

## Peers and Performance: How In-Group and Out-Group Comparisons Moderate Stereotype Threat Effects

Ronald A. Elizaga · Keith D. Markman

Published online: 31 October 2008  
© Springer Science + Business Media, LLC 2008

**Abstract** The present study examined how exposure to the performance of in-group and out-group members can both exacerbate and minimize the negative effects of stereotype threat. Female participants learned that they would be taking a math test that was either diagnostic or nondiagnostic of their math ability. Prior to taking the test, participants interacted with either an in-group peer (a female college student) or an out-group peer (a male college student) who had just taken the test, and learned that the student had either performed well or poorly on the test. Exposure to either an in-group or an out-group peer whose performance was consistent with the negative stereotype (a poor-performing female or a strong-performing male) exacerbated stereotype threat effects. In contrast, exposure to an in-group or an out-group peer whose performance challenged the negative stereotype (a strong-performing female or a poor-performing male) eliminated stereotype threat effects. These findings demonstrate that people can look to both in-group and out-group peers as sources of inspiration in the context of a negative stereotype.

**Keywords** Stereotype threat · In-group · Out-group · Social comparison · Performance

Sophistication in science and math has become increasingly important in our present age of technological advancement and complexity. In the continuing struggle for perceived equality among the sexes, however, women appear to be lagging behind in this domain. Throughout the past thirty years, high-school aged females have consistently achieved lower average math proficiency scores than their male counterparts (NCES 2000). Furthermore, by the twelfth grade, females as compared to males are less likely to enjoy math and to believe that they are good at math.

The observed divergence in math performance between males and females has fueled the development of the stereotype that women are inferior to men in

---

R. A. Elizaga · K. D. Markman (✉)  
Department of Psychology, Ohio University, 200 Porter Hall, Athens, OH 45701, USA  
e-mail: markman@ohio.edu

mathematical ability. Knowledge of such a belief is particularly likely to be acquired by individuals who are targets of the stereotype, and when such individuals find themselves in situations where the stereotype is relevant, they run the risk of behaving in a manner that will result in their being evaluated negatively (Spencer et al. 1999). The increased evaluation apprehension experienced by individuals in a stereotyped domain has been termed *stereotype threat* (Steele and Aronson 1995).

According to Steele and his colleagues, the activation of a negative stereotype can elicit a disruptive state that undermines performance and aspirations in stereotype-relevant domains (see Steele et al. 2002, for a review). For instance, research has demonstrated the detrimental effects of stereotype threat on female students' math test performance (Brown and Josephs 1999; Spencer et al. 1999) and African-American students' verbal test performance (Steele and Aronson 1995), in addition to a variety of other groups and performance domains (Croizet and Claire 1998; Gonzales et al. 2002; Levy 1996; Stone et al. 1999).

### Alleviating Stereotype Threat

Given the negative implications of these effects for stigmatized individuals' success, researchers have begun examining methods for combating stereotype threat. Useful strategies that have been investigated include providing individuals with a situational explanation for arousal and poor performance (Brown and Josephs 1999; Stone et al. 1999), providing instructions to view intelligence as a malleable trait (Aronson et al. 2002), and informing women about stereotype threat (Johns et al. 2005).

Of particular relevance to the present research are studies that have examined the ameliorative effects of exposure to positive role models (Blanton et al. 2000; Marx and Roman 2002; Marx et al. 2005; McIntyre et al. 2003). For instance, Marx et al. (2005, Experiment 3) had female participants read a bogus newspaper article describing a female student who was either very intelligent and excelled at math (positive social comparison target) or quite unintelligent and did not excel at math (negative social comparison target), and then take a math test. Relative to conditions in which the test was described as a reasoning exercise, participants for whom the test was described as being diagnostic of their math ability performed better on the math test after reading about the positive social comparison target, but performed worse on the test after reading about the negative social comparison target. According to these researchers, participants performed better under stereotype threat conditions when provided with positive (stereotype-disconfirming) information because knowing that an in-group member had done well in the stereotyped domain reduced concerns about the impression they were making in the testing situation.

*Challenging stereotypes by exposure to counter-stereotypic exemplars* Marx et al. (2005, Experiment 3) employed a measure of impression-related concerns (for example, "I am concerned about what others think of me") and found that scores on this measure partially mediated the relationship between the interactive effects of test diagnosticity (threat vs. no threat) and social comparison information (positive vs. negative) on math test performance. By definition, however, the evidence for partial

mediation implies that additional processes must also be at work here, and it is useful to speculate on what some of these might be.

Recently, a related literature has focused on the malleability of automatic stereotypes and prejudice in response to changes in the social context (Dasgupta and Greenwald 2001; Macrae et al. 1995; Wittenbrink et al. 2001). For instance, Dasgupta and Greenwald (2001) found that participants who were exposed to admired African-Americans and disliked European-Americans were subsequently less likely to express automatic race bias. Similarly, Dasgupta and Asgari (2004) showed that people who were exposed to pictures and biographies of famous women leaders were subsequently more likely to automatically associate women with leadership qualities.

Arguably, the effect of positive role model exposure on minimizing stereotype threat-induced performance decrements (Marx and Roman 2002; Marx et al. 2005; McIntyre et al. 2003) could be attributable to a process whereby counter-stereotypic exposure reduces the automaticity of stereotypical associations. In other words, exposing women to an in-group member who has performed well (counter-stereotypically) in a stereotype-relevant domain may decrease the accessibility of the negative stereotype and thereby negate the threat that had previously been aroused by receiving information about the diagnosticity of the test. In our view, such a mechanism might work in concert with, but also orthogonal to, concomitant decreases in impression-related concerns.

*Self-evaluative consequences of in-group and out-group comparisons* All of the work thus far that has examined the effects of exposure to positive and negative female role models on the moderation of stereotype threat has employed manipulations whereby participants are exposed to comparison information about a fellow *in-group* member (another woman). What, however, might be the effects of women's exposure to positive versus negative comparison information about an *out-group* member (a man) on subsequent performance in a stereotype-relevant domain?

Several studies have examined the self-evaluative consequences of comparisons to in-group and out-group members (Blanton et al. 2000; Brewer and Weber 1994; Brown et al. 1992; Mussweiler and Bodenhausen 2002). For instance, Blanton et al. (2000) gave performance feedback to a group of African-American female participants and then exposed them to either upward or downward social comparison information about the performance of a White or African-American female confederate. When the confederate was White, a contrast effect was observed such that participants reported higher state self-esteem in the downward than in the upward comparison condition. In contrast, an assimilation effect was observed when the confederate was African-American, such that participants reported higher state self-esteem in the upward than in the downward social comparison condition.

In related work, Mussweiler and Bodenhausen (2002) demonstrated assimilative effects on self-evaluations following in-group comparisons but contrastive effects following out-group comparisons. Drawing on the selective accessibility framework recently extended by Mussweiler and his colleagues (Mussweiler 2003; Mussweiler and Strack 2000), these researchers argued that in-group comparisons activate individuating knowledge indicating that the self is similar to the target. This results

in self-evaluations that are consistent with the implications of target-consistent self-knowledge. As a consequence, self-evaluations following in-group comparison are likely assimilated toward the target. On the other hand, out-group comparisons should tend to elicit contrast effects on self-evaluations because such comparisons activate category knowledge indicating that the self is different from the target.

From this perspective, the moderating influence of social comparisons on stereotype threat-related performance could stem, in part, from the consequences of enhancing the accessibility of a specific subset of knowledge about the self. Thus, exposing women to an in-group member who has performed *well* may enhance the accessibility of self-knowledge consistent with the possibility of performing well on the upcoming math test. This increased accessibility may then counteract the threat aroused by receiving information about the diagnosticity of the test and enhance the perceived attainability of a good performance (Lockwood and Kunda 1997).

Furthermore, exposure to an out-group member who has performed *poorly* should also counteract the threat, but by a somewhat different mechanism. To the extent that out-group comparisons render accessible category knowledge indicating that the self is different from the target, a dissimilarity-testing mechanism should be instigated (Mussweiler 2003). In the specific case of the downward out-group comparison, a woman may test the hypothesis that she is different from the poor-performing male target and thereby activate both individuating (“I’m OK at math”) and category knowledge (“Women aren’t *that* bad at math”) that effectively short-circuits the negative impact of stereotype threat on performance.

The goal of the present research is to initiate an integration of findings in the areas of stereotype malleability, the self-evaluative consequences of in-group and out-group comparisons, and stereotype threat. In case of point, stereotype malleability studies that feature exposure of participants to stereotypic and counter-stereotypic exemplars (Dasgupta and Greenwald 2001; Dasgupta and Asgari 2004) have not examined the effects of such exposure on subsequent behavior (task performance). Moreover, stereotype threat studies that have demonstrated the effects of comparisons with role models on subsequent math performance (Marx and Roman 2002; Marx et al. 2005; McIntyre et al. 2003) have only examined the consequences of in-group comparisons (other females).<sup>1</sup>

The present study, however, examines the effects of exposure to stereotypic and counter-stereotypic in-group (female) and out-group (male) exemplars on subsequent math performance. High school and college students are routinely bombarded with information pertaining to the academic abilities and accomplishments of their same- and opposite-sex peers. The present research examines the behavioral consequences of social comparisons to same- and opposite-sex peers who have performed in a manner that either confirms or disconfirms a negative stereotype about women’s math abilities.

---

<sup>1</sup> Some of this past research (McIntyre et al. 2003) has examined the consequences of out-group comparisons for male participants who compared themselves to female role models. However, the present study would be the first to examine the effects of female out-group comparisons (to male role models) on performance in stereotype threat situations.

## Study Overview

The present study investigated the effects of exposure to stereotype-confirming or stereotype-disconfirming information about a peer's math test performance on participants' subsequent math performance. Female participants either learned that they would be taking a test that was described as being diagnostic of math ability and was able to identify a person's mathematical strengths and weaknesses (*diagnostic* condition), or, they would be engaging in a reasoning exercise (*nondiagnostic* condition). This manipulation has successfully created a situation of stereotype threat in previous research (Marx et al. 2005; Steele and Aronson 1995). Prior to taking the test, however, participants interacted with either an in-group member (a *female* college student) or an out-group member (a *male* college student) who had just taken the test. From this interaction, participants learned that the student had either performed well (*strong-performer*) or poorly (*poor-performer*) on the test. Subsequently, participants took the test themselves.

It was predicted that exposure to either an in-group member or an out-group member whose performance was consistent with the negative stereotype (the female poor-performer or the male strong-performer) would elicit classic stereotype threat effects (poorer performance in the diagnostic condition relative to the nondiagnostic condition). In contrast, exposure to either an in-group member or an out-group member whose performance challenged the negative stereotype (the female strong-performer or the male poor-performer) would eliminate stereotype threat effects (no predicted difference in performance between the diagnostic and nondiagnostic conditions).

## Method

### Participants

One hundred sixty-five female participants were recruited from introductory psychology courses at a U.S. Midwestern university in partial fulfillment of a course requirement and randomly assigned to conditions of a 2 (*Test Diagnosticity*: diagnostic vs. nondiagnostic) × 2 (*Exemplar Sex*: female vs. male) × 2 (*Exemplar Performance*: strong vs. poor) between-subjects factorial design. A human subjects review committee approved the study prior to commencement.

### Materials

The math task comprised forty-five multiple-choice questions drawn from practice tests for the quantitative section of the Graduate Records Examination (GRE).<sup>2</sup> Most of the questions were word problems requiring algebraic calculations. This type of

<sup>2</sup> Twenty of the forty-five items came directly from stimulus materials developed by Schmader (2002), and twenty-five additional items were selected from an SAT preparation guide. This forty-five item test, somewhat lengthier than the tests typically employed by researchers in this area, was developed in order to examine additional hypotheses regarding the number of items attempted and skipped. Analyses, however, indicated no effects on these measures.

exam is suitable for this research given the results of a previous meta-analysis indicating that sex differences in math performance manifest only on tests requiring complex problem solving (Hyde et al. 1990; cf. Schmader 2002; Spencer et al. 1999).

## Procedure

Participants arrived and waited in the lobby outside of the laboratory. A male experimenter greeted each participant individually, escorted her into a separate waiting room, and announced that the study examined “working styles in problem solving.”

*Test description manipulation* Participants read instructions indicating that they were about to engage in a math task that would be used to examine problem-solving styles. In the *diagnostic* condition, the instructions indicated that the math task had been shown to be diagnostic of math ability and was also known to produce sex differences, whereas in the *nondiagnostic* condition, participants simply learned that the math task “was being evaluated across a large group of students” (Spencer et al. 1999).

*Exemplar sex and performance manipulations* In the *female exemplar* condition, the experimenter explained to the participant that she had been paired with a female student, whereas in the *male exemplar* condition, the experimenter explained that the participant had been paired with a male student. Participants were further told that the student with whom they had been paired was currently completing the experiment, and that the participant’s data were now needed in order to examine similarities and differences in working styles. The experimenter then left the room, ostensibly to check on the other student’s progress.

The experimenter returned five minutes later with either the male or female student with whom the participant supposedly had been paired. The student, a confederate of the experimenter, was told to wait while the experimenter checked to make sure that “all was well with the data.” When the experimenter left the room, the confederate casually initiated a conversation with the participant, asked if she had completed the math task yet, and proceeded to tell her how he (she) had performed. In the *strong performance* condition the confederate mentioned that he (she) had done well on the math task, scoring in the ninetieth percentile among the university’s undergraduates, whereas in the *weak performance* condition the confederate mentioned that he (she) had done poorly on the math task, scoring in only the tenth percentile of the university’s undergraduates.

About one minute after this conversation ended, the experimenter returned to escort the participant to a small computer cubicle, and the participant was then given twenty-five minutes to work on the math task, presented on MediaLab (Jarvis 2004) experimental software. Upon completion, participants were thoroughly debriefed and thanked for their participation.

## Results

Data from seven participants were eliminated, four because they appeared to have answered the questions randomly, two because they chose not to complete the

experiment, and one because she knew the confederate and thus voiced suspicion during debriefing. Analyses were thus performed on the data provided by the one hundred fifty-eight participants that remained. Fifty-three of these one hundred fifty-eight participants reported their scores on the quantitative section of the SAT, and analyses revealed no differences as a function of stereotype salience, exemplar sex, or exemplar performance, all  $p$ s > 0.15. Because only a minority of participants reported SAT scores, these scores were not used as a covariate in any subsequent analyses.

To examine the effect of comparison to stereotypic and counter-stereotypic exemplars on women's math test performance, a 2 (Test Diagnosticity) × 2 (Exemplar Sex) × 2 (Exemplar Performance) analysis of variance (ANOVA) was conducted on the number of test items answered correctly (see Marx et al. 2005). Results revealed a main effect of Test Diagnosticity,  $F(1, 150) = 4.49$ ,  $p = 0.04$ ,  $\eta^2 = 0.03$ , indicating that participants in the diagnostic condition answered fewer items correctly ( $M = 8.27$ ) than did participants in the nondiagnostic condition ( $M = 9.44$ ). Importantly, the analysis also revealed the predicted Diagnosticity × Sex × Performance interaction,  $F(1, 150) = 7.32$ ,  $p = 0.008$ ,  $\eta^2 = 0.05$  (see Table 1), and simple effects tests were subsequently performed on the female and male exemplar conditions in order to examine our specific hypotheses.

Focusing first on the female exemplar condition, participants exposed to the poor-performing (stereotypic) female exemplar answered fewer items correctly in the diagnostic condition ( $M = 7.00$ ) than in the nondiagnostic condition ( $M = 10.50$ ),  $t(137) = 2.88$ ,  $p = 0.005$ ,  $d = 1.04$ . In contrast, participants exposed to the strong-performing (counter-stereotypic) female exemplar did not differ with regard to the number of items answered correctly in the diagnostic ( $M = 9.40$ ) versus nondiagnostic ( $M = 9.10$ ) conditions,  $t < 1$ .

Analyses were then conducted on the male exemplar condition. Consistent with predictions, participants exposed to the strong-performing (stereotypic) male exemplar answered fewer items correctly in the diagnostic condition ( $M = 7.26$ ) than in the nondiagnostic condition ( $M = 9.43$ ),  $t(137) = 2.14$ ,  $p = 0.03$ ,  $d = 0.82$ . In contrast, participants exposed to the poor-performing (counter-stereotypic) male exemplar did not differ with regard to the number of items answered correctly in the diagnostic ( $M = 9.25$ ) versus nondiagnostic ( $M = 8.86$ ) conditions,  $t < 1$ .

**Table 1** Number of items answered correctly as a function of test diagnosticity, exemplar sex, and exemplar performance

Exemplar performance	Test diagnosticity			
	Nondiagnostic		Diagnostic	
	Poor	Strong	Poor	Strong
Female exemplar	10.50 (4.08)	9.11 (4.61)	7.00 (2.72)	9.40 (2.54)
Male exemplar	8.86 (4.53)	9.43 (2.63)	9.25 (3.81)	7.26 (3.97)

Cell means refer to the number of items answered correctly (out of 45). Standard deviations appear in parentheses

## Discussion

The present research examined the effects of exposure to stereotype-confirming or stereotype-disconfirming information regarding a peer's math test performance on female participants' subsequent math performance. Extending previous work that has demonstrated the moderating effects of exposure to female role models on math performance (Blanton et al. 2000; Marx and Roman 2002; Marx et al. 2005; McIntyre et al. 2003), the present work examined the effects of exposure to stereotypic and counter-stereotypic in-group (female) and out-group (male) exemplars on subsequent performance.

It was predicted that exposure to either an in-group member or an out-group member whose performance confirmed the negative stereotype about women's math abilities would elicit stereotype threat effects. Consistent with predictions, those exposed to either the female poor-performer or the male strong-performer answered fewer items correctly in the diagnostic condition than in the nondiagnostic condition.

Exposure to an in-group or out-group member whose performance *challenged* the negative stereotype, however, was predicted to eliminate stereotype threat effects. Again, consistent with predictions, no differences between the diagnostic and nondiagnostic conditions were observed for participants who were exposed to either the female strong-performer or the male poor-performer. In all, these results demonstrate the substantial moderating influence of peer-group comparisons on women's math performance when negative stereotypes are salient.

The results of the present study raise some intriguing theoretical questions for the social comparison literature. Stapel and Koomen (2001) recently extended a framework that describes the moderating influence of self-construal orientation on the evaluative consequences of social comparisons. According to this perspective, 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. The focus then is on how the self and comparison others are different in a manner that yields evaluative contrast. When a person's social self is more accessible, on the other hand, that person shifts into a "we" frame of mind and is likely to value being part of a group. As such, similarities with comparison others are emphasized in a manner that yields assimilation.

Extending this perspective, Marx et al. (2005) found that a collective self-construal orientation (Aron et al. 1992; Brewer and Gardner 1996; Gardner et al. 2002) is activated for those individuals targeted by stereotype threat. Moreover, because a stereotype threat situation enhances feelings of "we-ness," exposure to positive social comparison information minimizes the negative effects of stereotype threat on the performance of stereotyped individuals.

How might Marx et al.'s (2005) self-construal orientation perspective be applied toward the findings in the out-group exemplar conditions of the present study? According to Marx et al., stereotype threat conditions should heighten the accessibility of the collective self, thereby enhancing feelings of "we-ness" that pull for assimilative responses to in-group comparison information. The consequences of exposure to out-group comparison information, however, are less clear. If the individual remains in a "we" frame of mind under stereotype threat conditions, then perhaps exposure to out-group comparison information under these conditions activates a competing "they" representation.



This is perhaps most likely to occur following exposure to a superior out-group exemplar—the male strong-performer in the present study. Via an exclusion mechanism (Schwarz and Bless 1992; Stapel and Koomen 2001; see also Markman and McMullen 2003), contrastive effects on self-evaluations (Blanton et al. 2000) and performance are likely to obtain here, and the results of the present study do in fact provide evidence of contrast in the form of lowered performance following exposure to the strong-performing male under stereotype threat conditions.

Exposure to a poor-performing out-group exemplar, on the other hand, may only briefly activate a competing “they” representation. Subsequently, however, women whose collective self-construal has already been activated under stereotype threat conditions may come to commiserate with the poor-performing male and switch from an exclusionary to an inclusionary mind-set by thinking, “they’re just like us.” From Marx et al.’s perspective, such an “out-group inclusion” mechanism may result in a lowering of impression-related concerns that serve to minimize the negative effects of the stereotype on performance.

As described earlier, however, the selective accessibility framework extended by Mussweiler and his colleagues (Mussweiler 2003) offers a somewhat different account of the present findings. To the extent that out-group comparisons enhance the perception that the self is different from the target, a dissimilarity-testing mechanism should be instigated. Following exposure to the strong-performing male, the female participant may test the hypothesis that she is different from him, thereby activating both individuating (“I’m not good at math”) and category knowledge (“Women typically aren’t good at math”) that exerts detrimental effects on subsequent performance.

On the other hand, testing the hypothesis that she is different from the poor-performing male may activate individuating (“I’m OK at math”) and category knowledge (“Women aren’t *that* bad at math”) that short-circuits the negative impact of the stereotype on her performance. Future research might attempt to tease apart the self-construal and selective accessibility explanations by examining both the specific subset of knowledge and the more general mind-set (self-construal orientation) that is activated following exposure to in-group and out-group comparison information under both stereotype and nonstereotype threat conditions.

In addition, future work might be profitably directed toward understanding the extent to which the in-group and out-group exposure effects reported in this and other studies are due to the strengthening versus weakening of stereotypical associations in memory. As previously discussed, the work of Dasgupta and her colleagues (Dasgupta and Asgari 2004; Dasgupta and Greenwald 2001) has demonstrated how exposure to counter-stereotypic exemplars can weaken the automaticity of stereotypical associations. In the present study, exposure to stereotype-disconfirming exemplars (the female strong-performer and the male poor-performer) may have decreased the accessibility of the negative stereotype and thereby minimized the detrimental effects of stereotype threat on performance. From this perspective, the effects of exemplar exposure on subsequent performance are more attributable to varying levels of stereotype activation than to social comparison processes per se.

Regardless of the specific underlying processes that may be at work here, our study does demonstrate that people can look to members of meaningful in-groups

and out-groups as sources of inspiration in the context of a negative stereotype. Indeed, the results of the present study suggest that female students would do well to acknowledge that not all of their male peers perform at a superior level in math and science. In light of the stereotype regarding divergences in math ability between men and women, female students may selectively attend to evidence that confirms the stereotype (strong-performing male peers) and fail to attend to evidence that disconfirms the stereotype (average- or poor-performing male peers; cf. Klayman and Ha 1987). An initiative whereby teachers and guidance counselors refocus female students' attention on disconfirming information that promotes out-group inclusion (for example, "There's really no difference between us") might go a long way toward ameliorating the present math performance discrepancy between men and women.

## 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.
- Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, *38*, 113–125.
- 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.
- Brewer, M. B., & Gardner, W. L. (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 social comparisons. *Journal of Personality and Social Psychology*, *66*, 268–275.
- 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.
- 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.
- 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.
- Dasgupta, N., & Asgari, S. (2004). Seeing is believing: exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology*, *40*, 642–658.
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 800–814.
- 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.
- Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: a meta-analysis. *Psychological Bulletin*, *107*, 139–155.
- Jarvis, W. B. G. (2004). MediaLab [computer software]. New York: Empirisoft.
- Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: teaching stereotype threat as a means of improving women's math performance. *Psychological Science*, *16*, 175–179.
- Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*, 211–228.

- Levy, B. (1996). Improving memory in old age through implicit self-stereotyping. *Journal of Personality and Social Psychology*, *71*, 1092–1107.
- 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.
- Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). The dissection of selection in person perception: inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology*, *69*, 397–407.
- Markman, K. D., & McMullen, M. N. (2003). A reflection and evaluation model of comparative thinking. *Personality and Social Psychology Review*, *7*, 244–267.
- 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., & Muller, D. (2005). We can do it: the interplay of construal orientation and social comparisons under threat. *Journal of Personality and Social Psychology*, *88*, 432–446.
- McIntyre, R. B., Paulson, R. M., & Lord, C. G. (2003). Alleviating women's mathematics stereotype threat through salience of group achievements. *Journal of Experimental Social Psychology*, *39*, 83–90.
- Mussweiler, T. (2003). Comparison processes in social judgment: mechanisms and consequences. *Psychological Review*, *110*, 472–489.
- Mussweiler, T., & Bodenhausen, G. V. (2002). I know you are but what am I: self-evaluative consequences of judging in-group and out-group members. *Journal of Personality and Social Psychology*, *82*, 19–32.
- Mussweiler, T., & Strack, F. (2000). The "relative self": informational and judgmental consequences of self-evaluation. *Journal of Personality and Social Psychology*, *79*, 23–38.
- National Center for Education Statistics. (2000). *Trends in educational equity of girls and women* (NCES 2000-030). Washington, DC: U.S. Department of Education.
- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, *38*, 194–201.
- Schwarz, N., & Bless, H. (1992). Constructing reality and its alternatives: an inclusion/exclusion model of assimilation and contrast effects in social judgment. In L. L. Martin, & A. Tesser (Eds.), *The construction of social judgments* (pp. 217–245). Hillsdale: Erlbaum.
- Spencer, S. J., Steele, C. M., & Quinn, D. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*, 4–28.
- Stapel, D. A., & Koomen, W. (2001). I, we, and the effects of others on me: how self-construal level moderates social comparison effects. *Journal of Personality and Social Psychology*, *80*, 766–781.
- Steele, C. M., & 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: Academic.
- 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.
- Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, *81*, 815–827.