omnibus interaction is incorrect (Judd & McClelland, 1989).

⁴ When we conducted the same analyses by item a similar pattern of effects occurred ($F_s > 4.61$), hence our results were not driven by only one of the items in the scale.

Table 2
Mean (Standard Deviation) Variables as a Function of Test Description and Social Comparison Valence

Test	Social comparison valence		
	Positive	Negative	Control
Diagnostic test			
Stereotype threat score	3.20 (1.00)	3.25 (1.15)	3.18 (0.95)
Academic self-efficacy	4.83 (0.83)	2.54 (0.52)	3.00 (0.85)
Impression related concerns	2.54 (0.40)	4.69 (0.83)	4.47 (0.83)
Math performance score	13.83 (1.90)	9.38 (1.45)	10.60 (1.60)
Similarity score	5.33 (1.07)	4.46 (1.51)	
Nondiagnostic test			
Stereotype threat score	2.08 (0.67)	1.88 (0.67)	2.02 (0.61)
Academic self-efficacy	3.25 (0.62)	4.08 (0.76)	3.75 (0.62)
Impression related concerns	4.29 (0.54)	3.58 (0.86)	3.46 (0.66)
Math performance score	10.67 (1.97)	12.31 (2.25)	12.17 (1.70)
Similarity score	4.00 (1.04)	3.25 (0.62)	

stereotype threat situation leads to more overlap between the self and membership in the stereotyped group has been suggested in previous research on stereotype threat (Quinn & Spencer, 2000; Schmader, 2002; Spencer et al., 1999); however, it has never been tested explicitly as a function of stereotype threat. That is, we are unaware of any research that has measured the accessibility of the collective self immediately after targets have been reminded about the negative stereotype associated with their group (but see Cheryan & Bodenhausen, 2000; Shih et al., 1999, for research that specifically primed the participants' stereotyped identity).

It may be tempting to compare the current findings to earlier work on stereotype threat (Steele & Aronson, 1995; Experiment 3), which showed that Black participants tended to avoid stereotypic behaviors (e.g., playing basketball) and/or preferences (e.g., liking rap music) when they were in a situation of stereotype threat. We believe that this comparison may be unwarranted because the difference between our findings and those of Steele and Aronson (1995) can be explained in terms of specificity level. That is, in our experiment participants were asked to indicate a general identification with their stereotyped group membership (i.e., gender), whereas in Steele and Aronson's work the participants were asked about specific behaviors and preferences associated with their stereotyped identity. We suggest, therefore, that if Black participants were also asked to indicate general identification with their stereotyped group membership (i.e., Blacks), they too would show the same effects that we found with our female participants. Thus, identification with an individual's stereotyped identity may be increased while at the same time those individuals may attempt to distance themselves from specific stereotypic behaviors and preferences associated with their group. In other words, indicating stereotypic attributes and identifying with one's stereotyped group membership may be distinct processes. After all, saying "I do not like stereotypic behaviors associated with my stereotyped group" is very different from saying "Being a member of my group is not an important part of my self-image." In the end, it may be psychologically easier to distance oneself from behaviors associated with one's stereotyped identity (or at least indicate that this is the case) than it is to distance oneself from one's stereotyped identity.

It is interesting to note that on a more general level these data support research on self-categorization (Turner et al., 1987) and social identity theory (Tajfel & Turner, 1979) by showing that contextual factors can increase the accessibility of relevant social identities (see also Stapel, Reicher, & Spears, 1994). However, as we were examining the accessibility of different self-construals in a stereotype threat paradigm, our results do more than just show that different contexts can switch "on" or "off" particular self-categorizations or social identities. That is, the present results demonstrate that under stereotype threat conditions (i.e., female participants taking a diagnostic math test), compared with nonstereotype threat conditions (i.e., female participants taking a nondiagnostic math test), the relevant social identity (women) is salient. Although this is an important feature of stereotype threat theory, it has never been tested empirically until now.

Now that we have confirmed that stereotype threat (rather than nonstereotype threat) increases awareness of the collective self-construal orientation specific to the participants' stereotyped group membership, perhaps we can use this mindset to fight the negative effects of stereotype threat on its own territory—in the immediate testing situation. For example, on the basis of past research on social comparison, if a person's mindset is focused on the collective self, then that individual often assimilates group-based information into his or her self-concept (Brewer & Weber, 1994; Stapel & Koomen, 2001). So, given that stereotype threat activates a collective self-construal orientation for participants targeted by the stereotype, it seems logical to suggest that these same participants could use a fellow group member's success as a way to protect themselves from the negative effects of stereotype threat (e.g., Marx & Roman, 2002). More specifically, we expected that the enhanced test performance resulting from positive in-group social comparison information would be mediated by participants' impression-related concerns.

Experiment 3: Social Comparisons Under Threat

This third experiment was conducted to test whether social comparisons in a stereotyped domain led to assimilative behavior on participants' math performance because of their reduced impression-related concerns. Given that stereotype threat leads to a collective self-construal orientation, we should find assimilation among stereotyped participants because their mindset is focused on seeking out similarities between the self and comparison other, whereas in the nondiagnostic condition (in which the group-based stereotype is not activated), it should be more likely that we find less of an effect of social comparison information because participants' mindsets are focused on seeking out differences between the self and comparison other (Stapel & Koomen, 2001). We were also interested in the participants' perceptions of similarity with the comparison target. This similarity measure was used as another measure of "we-ness," in the sense that when a person is in a "we" frame of mind, that person is likely to value being part of a social unit and, thus, to accentuate similarities with close others (e.g., Aron et al., 1992; Brewer & Gardner, 1996; Schubert & Otten, 2002; Stapel & Koomen, 2001; Tajfel & Turner, 1979). Applying this logic to the current research, in stereotype threat situations participants should see more similarity between themselves and the comparison target, particularly when the comparison target is talented in math. After all, it is always nicer to feel similar to a

highly successful other compared with an unsuccessful other. Finally, we examined the mediating role of impression-related concerns on participants' math performance in stereotype threat and nonthreat conditions.

According to our theoretical framework (i.e., stereotype threat \rightarrow accessibility of the collective self \rightarrow assimilative social comparison effects), we expected that in the diagnostic condition participants would perform better on the math test when the social comparison target was described as highly talented in math (a positive social comparison) compared with the nondiagnostic condition. However, we predicted that when the target of the social comparison was described as not so talented in math (a negative social comparison), participants would perform worse on the math test in the diagnostic condition relative to the nondiagnostic condition. In other words, in the diagnostic conditions participants should show assimilation effects for both the positive and negative social comparison conditions such that positive social comparisons lead to better performance and negative social comparisons to worse performance. Finally, when no social comparison information is provided, the participants should perform better in the nondiagnostic than the diagnostic condition, as would be expected by stereotype threat theory.

In terms of the participants' impression-related concerns, we expected that they would feel less concern after making a positive social comparison in the diagnostic condition and more concern after making such a comparison in the nondiagnostic condition because in the former condition participants should not be as concerned about making a "bad impression" if they do not perform well on the test. That is, the knowledge that someone else from one's group (a similar other) has succeeded in the stereotyped domain may lessen the individual's unique concerns about the impression being made if she underperforms in a stereotype threat situation, which then could lead to higher test performance (i.e., mediate the beneficial effects of positive social comparisons under stereotype threat). However, in the negative social comparison and the control conditions, we anticipated the opposite result: more concern in the diagnostic condition and less concern in the nondiagnostic condition, as participants were not provided with positive in-group social comparison information in those conditions.

For our similarity measure, we hypothesized that in the positive social comparison conditions participants would feel more similar to the comparison target in the diagnostic condition compared with the comparison target in the nondiagnostic condition because the participants' collective self is more accessible in the former condition and the target is a more desirable comparison standard—that is, it is always nicer feeling similar to a successful other than an unsuccessful other. Because of this increased perceived similarity in the diagnostic condition, participants' impression-related concerns should be reduced, resulting in increased performance. In the negative social comparison conditions, we expected no differences in perceived similarity between the diagnostic and nondiagnostic conditions because the comparison target is not as talented in math and is thus a less favorable comparison standard. Moreover, even though the collective self is accessible in the diagnostic condition, participants should perceive less similarity with the negative comparison target because the target is not as appealing as the positive social comparison target.

Method

Participants and Design

Seventy-seven female Dutch undergraduates took part in exchange for course credit or pay. For this experiment we used a 3 (social comparison valence: positive, negative, control) \times 2 (test description: diagnostic, nondiagnostic) between-participants design.

Procedure

Participants came into the laboratory, where they were informed that the experiment was about scholastic ability and that they would complete a task related to student life as well as a task related to different aspects of scholastic ability. They were then given an envelope that contained all of the experimental materials except for the math test. We used the same stereotype threat manipulation from the previous two experiments. After learning about the test they would take, the participants were given social comparison information in the form of a fictitious newspaper article that described a fellow student from the participants' university.

Social comparison valence. We manipulated the valence of the social comparison information in a way that was modeled after Lockwood and Kunda (1997, p. 94; see also Marx & Roman, 2002; Stapel & Koomen, 2001). Participants in the social comparison conditions, but not in the control conditions, read a bogus newspaper article describing a female student (Boukje Timmer) who was either very intelligent and excelled in math or quite unintelligent and did not excel in math. The participants' task was to guess in which daily newspaper the article about the (positive or negative) social comparison target could have been published. Participants randomly assigned to the positive social comparison condition learned that Boukje was majoring in math and psychology, had taken six challenging math classes in college, and planned on earning a doctorate in quantitative psychology. Those participants randomly assigned to the negative social comparison condition learned that Boukje was majoring in literature and psychology, had taken two moderately challenging math classes in college, and planned on pursuing a career in acting. After reading the newspaper article and writing down their answers to the question of which newspaper it was published in, participants turned to the math test.

Math test performance. Participants had 20 min to complete the same 20-problem math test used in Experiment 2; however, in this experiment the participants actually took the math test. Performance scores could range from 0 to 20.

Similarity ratings. To assess perceptions of similarity with the comparison target (Boukje Timmer), we asked only those participants in the positive and negative social comparison conditions to respond to the statement, "How similar do you perceive Boukje Timmer to be to you?" This scale was anchored by 1 (*not at all similar*) and 7 (*extremely similar*).

Academic self-efficacy. We were also interested in how in-group social comparisons information would affect participants' feelings of academic self-efficacy. Accordingly, we asked participants to respond to the statement: "At this moment I think I have less scholastic abilities than others" (reverse coded). The scale was labeled on the endpoints by 1 (*not at all true*) and 7 (*extremely true*).

Impression-related concerns. To examine participants' impression-related concerns, we asked them to answer two statements: "I am concerned that I will be seen as a success or failure" and "I am concerned about what other people think of me" ($r = .34, p < .01$). The scale was anchored by 1 (*not at all true*) and 7 (*extremely true*).

Manipulation checks. To test for the effectiveness of the social comparison manipulation, participants were presented with the statement "How competent do you think Boukje Timmer is in math?," to which they responded on a 7-point scale with endpoints 1 (*not at all competent*) and 7 (*extremely competent*).

To assess the amount of stereotype threat the participants experienced, we used the Stereotype Threat Scale (Cronbach's $\alpha = .82$) from Experi-

ment 2. On completion of these measures, the participants were thanked and debriefed.

Results and Discussion

Manipulation Checks

Social comparison valence. The first manipulation check assessed whether participants correctly judged the valence of the social comparison information. To do this, we submitted the participants' ratings of Boukje Timmer's math competence to a 2 (social comparison valence) \times 2 (test description) ANOVA (see Table 2). For this analysis, we did not include the control condition. Results revealed the expected main effect for social comparison valence, $F(1, 45) = 203.60, p < .01, \eta = .90$, such that participants felt that the positive social comparison target ($M = 6.25$) was much more talented in math compared with the negative social comparison target ($M = 2.92$). No other effects were reliable ($F_s < 1.00$). This first manipulation check confirmed that we successfully manipulated the valence of the social comparison information.

Stereotype threat. Our second manipulation check focused on whether we created a situation of stereotype threat. To do this, we conducted a 3 (social comparison valence) \times 2 (test description) ANOVA on the participants' responses to the stereotype threat measure (see Table 2). We found the predicted main effect of test description, such that participants reported experiencing more stereotype threat in the diagnostic conditions ($M = 3.21$) compared with those participants in the nondiagnostic conditions ($M = 2.00$), $F(1, 71) = 37.49, p < .01, \eta = .59$. No other effects were reliable ($F_s < 1.00$). This second manipulation check demonstrated that again our test description manipulation was effective in creating a stereotype threat situation.

Main Analyses

Similarity ratings. Recall that in the first two experiments we showed that a stereotype threat situation leads to increased accessibility of the collective self, thus in this experiment, we should find that participants see more similarity between themselves and the positive social comparison target in the diagnostic condition compared with the nondiagnostic condition. To investigate how similar our participants perceived Boukje Timmer to be to them, we submitted their similarity ratings to a 2 (social comparison valence) \times 2 (test description) ANOVA (see Table 2). For this analysis we did not include the control condition because participants were not provided with any comparison information. There were main effects for test description, $F(1, 45) = 15.88, p < .01, \eta = .51$, and for social comparison valence, $F(1, 45) = 6.45, p < .02, \eta = .35$. Much to our surprise, the interaction was not reliable ($F < 1.00$). This result demonstrates that the pattern of means between the diagnostic and nondiagnostic conditions was similar for the positive and negative social comparison targets.

Academic self-efficacy. To examine the effect of social comparisons on participants' academic self-efficacy, we conducted a 2 (test description) \times 3 (social comparison valence) ANOVA. This analysis revealed a main effect for social comparison valence, $F(2, 71) = 7.95, p < .01, \eta = .43$. To interpret this omnibus main effect, we conducted a focused test comparing participants' academic self-efficacy in the positive social comparison condition

(+2) to participants in the negative social comparison (-1) and control conditions (-1). This contrast demonstrated that participants in the positive social comparison conditions ($M = 4.04$) indicated having higher levels of academic self-efficacy compared with the self-efficacy of participants in the other two conditions ($M = 3.34$), $F(1, 71) = 15.80, p < .01, \eta = .43$. The second contrast comparing the negative social comparison conditions ($M = 3.31$) to the control conditions ($M = 3.37$) showed no difference between them ($F < 1.00$). We also found a reliable omnibus interaction between social comparison valence and test description, $F(2, 71) = 31.86, p < .01, \eta = .69$. To interpret this interaction, we broke it down into a focused test comparing the difference between the diagnostic and nondiagnostic conditions in the positive social comparison conditions to the difference between the diagnostic and nondiagnostic conditions in the other two conditions. This single-degree-of-freedom interaction revealed that the omnibus interaction did, in fact, show that the difference between the positive social comparison conditions and the other two conditions was moderated by test description, $F(1, 71) = 59.94, p < .01, \eta = .68$. Specifically, participants in the positive social comparison condition had higher academic self-efficacy in the diagnostic condition ($M = 4.83$) compared with the nondiagnostic condition ($M = 3.25$), $F(1, 71) = 29.25, p < .01, \eta = .54$, but this effect was reversed for the negative social comparison, $F(1, 71) = 30.11, p < .01, \eta = .55$, and control conditions, $F(1, 71) = 7.32, p < .01, \eta = .31$ (i.e., higher academic self-efficacy in the nondiagnostic relative to the diagnostic conditions). The second single-degree-of-freedom interaction revealed that there was a bigger difference in academic self-efficacy between the diagnostic and nondiagnostic conditions for participants in the negative social comparison conditions compared with those participants in the control conditions, $F(1, 71) = 4.00, p = .05, \eta = .23$.

Impression-related concerns. To investigate the participants' impression-related concerns, we conducted a 3 (social comparison valence) \times 2 (test description) ANOVA (see Table 2). This analysis revealed a social comparison valence main effect, $F(2, 71) = 6.74, p < .01, \eta = .40$. The first focused test of this omnibus main effect showed that participants in the positive social comparison conditions ($M = 3.42$) expressed a lower level of concern than participants in the other two conditions ($M = 4.05$), $F(1, 71) = 12.74, p < .01, \eta = .39$. The second focused test demonstrated that participants' impression-related concerns in the negative social comparison ($M = 4.13$) and control ($M = 3.96$) conditions did not differ from each other ($F < 1.00$). We also found a reliable omnibus two-way interaction, $F(2, 71) = 31.62, p < .01, \eta = .69$. Tests of the single-degree-of-freedom interaction showed that the difference between the positive social comparison conditions and the other two conditions was moderated by test description, $F(1, 71) = 63.20, p < .01, \eta = .69$, showing that participants felt less concern in the diagnostic condition ($M = 2.54$), but more concern in the nondiagnostic condition ($M = 4.29$) after making a positive in-group comparison, $F(1, 71) = 35.61, p < .01, \eta = .58$, and the opposite effect occurred for the negative social comparison, $F(1, 71) = 15.52, p < .01, \eta = .42$, and control conditions, $F(1, 71) = 13.18, p < .01, \eta = .40$. Namely, when participants' collective self is *more* salient (i.e., in the diagnostic condition), positive social comparison information led to lowered concern, yet when the collective self is *less* salient (i.e., in the nondiagnostic condition),

positive social comparison information led to heightened concern. In contrast, participants' level of concern between the negative social comparison and control conditions was not moderated by our test description manipulation ($F < 1.00$).⁵

Math test performance. What effect does the in-group social comparison information have on female participants' math test performance? To address this question, we conducted a 3 (social comparison valence) \times 2 (test description) ANOVA on the participants' math test performance (see Table 2). Results revealed a main effect for social comparison valence, $F(2, 71) = 3.76, p < .03, \eta = .31$. Focused tests revealed that participants in the positive social comparison conditions ($M = 12.25$) performed better than participants in the other two conditions ($M = 11.12$), $F(1, 71) = 6.40, p < .02, \eta = .29$. The second contrast comparing the negative social comparison conditions ($M = 10.85$) to the control conditions ($M = 11.38$) revealed that these conditions did not differ from each other ($p = .23$). We also found a reliable omnibus interaction, $F(2, 71) = 19.06, p < .01, \eta = .59$. A focused test of this omnibus interaction demonstrated that the difference between the positive social comparison and the other two conditions was moderated by test description, $F(1, 71) = 36.36, p < .01, \eta = .58$. This contrast showed that participants performed better in the diagnostic condition ($M = 13.83$) after making a positive social comparison and worse after making such a comparison in the nondiagnostic condition ($M = 10.67$), $F(1, 71) = 18.09, p < .01, \eta = .45$, and that this effect was reversed for the negative social comparison, $F(1, 71) = 16.83, p < .01, \eta = .44$, and control conditions, $F(1, 71) = 4.96, p = .03, \eta = .26$ (i.e., better math test performance in the nondiagnostic relative to the diagnostic conditions). The second focused test revealed that participants' math test performance in the negative social comparison and control conditions was not moderated by test description ($p = .18$).

These results clearly show that positive social comparisons can be beneficial to participants' math test performance (lead to assimilation) in a stereotyped threat situation because participants' collective self was more accessible in comparison with the nondiagnostic conditions, in which participants' collective self was not as accessible. Given these results, the question now is what factors could account for these beneficial effects of positive social comparisons?

Mediated Moderation Analyses

On the basis of our theoretical framework and previously predicted findings, we began to explore whether the participants' impression-related concerns mediated the moderation effect of positive social comparisons on participants' math performance in stereotype threat situations. To do this, we used procedures recommended by Muller, Judd, and Yzerbyt (2004; see also Baron & Kenny, 1986; Wegener & Fabrigar, 2000).

As a first step, we tested whether our single-degree-of-freedom Test Description \times Social Comparison Valence interaction was a reliable predictor of participants' math test performance. This was indeed the case, $B = 0.90, F(1, 71) = 36.36, p < .01$, as shown earlier. The second mediational step examined whether the same single-degree-of-freedom interaction predicted participants' impression-related concerns. As demonstrated before, this interaction was reliable, $B = -0.47, F(1, 71) = 63.20, p < .01$. Our next

step examined whether participants' impression-related concerns predicted their math test performance, controlling for our experimental variables and their interactions. This analysis showed that impression-related concerns was a predictor, $B = -0.77, F(1, 70) = 7.08, p < .01$. Our final step investigated whether the beta associated with the focused Test Description \times Social Comparison Valence interaction was reduced once we controlled for participants' impression-related concerns.⁶ This analysis revealed that the interaction was still reliable, $B = 0.54, F(1, 70) = 7.56, p < .01$, however the beta was dramatically reduced from $B = 0.90$ to $B = 0.54$. Of importance, a Sobel (1982) test indicated that this reduction in betas was reliable, $Z = 2.52, p < .02$, thus supporting the partial mediating role of impression-related concerns on participants' math test performance in the diagnostic condition after making a positive social comparison.

On the whole, these results reveal how positive social comparisons can be beneficial (lead to assimilative math performance) in a stereotyped threat situation. Moreover, beyond our assimilative performance results, these positive social comparisons were associated with reduced impression-related concerns in the diagnostic condition, but only after the participant made a positive social comparison with someone who was perceived as similar. That is, positive group-based information may lessen the impression-related concerns of stereotyped targets (Blanton et al., 2002), because knowing that a fellow group member has done well in the stereotyped domain could likewise reduce targets' unique concerns about the impression they are making in the testing situation. What is probably even more important is that this is the first demonstration of how positive social comparisons can alleviate the concern participants may have about the impression they are making in a stereotype threat situation; thus, they perform better despite the awareness (as shown by their stereotype threat scores) of the negative gender stereotype alleging inferior math ability. Namely, these impression-related concerns and the knowledge about the stereotype associated with one's group might work independently.

The results from Experiment 3 clearly show that stereotyped participants perform better under stereotype threat conditions when they are provided with positive comparison information because this stereotype-disconfirming information reduces their impression-related concerns. Now the question is whether it is necessary for a comparison target to demonstrate success in a stereotyped domain, or does success in any domain result in the same relationship between positive comparisons and performance? In our opinion, finding an answer to this question would help demonstrate that positive social comparisons do indeed tell (at least implicitly) the perceiver "we can do it" in the specific domain, thereby boosting the perceiver's test performance.

Experiment 4: The Comparison Target's Domain of Success

As a way to develop our theoretical framework further, this final experiment focused on whether positive social comparisons need

⁵ When we conducted the same analyses by item a similar pattern of effects occurred ($F_s > 4.73$), hence our results were not driven by only one of the items in the scale.

⁶ This was the case because the Sobel (1982) test could not be used to test the omnibus (i.e., the two-degree-of-freedom) interaction.

Table 3
Mean (Standard Deviation) Variables as a Function of Test Description and Target of Comparison

Variable	Diagnostic test		Nondiagnostic test	
	Math target	Athlete target	Math target	Athlete target
Similarity score	5.09 (1.51)	3.55 (0.93)	3.79 (0.89)	3.43 (0.65)
Math test performance	13.64 (2.25)	11.91 (1.58)	10.71 (2.02)	12.07 (2.13)
Math competence	6.00 (1.26)	3.45 (1.45)	5.79 (1.37)	3.57 (1.55)
Athletic competence	3.50 (1.51)	5.55 (1.29)	3.29 (1.29)	5.50 (1.45)

to be domain specific or whether group-based success in any domain (e.g., athletics) would be just as enhancing for stereotyped individuals' math test performance. To this end, we provided participants with two extremely successful female comparison targets (who differed on their dimension of success: athletics or mathematics) and examined their effects on female participants' perceived similarity with the comparison target and their math test performance under stereotype threat and nonstereotype threat conditions.

For this experiment, we anticipated that the domain-relevant comparison target (math target) would lead to larger assimilative effects and thus higher math performance in the diagnostic condition compared with the domain-irrelevant comparison target (athlete target). No differences were expected in the nondiagnostic conditions. We also reasoned that in the diagnostic conditions the female participants would feel more similar to the math target relative to the athlete target. We say this for three reasons: first, because the participants' collective self should be more accessible in the diagnostic condition; second, because the math target is relevant to the situation (math); and third, because the comparison target is seen as someone who excels in an important and applicable domain. Participant's similarity ratings should not differ in the nondiagnostic conditions because their collective self is not as salient in this condition.

Method

Participants and Design

Fifty-one female Dutch undergraduates took part in exchange for course credit or pay. For this experiment we used a 2 (comparison target: math, athlete) \times 2 (test description: diagnostic, nondiagnostic) between-participants design.

Procedure

The procedure was nearly identical to the procedure used for Experiment 3 except in that participants were given social comparison information about a target who was very competent in either math or athletics. Following exposure to this comparison information, participants took a 20-problem math test under stereotype threat or nonstereotype threat conditions and, afterward, answered a question about the target's competence in math ("How competent do you think Boukje Timmer is in math?") and athletics ("How competent do you think Boukje Timmer is in athletics?"). Both questions could be answered on a 7-point scale with endpoints 1 (*not at all competent*) and 7 (*extremely competent*). Participants also responded, on a 1 (*not at all similar*) to 7 (*very similar*) scale, to a question about how similar they perceived the target to be to them: "How similar do you

perceive Boukje Timmer to be to you?" On completion of these measures the participants were thanked and debriefed.

Results and Discussion

Our analyses were conducted without the data from one female participant because outlier analyses revealed that she had uncommon studentized deleted residuals on relevant measures (Judd & McClelland, 1989; McClelland, 2000).⁷

Manipulation Checks

Math competence. To check whether participants perceived the math target as competent in math, we submitted the participants' math competence ratings to a 2 (comparison target) \times 2 (test description) ANOVA. Table 3 shows the expected main effect of comparison target, $F(1, 46) = 34.70, p < .01, \eta = .66$, demonstrating that the math target ($M = 5.88$) was perceived to be more competent in math than was the athlete target ($M = 3.52$). All other effects, $F_s < 1.00$.

Athletic competence. The manipulation check for athletic competence was analyzed using a 2 (comparison target) \times 2 (test description) ANOVA (see Table 3). Results yielded the anticipated main effect for comparison target, $F(1, 46) = 34.26, p < .01, \eta = .65$, such that the athlete target ($M = 5.52$) was judged as having more athletic competence relative to the math target ($M = 3.28$). All other effects, $F_s < 1.00$.

The results from these manipulation checks make it quite clear that the two comparison targets were perceived as different on the relevant dimension (math) as well as on the irrelevant dimension (athletics). Moreover, if we find differences between the two comparison targets, it would be difficult to conclude that the effects were driven by differences in the targets' competence levels (as the effect size estimates on both manipulation checks were nearly identical) and not by the relevance of the comparison target's success to the specific situation.

⁷ This index allows us to test whether an additional parameter is needed in the regression model to account for error in our predictions that is associated with a particular observation. Moreover, this index can be seen as an objective way to detect outliers given that it provides a t value for each observation. Put simply, an uncommon studentized deleted residual indicates whether "an observation is so extreme that it is unlike the other observations" in the data set (Judd & McClelland, 1989, p. 221; McClelland, 2000).

Main Analyses

Similarity. For this analysis, we submitted the participants' similarity scores to a 2 (comparison target) \times 2 (test description) ANOVA (see Table 3). Results revealed main effects for test description, $F(1, 46) = 6.04, p < .02, \eta = .34$, and for comparison target, $F(1, 46) = 10.82, p < .01, \eta = .44$. These main effects were qualified by the predicted Comparison Target \times Test Description interaction, $F(1, 46) = 4.22, p < .05, \eta = .29$.

Within the diagnostic conditions, we compared the participants' similarity scores in the math target condition ($M = 5.09$) to those in the athlete target condition ($M = 3.55$). This comparison revealed, as expected, that participants perceived themselves to be more similar to the math target than the athlete target when stereotype threat was activated, $F(1, 46) = 12.65, p < .01, \eta = .46$. Even though the collective self may be more accessible in the diagnostic condition, this did not increase the female participants' similarity scores as much because the comparison target was not relevant to a math test-taking situation, thus highlighting the critical role of target relevance in stereotyped social comparisons. Within the nondiagnostic conditions, we found no difference between the math target ($M = 3.79$) and athlete target ($M = 3.43$) conditions ($F < 1.00$), showing that when the collective self is less accessible, the participants' similarity scores are not as sensitive to the type of comparison target.

Math test performance. We examined the participants' math test performance using a 2 (comparison target) \times 2 (test description) ANOVA (see Table 3). Results showed a main effect for test description, $F(1, 46) = 5.76, p = .02, \eta = .33$. There was no main effect for comparison target ($F < 1.00$). We found, as hypothesized, a reliable two-way interaction, $F(1, 46) = 7.20, p = .01, \eta = .37$.

To examine this interaction, we compared the participants' math test performance in the diagnostic conditions. Results revealed that participants had higher math performance in the math target ($M = 13.64$) than in the athlete target condition ($M = 11.91$), $F(1, 46) = 4.04, p = .05, \eta = .28$. Of interest, in the nondiagnostic conditions participants in the athlete target condition ($M = 12.07$) performed slightly better than those participants in the math target condition ($M = 10.71$), $F(1, 46) = 3.18, p = .08, \eta = .25$. Despite not predicting any differences in the nondiagnostic condition (though one could make such a prediction, which in our opinion, would be consistent with our theoretical framework), this result suggests that when the personal self ("I") is relatively more salient than the collective self ("we") those positive comparison targets who are applicable to the immediate situation (taking a math test) harm more than help participants' math test performance (e.g., J. Brown et al., 1992; Festinger, 1954).

These results support our hypothesis that learning about a fellow group member who has succeeded in a domain relevant to the situation would lead to increased perceived similarity between the comparison target and the female participants in our experiment compared with when they learn about a fellow group member who has succeeded in a domain unrelated to the situation. It appears, therefore, that both shared gender group membership and success in the relevant domain are necessary conditions for perceived similarity to be increased. Furthermore, and of importance, we found that learning about the athlete target did not enhance female participants' math performance as much as did learning about a

math target. Specifically, we found that in the diagnostic conditions when the comparison target's success was unrelated to the immediate test-taking situation, participants did not benefit as much from this positive social comparison as they did when the comparison target's success was math related.

General Discussion

The four experiments presented highlight several novel contributions to stereotype threat and social comparison research. Together, the most important findings from our experiments can be summarized as follows:

1. For the targets of a negative stereotype, compared with those participants who were not targeted by the stereotype, the accessibility of the collective self ("we") was increased in a stereotype threat situation relative to non-stereotype threat situation.
2. This collective self-construal orientation ("we-ness") centered on the participants' stereotyped identity (e.g., women) and not on other aspects of the collective self that are not as relevant to the immediate situation and the associated stereotype (e.g., family, friends).
3. When an individual's construal orientation is centered on "we-ness" and the social comparison information is positive instead of negative, then this particular combination can serve as the switch that turns off the negative effects of stereotype threat, thereby leading to better performance under stereotype threat conditions compared with nonstereotype threat conditions.
4. When the collective self is salient and participants are presented with positive in-group social comparison information, they are less concerned about the impression they are making and, as a result, they perform better in a stereotype threat situation.
5. Learning about another group member's achievements in a domain related to the testing situation leads to increased perceived similarity and math test performance for those individuals who must contend with the negative stereotype relative to when the comparison target is entirely unrelated to the testing situation (cf. McIntyre, Paulson, & Lord, 2002).

In summary, the findings from these experiments not only increase our understanding of stereotype threat but also point to ways to eliminate it. Furthermore, these results advance social comparison theory by showing that the direction of social comparisons has very different outcomes depending on whether they are made in a stereotype threat situation or in a situation in which the stereotype is irrelevant. Indeed, these data have several practical implications for our understanding of stereotype threat and social comparison processes.

Implications for Stereotype Threat Research

Much of previous research on stereotype threat has focused on cataloguing the type of groups and domains in which the theory

applies. In contrast (though this research area is rapidly growing), not as much work has been conducted on the underlying mechanism of stereotype threat (but see Blascovich et al., 2001; Davies et al., 2002; Gonzales et al., 2002; Maass & Cadinu, 2003; O'Brien & Crandall, 2003; Schmader & Johns, 2003; Spencer et al., 1999; Wheeler & Petty, 2001) and how the threat can be alleviated or eliminated in the testing situation (Marx & Roman, 2002; McIntyre et al., 2002). Clearly understanding stereotype threat and the ways in which it can be reduced has important consequences for the targets of the stereotype.

The experiments reported here were designed to explore an aspect of stereotype threat that centered on what Steele and his colleagues (Steele, 1997; Steele et al., 2002) have argued is one of the core principles of the theory—a concern rooted in beliefs about one's group image. We believe that our approach to understanding stereotype threat is particularly advantageous in the sense that it focuses on processes that can easily alter the effects of stereotype threat rather than on the correlates (e.g., anxiety, emotional or physiological arousal) of stereotype threat, which may not be as easily altered in the immediate stereotype threat situation. By focusing on the correlates of stereotype threat, researchers are only able to show that the two constructs are related and not how to diminish the threat "online." Therefore, we contend that the most effective way to fight stereotype threat is to fight it at the group level because this level is argued to be the most applicable to the theory (Steele, 1997; Steele et al., 2002). Of importance, we are the first to use this perspective to understand as well as demonstrate how a collective self-construal orientation can be used to reduce the negative effects of stereotype threat. This group-based approach not only increases our knowledge about stereotype threat, but it also builds on social comparison research by showing that depending on whether social comparisons are made in a stereotyped or nonstereotyped domain, the targets of a negative stereotype could demonstrate assimilative performance effects or no substantial effect of the comparison information.

Implications for Social Comparison Research

Recall that when participants are in a situation that reminds them of a negative group-based stereotype, their membership in the stereotyped group (i.e., women) is more salient to them. Thus, because of this increased salience of their collective self, positive social comparison information leads to assimilative math test performance. This finding is supported by past research showing that social comparisons are more likely to lead to assimilation effects when an individual's mindset is focused on the collective rather than the personal. It is clear then that the distinction between the present results and most nonstereotyped social comparison research is that of context—that is, we used a context in which a stereotype about the perceivers' group was made accessible (a stereotype threat situation), therefore the targets' collective self was more salient to them than was their personal self, leading to assimilative behavioral effects (Stapel & Koomen, 2001; see also, Blanton et al., 2000; Blanton, 2001).

Our findings provide compelling evidence that the effects of positive social comparisons can be quite different depending on the situation in which they are made. For instance, returning to the Mary and Susan example discussed previously, it seems clear that if Mary is in a stereotype threat situation (when her group identity

is more salient) then she should shift to a "we" frame of mind, leading to assimilative behavior when making a positive social comparison. However, in nonstereotype threat situations (when her personal identity is more salient) Mary should be more oriented toward an "I" frame of mind, therefore not benefiting as much from positive group-based social comparison information. This latter characterization of the differences between stereotyped and nonstereotyped settings provided the background for our hypothesis that positive social comparisons can buffer the targets of a stereotype from the adverse effects of stereotype threat because their mindset is oriented more toward assimilation than contrast. These effects, in the end, could provide social comparison researchers with a new foundation on which to build the next generation of experiments on social comparison processes, namely experiments on social comparison processes in stereotyped versus nonstereotyped situations.

Coda

According to Festinger (1954) the primary goal of social comparisons is to acquire information about the self. However, what has been largely neglected in social comparison research is whether the social comparison was made in the context of a group-based stereotype or not. For example, in most cases when Damien makes an upward or positive social comparison with Derrick (who is a 6 ft 5 in., Portuguese-born male psychologist) this comparison should lead to contrastive self-evaluations for Damien (who is a 5 ft 7 in., German-born male psychologist) because Derrick is so much taller than Damien. Yet, if Damien compares himself to Derrick in the context of a negative stereotype about the height of European male psychologists (i.e., "They are all so tall that it is ridiculous"), then Derrick's height may lead to assimilative self-evaluations for Damien (i.e., "Actually, I am quite tall myself") because Damien's mindset should be focused on "we" rather than "I." The findings from our experiments clearly underscore this point by showing that positive social comparisons are beneficial to performance in stereotyped situations, whereas they are less helpful to performance in those situations in which the stereotype is irrelevant. Thus, the present research supports our main hypothesis that one important (but to date overlooked) determinant of the impact of social comparison on behavior is the definition of the social comparison context: Is the social comparison context stereotype relevant—yes or no?

References

- 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.
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.
- Blanton, H. (2001). Evaluating the self in the context of another: The three-selves model of social comparison assimilation and contrast. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton Symposium on the Legacy and Future of Social Cognition* (pp. 75–87). Mahwah, NJ: Erlbaum.
- 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.

- Blanton, H., Crocker, J., & Miller, D. T. (2000). The effects of in-group versus out-group social comparison on self-esteem in the context of a negative stereotype. *Journal of Experimental Social Psychology, 36*, 519–530.
- Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. M. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science, 12*, 225–229.
- 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.
- Brewer, M. B., & Weber, J. G. (1994). Self-evaluation effects of interpersonal versus intergroup comparison. *Journal of Personality and Social Psychology, 66*, 268–275.
- Brown, J. D., Novick, N. J., Lord, K. A., & Richards, J. M. (1992). When Gulliver travels: Social context, psychological closeness, and self-appraisals. *Journal of Personality and Social Psychology, 62*, 717–727.
- Brown, R. P., & Josephs, R. A. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. *Journal of Personality and Social Psychology, 76*, 246–257.
- Cheryan, S., & Bodenhausen, G. V. (2000). When positive stereotypes threaten intellectual performance: The psychological hazards of “model minority” status. *Psychological Science, 11*, 399–402.
- Collins, R. L. (1996). For better or worse: The impact of upward social comparisons on self-evaluation. *Psychological Bulletin, 119*, 51–69.
- Croizet, J.-C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin, 24*, 588–594.
- Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin, 28*, 1615–1628.
- Dijksterhuis, A., & van Knippenberg, A. (2000). Behavioral indecision: Effects of self-focus on automatic behavior. *Social Cognition, 11*, 55–74.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117–140.
- Gardner, W. L., Gabriel, S., & Hochschild, L. (2002). When you and I are “we,” you are not threatening: The role of self-expansion in social comparison. *Journal of Personality and Social Psychology, 82*, 239–251.
- 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.
- Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model-comparison approach*. San Diego, CA: Harcourt Brace Jovanovich.
- Leyens, J. P., Desert, M., Croizet, J.-C., & Darcis, C. (2000). Stereotype threat: Are lower status and history of stigmatization preconditions of stereotype threat? *Personality and Social Psychology Bulletin, 26*, 1189–1199.
- Lockwood, P., & Kunda, Z. (1997). Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology, 73*, 91–103.
- Luhtanen, R., & Crocker, J. (1992). A collective self-esteem scale: Self-evaluations of one’s social identity. *Personality and Social Psychology Bulletin, 18*, 302–318.
- Maass, A., & Cadinu, M. (2003). Stereotype threat: When minority members underperform. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 14, pp. 243–275). New York: Taylor & Francis Group.
- Marx, D. M., & Goff, P. A. (in press). Clearing the air: The effect of experimenter race on targets’ test performance and subjective experience. *British Journal of Social Psychology*.
- Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting female students’ math test performance. *Personality and Social Psychology Bulletin, 28*, 1185–1197.
- Marx, D. M., & Stapel, D. A. (2004). *Distinguishing stereotype threat from stereotype priming effects: The role of the collective self and threat-based concerns*. Unpublished manuscript, University of Groningen, The Netherlands.
- McClelland, G. H. (2000). Nasty data: Unruly, ill-mannered observations can ruin your analysis. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 393–411). New York: Cambridge University Press.
- McIntyre, R. B., Paulson, R. M., & Lord, C. G. (2002). Alleviating women’s mathematics stereotype through salience of group achievement. *Journal of Experimental Social Psychology, 39*, 83–90.
- Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2004). *When mediation is moderated and moderation is mediated*. Manuscript under review.
- Mussweiler, T., & Strack, F. (2000). The “relative self”: Informational and judgmental consequences of comparative self-evaluation. *Journal of Personality and Social Psychology, 79*, 23–38.
- O’Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women’s math performance. *Personality and Social Psychology Bulletin, 29*, 782–789.
- Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women’s generation of mathematical problem-solving strategies. *Journal of Social Issues, 57*, 55–71.
- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women’s math performance. *Journal of Experimental Social Psychology, 38*, 194–201.
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*, 440–452.
- Schubert, T. W., & Otten, S. (2002). Overlap of self, ingroup, and outgroup: Pictorial measures of self-categorization. *Self and Identity, 1*, 353–376.
- Shih, M., Pittinsky, T., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80–83.
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 290–312). Washington, DC: American Sociological Association.
- Spears, R., Gordijn, E. H., Dijksterhuis, A., & Stapel, D. A. (2004). Reaction in action: Intergroup contrast in automatic behavior. *Personality and Social Psychology Bulletin, 30*, 605–616.
- Spencer, S. J., Steele, C. M., & Quinn, D. (1999). Stereotype threat and women’s math performance. *Journal of Experimental Social Psychology, 35*, 4–28.
- Stangor, C., Carr, C., & Kiang, L. (1998). Activating stereotypes undermines task performance expectations. *Journal of Personality and Social Psychology, 75*, 1191–1197.
- 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.
- Stapel, D. A., Reicher, S. D., & Spears, R. (1994). Social identity, availability and the perception of risk. *Social Cognition, 12*, 1–17.
- Stapel, D. A., & Suls, J. (2004). Method matters: Effects of explicit versus implicit social comparisons on activation, behavior, and self-views. *Journal of Personality and Social Psychology, 87*, 860–875.
- Stapel, D. A., & Tesser, A. (2001). Self-activation increases social comparison. *Journal of Personality and Social Psychology, 81*, 742–750.
- 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.
- Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. *Journal of Personality and Social Psychology*, 77, 1213–1227.
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–48). Pacific Grove, CA: Brooks/Cole.
- 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, United Kingdom: Blackwell.
- Wegener, D. T., & Fabrigar, L. R. (2000). Analysis and design for non-experimental data: Addressing causal and non-causal hypothesis. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 412–450). New York: Cambridge University Press.
- Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin*, 127, 797–826.
- Wood, J. V. (1989). Theory and research concerning social comparisons of personal attributes. *Psychological Bulletin*, 106, 231–248.

Received January 8, 2004

Revision received October 27, 2004

Accepted October 28, 2004 ■

Members of Underrepresented Groups: Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write to Demarie Jackson at the address below. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In your letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, “social psychology” is not sufficient—you would need to specify “social cognition” or “attitude change” as well.
- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

Write to Demarie Jackson, Journals Office, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242.