




The Effects of Exposure to Positive Gender Stereotypes on Women's and Men's Performance in Counter-Stereotypical Tasks and Pursuit of Agentic and Communal Goals

Rotem Kahalon¹ , Nurit Shnabel¹, and Julia C. Becker²

¹The School of Psychological Sciences, Tel-Aviv University, Israel

²Institute of Psychology, Osnabrück University, Germany

Abstract: Two studies examined the effects of exposure to positive gender stereotypes on performance in counter-stereotypical domains and pursuit of agentic and communal goals. Exposure to stereotypes about women's communality (Study 1, $N = 108$) led to impaired math performance among women, regardless of their math identification. Exposure to stereotypes about men's agency (Study 2, $N = 129$) led to impaired performance in a test of socio-emotional ability among men high in domain identification. Moreover, among women with high math identification, exposure to the communality stereotype increased the pursuit of agentic goals. Among men, exposure to the agency stereotype tended to decrease the pursuit of communal goals. These results are consistent with accumulating evidence for the "dark side" of positive stereotypes, yet, for women, they also point to active attempts to counteract them.

Keywords: stereotype threat, women in STEM, gender stereotypes, positive stereotypes, interpersonal goals, gender roles

Gender segregation in professions and higher education, such that women are underrepresented in STEM fields (Science, Technology, Engineering, and Math) whereas men are underrepresented in HEED fields (Healthcare, Early Education, and Domestic Roles) is a major barrier to gender equality (e.g., Diekmann, Steinberg, Brown, Belanger, & Clark, 2017). The social psychological literature has pointed to negative gender stereotypes (e.g., that women are bad at math, or that men have low emotional intelligence) as one of the key mechanisms responsible for this segregation (Cundiff, 2018; Ellemers, 2018). Not only do negative stereotypes reduce women's and men's engagement in and identification with domains in which they are stigmatized (Hall, Schmader, & Croft, 2015; Kalokerinos, Kjelsaas, Bennetts, & von Hippel, 2017), they also impair their actual performance due to the experience of *stereotype threat*, namely concern about confirming negative stereotypes regarding their gender's inferior ability in stigmatized domains, which causes stress that ultimately undermines actual performance (Steele, 1997; for a review, see Spencer,

Logel, & Davies, 2016). Ironically, stereotype threat effects tend to be particularly pronounced among individuals with strong domain identification, who wish to succeed in the field in which their gender is stigmatized (e.g., Keller, 2007).

For women, stereotype threat effects can manifest in various domains (e.g., driving; Yeung & von Hippel, 2008), yet the vast majority of research has focused on math performance. This research has demonstrated that the mere awareness of negative stereotypes about women's math ability can interfere with the performance of female participants in math tests (Spencer, Steele, & Quinn, 1999; see Nguyen & Ryan, 2008 for a meta-analysis). Moreover, stereotype threat was associated with women's reduced interest in pursuing math-oriented occupations later on (Davies, Spencer, Quinn, & Gerhardstein, 2002) and with feelings of exhaustion and psychological burnout when already working in STEM-related fields (Hall et al., 2015).

Stereotype threat effects among men have been relatively understudied. The existing research revealed that exposure to either direct (Leyens, Désert, Croizet, & Darcis, 2000) or

subtle (Koenig & Eagly, 2005) reminders of the stereotype that men have low socio-emotional skills, impaired men's performance in tests of emotional abilities. Moreover, a field study among men working in female-dominated occupations revealed that these men were often experiencing stereotype threat, which in turn led to negative job attitudes and intentions to resign (Kalokerinos et al., 2017).

But what about positive gender stereotypes? As opposed to negative stereotypes, which are delegitimized in today's Western society, positive stereotypes are robustly used and considered legitimate and even complimentary (Kay, Day, Zanna, & Nussbaum, 2013; Mae & Carlston, 2005). Despite their flattery, positive stereotypes might have negative consequences because, like negative stereotypes, they ascribe traits to individuals based on their group affiliation (Czopp, Kay, & Cheryan, 2015). This leads the targets of such stereotypes to perceive positive comments about their group membership as prejudiced (Czopp, 2008). For example, female participants believed that, compared to a man who expressed no stereotypes, a man who expressed positive stereotypes about women (e.g., "women are nurturing") was also likely to endorse negative stereotypes about women (Siy & Cheryan, 2016) – suggesting that positive stereotypes signal to their targets an underlying negativity toward their group. If so, exposure to positive gender stereotypes might also lead to stereotype threat effects.

A recent study by Kahalon, Shnabel, and Becker (2018) confirmed this possibility, demonstrating that: (a) exposure to the stereotype about women's communality (according to which women are empathic, motherly, etc.) impaired the math performance of female participants with strong domain identification (i.e., who find math rewarding and wish to succeed in it; Smith & White, 2001), and (b) exposure to the stereotype about men's agency (according to which men are assertive, have natural leadership qualities, etc.) impaired the performance in a test of emotional skills of male participants with strong identification with the socio-emotional domain. These findings are consistent with the notion that stereotypes in general (Kervyn, Yzerbyt, & Judd, 2010), and gender stereotypes in particular (Eagly & Wood, 1999), are complementary: women are perceived as communal and warm but also as low on agency, whereas men are perceived as agentic but also as low on communality and warmth (Fiske, Cuddy, & Glick, 2007). Moreover, research on the "innuendo effect" (Fiske et al., 2015; Kervyn, Bergsieker, & Fiske, 2012) tells us that when people receive information about social targets that focuses on their positive traits, they infer complementary negative traits; for example, when learning that a given woman is very nice, people infer that she is not highly intelligent.

Although Kahalon et al.'s (2018) findings are consistent with the literature about the complementary nature of stereotype and the innuendo effect, they have been the sole demonstration to date that activation of positive gender stereotypes can produce stereotype threat effects. Thus, a main purpose of the present research was to test the replicability of these findings – which was critical for establishing the validity of the conclusions about the "darker side" of positive stereotypes. Replicating social psychological results is generally necessary to confirm the accuracy of the empirical findings, clarify the conditions under which an effect can be observed, and estimate the true effect size (Open Science Collaboration, 2012). Yet replication efforts are especially important with regard to stereotype threat effects, in light of recent concerns that – despite evidence that stereotype threat effects represent a real phenomenon (Flore & Wicherts 2015; Nguyen & Ryan 2008) – their size has been overestimated in the literature (i.e., stereotype threat effects are smaller than suggested by previous research; Flore & Wicherts, 2015).

Besides replicating previous findings, we tested some new hypotheses. Theoretically, stereotype threat effects are distinct from de-motivation processes; they occur due to the experience of stress and intrusive negative thoughts (Cadinu, Maass, Rosabianca, & Kiesner, 2005) and despite the motivation to do well (Schmader, Johns, & Forbes, 2008). Yet Kahalon et al.'s (2018) research did not tell us whether exposure to the positive aspects of the gender stereotype influenced women's and men's test motivation – which could be an alternative explanation (rather than stereotype threat) to the observed decrease in their performance. Finding that exposure to positive gender stereotypes impaired women's math performance and men's socio-emotional performance *without* changing their test motivation could provide further support for stereotype threat as the reason for performance decrements. Thus, the second aim of the present research was to examine women's and men's motivation to succeed in the test.

Finally, a third goal of the present study was to explore the effects of exposure to the communality and agency stereotypes on women's and men's interpersonal goals; that is, preferences for certain interpersonal outcomes or modes of conduct (Rokeach, 1973). Interpersonal goals can be organized within one conceptual space, defined by two orthogonal axes, in which each point is specified as a weighted mixture of agency and communion (Locke, 2000). High (vs. low) pursuit of agentic goals reflects the motivation to gain competence and power (vs. submissiveness and helplessness); high (vs. low) pursuit of communal goals reflects warmth and empathy toward others (vs. coldness and detachment).¹

¹ Although competence and power, or warmth and morality, reflect somewhat different constructs they can be subsumed under the broad "big two" categories of agency and communion (Abele & Wojciszke, 2013).

The pursuit of agentic and communal goals is influenced by situational cues. To illustrate, after being victimized by others, people pursue more agentic goals (e.g., being tough), whereas after victimizing others, people pursue more communal goals (e.g., being nice) (Aydin, Ullrich, Siem, Locke, & Shnabel, 2018). In the present study, thus, we examined how situational exposure to positive gender stereotypes affects the pursuit of agentic and communal interpersonal goals.

The existing literature offers two different (albeit not contradicting) predictions. First, theorizing on the “sweet persuasion” of positive stereotypes (Barreto & Ellemers, 2005) suggests that members of stigmatized groups often embrace and align their behavior with them. Hence, consistent with the notion that positive stereotypes are prescriptive (whereas negative stereotypes are descriptive; Heilman, 2001), participants should align their goals with the positive stereotypes to which they are exposed: exposure to stereotypes about women’s communality should increase female participants’ pursuit of communal goals, whereas exposure to stereotypes about men’s agency should increase male participants’ pursuit of agentic goals. If so, positive stereotypes may perpetuate traditional gender roles by leading women and men to fulfill their respective roles as “nice” or “assertive” in interpersonal interactions.

Second, the experience of stereotype threat may lead members of stigmatized groups who highly identify with the domain in which their group is stigmatized (e.g., female STEM students performing a math test, Pronin, Steele, & Ross, 2004; female business students performing a negotiation task; Kray, Thompson, & Galinsky, 2001) to distance themselves from the negative stereotype about their group or try to counteract it. If so, exposure to positive gender stereotypes may lead participants who are high in domain identification to endorse goals that would counteract the negative stereotypes about their group (which are activated along with the positive stereotypes). Thus, women who identify with math should pursue more agentic goals in response to the communality stereotype, and men who identify with the socio-emotional domain should pursue more communal goals in response to the agency stereotypes. Such patterns would indicate that the experience of stereotype threat – induced by exposure to positive gender stereotypes (Kahalon et al., 2018) – can lead to reactance responses (Brehm & Brehm, 1981), in an attempt to rebel against and challenge traditional gender roles (Kray et al., 2001).

The Present Research

The present research consisted of two studies, Study 1 conducted among women and Study 2 conducted among men,

which used a two-cell experimental design to test the following hypotheses:

Hypothesis 1 (H1): Exposure to positive stereotypes about one’s gender would lead to performance decrements in a counter-stereotypical task (math in Study 1, emotion recognition in Study 2) among women (Study 1) and men (Study 2) with high domain identification. Confirming this hypothesis would replicate Kahalon et al.’s (2018) previous findings.

Hypothesis 2 (H2): Exposure to positive gender stereotypes would not change women’s (Study 1) and/or men’s (Study 2) test motivation. Confirming this hypothesis would extend Kahalon et al.’s (2018) previous findings by ruling out reduced test motivation as an alternative explanation.

Hypothesis 3 (H3): Exposure to positive gender stereotypes would increase the pursuit of goals that are consistent with these positive stereotypes; namely, communal goals among women (Study 1) and agentic goals among men (Study 2). This exploratory hypothesis tested for changes in goal pursuit that *reinforce* traditional gender roles.

Hypothesis 4 (H4): Exposure to positive gender stereotypes would increase the pursuit of goals that are inconsistent with the negative gender stereotypes that “complement” these positive stereotypes; namely, increased pursuit of agentic goals among women (Study 1) and increased pursuit of communal goals among men (Study 2). This exploratory hypothesis tested for changes in goal pursuit that *challenge* traditional gender roles.

The study conforms to recognized ethical standards, data files can be accessed through the Open Science Framework (OSF) <https://osf.io/hktz3>.

Study 1

In Study 1, female participants were randomly assigned either to an experimental condition in which they were exposed to the positive stereotype about women’s communality, or to a control/no-stereotypes condition. They then completed a math test, a measure of their motivation to succeed in the test, and a circumplex measure of their pursuit of agentic and communal goals in interpersonal interactions. Based on our hypotheses, we expected the experimental condition to have (a) a negative effect on participants’ math performance – especially among those with

high math-identification, and (b) no effect on participants' motivation. We also explored whether the experimental condition would increase the pursuit of (a) communal goals, in alignment with the prescriptive nature of positive stereotypes (i.e., that women should be nice; Heilman, 2001), and/or (b) agentic goals, in an attempt to counteract negative expectations about women's agency (Kray et al., 2001) – especially among participants with high domain-identification (in line with Pronin et al., 2004). The study was pre-registered through the OSF <https://osf.io/pa9j5>.

Method

Participants

A power analysis using G*Power calculator (Faul, Erdfelder, Buchner, & Lang, 2009) revealed that 124 participants were needed to detect the small to medium effect size ($f^2 = .08$) observed by Kahalon et al. (2018), at a significance of 5% and power of 80%. Data collection was stopped after the recruitment of 115 participants, since there were no new sign-ups. Participants were undergraduate female students majoring in diverse disciplines (e.g., psychology, engineering, business). To avoid disproportional influence of extreme observations on our analysis (Osborne & Overbay, 2004), we excluded seven outliers based on our preregistered decision to exclude participants with high Cook's (1977) distance value (i.e., whose distance was greater than $4/n$; Bollen & Jackman, 1990). Thus, the sample included 108 participants.² A sensitivity analysis (Faul et al., 2009) for a sample of 108 participants and power of 80% at $\alpha = .05$, revealed that for a single predictor, in a multiple regression analysis, effects above $R^2 = .069$ will be reliably detected. The effect for the main predictor in Study 1 was $R^2 = .075$. All participants were Israeli Jewish, and their native tongue was Hebrew; 86.3% described themselves as heterosexual, and the rest as either lesbian (6.4%) or bisexual (7.3%); $M_{age} = 22.60$ ($SD = 2.11$).

Procedure and Materials

Participants were invited to a laboratory study on “academic tendencies in different domains.” All the materials were computerized. Participants first completed a 5-item measure of their math identification (adapted from Smith & White, 2001), which captured their interest and success in math (e.g., “It is highly likely that I will work in a math related field”; 1 = *strongly disagree* to 5 = *strongly agree*), $\alpha = .86$. To disguise the study's purpose, participants

completed additional filler questions about their identification with the domain of verbal ability.

Next, ostensibly presented as a verbal ability task, participants read short texts that constituted the experimental manipulation (see Appendix A in the OSF deposit <https://osf.io/hktz3/>). In both conditions, participants were presented with a symbol, followed by a short scientific explanation about the meaning of this symbol and three reading comprehension questions. This allowed us to disguise the real purpose of the manipulation as “a test of verbal abilities,” and activate the relevant stereotype in a subtle manner. The text in the control condition discussed the meaning of the @ sign. For example, it explained that the sign was originally used to represent a unit of weight (Amphora). The text in the communality stereotype condition discussed the meaning of the circle symbol in the Bender-Gestalt Test as representing femininity. For example, it was explained that this symbol represents women's “innate maternal qualities,” such as containment and natural sensitivity.

The texts were conceptually similar to those used by Kahalon et al. (2018), yet with slight modifications intended to improve the original manipulation. Specifically, the original control condition discussed the Yin-Yang symbol, which represents the complementarity of natural elements such as light and darkness; the original experimental condition first mentioned that the Bender-Gestalt test includes both a square and a circle symbols, representing the complementary roles of women and men, and then went on to discuss the circle symbol (explaining that it represents women's “innate maternal qualities,” such as sensitivity – as done in the present study). Thus, in the original study both conditions primed participants with the concept of complementarity, yet the experimental condition additionally reminded them with the positive stereotype about women's communality. A limitation of this approach was that it has remained unclear whether a reminder of the positive communality stereotype would have the same effect on women's math performance even when the concept of complementarity is not explicitly activated. The present research addressed this limitation by using a “cleaner,” unconfounded manipulation to activate the communality stereotype (for the importance of doing so, see Giner-Sorolla, Amodio, & Van Kleef, 2018).

Next, participants completed a math test (developed by Johns, Schmader, & Martens, 2005), which lasted 15 min and included 30 difficult yet solvable questions. Participants earned one point for each correct answer. They then completed the following measures.

² When all 115 participants were included in the analysis, the key main effect of the experimental condition on math performance became marginal ($B = -1.18$, $t = -1.86$, $p = .066$). For motivation and communal goals, all the effects remained nonsignificant, $ps > .114$. For agentic goals, the Condition \times Math Identification interaction remained significant ($B = 0.34$, $t = 2.68$, $p = .009$).

Test Motivation

We used a shortened version of the Student Opinion Scale (SOS; Sundre & Thelk, 2007), which assessed, using six 5-point items (1 = *strongly disagree* to 5 = *strongly agree*), the importance and effort participants placed on the test (e.g., “Doing well on this test was important to me”; “I engaged in good effort throughout this test”), $\alpha = .68$.

Circumplex Scales of Interpersonal Values (CSIV)

We used the shortened 32-items (instead of 64-item) version of the CSIV (Locke, 2000). The CSIV assesses a diversity of interpersonal goals (how people want to act or be perceived when interacting with others), reflecting all possible mixtures of agentic and communal tendencies, by asking respondents to rate the importance of various interpersonal outcomes or modes of conduct (0 = *not important to me* to 4 = *extremely important to me*). It consists of eight 4-item scales, such that each scale reflects a different circumplex octant: Agentic (+A; appearing self-confident), Agentic and Communal (+A+C; expressing oneself openly), Communal (+C; feeling closeness to and developing friendships with others), Submissive and Communal (–A +C; seeking others’ approval by complying with their opinions), Submissive (–A; avoiding making others angry by pleasing them), Submissive and Separate (–A–C; avoiding social embarrassment), Separate (–C; appearing detached, without revealing one’s thoughts and feelings), and Agentic and Separate (+A–C; having no interest in others’ opinions).

Testing the circumplex structure using multidimensional scaling (MDS; Gurtman & Pincus, 2000; detailed analysis is available at <https://osf.io/hktz3>) revealed that the data mapped onto two orthogonal dimensions (stress = .07). This allowed us to calculate Agentic and Communal vectors, which reflect the pursuit of agency (vs. submissiveness) and communion (vs. separateness), respectively (Locke, 2000).

Threat Appraisal

As a manipulation check, adapted from Marx (2012), three 7-point items (1 = *not at all* to 7 = *very much*) evaluated participants’ experience of stereotype threat (e.g., “I worry that my ability to perform well on math tests is affected by my gender”). After the exclusion of one item ($\alpha = .53$ for the 3-item scale),³ $\alpha = .84$. Note that although a manipulation check should ideally be employed immediately after the manipulation, we decided to employ it only after measuring the dependent variables because we were concerned that

the blatant wording of the threat-appraisal items would reveal the real purpose of the study.

Finally, participants completed a short demographic questionnaire that assessed their pre-existing math ability by asking about their Psychometric score (the Israeli equivalent to the SAT). They were then thanked and debriefed.

Results

Descriptive statistics and correlations for all variables are presented in Table 1.

Threat Appraisal

A *t*-test for independent samples revealed that, as intended, women in the experimental condition reported experiencing a higher level of stereotype threat compared to participants in the control condition, $t(98) = 2.10$, $p = .039$.

Test Motivation

We conducted a regression analysis in which the predictors were the experimental condition, domain identification and their interaction. As expected, none of the effects or interactions reached significance, $ps > .187$. In addition, the data were analyzed by estimating a Bayes factor (BF) using Bayesian Information Criteria (BIC; Wagenmakers, 2007). This procedure examines the fit of the data under the null hypothesis, compared to the alternative hypothesis (such that $BF_{01} > 1$ suggest that there is a support for the null hypothesis). The estimation of the Bayes factor for condition and the interaction between condition and math identification, suggested that the data were in favor of the null hypothesis, $BF_{01range} = 2.45\text{--}17.97$ (Jarosz & Wiley, 2014).

Math Performance

To test our main hypothesis, we conducted a regression analysis in which the predictors were the experimental condition, math identification, and their interaction. Consistent with previous research (Kahalon et al., 2018) participants’ preexisting math ability (psychometric score) was used as a covariate, to isolate the unique effect of the experimental manipulation on their math performance. As seen in Table 2, consistent with previous research (e.g., Steinberg, Okun, & Aiken, 2012), preexisting math ability and math identification predicted better performance. Importantly, participants’ math performance in the communality condition was significantly worse than in the control condition. The Condition \times Math Identification interaction was nonsignificant.⁴

³ The test remained significant even when the 3-item scale was used, $t(96) = -2.23$, $p = .028$.

⁴ The main effect of condition remained significant, $B = -1.83$, $t = -2.82$, $p = .006$, such that participants’ math performance in the experimental condition ($M = 7.70$, $SD = 3.14$) was significantly worse than in the control condition ($M = 9.25$, $SD = 3.82$), even without using pre-existing math ability as a covariate.

Table 1. Descriptive statistics and correlations for Study 1 variables

		Communality condition	Control condition						
	Range	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	1	2	3	4	5	6
1. Threat appraisal	1.00–6.50	1.90 (1.42)	1.39 (1.07)	–					
2. Test motivation	2.50–5.00	3.81 (0.66)	3.96 (0.53)	–.06	–				
3. Math performance	1.00–18.00	7.70 (3.14)	9.25 (3.82)	.03	.29**	–			
4. Agentic goals	–1.75–1.00	0.14 (0.67)	0.10 (0.71)	–.05	–.01	.03	–		
5. Communal goals	–1.12–2.98	1.25 (0.72)	1.31 (0.69)	–.02	.71	–.08	.21*	–	
6. Math identification	1.00–5.00	3.90 (0.85)	3.65 (0.90)	–.01	.06	.28**	.10	–.07	–

Notes. *N* = 108 female participants. For threat appraisal: the scale ranged from 1 to 7. For test motivation and math identification: the scales ranged from 1 to 5. For agentic and communal goals: the scales ranged from –4 to 4. For math performance: the scale ranged from 0 to 30. **p* < .05; ***p* < .01.

Table 2. Results of regression analysis on performance in the Math Test.

	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>	95% CI	
						LL	UL
Intercept	–2.727	2.351		–1.160	.249	–7.390	1.937
Preexisting Math Ability	0.018	0.003	.426	5.231	.000	0.011	0.025
Exposure to Communality Stereotype (vs. control)	–1.960	0.581	–.276	–3.374	.001	–3.112	–0.808
Math Identification	1.469	0.435	.381	3.377	.001	0.606	2.332
Exposure to Communality Stereotype × Math Identification	–0.826	0.632	–.147	–1.307	.194	–2.082	0.428

Notes. *N* = 108 female participants. $R^2 = .33$, $F_{\text{change}}(4, 102) = 12.68$, $p < .001$.

Interpersonal Goals

We conducted two regression analyses: one with participants' communal goals as the outcome variable and one with agentic goals as the outcome. The predictors in both models were the experimental condition, math identification, and their interaction. For the communion vector, none of the main effects nor the interaction were significant, $ps > .564$. For the agency vector, the effects of condition and math identification were non-significant, $ps > .306$, whereas their interaction, illustrated in Figure 1, was significant, $\beta = .36$, $t = 2.52$, $p = .013$. The region of significance, calculated using Preacher, Curran, and Bauer's (2006) online calculator, was $Z_{\text{identification}} > 1.07$, indicating that for participants whose standardized level of math identification was higher than 1.07 exposure to the communality stereotype significantly increased the pursuit of agentic goals, and $Z_{\text{identification}} < -1.28$, indicating that, for participants whose level of math identification was lower than –1.28, exposure to the communality stereotype significantly decreased the pursuit of agentic goals.

Discussion

Study 1 revealed that the math performance of female participants who were exposed to the positive stereotype about women's communality was worse than that of participants in the control/no-stereotype condition. This result cannot

be explained by preexisting differences in participants' math ability, which were controlled for. Whereas in previous research of Kahalon et al. (2018) the negative effect of exposure to the communality stereotype emerged only among participants with high domain identification, Study 1 revealed an even more robust effect, which did not depend on participants' level of math identification. This result underscores the potentially adverse consequences for women of exposure to the seemingly positive stereotype about their "innate" warmth and communal nature.

In line with the assumption that the performance decrements were caused by stereotype threat, the manipulation check revealed that participants had a higher threat appraisal in the experimental compared to the control condition. Moreover, exposure to the communality stereotype did not affect participants' motivation to do well in the test. Stereotype threat is known to occur without a decrease in participants' wish and effort to succeed (Jamieson & Harkins, 2007; Keller, 2007), that is, it is not a motivational phenomenon. Therefore, this finding is consistent with stereotype threat as underlying the observed performance decrements – ruling out de-motivation as an alternative explanation.

The pattern of results for the interpersonal goals, as observed in the analysis of regions of significance, suggests that exposure to the communality stereotype led to an increase in pursuit of agentic goals among women with relatively high math identification, and to a decrease among

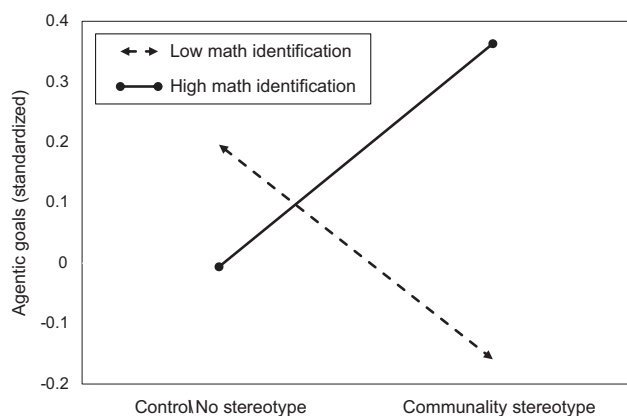


Figure 1. The effect of exposure to the communality stereotype (Study 1, $N = 108$) on the pursuit of agentic goals among women with high math identification (+1 SD above average), $B = 0.37$ ($SE = 0.19$), $t = 1.92$, $p = .057$, versus low math identification (−1 SD below average), $B = -0.36$ ($SE = 0.20$), $t = -1.80$, $p = .076$.

women with relatively low identification. The results for participants high in math identification can be interpreted as stemming from stereotype reactance (Kray et al., 2001). The pattern of results for women with low domain identification was not predicted a priori and therefore should be treated cautiously. If replicated in future research, it would suggest that the communality stereotype reinforces traditional gender roles not only through impairing women's math performance but also by leading some women to behave more submissively.

Study 2

Study 2 examined the effects of exposure to a reminder of the positive stereotype about men's agency on men's performance in a test of socio-emotional abilities; their motivation to do well in this test; and their pursuit of agentic and communal interpersonal goals. We expected that exposure to the positive stereotype about men's agency would (a) impair men's performance in a test of socio-emotional abilities – especially among those high in domain identification, and (b) have no effect on participants' test motivation. We also explored whether exposure to the agency stereotype would increase men's pursuit of agentic interpersonal goals, in alignment with the prescriptive positive stereotype about their gender, and/or increase their pursuit of communal interpersonal goals, in an attempt to counteract the negative stereotype about men's communality.

Before we move on to describe the method and results, for the sake of transparency in reporting, we disclose that we ran Study 2 twice. In the first time, to facilitate data collection, we ran an online study, which was preregistered along with Study 1 (<https://osf.io/pa9j5>). The procedure was identical to that specified below with one exception – participants ($N = 132$) completed the study in their homes rather than in the laboratory. This study failed to replicate Kahalon et al.'s (2018) findings: the effects of exposure to the agency stereotype, domain identification, and their two-way interaction on participants' performance in the emotion recognition test were nonsignificant, $ps > .196$ (detailed description and dataset are available in <https://osf.io/hktz3>). The manipulation check indicated that participants in both the experimental and the control conditions had similar levels of threat appraisal, $Ms = 1.72$ and 1.70 , $t = 0.13$, $p = .896$, suggesting that we failed to induce stereotype threat in the positive stereotype condition. This finding is consistent with previous reports that stereotype threat effects are difficult to induce online (Finnigan & Corker, 2016). Hence, we decided to run Study 2 again, this time in the laboratory. We preregistered this study through the OSF <https://osf.io/uj5av>.

Method

Participants

Based on the effect size observed by Kahalon et al. (2018), a power analysis using G*Power calculator revealed that 137 participants were needed to detect the small to medium effect size ($f^2 = .09$) at a significance of 5% and power of 80%. We managed to recruit 136 male participants through advertisements placed around the campus and in social networks. Students were compensated by either course credit or money (30 NIS). Seven outliers were excluded based on the preregistered decision to exclude observations whose Cook's (1977) distance is greater than $4/n$ (Bollen & Jackman, 1990). The final sample comprised of 129 participants,⁵ $M_{age} = 25.20$ ($SD = 3.53$). A sensitivity analysis (Faul et al., 2009) for a sample of 129 participants and power of 80% at $\alpha = .05$, revealed that for a single predictor, in a multiple regression analysis, effects above $R^2 = .058$ will be reliably detected. The effect for the interaction in Study 2 was $R^2 = .050$. All participants were Israeli men; their native tongue was Hebrew (98.5%) or other (1.5%). Most participants (86%) described themselves as heterosexual, and the rest as homosexual (10%), bisexual (2%), or not wanting to indicate their sexual orientation (2%).

⁵ When all 136 participants were included in the analysis, the key interaction between the condition and domain identification on test performance remained significant ($B = -1.10$, $t = -2.52$, $p = .026$). For motivation and agentic goals, all the effects remained nonsignificant, $ps > .255$. The main effect for communal goals became marginal ($B = -0.20$, $t = -1.67$, $p = .098$).

Table 3. Descriptive statistics and correlations for Study 2 variables

	Range	Agency condition	Control condition	1.	2.	3.	4.	5.	6.
		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)						
1. Threat appraisal	1.00–4.67	2.12 (0.95)	1.66 (0.76)	–					
2. Test motivation	1.80–5.00	4.02 (0.71)	4.08 (0.61)	–.04	–				
3. Emotion recognition test performance	11.00–23.00	17.58 (2.52)	17.82 (2.32)	–.12	.03	–			
4. Agentic goals	–1.25–1.94	0.31 (0.55)	0.32 (0.61)	–.25**	.10	.17*	–		
5. Communal goals	–1.56–2.68	0.93 (0.76)	1.17 (0.69)	–.20*	.07	.19*	.05	–	
6. Domain identification	1.50–5.00	4.09 (0.72)	4.21 (0.57)	–.06	.27**	–.02	.07	.25**	–

Notes. *N* = 129 male participants. For threat appraisal: the scale ranged from 1 to 7. For test motivation and domain identification: the scales ranged from 1 to 5. For agentic and communal goals: the scales ranged from –4 to 4. For the emotion recognition test (DANVA2) performance: the scale ranged from 0 to 24. **p* < .05; ***p* < .01.

Most of them (98%) were students, majoring in different disciplines (e.g., psychology, exact sciences, law).

Procedure and Materials

The procedure generally matched that of Study 1. Participants first completed a demographic questionnaire, which included the 4-item 5-point measure of their domain identification, adapted from Smith and White (2001) to the domain of socio-emotional abilities (e.g., “It is highly likely that my future career will involve understanding the feelings and expressions of other people”; $\alpha = .69$).

Next, ostensibly presented as a verbal ability task, participants read the short texts that constituted the experimental manipulation (see Appendix B in the OSF deposit <https://osf.io/hktz3/>). The control condition was identical to that used in Study 1. The agency-stereotype condition was similar to the communality condition used in Study 1, except that it discussed the square symbol in the Bender-Gestalt test, which represents the qualities of masculinity. For example, it was explained that the angled shape of the square represents men’s typical traits – such as assertiveness, strength, and ambitiousness – which makes them especially suitable for leadership positions.

As the primary outcome variable, participants completed the Adult Facial Expressions subtest from the Diagnostic Analysis of Nonverbal Accuracy 2 (DANVA2; Nowicki & Duke, 1994), a well-established measure of the ability to apprehend others’ emotions, which is a key component of socio-emotional intelligence (Mayer, Salovey, & Caruso, 2004). The DANVA2 examines emotion recognition and consists of 24 facial photographs. For each photograph, participants are asked to choose the correct feeling expressed in it, out of four options. Participants earned one point for each correct answer.

Next, participants completed the measures of test motivation, $\alpha = .76$ (one item was excluded due to low reliability, $\alpha = .67$ for the 6-item scale); the CSIV (the 32-item version), and threat appraisal, $\alpha = .52$. Testing the circumplex structure of the CSIV using multidimensional scaling revealed

that the data mapped into two different dimensions (stress = .036), allowing us to calculate the Agentic and Communal vectors.

Results

Descriptive statistics and correlations for all variables are presented in Table 3.

Threat Appraisal

A *t*-test for independent samples revealed that, as intended, participants reported experiencing a higher level of stereotype threat in the experimental compared to the control condition, $t(127) = 3.04$, $p = .003$.

Test Motivation

A regression analysis was conducted, in which the experimental condition, domain identification and their two-way interaction were the predictors. As expected, none of the effects or interactions reached significance, $ps > .148$. Using Bayesian statistic to compare the fit of the data under the null hypothesis compared to the alternative hypothesis, suggested that the estimation of the Bayes factor for the experimental condition was in favor of the null hypothesis, $BF_{01} = 4.75$, yet the estimation of the Bayes factor for the interaction between condition and domain identification was close to 1, indicating data insensitivity (support for neither hypothesis; Dienes, 2014), $BF_{01} = .73$.

Emotion Recognition Test Performance

We conducted a regression analysis in which condition, domain identification and their two-way interaction were the predictors. As seen in Table 4, domain identification marginally predicted better performance. Consistent with Kahalon et al.’s (2018) findings, the effect of the experimental condition was not significant, while the Condition \times Math Identification interaction reached significance. The region of significance was $Z_{\text{identification}} > .74$, indicating that for participants whose standardized level of domain

Table 4. Results of regression analysis on performance in the Emotion Recognition Test

	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>	95% CI	
						LL	UL
Intercept	17.69	.30		58.38	.000	17.09	18.29
Exposure to agency stereotype (vs. control)	−0.10	.42	−.02	−0.23	.815	−0.94	0.74
Domain identification	0.71	.39	.26	1.83	.069	−0.06	1.48
Exposure to agency stereotype × Domain identification	−1.26	.50	−.35	−2.53	.013	−2.24	−0.28

Notes. *N* = 129 male participants. $R^2 = .05$, $F_{\text{change}}(3, 125) = 2.27$, $p = .084$.

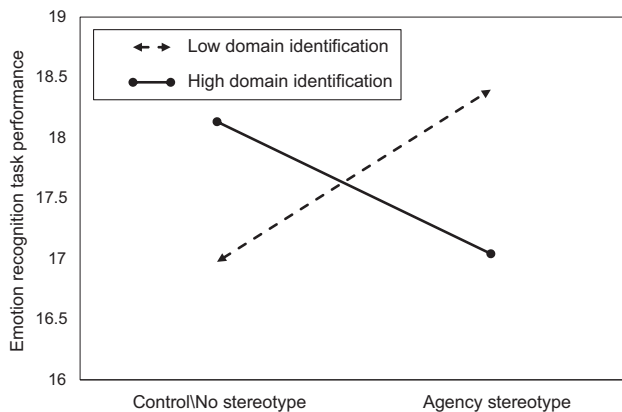


Figure 2. The effect of exposure to the agency stereotype (Study 2, $N = 129$) on performance in an emotion recognition test among men with high domain identification (+1 SD above average), $B = -1.36$ ($SE = 0.61$), $t = -2.24$, $p = .027$, versus low domain identification (−1 SD below average), $B = 1.16$ ($SE = 0.70$), $t = 1.66$, $p = .100$.

identification was higher than .74, exposure to the agency stereotype led to significantly worse test performance, and $Z_{\text{identification}} < -1.52$, indicating that for participants whose level of domain identification was lower than −1.52, exposure to the agency stereotype led to significantly improved test performance. Figure 2 shows the obtained pattern of results.

Interpersonal Goals

We conducted two regression analyses, such that the outcome variable was participants' communal goals in the first analysis, and agentic goals in the second. The predictors in both models were condition, domain identification, and their interaction. For the communion vector, the effect of the experimental condition was marginal, $B = -0.22$, $\beta = -.13$, $t = -1.77$, $p = .079$, such that participants in the agency-stereotype condition reported pursuing marginally less communal goals than in the control condition (see Table 3). The effects of domain identification and the Identification × Condition interaction were not significant, $\beta_s < .15$, $p_s > .269$. For the agency vector, neither the model, $F(3, 125) = 0.23$, $p = .879$, nor any of the effects, $p_s > .505$, reached significance.

Discussion

Replicating previous findings, Study 2 revealed that when male participants were exposed to the positive stereotype about men's agency in the laboratory (but, admittedly, not online), they were induced with stereotype threat, leading to impaired performance in an emotion recognition test among those with high domain identification (i.e., men who care about their socio-emotional abilities). In addition, exposure to the agency stereotype did not result in decreased test motivation, thus supporting stereotype threat, rather than demotivation, as the reason for the observed decrements in the emotion recognition test.

Exposure to the agency stereotype did not influence participants' pursuit of agentic goals, yet there was a marginal decrease in their pursuit of communal goals. If replicates in future studies, this means that exposure to the positive stereotype about men's agency reinforces traditional gender roles not only by undermining (some) men's socio-emotional performance, but also by decreasing men's efforts to be nice and cooperative – traditionally feminine traits – in interpersonal interactions.

General Discussion

Can positive stereotypes have detrimental effects? Yes, they can. Replicating Kahalon et al.'s (2018) findings, women who were exposed to the positive communality stereotype about their gender had worse math performance compared to women in the control/no-stereotype condition, and men who were exposed to the positive agency stereotype about their gender had worse performance in an emotion recognition test compared to men in the control/no-stereotype condition. For men, this effect was evident among participants with high identification with the socio-emotional domain, while for women, this effect was even more robust – occurring for all participants regardless of their level of math identification. In both studies, exposure to these positive stereotypes increased participants' threat appraisal, with no conclusive evidence for a reduction in their motivation to perform well. These findings establish stereotype

threat as the cause underlying the observed performance decrements.

Our results point to the importance of considering how subtle social psychological mechanisms, such as positive gender stereotypes, perpetuate gender inequality. Such subtle mechanisms are often harder to recognize than more overtly hostile mechanisms (Barreto & Ellemers, 2005), yet they have far reaching implications. Identifying the subtle barriers for women in STEM is important for both STEM fields, which are losing out potential talented workers and innovators, as well as for women, as STEM jobs offer higher salaries than non-STEM jobs (Langdon, McKittrick, Beede, Khan, & Doms, 2011). The same is true for men who are more communally oriented. These men can benefit from entering HEED fields, as it can improve their well-being (Le, Impett, Kogan, Webster, & Cheng, 2013; Sheldon & Cooper, 2008), and greater gender diversity in HEED fields could benefit society as a whole (Croft, Schmader, & Block, 2015).

Besides performance in counter-stereotypical domains, we also explored whether exposure to positive gender stereotypes affect women's and men's goals in interpersonal interactions. Perhaps surprisingly, women did not increase their pursuit of communal goals when exposed to the communality stereotype. That is, they did not follow the prescription that they should be "nice" and caring for others. Moreover, among women with relatively high math identification, exposure to the communality stereotype led to an increase in pursuit of agentic goals. Possibly, this finding suggests that women who highly identify with a counter-stereotypical domain try to counteract not only negative (Kray et al., 2001) but also positive gender stereotypes – which subjugate women in a seemingly benevolent way, by putting them on a pedestal (Glick & Fiske, 2001). Notably, although women's pursuit of agentic goals can ultimately challenge traditional gender roles, it comes with a personal cost, as women are often socially penalized for pursuing agency (Phelan, Moss-Racusin, & Rudman, 2008).

Among men, corresponding to the findings among women, we did not find evidence that exposure to the agency stereotype led to alignment with the prescription that they should be strong and assertive. However, compared to men in the control condition, men in the agency stereotype condition reported marginally less communal goals – which are traditionally associated with femininity. Besides examining whether this unexpected result replicates, future research may further explore if exposure to the agency stereotype leads men to detach themselves from other things that are associated with womanhood, such as feminine behaviors (e.g., caring and nurturing of the environment; Brough, Wilkie, Ma, Isaac, & Gal, 2016), feminine men (Glick, Gangl, Gibb, Klumpner, & Weinberg, 2007), and products associated with women (White & Dahl, 2006).

Limitations

The main limitation of the present research is that it did not include a comparison group of male participants in Study 1, and female participants in Study 2. We decided not to include these comparison groups because there is no theoretical ground to assume that exposure to the stereotype about women's communality would affect men's math performance, or that exposure to the stereotype about men's agency would affect women's socio-emotional performance. If anything, such exposure can lead to a stereotype lift (Walton & Cohen, 2003) – a performance boost that occurs due to a downward comparison to a stigmatized out-group. Still, we admit that without the comparison to the non-stigmatized gender in both studies we cannot establish with full confidence that the effects of exposure to the communality and agency stereotypes are indeed unique to women and men (respectively).

Another issue that should be taken in consideration is that we observed stereotype threat effects in the studies conducted in a laboratory setting, but not in the online study. Future research should systematically examine our assumption that stereotype threat effects are more difficult to induce online than in the laboratory, by randomly assigning participants to complete on online versus a laboratory study. Such examination is especially important due to the rapid growth in Internet studies, and the reported consistency of Internet studies with findings obtained using traditional methods (Gosling, Vazire, Srivastava, & John, 2004). It can also point to a potential route for reducing stereotype threat effects; namely, by letting examinees take online tests when possible.

Implication and Future Directions

The present research provides evidence for the idea that not only negative stereotypes, but also positive stereotypes, have the potential to reinforce gender inequality as they might prevent women and men to use their full potential in non-stereotypical domains. Still, our findings also suggest that some women try to counteract these stereotypes, perhaps because they are intuitively aware of their implied negativity (Siy & Cheryan, 2016). Future research could examine the effectiveness of women's strategies for coping with positive stereotypes, and identify means in which the backlash against women's agentic behavior (O'Neill & O'Reilly, 2011) can be minimized.

Our results are also relevant for interventions using group affirmation techniques (emphasizing positive ingroup dimensions), which were found to be helpful in improving stereotyped groups' well-being, motivation and performance (Derks, Van Laar, & Ellemers, 2009; Van Laar, Derks, & Ellemers, 2013). Practitioners who use such

interventions should be careful not to affirm group members' positive identity through the use of positive stereotypes, in order not to arouse the negative outcomes of stereotype threat or undesirable changes in group members' goal pursuit (as we found among men).

Future research is also needed in order to examine the effects of positive stereotype on participants from collectivistic countries. Because collectivistic cultures place a greater focus on relationships, which in turn undermines universal stereotyping principles (Cuddy et al., 2009), it is possible that they will be less affected by positive stereotypes (in compare to participants from individualistic countries).

Finally, as positive stereotypes are relevant to other groups as well, future research could examine the effects of positive stereotypes on racial, ethnic or religious groups. For example, it could test whether exposure to the positive stereotype about Blacks' natural athletic ability (Stone, 2002) impairs Black participants' performance in intelligence tests. Extending our understanding of the effects of positive stereotypes and the way in which people react to them, is critical for eradicating intergroup achievement gaps and restriction to social roles.

References

- Abele, A. E., & Wojciszke, B. (2013). The Big Two in social judgment and behavior. *Social Psychology*, 44, 61–62. <https://doi.org/10.1027/1864-9335/a000137>
- Aydin, A. L., Ullrich, J., Siem, B., Locke, K. D., & Shnabel, N. (2018). The effect of social class on agency and communion: Reconciling identity-based and rank-based perspectives. *Social Psychological and Personality Science*. Advance online publication. <https://doi.org/10.1177/1948550618785162>
- Barreto, M., & Ellemers, N. (2005). The burden of benevolent sexism: How it contributes to the maintenance of gender inequalities. *European Journal of Social Psychology*, 35, 633–642. <https://doi.org/10.1002/ejsp.270>
- Bollen, K. A., & Jackman, R. W. (1990). Regression diagnostics: An expository treatment of outliers and influential cases. In J. Fox & J. S. Long (Eds.), *Modern methods of data analysis* (pp. 257–291). Newbury Park, CA: Sage.
- Brehm, S. S., & Brehm, J. W. (1981). *Psychological Reactance: A theory of freedom and control*. New York, NY: Academic Press.
- Brough, A. R., Wilkie, J. E., Ma, J., Isaac, M. S., & Gal, D. (2016). Is eco-friendly unmanly? The green-feminine stereotype and its effect on sustainable consumption. *Journal of Consumer Research*, 43, 567–582. <https://doi.org/10.1093/jcr/ucw044>
- Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science*, 16, 572–578. <https://doi.org/10.1111/j.0956-7976.2005.01577.x>
- Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19, 15–18. <https://doi.org/10.1080/00401706.1977.10489493>
- Croft, A., Schmader, T., & Block, K. (2015). An underexamined inequality: Cultural and psychological barriers to men's engagement with communal roles. *Personality and Social Psychology Review*, 19, 343–370. <https://doi.org/10.1177/1088868314564789>
- Cuddy, A. J., Fiske, S. T., Kwan, V. S., Glick, P., Demoulin, S., Leyens, J. P., ... Htun, T. T. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology*, 48, 1–33. <https://doi.org/10.1348/014466608X314935>
- Cundiff, J. L. (2018). Subtle barriers and bias in STEM: How stereotypes constrain women's STEM participation and career progress. In M. R. Lowery & J. T. Nadler (Eds.), *The war on women in the United States: Beliefs, tactics, and the best defenses* (pp. 116–156). Santa Barbara, CA: ABC-CLIO.
- Czopp, A. M. (2008). When is a compliment not a compliment? Evaluating expressions of positive stereotypes. *Journal of Experimental Social Psychology*, 44, 413–420. <https://doi.org/10.1016/j.jesp.2006.12.007>
- Czopp, A. M., Kay, A. C., & Cheryan, S. (2015). Positive stereotypes are pervasive and powerful. *Perspectives on Psychological Science*, 10, 451–463. <https://doi.org/10.1177/1745691615588091>
- Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin*, 28, 1615–1628. <https://doi.org/10.1177/014616702237644>
- Derks, B., Van Laar, C., & Ellemers, N. (2009). Working for the self or working for the group: How self-versus group affirmation affects collective behavior in low-status groups. *Journal of Personality and Social Psychology*, 96, 183–202. <https://doi.org/10.1037/a0013068>
- Diekmann, A. B., Steinberg, M., Brown, E. R., Belanger, A. L., & Clark, E. K. (2017). A goal congruity model of role entry, engagement, and exit: Understanding communal goal processes in STEM gender gaps. *Personality and Social Psychology Review*, 21, 142–175. <https://doi.org/10.1177/1088868316642141>
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5, 781. <https://doi.org/10.3389/fpsyg.2014.00781>
- Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist*, 54, 408–423. <https://doi.org/10.1037/0003-066X.54.6.408>
- Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology*, 69, 275–298. <https://doi.org/10.1146/annurev-psych-122216-011719>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Finnigan, K. M., & Corker, K. S. (2016). Do performance avoidance goals moderate the effect of different types of stereotype threat on women's math performance? *Journal of Research in Personality*, 63, 36–43. <https://doi.org/10.1016/j.jrp.2016.05.009>
- Fiske, S. T., Bergsieker, H., Constantine, V., Dupree, C., Holoiu, D. S., Kervyn, N., ... Swencionis, J. (2015). Talking up and talking down: The power of positive speaking. *Journal of Social Issues*, 71, 834–846. <https://doi.org/10.1111/josi.12152>
- Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11, 77–83. <https://doi.org/10.1016/j.tics.2006.11.005>
- Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A

- meta-analysis. *Journal of School Psychology*, 53, 25–44. <https://doi.org/10.1016/j.jsp.2014.10.002>
- Glick, P., Gangl, C., Gibb, S., Klumpner, S., & Weinberg, E. (2007). Defensive reactions to masculinity threat: More negative affect toward effeminate (but not masculine) gay men. *Sex Roles*, 57, 55–59. <https://doi.org/10.1007/s11199-007-9195-3>
- Hall, W. M., Schmader, T., & Croft, E. (2015). Engineering exchanges: Daily social identity threat predicts burnout among female engineers. *Social Psychological and Personality Science*, 6, 528–534. <https://doi.org/10.1177/1948550615572637>
- Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, 57, 657–674. <https://doi.org/10.1111/0022-4537.00234>
- Giner-Sorolla, R., Amodio, D., & Van Kleef, G. (2018). Three strong moves to improve research and replications alike. *Behavioral and Brain Sciences*, 41, E130. <https://doi.org/10.1017/S0140525X18000651>
- Glick, P., & Fiske, S. T. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist*, 56, 109–118. <https://doi.org/10.1037/0003-066X.56.2.109>
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, 59, 93–104. <https://doi.org/10.1037/0003-066X.59.2.93>
- Gurtman, M. B., & Pincus, A. L. (2000). Interpersonal adjective scales: Confirmation of circumplex structure from multiple perspectives. *Personality and Social Psychology Bulletin*, 26, 374–384. <https://doi.org/10.1177/0146167200265009>
- Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology*, 93, 544–564. <https://doi.org/10.1037/0022-3514.93.4.544>
- Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, 7, 2–9. <https://doi.org/10.7771/1932-6246.1167>
- 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. <https://doi.org/10.1111/j.0956-7976.2005.00799.x>
- Kahalon, R., Shnabel, N., & Becker, J. C. (2018). Positive stereotypes, negative outcomes: Reminders of the positive components of complementary gender stereotypes impair performance in counter-stereotypical tasks. *British Journal of Social Psychology*, 57, 482–502. <https://doi.org/10.1111/bjso.12240>
- Kalokerinos, E. K., Kjelsaas, K., Bennetts, S., & von Hippel, C. (2017). Men in pink collars: Stereotype threat and disengagement among male teachers and child protection workers. *European Journal of Social Psychology*, 47, 553–565. <https://doi.org/10.1002/ejsp.2246>
- Kay, A. C., Day, M. V., Zanna, M. P., & Nussbaum, A. D. (2013). The insidious (and ironic) effects of positive stereotypes. *Journal of Experimental Social Psychology*, 49, 287–291. <https://doi.org/10.1016/j.jesp.2012.11.003>
- Keller, J. (2007). Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' maths performance. *British Journal of Educational Psychology*, 77, 323–338. <https://doi.org/10.1348/000709906X113662>
- Kervyn, N., Bergsieker, H. B., & Fiske, S. T. (2012). The innuendo effect: Hearing the positive but inferring the negative. *Journal of Experimental Social Psychology*, 48, 77–85. <https://doi.org/10.1016/j.jesp.2011.08.001>
- Kervyn, N., Yzerbyt, V., & Judd, C. M. (2010). Compensation between warmth and competence: Antecedents and consequences of a negative relation between the two fundamental dimensions of social perception. *European Review of Social Psychology*, 21, 155–187. <https://doi.org/10.1080/13546805.2010.517997>
- Koenig, A. M., & Eagly, A. H. (2005). Stereotype threat in men on a test of social sensitivity. *Sex Roles*, 52, 489–496. <https://doi.org/10.1007/s11199-005-3714-x>
- 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. <https://doi.org/10.1037/0022-3514.80.6.942>
- Langdon, D., McKittrick, G., Beede, D., Khan, B., & Doms, M. (2011). *STEM: Good jobs now and for the future*. Washington, DC: US Department of Commerce. Economics and Statistics Administration. Retrieved from <https://www.commerce.gov/doc/office-under-secretary-economic-affairs/1139/-78>
- Le, B. M., Impett, E. A., Kogan, A., Webster, G. D., & Cheng, C. (2013). The personal and interpersonal rewards of communal orientation. *Journal of Social and Personal Relationships*, 30, 694–710. <https://doi.org/10.1177/0265407512466227>
- Leyens, J. P., Désert, 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. <https://doi.org/10.1177/0146167200262002>
- Locke, K. D. (2000). Circumplex scales of interpersonal values: Reliability, validity, and applicability to interpersonal problems and personality disorders. *Journal of Personality Assessment*, 75, 249–267. https://doi.org/10.1207/S15327752JPA7502_6
- Mae, L., & Carlston, D. E. (2005). Hoist on your own petard: When prejudiced remarks are recognized and backfire on speakers. *Journal of Experimental Social Psychology*, 41, 240–255. <https://doi.org/10.1016/j.jesp.2004.06.011>
- Marx, D. M. (2012). Differentiating theories: A comparison of stereotype threat and stereotype priming effects. In M. Inzlicht & T. Schmader (Eds.), *Stereotype threat: Theory, process, and application* (pp. 124–140). New York, NY: Oxford University Press.
- Mayer, J. D., Salovey, P., & Caruso, D. R. (2004). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry*, 15, 197–215. https://doi.org/10.1207/s15327965pli1503_02
- Nguyen, H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, 93, 1314–1334. <https://doi.org/10.1037/a0012702>
- Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior*, 18, 9–35. <https://doi.org/10.1007/BF02169077>
- O'Neill, O. A., & O'Reilly, C. A. III (2011). Reducing the backlash effect: Self-monitoring and women's promotions. *Journal of Occupational and Organizational Psychology*, 84, 825–832. <https://doi.org/10.1111/j.2044-8325.2010.02008.x>
- Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, 7, 657–660. <https://doi.org/10.1177/1745691612462588>
- Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research and Evaluation*, 9, 1–12.
- Phelan, J. E., Moss-Racusin, C. A., & Rudman, L. A. (2008). Competent yet out in the cold: Shifting criteria for hiring reflect backlash toward agentic women. *Psychology of Women Quarterly*, 32, 406–413. <https://doi.org/10.1111/j.1471-6402.2008.00454.x>

- Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, 31, 437–448. <https://doi.org/10.3102/10769986031004437>
- Pronin, E., Steele, C. M., & Ross, L. (2004). Identity bifurcation in response to stereotype threat: Women and mathematics. *Journal of Experimental Social Psychology*, 40, 152–168. [https://doi.org/10.1016/S0022-1031\(03\)00088-X](https://doi.org/10.1016/S0022-1031(03)00088-X)
- Rokeach, M. (1973). *The nature of human values*. New York, NY: Free Press.
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115, 336–356. <https://doi.org/10.1037/0033-295X.115.2.336>
- Sheldon, K. M., & Cooper, M. (2008). Goal striving within agentic and communal roles: Separate but functionally similar pathways to enhanced well-being. *Journal of Personality*, 76, 415–447. <https://doi.org/10.1111/j.1467-6494.2008.00491.x>
- Siy, J. O., & Cheryan, S. (2016). Prejudice masquerading as praise: The negative echo of positive stereotypes. *Personality and Social Psychology Bulletin*, 42, 941–954. <https://doi.org/10.1177/0146167216649605>
- Smith, J. L., & White, P. H. (2001). Development of the domain identification measure: A tool for investigating stereotype threat effects. *Educational and Psychological Measurement*, 61, 1040–1057. <https://doi.org/10.1177/00131640121971635>
- Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology*, 67, 415–437. <https://doi.org/10.1146/annurev-psych-073115-103235>
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35, 4–28. <https://doi.org/10.1006/jesp.1998.1373>
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613–629. <https://doi.org/10.1037/0003-066X.52.6.613>
- Steinberg, J. R., Okun, M. A., & Aiken, L. S. (2012). Calculus GPA and math identification as moderators of stereotype threat in highly persistent women. *Basic and Applied Social Psychology*, 34, 534–543. <https://doi.org/10.1080/01973533.2012.727319>
- Stone, J. (2002). Battling doubt by avoiding practice: The effects of stereotype threat on self-handicapping in white athletes. *Personality and Social Psychology Bulletin*, 28, 1667–1678. <https://doi.org/10.1177/014616702237648>
- Sundre, D. L., & Thelk, A. D. (2007). *The Student Opinion Scale (SOS): A measure of examinee motivation.. Test manual*. Harrisonburg, VI: Center for Assessment and Research Studies, James Madison University.
- Van Laar, C., Derks, B., & Ellemers, N. (2013). Motivation for education and work in young Muslim women: The importance of value for ingroup domains. *Basic and Applied Social Psychology*, 35, 64–74. <https://doi.org/10.1080/01973533.2012.746609>
- Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, 14, 779–804. <https://doi.org/10.3758/BF03194105>
- Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, 39, 456–467. [https://doi.org/10.1016/S0022-1031\(03\)00019-2](https://doi.org/10.1016/S0022-1031(03)00019-2)
- White, K., & Dahl, D. W. (2006). To be or not be? The influence of dissociative reference groups on consumer preferences. *Journal of Consumer Psychology*, 16, 404–414. https://doi.org/10.1207/s15327663jcp1604_11
- Yeung, N. C. J., & von Hippel, C. (2008). Stereotype threat increases the likelihood that female drivers in a simulator run over jaywalkers. *Accident Analysis & Prevention*, 40, 667–674. <https://doi.org/10.1016/j.aap.2007.09.003>

History

Received March 4, 2019

Revision received May 28, 2019

Accepted May 28, 2019

Published online August 20, 2019

Acknowledgments

We hereby declare that we have no potential conflicts of interest with respect to the authorship or the publication of this article.

Authorship

All authors were involved in all parts of the research.

Open Data

We preregistered these studies through the OSF <https://osf.io/uj5av>, <https://osf.io/pa9j5>, and <https://osf.io/hktz3>.

ORCID

Rotem Kahalon

 <https://orcid.org/0000-0002-4718-6119>

Rotem Kahalon

The School of Psychological Sciences

Tel-Aviv University

Ramat Aviv

Tel-Aviv 69978

Israel

rotemkah@mail.tau.ac.il