Women's Representation in Science Predicts National Gender–Science Stereotypes: Evidence From 66 Nations

Article in Japanese Journal of Educational Psychology · October 2014
DOI: 10.1037/edu0000005

CITATIONS
11
READS
415

3 authors, including:

David Miller
Northwestern University
9 PUBLICATIONS 132 CITATIONS

Marcia C Linn
University of California, Berkeley
295 PUBLICATIONS 11,366 CITATIONS

All in-text references underlined in blue are linked to publications on ResearchGate, letting you access and read them immediately.

Available from: David Miller
Retrieved on: 07 September 2016
Women’s Representation in Science Predicts National Gender-Science Stereotypes: Evidence From 66 Nations

David I. Miller and Alice H. Eagly
Northwestern University

Marcia C. Linn
University of California, Berkeley

In the past 40 years, the proportion of women in science courses and careers has dramatically increased in some nations but not in others. Our research investigated how national differences in women’s science participation related to gender-science stereotypes that associate science with men more than women. Data from ~350,000 participants in 66 nations indicated that higher female enrollment in tertiary science education (community college or above) related to weaker explicit and implicit national gender-science stereotypes. Higher female employment in the researcher workforce related to weaker explicit, but not implicit, gender-science stereotypes. These relationships remained after controlling for many theoretically relevant covariates. Even nations with high overall gender equity (e.g., the Netherlands) had strong gender-science stereotypes if men dominated science fields specifically. In addition, the relationship between women’s educational enrollment in science and implicit gender-science stereotypes was stronger for college-educated participants than participants without college education. Implications for instructional practices and educational policies are discussed.

Keywords: diversity, gender, science education, science workforce, stereotypes

Supplemental materials: http://dx.doi.org/10.1037/edu0000005.supp

Pervasive stereotypes associating science with men emerge early in development (Chambers, 1983; Steffens, Jelenec, & Noack, 2010) and exist across cultures (Nosek et al., 2009). Over 40 years ago, Chambers asked nearly 5,000 American and Canadian children to draw a picture of a scientist, and only 28 children (0.6%) depicted a woman scientist. Although most children still associate science with men, these associations may have weakened over time at least in the United States (Fralick, Kearn, Thompson, & Lyons, 2009; Milford & Tippett, 2013). For example, in one recent study (Farland-Smith, 2009), 35% of American children depicted a woman scientist. These changes in stereotypes mirror women’s increasing participation in science in the United States.

Women’s representation in science courses and careers has dramatically increased in some nations but not in others. Our research investigated how national differences in women’s science participation related to gender-science stereotypes that associate science with men more than women. Data from ~350,000 participants in 66 nations indicated that higher female enrollment in tertiary science education (community college or above) related to weaker explicit and implicit national gender-science stereotypes. Higher female employment in the researcher workforce related to weaker explicit, but not implicit, gender-science stereotypes. These relationships remained after controlling for many theoretically relevant covariates. Even nations with high overall gender equity (e.g., the Netherlands) had strong gender-science stereotypes if men dominated science fields specifically. In addition, the relationship between women’s educational enrollment in science and implicit gender-science stereotypes was stronger for college-educated participants than participants without college education. Implications for instructional practices and educational policies are discussed.

Keywords: diversity, gender, science education, science workforce, stereotypes

Supplemental materials: http://dx.doi.org/10.1037/edu0000005.supp

Pervasive stereotypes associating science with men emerge early in development (Chambers, 1983; Steffens, Jelenec, & Noack, 2010) and exist across cultures (Nosek et al., 2009). Over 40 years ago, Chambers asked nearly 5,000 American and Canadian children to draw a picture of a scientist, and only 28 children (0.6%) depicted a woman scientist. Although most children still associate science with men, these associations may have weakened over time at least in the United States (Fralick, Kearn, Thompson, & Lyons, 2009; Milford & Tippett, 2013). For example, in one recent study (Farland-Smith, 2009), 35% of American children depicted a woman scientist. These changes in stereotypes mirror women’s increasing participation in science in the United States.
Rudman, Buettner, & McLean, 2013). Hence, stereotypes are formed and changed, in part, by repeatedly observing members of different social groups in role-linked activities. This theoretical framework can also help to explain why stereotypes about other social groups vary across nations. For instance, consistent with social role theory, stereotypes about older adults’ incompetence were weaker in nations where more older adults participated in paid and volunteer work; this cross-national relationship remained even after controlling for national differences in older adults’ cognitive abilities (Bowen & Skirbekk, 2013).

Multiple observations of counterstereotypic women across diverse contexts, such as directly in science courses and indirectly in televisions shows, are critical to changing stereotypes (Eagly & Wood, 2012; Koenig & Eagly, in press; Wood & Eagly, 2012). People need multiple, mutually reinforcing examples to see counterstereotypic individuals as evidence of trends. Otherwise, sparse counterstereotypic examples can be dismissed as atypical through a process called subtyping (Bigler & Liben, 2006; Richards & Hewstone, 2001). For instance, individual women scientists could be perceived as having followed unusual paths to science and exerted exceptional effort to succeed (Smith, Lewis, Hawthorne, & Hodges, 2013). These stereotyping processes may explain why experimental studies have revealed that exposure to successful women engineers and mathematicians have not consistently weakened gender-STEM stereotypes (Ramsey, Betz, & Sekaquaptewa, 2013; Steinke et al., 2007; Stout, Dasgupta, Hunsinger, & Macmanus, 2011; Young et al., 2013).

For instance, in Stout et al.’s Study 3, intended STEM majors (n = 100, 47% women) took a 3-month calculus course from a professor and teaching assistant who were either both male or both female. Although taking the calculus course from female instructors increased female students’ implicit identification with mathematics, the gender of the course instructors had no observable effect on gender-math stereotypes. Such short-term interventions may be insufficient to override pervasive, everyday experiences linking math-intensive science fields with men. For instance, male students outnumbered female students by three to one in the calculus course taken by Stout et al.’s participants. In such contexts, sparse examples of female math professors may have been subtyped and seen as atypical. Moreover, taking a STEM course from a female rather than male professor can even strengthen gender-science stereotypes if students do not view the professor as similar to themselves (Young et al., 2013).

Even students in female-dominated science majors could still strongly associate science with men. For instance, although women currently earn 60% of biology bachelor’s degrees in the United States (National Science Board, 2014), biology majors would likely encounter other stereotype-consistent evidence. This evidence could include the preponderance of men among biology faculty (Ceci, Williams, & Barnett, 2009) or students in required STEM courses in other fields such as physics (Barone, 2011). Moreover, students could form separate stereotypes about biologists while maintaining their belief that science is generally associated with men (Richards & Hewstone, 2001). Such conflicting experiences suggest that gender-science stereotypes would likely vary in nuanced ways across students’ field of study. For instance, one large correlational study (n = 100,000) revealed that, compared with physical science majors, biological science majors reported weaker explicit gender-science stereotypes but still implicitly associated science with men to the same extent (Smyth & Nosek, 2013). Furthermore, pervasive cultural images associating science with men fuel stereotyping processes for students in all academic disciplines. Archetypes of White male scientists are present in diverse cultural artifacts such as television shows (Long et al., 2010), movies (Flicker, 2003), national news reports (Chimba & Kitzinger, 2010; Shachar, 2000), science textbooks (Bazler & Simonis, 1991; Brotman & Moore, 2008), and even advertisements in the journal Science (Barbercheck, 2001). Such shared cultural experiences likely disseminate and reinforce stereotypes about gender in general (Furnham & Paltzer, 2010; Kimball, 1986) and women in science specifically (Steinke, 2013).

Comparing gender-science stereotypes across nations could help reveal the impact of such varied cultural experiences. In one such effort, Nosek et al. (2009) found that nations with stronger implicit gender-science stereotypes also had larger national gender differences favoring boys in science and mathematics achievement. The authors suggested that this result reflected a bidirectional relationship in which stereotypes influence achievement and achievement influences stereotypes. We built on this prior research by investigating how women’s participation in science relates to cross-national differences in gender-science stereotypes. Our focus on participation in science extends Nosek et al.’s study because women’s participation in science does not necessarily reflect gender differences in science achievement (Riegle-Crumb, King, Grodsky, & Muller, 2012). When more women enter science, people can observe counterstereotypic women across diverse contexts such as in science classes and news articles, especially if these changes occur across multiple science fields. These diverse observations can then influence stereotypes, as predicted by social role theory (Eagly & Wood, 2012; Wood & Eagly, 2012). To test these predictions, our study analyzed two aspects of women’s participation in science: percentage of women among (a) all science majors (community college or above) and (b) employed researchers. Many participants in our mostly college-educated sample likely had direct repeated exposure to women and men enrolled as science majors; direct exposure to employed researchers was perhaps more limited.

We investigated how women’s participation in science related to both implicit and explicit measures of gender-science stereotypes. Consistent with contemporary theorizing about dual processes in social cognition (Sherman, Gawronksi, & Trope, 2014), the implicit measure assessed aspects of stereotyping that are generally more automatic and less conscious, whereas the explicit measure assessed those aspects that emerge as conscious knowledge that is willingly reported (Nosek, Hawkins, & Frazier, 2011). Empirical findings have generally supported the interpretation that these measures assess related, but distinct, constructs. For instance, explicit and implicit attitude measures often significantly, but weakly, correlate with each other (Greenwald, Pachman, Uhlmann, & Banaji, 2009; Nosek et al., 2007). Moreover, both measures often add incremental validity when predicting behavioral outcomes such as discrimination (Greenwald et al., 2009; but see Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013).

Gawronski and Bodenhausen’s (2006, 2011) associative-propositional model provides a theoretical account for why explicit and implicit measures should often differ. According to this model, implicit measures reflect the activations of associations in memory, whereas explicit measures reflect the outcomes of propositional processes. For instance, a person could automatically asso-
ciate Black people with negative attributes such as violent crime but reject the proposition that “I dislike Black people.” That person would therefore show negative bias toward Black people on an implicit attitude measure, but not on an explicit measure. Also consistent with this theoretical model, different types of counterstereotypic exposure may be necessary to change implicit versus explicit stereotypes. Specifically, repeated counterstereotypic exposure would be critical to changing implicit stereotypes, which reflect associations learned from repeated pairings of stimuli representing two concepts (e.g., science and male). In contrast, brief exposure to propositional information (e.g., statistics about women’s representation in science) could change explicit stereotypes. For instance, a person could learn that women earn half of the U.S.’s chemistry bachelor’s degrees (National Science Board, 2014) and readily incorporate that information into explicit responses (e.g., answering a questionnaire item asking how much that person associates chemistry with men or women).

To explore these ideas, we analyzed four relationships between gender-science stereotypes and women’s participation in science by crossing two types of women’s participation (in educational enrollment and in the workforce) with two types of gender-science stereotypes (explicit and implicit). Our critical hypothesis was that a higher participation of women in science would relate to weaker national-level gender-science stereotypes, consistent with social role theory. The associative-propositional model would additionally predict that, compared with explicit stereotypes, implicit stereotypes should relate more strongly to repeated counterstereotypic exposure. As a proxy for this repeated exposure to women in STEM fields, we used participants’ level of education (e.g., college-educated vs. some or no college). In nations with a high percentage of women among science majors, college-educated individuals would have frequently encountered examples of female science majors during college.

Method

Sample

The 66 nations included in our focal analyses (see Figure 1) represented ~350,000 participants who self-selected into our sample by completing stereotype measures on a widely distributed website called Project Implicit (see Nosek et al., 2009). These nations met the requirements of (a) a minimum sample size of \( n > 50 \) and (b) populations of more than 5% Internet users during the time of stereotype data collection (years 2000–2008). The Results section explains the rationale for these selection criteria and reports results across alternate criteria. In an average national sample, 50% of participants had a college degree or higher, and 79% had some college or higher. Therefore, most participants likely had direct, repeated exposure to the representation of women among college science majors. Also, in an average national sample, 60% of participants were women, and the average age was 27 years (\( SD = 11 \) years within nations).

Measures

Explicit gender-science stereotypes. For the explicit stereotype measure, participants rated “how much you associate science with males or females” on a 5-point or 7-point scale\(^1\) ranging from strongly male to strongly female. This same question was repeated replacing “science” with “liberal arts” to serve as a comparison measure of stereotypes in an alternate academic domain. These questions were worded to correspond to the implicit measure (see below) and definition of gender-science stereotypes (i.e., associations connecting science with men more than women). These questions therefore did not ask about gender stereotypes regarding science-related abilities and interests (e.g., “Do you think males or females are more interested in science?”); such wording would have addressed gender stereotypes about science-related attributes rather than participants’ more general associations between science and gender.

Single-item measures such as our study’s explicit measure sometimes have lower reliabilities than multiple-item measures and therefore can underestimate relationships. Hence, to the extent that our explicit measure was unreliable, it would have provided conservative tests of hypotheses regarding explicit stereotypes. However, compared with multiple-item measures, single-item measures often have equal reliability and validity for assessing psychosocial constructs such as attitudes (Bergkvist & Rossiter, 2007; Fishbein & Ajzen, 1974), job satisfaction (Wanous, Reichers, & Hudy, 1997), and math anxiety (Núñez-Peña, Guilerà, & Suárez-Pellicioni, in press).

Implicit gender-science stereotypes. For the implicit measure, participants completed a gender-science Implicit Association Test (IAT; for an overview of the IAT methodology, see Greenwald et al., 2009). As described by Nosek et al. (2009), this computerized task recorded how quickly participants associated science with males. Participants categorized words representing the categories of male (boy, father, grandpa, husband, male, man, son, uncle), female (aunt, daughter, female, girl, grandma, mother, wife, woman), science (astronomy, biology, chemistry, engineering, geology, math, physics, math), and liberal arts (arts, English, history, humanities, literature, music, philosophy). These 30 words were presented one at a time, and participants categorized them by pressing one of two keyboard keys; one response key was on the left side of the keyboard and the other was on the right. The response keys were paired stereotypically for some trials (e.g., participant presses the \( e \) key for male and science words, and \( i \) key for female and liberal arts words) and counterstereotypically for other trials (e.g., participant presses the \( e \) key for female and science words). Participants responded faster when the keys were paired stereotypically than counterstereotypically by an average of ~100–150 milliseconds (Nosek, Banaji, & Greenwald, 2002). This response time difference was interpreted as evidence of implicit gender-science stereotypes.

Participants were given unlimited time to make a response for each word, but were instructed to go as fast as possible. The precision of these reaction times was limited by the clock rates of

\(^1\) The response categories for the 5-point scale (strongly male, somewhat male, neither male nor female, somewhat female, strongly female) and the 7-point scale (strongly male, moderately male, somewhat male, neither male nor female, somewhat female, moderately male, strongly female) were similar. These response categories were converted to a numeric scale by assigning neither male nor female to a value of 0 and assuming equal numeric spacing between the ordinal response categories. Male responses were given positive scores, and female responses were given negative scores. We standardized the variances of 5-point and 7-point scales to both be 1 before using the scales to compute national averages.
facilitate comparison across the two stereotype measures, each (e.g., an explicit response of "neither male nor female"). To
tions, and scores of 0 indicated neutral gender–science associations
positive scores indicated male–science associa-
times to compute an
was divided by each individual's standard deviation of reaction
between stereotype-consistent and stereotype-inconsistent blocks
trials in any one of the practice blocks). These data quality stan-
test trials, 25% of trials in any one of the critical blocks, 35% of
errors (i.e., made errors on more than 30% of trials across all the
minimize the impact of careless responding, participants' IAT
responses in that response block plus a 600-ms penalty. To help
process the IAT data. Individual trial response times faster than
400 ms or slower than 10,000 ms were removed. Response times
for trials with errors (i.e., participant presses the wrong response
key for the presented word) was replaced with the mean of correct
responses in that response block plus a 600-ms penalty. To help
minimize the impact of careless responding, participants' IAT
scores were disqualified if participants consistently made many
errors (i.e., made errors on more than 30% of trials across all the
critical blocks, 40% of trials in any one of the critical blocks, 40%
of trials across all the practice blocks, and/or 50% of trials in any
one of the practice blocks) or consistently responded too quickly
(i.e., responded faster than 300 ms on more than 10% of the total
test trials, 25% of trials in any one of the critical blocks, 35% of
trials in any one of the practice blocks). These data quality stan-
ards disqualified 9% of IAT scores. The reaction time difference
between stereotype-consistent and stereotype-inconsistent blocks
was divided by each individual’s standard deviation of reaction
times to compute an IAT D score (Greenwald et al., 2003).

Scoring of stereotype measures. For both explicit and im-
PLICIT stereotype measures, positive scores indicated male–science
associations, negative scores indicated female–science associations,
and scores of 0 indicated neutral gender–science associations
(e.g., an explicit response of "neither male nor female"). To
facilitate comparison across the two stereotype measures, each
measure’s raw scores were standardized by dividing by the stan-
standard deviation of all individual scores across the globe. These
standardized scores are identical to z-scores if z-scores were com-
puted without first subtracting the population mean. Hence, for
both stereotype measures, a standardized score of 0.5 represented
a response that differed 0.5 standard deviations in the male direc-
tion from neutral gender–science associations, with standard de-
viation representing variability across individuals. This approach
has the advantage that the magnitude of stereotypes can be inter-
preted in Cohen's d effect size units (for an example meta-analytic
application, see Koenig, Eagly, Mitchell, & Ristikari, 2011,
masculinity-femininity paradigm). Hence, national averages ex-
ceeding 0.5 can be considered moderate to large.

Women’s representation in science. Two indicators of wom-
en’s representation in science were downloaded from UNESCO’s
website (stats.uis.unesco.org): the percentage of women among
individuals (a) enrolled in tertiary science education and (b) em-
ployed as researchers. Both indicators were based on head counts.
Statistics by field of science (e.g., life vs. physical sciences) were
generally less available. The composite measure for women’s
representation in the researcher workforce combined statistics
across sectors of employment: business enterprise, government,
higher education, and private nonprofit. Although this measure
aggregated researcher statistics across many fields, the composite
measure correlated highly with the specific, but less available,
measure for natural sciences (r = .86, p < .0001, n = 28). Our
central results were similar when using the aggregated or disag-
gregated measure. Consistent with prior analyses (Else-Quest,
Hyde, & Linn, 2010; Reilly, 2012), we therefore focused on the
more available, aggregated statistics to maximize both statistical
power and the diversity of nations in our analyses. We averaged all
available statistics for the years of stereotype data collection
(2000–2008), or if those data were not available, then for the 4
years before and after data collection.

Other national indicators. In addition to using women’s
representation in science to predict gender-science stereotypes,
multiple regression analyses included 25 other national attributes
as covariates. These covariates included broad and domain-
specific indicators of gender equity, gender differences in science
achievement, Hofstede’s cultural dimensions, human develop-
ment, prevalence of scientists, world region, and sample demo-
Results

Averaged across the nations, explicit and implicit measures indicated strong associations of science with men (Ms = 0.99 and 0.98, respectively, based on random-effects weighting). The magnitude of these stereotypes was large in all nations. For instance, 90% of national averages for explicit and implicit measures fell within the ranges 0.78–1.20 and 0.76–1.20, respectively, which were estimated using the between-nation heterogeneity (both rs = 0.13) that adjusts for within-nation sampling variance. As shown in Figure 2, stereotypes were large even in nations such as Argentina and Bulgaria where women were approximately half of the nation’s science majors and employed researchers. However, the between-nation heterogeneity was significant (both ps < .0001) and substantial relative to sampling error (only 3%–4% of observed heterogeneity could be attributed to within-nation sampling variance). This heterogeneity suggests that national attributes (e.g., women’s representation in science) may explain differences in observed national averages. In addition, explicit and implicit measures correlated weakly among individuals within nations (r = .19, p < .0001, based on random-effects weighting) and across nations (based on national averages, r = .35, p = .004, N = 66 nations), suggesting that some national attributes may differently predict explicit versus implicit stereotypes.

Does Women’s Representation in Science Predict National Gender-Science Stereotypes?

As shown in Figure 2, higher female enrollment in tertiary science education predicted weaker national averages of explicit (Panel a, p = .0006) and implicit (Panel c, p = .0002) gender-science stereotypes. Higher female employment in the researcher workforce predicted weaker explicit (Panel b, p = .0004) but not implicit (Panel d, p = .88) stereotypes. Additionally, the difference between women’s representation in science education versus researcher workforce predicted implicit stereotypes (p = .006), but not explicit stereotypes (p = .55). This last result established that Panel c’s regression coefficient significantly differed from Panel d’s and that Panel a’s and Panel b’s were both significant but did not differ from each other.

What might explain the exception in which women’s employment in the researcher workforce did not predict implicit stereotypes (Panel d)? As suggested earlier, repeated counterstereotypic exposure is critical to changing implicit associations between

---

Procedure

Participants found the Project Implicit website mainly through links from other websites, media coverage, search engines, and word of mouth (Nosek et al., 2002). The website was available in 17 different languages and hosted on various web servers across the world. Participants choose the gender-science task from a list of five to 12 topics (e.g., implicit age attitudes, implicit racial attitudes). Participants therefore self-selected into the sample by having Internet access, learning about the Project Implicit website, visiting the website, and choosing the gender-science task. The Results and Limitations sections consider the influence of possible self-selection biases. The explicit stereotype measure, implicit stereotype measure, and a brief demographics questionnaire (e.g., about participants’ gender, nationality) were completed in counterbalanced order. The gender-science task required approximately 10 min to fully complete. We analyzed data from participants who had indicated their nationality and had usable data for at least one of the two gender-science stereotype measures (see Nosek et al., 2009, for description of the data cleaning procedures for the implicit measure).

Data Analysis

Our analysis addressed three questions: (a) Does women’s participation in science predict national explicit and implicit gender-science stereotypes? If so, how robust are these relationships across criteria for including nations? (b) Can other variables alternatively explain these relationships? (c) Are gender-science stereotypes better predicted by women’s representation in science or gender differences in science achievement? Unless otherwise noted, all analyses used mixed-effects meta-regression models, which assumed that national averages were combinations of fixed effects of predictor variables (e.g., women’s representation in science), between-nation heterogeneity, and within-nation sampling variance (Borenstein, Hedges, Higgins, & Rothstein, 2009). The metafor package in the statistical software R (Viechtbauer & Cheung, 2010) identified potential outliers using a diagnostic (DFITS) of a nation’s influence on the overall regression model. Nations were considered outliers if their |DFITS| > 1, a rule of thumb useful to previous researchers (e.g., Cohen, Cohen, West, & Aiken, 2003; Nosek et al., 2009). Our raw data and analysis scripts are available from the first author.

---

2 Prior research has generally revealed that the order of administration (i.e., explicit or implicit measure first) does not substantially affect measurement of stereotypes at least for Project Implicit samples (Nosek et al., 2005, Study 3). Moreover, we found similar results when separating analyses by order of administration. For instance, the relationships reported in Figure 2 never differed by order of administration (all ps > .46).

3 We also reanalyzed explicit stereotypes using difference scores that resembled those for the implicit measure: individuals’ male–female associations for science minus for liberal arts. National averages of these difference scores marginally related to female science enrollment (p = .06) and significantly related to female researcher employment (p = .01). These p values, which were higher compared with Panels a’s and b’s values, suggested that including the contrast category of liberal arts introduced some construct-irrelevant variance. However, because these relationships were still significant or marginally so, these results cannot explain why Panel d’s relationship with the implicit measure (which included a contrast category by design of the implicit measure) was not significant.
science and men. Notably, these mostly college-educated participants likely had less exposure to people employed as researchers than to science majors in universities, perhaps explaining why Panel d’s relationship was not significant. To test this explanation, we investigated a corollary hypothesis: Panel c’s relationship between women’s science enrollment and implicit stereotypes should also be weaker among individuals less exposed to science majors than among those with more exposure. Additional analyses supported this hypothesis. As shown in Figure 3, Panel c’s relationship between implicit stereotypes and women’s enrollment was about half as strong for participants who had never attended college than for college-educated participants ($p < .001$), based on two-level hierarchical linear models (Raudenbush & Bryk, 2002). Presumably, participants without college education had less repeated exposure to female and male science majors. In contrast, relationships with explicit stereotypes (Panels a and b) did not differ by participants’ level of education (all $p$s > .10). Finally, all significant relationships (Panels a–c) were approximately twice as strong for female than male participants (see Figure S1), consistent with other evidence that women are more sensitive to changes in gender diversity in STEM fields (Inzlicht & Ben-Zeev, 2000; Young et al., 2013). These differences by participant gender, however, were not as robust as differences by college education or the central findings in Figure 2 (see next section, Footnote 2).

### How Robust Are Results Across Criteria for Selecting Nations?

Self-selected Internet samples such as ours have limited representativeness of national populations (Yeager et al., 2011). Consistent with other research (e.g., Lippa, Collaer, & Peters, 2010), we therefore selected nations on the basis of two variables (sample size and the population’s percentage of Internet users) to maximize the likelihood of producing reasonably precise and representative national-level estimates. Rather than using a single criterion, we report results across many choices of selection criteria, as advocated by Simmons, Nelson, and Simonsohn (2011). Results in Figure 2 were robust across 36 choices in selection criteria based on minimum sample size ($n > 1$, $n > 10$, $n > 25$, $n > 50$, $n > 100$, $n > 200$) and percentage of Internet users (>0%, >1%, >5%, >10%, >25%, >50%). Across criteria, results were consistently replicated for the significant relationships in Panel a (all $p$s < .005), Panel b ($p < .05$ in 86% of cases), and Panel c ($p < .05$ in 86% of cases), as well as for the nonsignificant relationship in Panel d (all $p$s > .28). For Panels a–c, all relationships were in the predicted direction. Furthermore, consistent with results presented in the last section, Panel c’s estimated relationship was always more than 50% stronger for individuals with a bachelor’s degree
compared with those who never attended college ($p < .05$ in 72% of cases). Also consistent with results presented earlier, Panel a’s and b’s estimated relationships never differed by college education (all $p > .098$). Finally, Figure 2’s relationships were also robust to exclusion of outliers. For instance, across selection criteria, Panel c’s relationship was significant in 86% versus 78% of cases when including versus excluding outliers, respectively. Romania was an outlier in Figure 2’s Panel c and therefore was excluded from that panel and subsequent analyses of that relationship; results were similar with and without the outlier. This robustness across selection criteria strengthens our central findings.

**Can Covariates Explain Relationships Between Gender Diversity and Stereotypes?**

Multiple regression models tested whether other national attributes could have accounted for Figure 2’s relationships between women’s representation in science and gender-science stereotypes. Closely following Bryk and Thum’s (1989) analytic approach, we first developed separate regression models that each contained only one group of covariates (e.g., composite indices of gender equity). These initial models helped identify specific covariates that were most related to stereotypes. Consistent with Bryk and Thum, a composite model then included those covariates that significantly predicted stereotypes in the initial models. This approach maximized statistical power while investigating a wide range of covariates.

Multiple regression analyses generally indicated that (a) covariates such as national gender equity did not independently predict implicit or explicit gender-science stereotypes and (b) inclusion of covariates did not nullify relationships between women’s representation in science and these stereotypes (see Table S1 for detailed results). For example, two widely used composite indices of national gender equity—the Gender Empowerment Measure and Gender Gap Index—did not independently predict explicit or implicit gender-science stereotypes (all $p > .38$). When controlled for these measures, all relationships between women’s science participation and gender-science stereotypes that were previously significant (see Figure 2, Panels a–c) remained significant (all $p < .002$). The Netherlands was a particularly dramatic example of composite equity indices not predicting gender-science stereotypes. Despite scoring high on composite indices of gender equity, this nation (sample size $n \sim 3,000$) had the strongest explicit and second strongest implicit gender-science stereotypes among the nations in Figure 1. This seemingly paradoxical result, however, makes sense because of high domain-specific sex segregation in the Netherlands, whereby male scientists outnumbered

---

**Figure 3.** Moderation of cross-national relationships by participant’s level of college education. The $p$ values concern differences in the regression slopes, and “Slope ratio” is the slope for college-educated participants divided by the slope for participants with some or no college.
female scientists nearly four to one in both employment and educational enrollment.

Furthermore, indicating discriminant validity, the percent of women among science majors or researchers did not predict explicit stereotypes about liberal arts (all \( p > .05 \)). Women’s representation in science therefore did not predict gender stereotypes that are not related to science. Additionally, average explicit stereotypes for liberal arts and science were generally not related across nations (e.g., \( r = .09 \) among the 66 nations in Figure 1). In summary, covariate and discriminant validity analyses together support the domain specificity of relationships between women’s representation in science and national gender-science stereotypes.

**How Do Achievement Differences, Compared With Gender Diversity, Relate to Stereotypes?**

Nosek et al. (2009) presented evidence that gender differences in science achievement related to national implicit gender-science stereotypes (see also Hamamura, 2012; Pope & Sydnor, 2010). Our covariate analyses, however, revealed that these achievement differences did not independently relate to stereotypes after controlling for women’s enrollment in science education. Hence, although both gender differences in achievement and in enrollment sometimes related to cross-national differences in gender-science stereotypes, gender differences in enrollment may be more relevant to explaining differences in stereotypes. To investigate further, we compared the strength of stereotype–achievement relationships across time, selection criteria, participant gender, inclusion of covariates, and international data sources (for further detail, see the supplemental materials).

Consistent with Nosek et al. (2009), stereotype–achievement relationships were found in data from the Trends in International Mathematics and Science Study (TIMSS), which focuses on assessing what students learn in science classrooms. However, these results for TIMSS were somewhat inconsistent over time (e.g., not replicated in the year 2007), as shown in the top-left corner of Table 1. Averaging across four testing administrations helped to identify overall trends. For instance, indicating some robustness, time-averaged gender differences in TIMSS science achievement significantly related to implicit gender-science stereotypes in 39% of cases of selection criteria after excluding one influential outlier. These cross-national relationships were somewhat more robust for the stereotypes of female than male participants (see bottom-left corner of Table 1). For instance, time-averaged TIMSS gender differences related to women’s implicit stereotypes in 58% of cases of selection criteria after excluding one influential outlier. When controlled for women’s enrollment in science education, however, this relationship remained significant in only 8% of cases (and in the predicted direction in 89% of cases), whereas women’s enrollment continued to significantly predict stereotypes in 67% of cases (see Tables S2–S6 for more detailed results). Finally, our analysis identified another novel finding that relationships between achievement gender differences and stereotypes were generally not found in data from the Programme for International Student Assessment (PISA), which focuses more on assessing how well students apply science to everyday contexts than does TIMSS (Else-Quest et al., 2010; but see Fensham, 2008). See right half of Table 1 for results for PISA. Hence, achievement differences independently predicted stereotypes in some cases when specifically analyzing women’s implicit stereotypes and TIMSS (not PISA) data. However, evidence for this relationship was considerably less robust than for relationships between gender-science stereotypes and women’s representation in science.

**Discussion**

Results indicated robust relationships between women’s representation in science and national gender-science stereotypes, defined as associations connecting science with men more than women. These relationships tended to be stronger for female participants and remained after controlling for many covariates such as national gender equity. Even nations with high overall gender equity had strong gender-science stereotypes if men dominated science fields specifically (see also Charles & Bradley,
In support of the specificity to science fields, women’s representation in science did not predict explicit gender stereotypes about liberal arts. Furthermore, compared with gender differences in science achievement (Nosek et al., 2009), women’s representation in science more robustly predicted explicit and implicit stereotypes.

Women’s representation in science predicted national gender-science stereotypes in three of four cases. As an informative boundary condition, women’s employment in the researcher workforce predicted only explicit, but not implicit, gender-science stereotypes. This result suggests that repeated and varied exposures to counterstereotypic women may be necessary to stably change implicit gender-science stereotypes, consistent with broader literature on implicit social cognition (Gawronski & Bodenhausen, 2006, 2011). The implicit stereotypes of these mostly college-educated participants thus likely related more to their frequent exposure to female and male science majors and less to their rarer exposure to female and male employed researchers. If this reasoning is valid, then cross-national relationships between implicit stereotypes and gender diversity among science majors should also be weaker among participants who never attended college. Our analyses supported these predictions.

Repeated counterstereotypic exposure may be less critical to changing explicit stereotypes because they also respond to more abstract propositional information such as statistics about women’s representation in science (Gawronski & Bodenhausen, 2006, 2011). Consistent with these hypotheses, relationships between gender diversity and explicit stereotypes were similar for participants with and without college education, even though participants without college education likely had less repeated exposure to female and male science majors. These results align with other findings that people often are highly accurate in explicitly estimating gender compositions of occupations. For instance, in one study, undergraduates had high accuracy across 80 occupations, despite little direct exposure to women and men in those occupations (Cejka & Eagly, 1999).

These cross-national findings extend previous research investigating the psychological effects of encountering female role models in STEM fields (Dasgupta, 2011; González de San Román & de la Rica Goirichelay, 2012; Riegle-Crumb & Moore, 2014). Such role models show promise for weakening stereotypes, especially among female students (Bellock et al., 2010; Galdi et al., 2014; although see Lent, Bruder, & Sedikides, 2009) and students who strongly identify with the role models (Young et al., 2013). However, these counterstereotypic examples could be subtyped because they occur along with pervasive stereotypic evidence from the broad cultural environment (Richards & Hewstone, 2001; Stout et al., 2011). Changes in broader cultural environments such as women’s increasing representation in science fields in the United States (Hill et al., 2010) might have stronger, more robust effects on gender-science stereotypes. Role models may be one of the first steps in changing stereotypes over time, especially because female STEM peers and mentors can help protect girls and women against the negative effects of current stereotypes (Dasgupta, 2011; Stout et al., 2011). Cultural stereotypes could then change as more women enter STEM fields and gender compositions change at the national level (Beaman, Chattopadhyay, Duflo, Pande, & Topalova, 2009). Future research can help understand how individual differences in counterstereotypic exposure contribute to these cultural trends. For instance, our analysis of the relationship between individuals’ educational attainment and gender-science stereotypes could be extended by measuring how closely college-educated individuals identified with female science peers and professors (Young et al., 2013).

Throughout this article, we have primarily considered how gender diversity might influence stereotypes. However, as noted earlier, because the data are correlational, a bidirectional relationship is plausible. For instance, as Nosek et al. (2009) suggested, implicit stereotypes could cause women to underperform on science achievement tests because of a phenomenon known as stereotype threat (Schmader, Johns, & Forbes, 2008; Walton & Spencer, 2009). This lower achievement could, in turn, limit women’s access to science fields. Consistent with this reasoning, women’s implicit gender-science stereotypes related to male advantages in the TIMSS test, which was designed to assess students’ learning of science curriculum. However, this evidence was less robust than our central findings relating stereotypes and women’s representation in science. Stereotypes could also influence women’s representation in science through other factors such as women’s identification with STEM fields (Dasgupta, 2011; Nosek & Smyth, 2011). Both causal directions between gender composition in science and gender-science stereotypes are thus plausible, although gender composition likely influences stereotypes more directly than stereotypes influence gender composition. The impact of stereotypes on gender compositions would be mediated over many years as women enroll in STEM courses and seek employment in STEM fields, whereas the impact of gender compositions on stereotypes can be more immediate (Lenton et al., 2009).

Furthermore, some of our study’s results would be difficult to explain if gender composition did not influence stereotypes in some way. For instance, if the gender composition of science majors in college did not affect stereotypes, then stereotypes of individuals with and without college education should not differ. Another alternative hypothesis is that individuals with and without college education might differ on average if other correlated individual-level variables (e.g., age or socioeconomic status) influence stereotypes. However, our data supported neither hypothesis because college education predicted stronger implicit stereotypes, but only in nations where men dominated science majors (see Figure 3). In contrast, college education predicted weaker implicit stereotypes in nations where women dominated science majors. Compared with those alternative hypotheses, the associative-propositional model (Gawronski & Bodenhausen, 2006, 2011) can more parsimoniously account for the cross-level interactions with women’s representation in science, as discussed earlier.

**Limitations**

Our correlational design revealed the possible impact of cultural environments, but experimental manipulations offer greater potential for causal inference. However, the effects of experimental manipulations may be weakened by broader sociocultural messages (e.g., Stout et al., 2011). Hence, the effects of cultural environments are inherently challenging to study because they generally cannot be experimentally manipulated. Although investigating changes over time could strengthen cross-cultural analyses...
such as ours (Brandt, 2011), this study’s time period of data collection was too small (2000–2008) to meaningfully test for such longitudinal changes.

This study used self-selected Internet samples, which have limited representativeness of national populations (Yeager et al., 2011), especially if the percentage of Internet users is low. These concerns were somewhat lessened because our central results were robust across a wide range of minimum percentages of Internet users. Nevertheless, participants also self-selected into our study by finding the Project Implicit website and choosing the gender-science task. For instance, participants especially interested in gender issues might have been more likely to choose the gender-science task. Self-selection is therefore a general methodological concern. However, it is unclear how much self-selection affected our specific empirical findings (e.g., the regression slopes in Figure 2).

Another limitation was that available statistics did not permit analysis of how results might have differed by field of science (e.g., physics vs. biology). Future research should address this issue because both stereotypes and gender diversity vary substantially by field (Nosek & Smyth, 2011; Smyth & Nosek, 2013). However, as noted in the introduction, our analyses that averaged across science fields may be justified because female dominance in only one field (e.g., biology) could be regarded as an exception to the usual pattern of male dominance in science (Richards & Hewstone, 2001). Other limitations were that explicit stereotypes were measured by a single survey item, participants without college education were underrepresented in our sample, nations characterized as low in human development and/or had low Internet usage rates were underrepresented in our sample, and the number of nations was small (N = 66) even if the number of participants was large. However, in defense of our findings, they proved to be robust despite these limitations.

**Educational Implications and Future Research**

Our results indicated that participants across 66 nations strongly associated science with men more than women, including in nations where women were approximately half of the nation’s science majors and employed researchers (see also Nosek et al., 2009). Hence, across the world, gender-science stereotypes present concerns for science educators and students to the extent that these associations affect the experiences of women and men pursuing science degrees and occupations. For instance, such stereotypes negatively impact women by causing underachievement in introductory undergraduate STEM courses (Miyake et al., 2010), disidentification with and negative attitudes toward science (Good, Rattan, & Dweck, 2012; Nosek & Smyth, 2011; Steffens et al., 2010), and gender discrimination (Rueben, Sapienza, & Zingales, in press). Despite their ubiquity, gender-science stereotypes also demonstrated cultural variability and therefore potential for change. Yet, gender-science stereotypes were still strong even in nations with small gender differences in overall science participation.

Although it is not yet clear how best to weaken these stereotypes, a number of promising strategies can be explored. Science educators could help weaken stereotypes by highlighting diverse examples of female scientists. Presenting single or infrequent examples of female scientists will likely not substantially change gender-STEM stereotypes (Stout et al., 2011), especially if such women are presented as token examples (Shachar, 2000). A more effective strategy to weaken stereotypes could be to integrate many examples of female scientists as part of teachers’ normal classroom instruction. For instance, teachers could motivate the learning of specific scientific concepts by discussing how they relate to the research of currently practicing female and male scientists (Linn & Eylon, 2011). Related prior research has indicated the benefits of integrating narrative information about scientists into instruction (Arya & Maul, 2012). For instance, in one experimental study (Hong & Lin-Siegler, 2012), learning how scientists struggled in their research increased students’ interest in the science lesson and students’ content understanding (e.g., about Newtonian mechanics). Learning how both female and male scientists struggle could also help protect female students against the negative effects of gender-STEM stereotypes (Asgari, Dasgupta, & Stout, 2012; Good et al., 2012). Future research should extend these approaches to understand how repeated examples of female scientists might weaken gender-STEM stereotypes over time.

Our study might also have implications for social policies such as affirmative action. For the recent U.S. Supreme Court case *Fisher v. Texas*, social psychologists prepared an amicus brief outlining the implications of stereotype threat for affirmative action (Brief of Experimental Psychologists, 2012). The brief argued that diversity in college populations helps minorities reach their maximum potential because ingroup peers can inoculate minorities against the negative effects of cultural stereotypes (Dasgupta, 2011; Murphy, Steele, & Gross, 2007; Richman, vanDellen, & Wood, 2011). Increasing the diversity of college populations might also change underlying stereotypes about science fields. Future research should investigate this possibility. These efforts to weaken stereotypes could then have cascading influence by encouraging more women to pursue and excel in fields in which they have been historically underrepresented.

**References**


This document is copyrighted by the American Psychological Association or one of its allied publishers.

This is intended solely for the personal use of individual users and may not be duplicated.

GENDER DIVERSITY AND SCIENCE STEREOTYPES

National Academies of Sciences, 107, 1060–1063. doi:10.1073/pnas.0910967107


(Appendix follows)
Appendix

Covariates Included in Multiple Regression Analyses

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Composite indices of gender equity</strong></td>
<td></td>
</tr>
<tr>
<td>GGI</td>
<td>Gender Gap Index. Based on four subindices for gender gaps in economic participation and opportunity, educational attainment, political empowerment, and health/survival.</td>
</tr>
<tr>
<td>GEM</td>
<td>Gender Empowerment Measure. Based on gender gaps in earned income; women’s representation in parliament; and women’s employment in managerial, professional, and technical occupations.</td>
</tr>
<tr>
<td><strong>Domain-specific gender equity</strong></td>
<td></td>
</tr>
<tr>
<td>GGI_eco</td>
<td>Economic subindex of Gender Gap Index. Based on gender gaps in labor force participation rates, wage equality for similar work, earned income, and high labor employment.</td>
</tr>
<tr>
<td>GGI_edu_log</td>
<td>Education subindex of Gender Gap Index. Based on gender gaps in literacy rates and enrollment rates in primary, secondary, and tertiary education. Reflected and log-transformed to reduce negative skew.</td>
</tr>
<tr>
<td>TertArtsF</td>
<td>Percentage of women among liberal arts majors (humanities and arts).</td>
</tr>
<tr>
<td>TertTeachF</td>
<td>Percentage of teachers in tertiary education who are women.</td>
</tr>
<tr>
<td><strong>Achievement differences</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Cultural dimensions</strong></td>
<td></td>
</tr>
<tr>
<td>PowerDist</td>
<td>Power Distance. Represents “the extent to which the less powerful members of institutions and organizations within a country expect and accept that power is distributed unequally” (Hofstede et al., 2010, p. 61).</td>
</tr>
<tr>
<td>UncertAvoid</td>
<td>Uncertainty Avoidance. Represents “the extent to which the members of a culture feel threatened by ambiguous or unknown situations” (Hofstede et al., 2010, p. 191).</td>
</tr>
<tr>
<td>MascFem</td>
<td>Masculinity minus Femininity. Masculinity represents “when emotional gender roles are clearly distinct: Men are supposed to be assertive, tough, and focused on material success, whereas women are supported to be more modest, tender, and concerned with the quality of life,” whereas femininity represents “when emotional gender roles overlap” (Hofstede et al., 2010, p. 140).</td>
</tr>
<tr>
<td>IndivCollect</td>
<td>Individualism minus Collectivism. Individualism represents “societies in which the ties between individuals are loose,” whereas collectivism represents “societies in which people from birth onward are integrated into strong, cohesive in-groups” (Hofstede et al., 2010, p. 92).</td>
</tr>
<tr>
<td>Atheism_log</td>
<td>Percentage of population that does not believe in a God. Log-transformed to reduce positive skew.</td>
</tr>
<tr>
<td>HDI_log</td>
<td>Human Development Index. Based on life expectancy at birth, mean years of schooling, expected years of schooling, and gross national income per capita. Reflected and log-transformed to reduce negative skew.</td>
</tr>
<tr>
<td>IQ</td>
<td>Nation’s average IQ.</td>
</tr>
<tr>
<td>Prevalence of scientists</td>
<td>Number of employed researchers (based on head counts) per one million people. Log-transformed to reduce positive skew.</td>
</tr>
<tr>
<td>Rsrcher_log</td>
<td>Percentage of tertiary students in science.</td>
</tr>
<tr>
<td>TertSciPct</td>
<td>Dummy code comparing nations in Asia with nations in the Americas.</td>
</tr>
<tr>
<td>World region</td>
<td>Dummy code comparing nations in Europe with nations in the Americas.</td>
</tr>
<tr>
<td>Other</td>
<td>Dummy code comparing nations in other world regions (Africa; Oceania) with nations in the Americas. These other regions were combined into one dummy code because of their low frequency in our sample of nations.</td>
</tr>
<tr>
<td>Sample characteristics</td>
<td>Average trial latency collapsed across experimental conditions of the implicit measure.</td>
</tr>
<tr>
<td>prct_male</td>
<td>Percentage of men in the stereotype sample.</td>
</tr>
<tr>
<td>prct_college</td>
<td>Percentage of stereotype sample with bachelor’s degree or higher.</td>
</tr>
<tr>
<td>age_mean</td>
<td>Average age of the stereotype sample.</td>
</tr>
<tr>
<td>corr_iatexp</td>
<td>Correlation between implicit and explicit gender-science stereotypes.</td>
</tr>
</tbody>
</table>

Note. TIMSS = Trends in Mathematics and Science Study; PISA = Programme for International Student Assessment.