Available online at www.sciencedirect.com
SCIENCE DIRECT®
ACADEMIC PRESS
Journal of Experimental Social Psychology 39 (2003) 456–467
Journal of Experimental Social Psychology
www.elsevier.com/locate/jesp

Stereotype Lift

Gregory M. Walton* and Geoffrey L. Cohen

Department of Psychology, Yale University, 2 Hillhouse Avenue, New Haven, CT 06520, USA

Received 25 September 2001; revised 9 October 2002

Abstract

When a negative stereotype impugns the ability or worth of an outgroup, people may experience *stereotype lift*—a performance boost that occurs when downward comparisons are made with a denigrated outgroup. In a meta-analytic review, members of non-stereotyped groups were found to perform better when a negative stereotype about an outgroup was linked to an intellectual test than when it was not ($d = .24$, $p < .0001$). Notably, people appear to link negative stereotypes to evaluative tests more or less automatically. Simply presenting a test as diagnostic of ability was thus sufficient to induce stereotype lift. Only when negative stereotypes were *explicitly* invalidated or rendered irrelevant to the test did the lift effect disappear.

Keywords: Stereotype; Stereotype threat; Performance; Downward comparison; White privilege; Academic achievement; Prejudice; Race; Gender

> "I propose...to examine the impact of notions of racial hierarchy, racial exclusion, and racial vulnerability and availability on non-blacks... The scholarship that looks into the mind, imagination, and behavior of slaves is valuable. But equally valuable is a serious intellectual effort to see what racial ideology does to the mind, imagination, and behavior of masters."
>
> Toni Morrison, *Playing in the Dark*, pp. 11–12

Stereotype lift is the performance boost caused by the awareness that an outgroup is negatively stereotyped. People may benefit from stereotype lift when the ability or worth of an outgroup is explicitly called into question. But they may also benefit even when there is no specific reference to a stereotyped outgroup, if the performance task is linked to a widely known negative stereotype. Because stereotypes about the intellectual abilities of different social groups are pervasive in American society (Devine, 1989), people may link negative stereotypes to intellectual tests more or less automatically. As research on *stereotype threat* suggests (Spencer, Steele, & Quinn, 1999; Steele & Aronson, 1995; for a review see Steele, 1997), negative stereotypes about women's ability in math, and about racial minorities' intellectual abilities in general, are so ubiquitous that evaluative tests can trigger among members of these groups the fear that, should they do poorly, they could confirm a negative stereotype about their gender or racial group. This fear, in turn, can cause their performance to suffer (see also Cohen & Steele, 2002; Gonzales, Blanton, & Williams, 2002). We argue that negative stereotypes are linked to intellectual tests in the minds of members of non-stereotyped groups too (e.g., men and Whites), but that here their effect is to enhance performance rather than to undermine it.

How do negative outgroup stereotypes improve performance? They do so, we suggest, by encouraging downward social comparisons with a denigrated outgroup (see Blanton, Buunk, Gibbons, & Kuyper, 1999; see also Fein & Spencer, 1997). By comparing themselves with a socially devalued group, people may experience an elevation in their self-efficacy or sense of personal worth (Bandura, 1986), which may, in turn, improve performance. Particularly for difficult tests where one must persist in the face of frustration, the extra boost in feelings of efficacy and worth may be important to maintaining confidence and motivation. Knowing that another group is stereotyped as inferior to their own, people may also expect to be viewed with respect rather than suspicion (see Cohen, Steele, &

* Corresponding author. Fax: 1-203-432-7172.
E-mail address: gregory.walton@yale.edu (G.M. Walton).

0022-1031/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0022-1031(03)00019-2

Ross, 1999; Walton & Cohen, 2003). Non-stereotyped test-takers may feel that even failure is unlikely to bring them the low status or social rejection faced by the outgroup, and this assurance may help them to maintain optimal performance (see also Baumeister, Twenge, & Nuss, 2002). In sum, stereotype-inspired social comparison may alleviate the self-doubt, anxiety, and fear of rejection that could otherwise hamper performance on important intellectual tests (see also Sarason, 1991).

As past research suggests, certain individuals are more likely than others to engage in downward social comparison. For example, people who value a given achievement domain as a source of self-esteem (who are *identified*; Steele, 1997) may view the possibility of failure on a test as particularly self-threatening and may thus welcome the opportunity to buttress their self-worth by comparing themselves with a stereotyped outgroup (see Fein & Spencer, 1997; Hogg, 2000; Tesser, 1988; Wills, 1981). In addition, people who are prejudiced may also be particularly likely to experience stereotype lift, because they view the outgroup more negatively than do people low in prejudice (Wills, 1981).

We provide evidence for stereotype lift using meta-analysis. Relevant studies encompass those that assess the performance of members of a non-stereotyped group in at least two conditions—one condition where a negative outgroup stereotype is linked or made *relevant* to a test, and another condition where such a stereotype is not linked or made *irrelevant* to that test. The data are drawn largely from the non-stereotyped participants used as controls in stereotype threat research. We statistically combine results from these studies to assess whether performance on the part of non-stereotyped people is higher in the stereotype-relevant condition than in the stereotype-irrelevant condition.

We have argued that people link negative stereotypes to intellectual tests more or less automatically, and that they thus experience stereotype lift even when there is no explicit reference to an outgroup. To test this idea, we examine the size of the lift effect as a function of the types of experimental manipulations used in different studies. Almost all the studies included in our sample used an evaluative test in the stereotype-relevant condition, that is, one presented as *diagnostic of ability* (the two exceptions were Sekaquaptewa & Thompson, 2002, Study 2, and Steele & Aronson, 1995, Study 4). If people automatically link such evaluative tests to negative stereotypes, then there should be little if any additional benefit of making that link explicit in the stereotype-relevant condition. That is, studies that explicitly validate the stereotype (e.g., by stating that women perform worse than men on a math test) should show no greater lift effect than studies that do not explicitly validate the stereotype (e.g., by simply presenting a test as diagnostic of math ability). Making the link between the test and the stereotype explicit should have little if any effect if that link is already implicit. Rather, what should be predictive of stereotype lift is whether the *stereotype-irrelevant* condition adequately *refutes* the assumed link between the test and the negative stereotype—for example, by presenting the test as *not* diagnostic of ability or by stating that the test yields *no* group differences in performance.

Method

Retrieval of studies and inclusion criteria

To retrieve relevant studies, we first conducted a January 2002 search of the *PsychINFO* database using the words "stereotype," "threat," and "performance" or "test." Second, we solicited additional studies by emailing the discussion list of the Society for Personality and Social Psychology (*spsp@stolaf.edu*). Finally, we emailed the first author of every study obtained through each of these two methods.

Our inclusion criteria required that studies assess the test performance of members of a real-world, self-identified, and non-stereotyped group (e.g., men and Whites). Participants had to be randomly assigned to one of at least two conditions—one "stereotype-relevant" condition and one "stereotype-irrelevant" condition. The performance test had to be linked to a negative stereotype more in the stereotype-relevant condition than in the stereotype-irrelevant condition. The manipulation could be accomplished through explicit instructions that implied the relevance or validity (or irrelevance or invalidity) of a negative stereotype. Alternatively, the manipulation could be accomplished through features of the test-taking environment that could increase (or decrease) the perceived relevance of a stereotype. For example, male students completing a math test could plausibly infer from the presence of predominantly female test-takers that one purpose of the study involved examining gender differences in math. Studies that used implicit manipulations (e.g., with subliminal stimuli) or manipulations embedded in tasks ostensibly unrelated to the test were thus excluded. Such studies either (a) activate the content of the stereotype without linking that content to a specific outgroup (e.g., Bargh, Chen, & Burrows, 1996; see Steele, Spencer, & Aronson, 2002), or (b) activate the stereotype without making it psychologically relevant to the test.

Finally, because stereotype lift is assumed to alleviate the doubt, anxiety, or fear of rejection that accompanies the threat of failure, it was also required that each included study use a difficult test rather than an easy one. Ultimately, 43 studies meeting these criteria were identified.[1] For each study, we calculated the standard deviation (*SD*), the size of the stereotype lift effect (*d*), the *t* test statistic (*t*), and the *p* value corresponding to the *t* test.[2]

We also used a coding system to test the claim that people link intellectual tests to negative stereotypes automatically. If this claim is accurate, then studies that use a stereotype-relevant treatment condition where an evaluative test is explicitly linked to a negative stereotype will *not* yield a greater lift effect than studies that use a stereotype-relevant treatment condition where the test is merely presented as evaluative in nature. Adding further relevance to the stereotype will have little if any effect if its relevance is assumed. By contrast, studies that use a stereotype-irrelevant condition where the link between the test and the stereotype is refuted (either by presenting the test as non-diagnostic of ability or by presenting it as insensitive to group differences) will yield a substantially larger lift effect than studies that use a stereotype-irrelevant condition where that link is not refuted.

To test our reasoning, we coded each study on two dichotomous scales to assess: (a) whether the stereotype-relevant condition reinforced the link between the test and the stereotype, and (b) whether the stereotype-irrelevant condition refuted that link. The stereotype-relevant condition was coded as a *1* if the link between the test and the stereotype was explicit and thus reinforced (i.e., if the test was presented not only as being diagnostic of ability, but also as yielding group differences in performance) and as a *0* if that link was implicit (e.g., if the test was presented as only diagnostic of ability). The stereotype-irrelevant control condition was coded as a *1* if the link between the test and the stereotype was refuted (e.g., if the test was presented as *not* diagnostic of ability or if it was characterized as *not* yielding group differences in performance) and as a *0* if that link was not refuted (i.e., if the test was presented as diagnostic of ability).

For example, studies that manipulated the perceived diagnosticity of the test (e.g., Steele & Aronson, 1995, Studies 1 and 2) were coded as a *0* for the stereotype-relevant condition (because the test in this condition was presented as diagnostic of ability but was not said to yield group differences in performance) and as a *1* for the stereotype-irrelevant condition (because the test in this condition was presented as non-diagnostic of ability and hence as irrelevant to a negative stereotype). In contrast, studies that manipulated the perception that men outperformed women on a math test (e.g., Spencer et al., 1999) were coded as a *1* for the stereotype-relevant condition (because the test in this condition was said to yield gender differences) and as a *1* for the stereotype-irrelevant condition (because the test in this condition was presented as yielding *no* gender differences and hence as irrelevant to a negative stereotype). Finally, studies that manipulated stereotype relevance by asking participants to note their race or gender before an evaluative test (i.e., Anderson, 2002; Stricker, 1998; Stricker & Ward, 1998) were coded as a *0* for the stereotype-relevant condition (because the test in this condition was not said to yield group differences in performance) and as a *0* for the stereotype-irrelevant condition (because the test in this condition was presented as diagnostic of ability and hence as relevant to a negative stereotype).

Results

Overview of sample

The key characteristics of each study included in the meta-analysis are presented in Table 1. The summary statistics and condition codes are presented in Table 2. As noted in Table 2, 30 of the 43 studies showed the predicted pattern of results (in four studies the results were statistically significant). The stereotype-relevant condition yielded higher performance on the part of members of non-stereotyped groups than did the stereotype-irrelevant condition. The effect sizes were normally distributed ($\text{skewness} = .06$, $SE = .36, ns$; $\text{kurtosis} = .26$, $SE = .71, ns$).

Overall tests of homogeneity of effect sizes and statistical significance

We followed procedures outlined by Hedges and Olkin (1985) to calculate all meta-analytic statistics. The homogeneity test assesses whether the sampled effect sizes are likely to derive from the same population. They were not, $Q_T = 59.65 = \chi^2_{42}$, $p = .038$, indicating variance in effect sizes among the sampled studies (a point to

[1] A number of studies failed to meet our criteria. Ten studies had no condition that exposed participants to a negative outgroup stereotype without simultaneously exposing them to a positive ingroup stereotype (Cheryan & Bodenhausen, 2000; Kray, Thompson, & Galinsky, 2001, Studies 1, 2, 3, and 4; Leyens, Desert, Croizet, & Darcis, 2000; Shih, Pittinsky, & Ambady, 1999, Studies 1 and 2; Stone, Lynch, Sjomeling, & Darley, 1999, Studies 1 and 2). Seven studies had no theoretically appropriate stereotype-irrelevant control group (Aronson et al., 1999, Studies 1 and 2; Brown & Josephs, 1999, Studies 1, 2, and 3; Quinn & Spencer, 2001, Study 1; Walsh, Hickey, & Duffy, 1999, Study 2). Six studies used manipulations that were either implicit or embedded in tasks ostensibly unrelated to the dependent measure (Davies, Spencer, Quinn, & Gerhardstein, 2001, Studies 2 and 3; Dijksterhuis & van Knippenberg, 1998, Study 3; Levy, 1996, Study 2; Wheeler, Jarvis, & Petty, 2001, Studies 1 and 2). One study used a test irrelevant to a negative outgroup stereotype (Sekaquaptewa & Thompson, 2002, Study 1).

[2] $SD = \text{SQRT}[[(n1-1)SD1\hat{}2 + (n2-1)SD2\hat{}2]/(n1+n2-2)]$; $d = (M1-M2)/SD$; $t = d * \text{SQRT}[(n1*n2)/(n1+n2)]$, where $n1$ and $n2$, $SD1$ and $SD2$, and $M1$ and $M2$, refer to the sample size, the standard deviation of test performance, and the mean test performance, in the stereotype-relevant condition and in the stereotype-irrelevant condition.

Table 1
Characteristics of included studies

Study No.	Study	Participants	Stereotype	Manipulation of stereotype relevance	Dependent measure
1	Anderson (2002)	Male students	Gender	Before ability-diagnostic test, participants either indicated race or not	Math and science general education instrument
2	Blascovich, Spencer, Quinn, and Steele (2001)	White students	Race	Test characterized either as diagnostic of "intelligence" or as "culturally unbiased" (i.e., yielding no race differences)	Remote Associates Test
3	Broadnax, Crocker, and Spencer (1997)	White students	Race	Test characterized either as an "excellent indicator of English competence" or as a "culture fair" test yielding "no race differences"	English literature test
4	Brown, Steele, and Atkins (2001), Study 1	White students	Race	Test characterized either as diagnostic of verbal skills or as "racially and ethnically unbiased"	Verbal GRE
5	Brown et al. (2001), Study 2	White students	Race	Test characterized either as diagnostic of verbal skills or as "racially and ethnically unbiased"	Verbal GRE
6	Croizet and Claire (1998)	High SES students	SES	Test characterized either as "diagnostic" or as "non-diagnostic" of "intellectual ability"	GRE-like verbal questions
7	Danso and Esses (2001), Pilot Study	White students	Race	Either Black or White experimenter administered ability-diagnostic test	Necessary Arithmetic Operations
8	Danso and Esses (2001), Main Study	White students	Race	Either Black or White experimenter administered ability-diagnostic test	Necessary Arithmetic Operations
9	Davies et al. (2001), Pilot Study	Male students	Gender	Test characterized either as "diagnostic" or as "non-diagnostic" of mathematical ability	Math GRE
10	Ewing and Smith (2001)	Young students	Age	Participants either told aging is related to memory loss or told aging is unrelated to memory loss	Weschler Adult Intelligence Scale-III
11	Foels (2000)	Male students	Gender	Test characterized either as a math test or as a math test yielding "no gender differences"	Math GMAT
12	Gonzales et al. (2002), analyses by gender	Male students	Gender	Test characterized either as "diagnostic" or as "non-diagnostic" of mathematical and spatial abilities	Wonderlic Personnel Test
13	Gonzales et al. (2002), analyses by race	White students	Race	Test characterized either as "diagnostic" or as "non-diagnostic" of mathematical and spatial abilities	Wonderlic Personnel Test
14	Inzlicht and Ben-Zeev (2000), Study 2	Male students	Gender	Groups composed either of one man and two women or of three men completed ability-diagnostic test	Math GRE
15	Josephs, Newman, Brown, and Beer (2001), Study 1	Male students	Gender	Before ability-diagnostic test, participants either completed questions about gender and math or completed questions about school	Math GRE
16	Keller (2002)	Male adolescents	Gender	Test characterized either as yielding "gender differences" or as yielding no gender differences	26 math problems
17	Keller and Dauenheimer (2002)	Male adolescents	Gender	Test characterized either as yielding "gender differences" or as yielding no gender differences	26 math problems
18	Martens, Johns, Greenberg, and Schimel (2002)	Male students	Gender	Test either said to assess "math and reasoning abilities" or said to be a pilot test for future studies	Math GMAT
19	Marx and Roman (2002), Study 1	Male students	Gender	Either female or male experimenter administered ability-diagnostic test	Math GRE
20	Marx and Roman (2002), Study 2	Male students	Gender	Either math-incompetent or math-competent female experimenter administered ability-diagnostic test	Math GRE

Table 1 (*continued*)

Study No.	Study	Participants	Stereotype	Manipulation of stereotype relevance	Dependent measure
21	McKay, Doverspike, Bowen-Hilton, and Martin (2002)	White students	Race	Test characterized either as "diagnostic" or as "non-diagnostic" of intellectual ability	Raven Advanced Progressive Matrices
22	Quinn and Spencer (2001) Study 2	Male students	Gender	Test characterized either as "math test" or as "gender-fair" math test (i.e., yielding no gender differences)	Math SAT
23	Salinas (1998), Study 1	White students	Race	Midway through the ability-diagnostic test, participants either had the opportunity to label the test as biased or did not have that opportunity	Verbal GRE
24	Salinas (1998), Study 2	White students	Race	Midway through the ability-diagnostic test, participants either had the opportunity to label the test as biased or did not have that opportunity	Verbal GRE
25	Schmader (2002)	Male students	Gender	Study said to assess either gender differences in math ability or individual differences in math ability	Math GRE
26	Schultz, Baker, Herrera, and Khazian (2002), Study 1	White students	Race	Study said to assess either racial differences in verbal skills or individual differences in verbal skills	Verbal GRE
27	Schultz et al. (2002), Study 2	White students	Race	Study said to assess either racial differences in verbal skills or individual differences in verbal skills	Verbal GRE
28	Sekaquaptewa and Thompson (2001)	Male students	Gender	Test characterized either as a "traditional" math test or as "special kind of math material" (i.e., yielding no gender differences)	12 math and math-related questions
29	Sekaquaptewa and Thompson (2002), Study 2	White students	Race	Groups composed either of one White and three Blacks or of four Whites completed non-diagnostic test	16 questions about classifications of animal species
30	Spencer (1993), Study 4	Male students	Gender	Test characterized either as yielding "gender differences" or as yielding "no gender differences"	Math GRE
31	Spencer (1993), Study 5	Male students	Gender	Test characterized either as a math test, as a math test yielding "gender differences," or as a math test yielding "no gender differences"	Math GRE
32	Spencer, Iserman, Davies, and Quinn (2002)	Male students	Gender	Test characterized either as "diagnostic" of math ability or as yielding "no gender differences"	Math GRE under cognitive load
33	Spencer et al. (1999), Study 2	Male students	Gender	Test characterized either as yielding "gender differences" or as yielding "no gender differences"	Math GRE
34	Spencer et al. (1999), Study 3	Male students	Gender	Test characterized either as yielding "gender differences" or as yielding "no gender differences"	Math GMAT
35	Steele and Aronson (1995), Study 1	White students	Race	Test characterized either as "diagnostic" or as "non-diagnostic" of verbal ability	Verbal GRE
36	Steele and Aronson (1995), Study 2	White students	Race	Test characterized either as "diagnostic" or as "non-diagnostic" of verbal ability	Verbal GRE
37	Steele and Aronson (1995), Study 4	White students	Race	Before non-diagnostic test, participants either indicated race or not	Verbal GRE
38	Sternberg et al. (2002), Study 1	Male students	Gender	Either men either said to perform better than women, or gender not mentioned; or women said to perform better than men, or men and women said to perform equally well	Math GRE

39	Sternberg et al. (2002), Study 2	Male students	Gender	Either men either said to perform better than women, or gender not mentioned; or women said to perform better than men, or men and women said to perform equally well	Math GRE
40	Stricker (1998), analyses by gender	Male students	Gender	Before ability-diagnostic test, participants either indicated gender or not	AP Calculus AB Examination
41	Stricker (1998), analyses by race	White students	Race	Before ability-diagnostic test, participants either indicated race or not	AP Calculus AB Examination
42	Stricker and Ward (1998), analyses by gender	Male students	Gender	Before ability-diagnostic test, participants either indicated gender or not	Two math Computerized Placement Tests
43	Stricker and Ward (1998), analyses by race	White students	Race	Before ability-diagnostic test, participants either indicated race or not	Two math and two verbal Computerized Placement Tests

which we return shortly). After correcting for small sample sizes (Hedges & Olkin, 1985), we conducted the initial meta-analytic test to assess the overall effect size and significance level of stereotype lift. The effect size was $d = .10$, and it differed significantly from $d = 0$, $z = 3.40$, $p < .01$. We also calculated the fail-safe n (Rosenthal, 1984), and found that an additional 122 studies, each one yielding a null stereotype lift effect (i.e., $d = 0$), would be required to render this primary result non-significant. However, the heterogeneity in effect sizes suggests that one or more variables moderate stereotype lift.

The moderating role of stereotype relevance

As discussed in the Introduction, one potential moderator involves whether the stereotype-irrelevant condition severs the link between the test and the stereotype. If people link evaluative tests to negative stereotypes more or less automatically, studies will yield a between-condition difference only if the stereotype-irrelevant condition refutes either the content of the stereotype or its relevance to the test. To test this idea, we conducted an internal analysis using the two dichotomous codes assigned to each study.

As predicted, the size of the stereotype lift effect did not differ between the two types of stereotype-relevant treatment conditions, $Q_B = \chi^2_1 = .14, ns.$[3] Studies that made the link between the test and the stereotype explicit (i.e., by stating that an ability-diagnostic test showed group differences in performance) yielded roughly the same effect size ($d = .11$) as studies that simply presented the test as diagnostic of ability ($d = .07$). The link between negative outgroup stereotypes and evaluative tests is thus so strong that there is no additional benefit of making that link explicit. By contrast, there was a significant effect involving the two types of stereotype-irrelevant control conditions, $Q_B = \chi^2_1 = 8.66$, $p = .003$. Studies that refuted either the validity of the stereotype or its relevance to the test yielded a stronger lift effect ($d = .22$) than studies that did not ($d = .00$). Whether the treatment condition reinforces the relevance of a negative stereotype beyond what is normally assumed for the typical evaluative test does not predict larger lift effects. On the contrary, what is predictive is whether (in the control condition) the link between the test and the stereotype is refuted.

The moderating role of identification

We found tentative evidence that people who are identified with the performance domain benefit most from

[3] For these internal analyses, we excluded the two studies that presented the test as non-evaluative in both conditions (i.e., Sekaquaptewa & Thompson, 2002, Study 2; Steele & Aronson, 1995, Study 4).

Table 2
Sample summary statistics

Study #	*M*1[a]	*M*2[b]	*SD*	*d*	*N*	*t*	*p*	S-R[c]	S-IR[d]
1	52.98	56.80	11.03	−.35	344	−3.21	<.01	0	0
2	1.40	1.40	.90	0	20	0	1.00	0	1
3	11.32	10.61	3.39	.21	37	.64	.53	0	1
4	.56[e]	.52[e,f]	.20	.23	49	.74	.46	0	1
5	11.11	9.79	3.73	.35	23	.85	.41	0	1
6	11.25	10.28	2.80	.35	64	1.39	.17	0	1
7	12.70	11.00	3.36	.51	95	2.47	.02	0	0
8	13.16	11.94	3.28	.37	100	1.86	.07	0	0
9	.26[e]	.25[e]	.14	.07	52	.25	.80	0	1
10	12.57	11.28	1.87	.69	52	2.49	.02	1	1
11	8.71	8.15	2.58	.22	27	.56	.58	0	1
12	26.39	26.57	4.39	−.04	60[g]	−.16	.87	0	1
13	29.08	27.97	7.07	.16	60[g]	.61	.55	0	1
14	.67[e]	.66[e]	.17	.06	36	.18	.86	0	0
15	15.16	15.05	2.87	.04	75	.17	.87	0	0
16	16.30	16.00	3.77	.08	30[h]	.21	.84	1	1
17	14.20	13.40	3.64	.22	39	.69	.50	1	1
18	8.03	7.21	2.12	.39	27	.99	.33	0	1
19	13.56	15.02	4.37	−.33	20	−.75	.46	0	0
20	15.90	12.79	3.25	.96	20	2.14	.05	0	1
21	20.70	18.68	6.51	.31	42	1.00	.32	0	1
22	8.17	6.03	3.50	.61	14	1.14	.27	0	1
23	–	–	4.50	−.22[i]	19	−.47	.64	0	0
24	–	–	3.42	.18[i]	144	1.07	.29	0	0
25	5.20	6.15	2.13	−.45	33	−1.28	.21	1	0
26	8.43	8.67	2.60	−.09	116	−.50	.62	1	0
27	8.73	8.90	2.86	−.06	80	−.27	.79	1	0
28	14.54	14.54	3.26	0	77	0	1.00	0	1
29	24.48	21.95	5.45	.46	40	1.47	.15	0	1
30	31.50	28.50	20.25	.15	30	.57	.57	1	1
31	25.50[f]	27.00	16.11	−.09	45	−.29	.77	[j]	1
32	18.57	17.74	2.93	.38	29	1.01	.32	0	1
33	26.70	18.90	17.25	.45	24	1.57	.12	1	1
34	21.20	18.50	12.95	.21	31	.58	.57	1	1
35	11.84	12.32	3.79	−.13	38	−.39	.70	0	1
36	10.64	9.13	3.40	.44	20	.99	.33	0	1
37	9.52	6.83	4.01	.67	22	1.57	.13	0	1
38	20.79[f]	23.66[f]	5.63	−.51	72	−2.16	.03	[j]	1
39	20.20[f]	17.70[f]	5.98	.42	222	3.11	<.01	[j]	1
40	40.44	38.08	20.22	.12	154[g,k]	.72	.47	0	0
41	40.27	40.11	20.22	.01	154[g,k]	.05	.96	0	0
42	–	–	–	.13[l]	583[g]	1.54	.12	0	0
43	–	–	–	−.02[l]	612[g]	−.25	.81	0	0

[a] Stereotype-relevant condition.
[b] Stereotype-irrelevant condition.
[c] Stereotype-relevant condition code.
[d] Stereotype-irrelevant condition code.
[e] Percent correct responses.
[f] Combines data from two experimental conditions on the basis of sample size.
[g] Because identical or overlapping groups of participants were included in analyses of gender and race, both analyses are weighted by .5.
[h] Only participants identified with math included.
[i] Effect size calculated from relative change in performance after administration of experimental manipulation halfway through test.
[j] Stereotype-relevant condition could not be clearly assigned to any one coding condition because it was operationalized using multiple methods.
[k] Unit of statistical analysis was classroom not individual.
[l] Effect size calculated by averaging the effect sizes of multiple independent tests completed by overlapping samples.

stereotype lift (Wills, 1981). There was a trend in the predicted direction for studies that preselected participants for high levels of identification with the subject matter being tested (or that preselected on the basis of a relevant criterion variable such as SAT score) to show a somewhat larger lift effect ($d = .20$) than studies that did not ($d = .07$), $Q_B = \chi^2_1 = 2.05$, $p = .15$. Like stereotype threat (Steele, 1997), stereotype lift seems to have a larger effect on people who care most about doing well. While stereotype threat exacerbates pressures felt by people who have invested their self-worth in the domain of evaluation, stereotype lift alleviates those pressures.

Estimation of effect size and clinical significance

The results indicate that people link intellectual tests to negative stereotypes more or less automatically. Only studies that refute that link in the stereotype-irrelevant condition thus provide an accurate estimate of the stereotype lift effect size. We recalculated the meta-analytic statistics using only these studies ($n = 28$). The lift effect was robust, $d = .24$, highly significant, $z = 4.04$, $p < .0001$, and yielded a fail-safe *n* of 132. The homogeneity statistic was not significant, $Q_T = \chi^2_{27} = 24.63$, *ns*, indicating the consistency of the lift effect in this sample of studies.

Stereotype lift can help to explain racial and gender differences in standardized test performance. Several studies included in the meta-analysis used modified SAT tests (or similar standardized tests) as dependent measures. The standard deviation on each section of the SAT is approximately 112 points (College Board, 2001). The obtained lift effect size ($d = .24$) thus indicates that stereotype lift improves performance on each stereotype-relevant section of the SAT by approximately 25 points. Stereotype lift produces a 50-point advantage for White men—a performance boost that, at the most selective colleges, could make the difference between rejection and acceptance.[4]

Comparative analysis of stereotype threat

For exploratory purposes, we performed the same analyses on the effect sizes obtained for stereotype threat. The overall threat effect was significant, $d = .29$, $z = 10.44$, $p < .0001$. The stereotype-relevant condition yielded *lower* performance on the part of members of stereotyped groups than did the stereotype-irrelevant condition. The heterogeneity test was also significant, $Q_T = 84.98 = \chi^2_{40}$, $p < .0001$, suggesting that one or more variables moderate stereotype threat.

Like stereotype lift, the stereotype threat effect was larger among studies that refuted the link between the test and the stereotype in the stereotype-irrelevant condition ($d = .45$) than among studies that did not ($d = .20$), $Q_B = \chi^2_1 = 9.61$, $p = .002$. This result indicates that, like their non-stereotyped peers, Blacks, women, and members of other historically excluded groups link evaluative tests to negative stereotypes more or less automatically. In contrast to stereotype lift, however, the stereotype threat effect was also larger among studies that reinforced the link between the test and the stereotype in the stereotype-relevant condition ($d = .57$) than among studies that did not ($d = .29$), $Q_B = \chi^2_1 = 5.84$, $p = .016$, although this result should be viewed tentatively because it is driven entirely by five studies with disproportionately large sample sizes.[5] Both targets and non-targets thus link intellectually evaluative tests to negative stereotypes (Steele, 1997). Moreover, like stereotype lift, stereotype threat was larger among studies that selected students who were identified with the performance domain ($d = .68$) than among studies that did not ($d = .22$), $Q_B = \chi^2_1 = 22.04$, $p < .0001$, a result consistent with the idea that students at the vanguard of achievement suffer stereotype threat most (see Aronson et al., 1999; Spencer et al., 1999; Steele, 1997).

Because intellectual tests appear to be linked to negative stereotypes more or less automatically, we calculated the stereotype threat effect size using only studies that refuted that link in the stereotype-irrelevant condition ($n = 28$). The effect was robust and significant,

[4] Additional analyses were undertaken to rule out alternative explanations and to test other potential moderators of the lift effect. Stereotype lift does not appear to be due to heightened effort on evaluative tests relative to non-evaluative ones. When analyses were confined to studies that presented the test as equally evaluative in both conditions (i.e., equally diagnostic of ability, $n = 35$), the overall lift effect was very consistent with that obtained in the entire sample, $d = .09$, $z = 2.68$, $p = .007$; $Q_T = \chi^2_{34} = 61.01$, $p = .003$. The lift effect was again larger among studies using stereotype-irrelevant control conditions that refuted the link between the test and the stereotype ($d = .23$) than among studies that did not ($d = .00$), $Q_B = \chi^2_1 = 7.91$, $p = .005$. It also seemed possible that statistical dependency among studies from the same laboratory might violate the assumption of independence. Twenty laboratories are represented in the sample (i.e., studies reported by the same researchers or by overlapping groups of researchers). When laboratory was used as the unit of analysis, the overall lift effect was again very consistent with that obtained when study was the unit of analysis, $d = .11$, $z = 3.34$, $p = .001$; $Q_T = \chi^2_{19} = 36.86$, $p = .008$. Once more, the lift effect was larger among studies using stereotype-irrelevant control conditions that refuted the link between the test and the stereotype ($d = .24$) than among studies that did not ($d = .02$), $Q_B = \chi^2_1 = 7.93$, $p = .005$. Two final moderators were also assessed. Neither publication status, $Q_B = \chi^2_1 = 3.04$, $p > .05$, nor type of stereotype (i.e., racial stereotypes or gender stereotypes), $Q_B = \chi^2_1 = 1.50$, $p > .20$, predicted stereotype lift.

[5] These five studies (Anderson, 2002; Stricker, 1998 analyses by gender and by race; Stricker & Ward, 1998, analyses by gender and by race) did not reinforce the link between the test and the stereotype in the stereotype-relevant condition and they yielded small stereotype threat effects (*d*s = .00 to .19). On average, they used samples sizes 10 times larger than those of the other studies, and thus may have overdetermined the mean effect size for this category of studies. To address this problem, we assessed the effect of the two types of stereotype-relevant conditions using an independent samples *t* test (an analysis insensitive to the sample size used to derive each effect size). In this analysis, there was no difference in the mean effect size observed in the two types of stereotype-relevant conditions, $t(34) = .25$, *ns*. By contrast, the *t* test analysis largely revealed the same moderational patterns found using the meta-analytic statistics reported earlier. There was a trend in the predicted direction such that the type of stereotype-irrelevant condition predicted stereotype threat, $t(37) = 1.45$, $p = .15$. Likewise, while there was no difference in the mean stereotype lift effect size observed in the two types of stereotype-relevant conditions, $t(36) = .35$, *ns*, there was a difference in the mean effect size observed in the two types of stereotype-irrelevant conditions, $t(39) = 2.67$, $p = .011$.

$d = .48$, $z = 8.78$, $p < .0001$; $Q_T = 54.94 = \chi^2_{27}$, $p = .001$. In summary, both stereotype lift and stereotype threat are statistically reliable, with stereotype lift being roughly half the magnitude of stereotype threat.

Discussion

Stereotype lift causes people to perform better in contexts where the ability or worth of an outgroup is impugned. The predicted pattern was found in studies where the relevance of a stereotype was manipulated through the perceived existence of group differences in performance on an administered exam (e.g., Spencer et al., 1999), through the purported diagnostic or evaluative nature of a test (e.g., Croizet & Claire, 1998), and through situational cues indicating the relevance of a negative stereotype to a non-evaluative test (i.e., Sekaquaptewa & Thompson, 2002, Study 2; Steele & Aronson, 1995, Study 4).

Whites, men, and wealthy people link intellectual tests to negative stereotypes automatically. In fact, the link between evaluative tests and negative stereotypes is so strong that there is no additional benefit to performance when that link is made explicit. Only when this link is refuted—for example, when men are told that a math test yields *no* gender differences—does stereotype lift disappear. The culturally ingrained assumption is that evaluative tests yield differences as a function of race, gender, or class (Steele, 1997), and this assumption facilitates performance for members of non-stereotyped groups.

From evidence of moderation to evidence of mediation

If stereotype lift results from downward comparison with a denigrated outgroup, then people who are who are apt to engage in downward comparison should benefit most. People who are identified with the skill under evaluation should be more likely to view the prospect of failure as self-threatening and thus be motivated to buttress their self-worth through downward comparison (Fein & Spencer, 1997; see also Hogg, 2000; Tesser, 1988; Wills, 1981). Some support for this prediction was found. There was a trend for studies that selected participants identified with the subject matter being tested to yield larger lift effects than studies that did not.

Because prejudiced people are more likely to engage in downward comparison with a stereotyped outgroup (Wills, 1981), they should also be more likely to benefit from stereotype lift. Three studies in our sample assessed individual differences in prejudice or in related measures, and each result was consistent with this prediction. Danso & Esses (2001) found that people high in social dominance orientation—the motivation to protect the social hierarchy—performed better than people low in social dominance orientation in the presence of a Black experimenter but not in the presence of a White experimenter. Likewise, Josephs et al. (2001) found that men high in testosterone (a physiological marker of dominance orientation) performed marginally better on a math test than men low in testosterone after being primed on negative gender stereotypes. Finally, Schultz, Baker, Herrera, and Khazian (2002) reported that White students' racial prejudice correlated positively with performance on a test portrayed as assessing racial differences but that racial prejudice did not correlate with performance on a test portrayed as assessing individual differences. Stereotype lift particularly benefits people who believe either in the validity of negative stereotypes, or in the legitimacy of group-based hierarchy.

If stereotype lift is driven by downward comparison, then other variables beyond identification and prejudice should moderate its effect. For example, people low in self-esteem appear to be particularly likely to make downward comparisons to protect their self-image (Aspinwall & Taylor, 1993; Smith & Insko, 1987; Wills, 1981, 1991), and they may thus experience stereotype lift more. Individuals low in self-certainty may also be more likely to experience stereotype lift, because their less stable self-concepts make them more susceptible to the effects of social comparison (Bui & Pelham, 2000; Pelham & Wachsmuth, 1995). If stereotype lift is mediated by downward social comparison, it should also produce specific psychological outcomes, including cognitive activation of the relevant stereotype and elevations in self-efficacy or in perceived social acceptance.

Stereotype lift is unlikely to be stereotype susceptibility

Stereotype lift appears similar to *stereotype susceptibility* (Shih et al., 1999), the performance boost caused by activation of a positive ingroup stereotype. In contrast to stereotype susceptibility, however, stereotype lift is triggered not by a positive stereotype about one's own group but by a negative stereotype about another group. While stereotype susceptibility focuses on groups targeted by positive stereotypes (e.g., Asians completing a math test), stereotype lift focuses on groups that are *non-stereotyped*. Whites, men, and other majority groups are considered normal and typical in most sectors of society (Miller, Taylor, & Buck, 1991). They are thus less likely to be targets of either negative stereotypes or positive ones (Miller et al., 1991). Indeed, Aronson et al. (1999, p. 40) report that among a large, diverse sample of undergraduates asked to describe stereotypes about various groups, not a single participant mentioned a stereotype about the intellectual abilities of Whites. Instead, they cited stereotypes about Blacks, Asians, and women. Thus, we think that stereotype lift is driven *not* by positive stereotypes about Whites, men, and wealthy people, but by *negative* stereotypes about Blacks, women, and poor people. This conclusion is buttressed by the result (noted above) that lift effects are larger among people high in outgroup prejudice (Schultz et al., 2002; see also Danso & Esses, 2001; Josephs et al., 2001).

Implications of stereotype lift

Stereotype lift has at least four implications. First, it may improve intellectual attainment on evaluative tests that scrupulously avoid reference to race or gender. Second, stereotype lift may improve performance among many social groups rather than only one, insofar as each group has higher status than the stereotyped outgroup. Third, stereotype lift implies ingroup advantage indirectly rather than directly, and it is thus unlikely to cause the ironic, performance-debilitating effects that sometimes accompany the direct invocation of a positive ingroup stereotype (see Cheryan & Bodenhausen, 2000). Fourth, although the effects of stereotype lift may be subtle on any given test, its impact on the achievement of the nonstereotyped may be dramatic when its effects accumulate either within a large group of test-takers or across numerous performance opportunities for a single individual.

Stereotype lift complements stereotype threat by providing further evidence for the social-psychological origins of the Black–White test score gap and other group-based differences in achievement (see Jencks & Phillips, 1998; Steele, 1997). Because negative stereotypes are embedded in the social representation of standardized tests, these tests disadvantage certain groups and advantage others.

Acknowledgments

This research was supported by a research grant from the URO at Stanford University to Gregory Walton. We thank Joseph Brown, Peter Hegarty, Julio Garcia, Felicia Pratto, Claude Steele, Ewart Thomas, and Vanessa Robin Weersing for their valuable assistance and insightful comments. We also thank Keisha Burdick, Victoria Brescoll, Stacy Fambro, Elizabeth Levy Paluck, Aaron Sackett, Eric Uhlmann, Shirley Wang and others in the Department of Psychology at Yale University who provided helpful suggestions.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Anderson (2002). Stereotype threat: The effects of gender identification on standardized test performance. Unpublished doctoral dissertation, James Madison University, Harrisonburg, VA.

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When White men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*, 29–46.

Aspinwall, L. G., & Taylor, S. E. (1993). Effects of social comparison direction, threat, and self-esteem on affect, self-evaluation, and expected success. *Journal of Personality and Social Psychology, 64*, 708–722.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ, USA: Prentice-Hall.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71*, 230–244.

Baumeister, R. F., Twenge, J. M., & Nuss, C. K. (2002). Effects of social exclusion on cognitive processes: Anticipated aloneness reduces intelligent thought. *Journal of Personality and Social Psychology, 83*, 817–827.

Blanton, H., Buunk, B. P., Gibbons, F. X., & Kuyper, H. (1999). When better-than-others compare upward: Choice of comparison and comparative evaluation as independent predictors of academic performance. *Journal of Personality and Social Psychology, 76*, 420–430.

*Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science, 12*, 225–229.

*Broadnax, S., Crocker, J., & Spencer, S. (1997). African-Americans and academic performance: The mediational role of stereotype vulnerability. Unpublished manuscript, State University of New York at Buffalo.

*Brown, J. L., Steele, C. M., & Atkins, D. (2001). Performance expectations are not a necessary mediator of stereotype threat in African American verbal test performance. Unpublished manuscript, University of Washington, Seattle.

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.

Bui, K. V. T., & Pelham, B. W. (2000). Cognitive and affective reactions to social comparison. *Journal of Social Behavior and Personality, 14*, 569–583.

Cheryan, S., & Bodenhausen, G. V. (2000). When positive stereotypes threaten intellectual performance: The psychological hazards of "model minority" status. *Psychological Science, 11*, 399–402.

Cohen, G. L., & Steele, C. M. (2002). A barrier of mistrust: How negative stereotypes affect cross-race mentoring. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 305–331). San Diego: Academic Press.

Cohen, G. L., Steele, C. M., & Ross, L. D. (1999). The mentor's dilemma: Providing critical feedback across the racial divide. *Personality and Social Psychology Bulletin, 25*, 1302–1318.

The College Board (2001). 2001 College-bound seniors: A profile of SAT Program test takers. Retrieved September 5, 2001. Available from: http://www.collegeboard.org/toc/html/tocresearch000.html.

*Croizet, J.-C., & Claire, T. (1998). Extending the concept of stereotype and threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin, 24*, 588–594.

*Danso, H. A., & Esses, V. M. (2001). Black experimenters and the intellectual test performance of White participants: The tables are turned. *Journal of Experimental Social Psychology, 37*, 158–165.

*Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2001). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin, 28*, 1615–1628.

Devine, P. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5–18.

Dijksterhuis, A., & van Knippenberg, A. (1998). The relation between perception and behavior, or How to win a game of Trivial Pursuit. *Journal of Personality and Social Psychology, 74*, 865–877.

*Ewing, V., & Smith, R. (2001). Negative stereotypes about older adults may be beneficial to young adults. Poster session presented at the annual meeting of the American Psychological Society, Toronto, Canada.

Fein, S., & Spencer, S. J. (1997). Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology, 73*, 31–44.

*Foels, R. (2000). Disidentification in the face of stereotype threat. Poster session presented at the annual meeting of the Society for Personality and Social Psychology, Nashville, TN.

*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.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego: Academic Press.

Hogg, M. A. (2000). Social identity and social comparison. In J. Suls, & L. Wheeler (Eds.), *Handbook of social comparison: Theory and research* (pp. 401–422). New York: Kluwer Academic/Plenum Publishers.

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

Jencks, C., & Phillips, M. (Eds.). (1998). *The Black–White test score gap*. Washington, DC: Brookings Institution.

*Josephs, R. A., Newman, M. L., Brown, R. P., & Beer, J. M. (2001). Using the relationship between status and testosterone to explain stereotype-based sex differences in cognitive performance. Manuscript in preparation, University of Texas at Austin.

*Keller, J. (2002). Stereotype threat and women's math performance in an everyday setting: The power of stereotype threat reduction in high school classrooms. Manuscript submitted for publication, University of Mannheim, Germany.

*Keller, J., & Dauenheimer, D. (2002). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. Manuscript submitted for publication, University of Mannheim, Germany.

Kray, L. J., Thompson, L., & Galinsky, A. (2001). Battle of the sexes: Gender stereotype confirmation and reactance in negotiations. *Journal of Personality and Social Psychology, 80*, 942–958.

Levy, B. (1996). Improving memory in old age through implicit self-stereotyping. *Journal of Personality and Social Psychology, 71*, 1092–1107.

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.

*Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2002). Combating stereotype threat: The effect of self-affirmation and emotional expression on women's math performance. Unpublished manuscript, University of Arizona, Tucson.

*Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin, 28*, 1183–1193.

*McKay, P. F., Doverspike, D., Bowen-Hilton, D., & Martin, Q. D. (2002). Stereotype threat effects on the Raven Advanced Progressive Matrices scores of African-Americans. *Journal of Applied Social Psychology, 32*, 767–787.

Miller, D. T., Taylor, B., & Buck, M. L. (1991). Gender gaps: Who needs to be explained. *Journal of Personality and Social Psychology, 61*, 5–12.

Pelham, B. W., & Wachsmuth, J. O. (1995). The waxing and waning of the social self: Assimilation and contrast in social comparison. *Journal of Personality and Social Psychology, 69*, 825–838.

*Quinn, D., & Spencer, S. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues, 57*, 55–71.

Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills: Sage Publications.

*Salinas, M. F. (1998). Stereotype threat: The role of effort withdrawal and apprehension on the underperformance of Mexican Americans. Unpublished doctoral dissertation, The University of Texas at Austin.

Sarason, I. (1991). Anxiety, self-preoccupation and attention. In R. Schwarzer, & R. A. Wicklund (Eds.), *Anxiety and self-focused attention*. New York: Plenum Press.

*Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38*, 194–2001.

*Schultz, P. W., Baker, N., Herrera, E., & Khazian, A. (2002). Stereotype threat among Hispanic-Americans and the moderating role of ethnic identity. Manuscript in preparation, California State University, San Marcos.

*Sekaquaptewa, D., & Thompson, M. (2001). Performance under scrutiny: The additive effects of solo status and stereotype threat on women's public performance. Manuscript submitted for publication, University of Michigan, Ann Arbor.

*Sekaquaptewa, D., & Thompson, M. (2002). The differential effects of solo status on members of high and low status groups. *Personality and Social Psychology Bulletin, 28*, 694–707.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80–83.

Smith, R. H., & Insko, C. A. (1987). Social comparison choice during ability evaluation: The effects of comparison publicity, performance feedback, and self-esteem. *Personality and Social Psychology Bulletin, 13*, 111–122.

*Spencer, S. J. (1993). The effect of stereotype vulnerability on women's math performance. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

*Spencer, S., Iserman, Davies, P., & Quinn, D. (2002). Unpublished raw data, University of Waterloo, Ontario, Canada.

*Spencer, S., Steele, C. M., & Quinn, D. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4–28.

*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 threat 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. *Advances in Experimental Social Psychology, 34*, 379–440.

*Sternberg, R. J., Jarvin, L., Leighton, J., Newman, T., Moon, T., Callahan, C., & Grigorenko, E. L. (2002). Girls can't do math?: The disidentification effect and gifted high school students' math performance. Unpublished manuscript, Yale University, New Haven, CT.

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.

*Stricker, L. J. (1998). *Inquiring about examinee's ethnicity and sex: Effects on AP calculus AB examination performance*. (College Board Rep. 98-1; ETS Research Rep. No. 98-5). New York: College Entrance Examination Board.

*Stricker, L. J., & Ward, W. C. (1998). *Inquiring about examinee's ethnicity and sex: Effects on computerized placement tests performance*. (College Board Rep. 98-2; ETS Research Rep. No. 98-9). New York: College Entrance Examination Board.

Tesser, A. (1988). Toward a self-evaluation maintenance model of social behavior. *Advances in Experimental Social Psychology, 21*, 181–227.

Walsh, M., Hickey, C., & Duffy, J. (1999). Influence of item content and stereotype situation on gender differences in mathematical problem solving. *Sex Roles, 41*, 219–240.

Wheeler, S. C., Jarvis, W. B. G., & Petty, R. E. (2001). Think unto others: The self-destructive impact of negative racial stereotypes. *Journal of Experimental Social Psychology, 37*, 173–180.

Walton, G. M., & Cohen, G. L. (2003). Belongingness and social representations of computer science. Manuscript in preparation, Yale University, New Haven, CT.

Wills, T. A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin, 90*, 245–271.

Wills, T. A. (1991). Similarity and self-esteem in downward comparison. In J. Suls, & T. A. Wills (Eds.), *Social comparison: Contemporary theory and research*. Hillsdale, NJ: Erlbaum.

Further reading

Morrison, T. (1992). *Playing in the dark: Whiteness and the literary imagination*. Cambridge, MA: Harvard University Press.