



National differences in gender-science stereotypes predict national sex differences in science and math achievement

Citation

Nosek, Brian A., Frederick L. Smyth, N. Sriram, Nicole M. Lindner, Thierry Devos, Alfonso Ayala, Yoav Bar-Anan, et al. 2009. National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences* 106, no. 26: 10593–10597. doi:10.1073/pnas.0809921106.

Published Version

doi:10.1073/pnas.0809921106

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:27414715>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

National differences in gender–science stereotypes predict national sex differences in science and math achievement

Brian A. Nosek^{a,1}, Frederick L. Smyth^a, N. Sriram^a, Nicole M. Lindner^a, Thierry Devos^b, Alfonso Ayala^c, Yoav Bar-Anan^{a,2}, Robin Bergh^d, Huajian Cai^e, Karen Gonsalkorale^f, Selin Kesebir^a, Norbert Maliszewski^g, Félix Neto^h, Eero Olliⁱ, Jaihyun Park^j, Konrad Schnabel^k, Kimihiro Shiomura^l, Bogdan Tudor Tulbure^m, Reinout W. Wiersⁿ, Mónika Somogyi^o, Nazar Akrami^d, Bo Ekehammar^d, Michelangelo Vianello^p, Mahzarin R. Banaji^q, and Anthony G. Greenwald^r

^aDepartment of Psychology, University of Virginia, Charlottesville, VA 22904-4400; ^bDepartment of Psychology, San Diego State University, San Diego, CA 92182-4611; ^cElectoral Institute of Veracruz, Xalapa, 91000, Mexico; ^dDepartment of Psychology, Uppsala University, SE-751 42 Uppsala, Sweden; ^eDepartment of Psychology, Sun Yat-Sen University, Guangzhou 510275, China; ^fSchool of Psychology, University of Sydney, Sydney NSW 2006, Australia; ^gDepartment of Psychology, University of Warsaw, 00-927, Warsaw, Poland; ^hDepartment of Psychology, University of Porto, 4099-002 Porto, Portugal; ⁱEquality and Anti-Discrimination Ombud, N-0031 Oslo, Norway; ^jDepartment of Psychology, Baruch College—The City University of New York, New York, NY 10010; ^kDepartment of Psychology, Humboldt University, 10099 Berlin, Germany; ^lDepartment of Clinical Social Services, Iwate Prefectural University, 152-52 Takizawa-aza-sugo, Takizawa, Iwate 020-0193, Japan; ^mDepartment of Psychology, Transilvania University of Brasov, 500036, Brasov, Romania; ⁿDepartment of Psychology, University of Amsterdam, 1012 ZA Amsterdam, The Netherlands; ^oDepartment of Psychology, Eötvös Loránd University, Budapest 1053, Hungary; ^pDepartment of Psychology, University of Padua, 35122 Padua, Italy; ^qDepartment of Psychology, Harvard University, Cambridge, MA 02138; and ^rDepartment of Psychology, University of Washington, Seattle, WA 98195

Edited by Claude M. Steele, Stanford University, Stanford, CA, and approved April 24, 2009 (received for review October 6, 2008)

About 70% of more than half a million Implicit Association Tests completed by citizens of 34 countries revealed expected implicit stereotypes associating science with males more than with females. We discovered that nation-level implicit stereotypes predicted nation-level sex differences in 8th-grade science and mathematics achievement. Self-reported stereotypes did not provide additional predictive validity of the achievement gap. We suggest that implicit stereotypes and sex differences in science participation and performance are mutually reinforcing, contributing to the persistent gender gap in science engagement.

Implicit Association Test | culture | social psychology | implicit social cognition

The gender gap in interest, participation, and performance in science is well known and the subject of intense scrutiny. World-wide, for example, 8th-grade boys show significantly greater achievement than girls in science (1, 2). Observations of such differences have reinforced the view that boys are “naturally” better equipped to excel in science and mathematics (3). However, the size of the sex gap varies, representing a challenge to that nativist position. For instance, in the 2003 Trends in International Mathematics and Science Study (TIMSS) (1, 2) among 34 nations, there was substantial variability in the size of the sex difference, and 8th-grade girls in 3 nations significantly outperformed boys in science. For the same sample there was no overall sex difference in mathematics achievement, with girls significantly outperforming boys in 7 nations, and boys significantly outperforming girls in 5 nations. Beyond data from 8th graders, a recent review across age groups found that the U.S. sex gap in math performance has been declining over time (4), and another study reported that the size of the sex gap in math performance across countries was related to national indicators of gender egalitarianism (5). This variability across time and place suggests that sex differences in math and science achievement are shaped by socio-cultural factors (4–10).

Stereotypes that men are naturally more talented and interested in math and science are thought to influence the science, technology, engineering, and math aspirations and achievements of boys and girls, men and women (11, 12, 14–16). For example, women who endorse such stereotypes also report less interest in math and science, and are less likely to pursue a math or science degree (17, 18). Also, reminding women of the “math = male” stereotype, or just unobtrusively highlighting their gender, is

sufficient to weaken their performance on a subsequent math or engineering examination compared with a control group (19–22). This phenomenon, termed social identity threat, is thought to occur via increased anxiety, and increased cognitive load created by such anxiety, that one’s own behavior will potentially confirm a stereotype about one’s group (23–25).

These examples illustrate that stereotypes can influence individual performance in math and science domains. The reverse causal scenario can also occur. People are sensitive to covariation in their environment and learn easily by observation or the testimony of others (10, 26–29). In the case of sex differences in math and science, the vastly greater presence of men, especially in the highly visible top echelons of these fields, is likely to be noticed and the covariation acquired (30–33). In one study, female science majors who saw a science conference video with 75% male participants (akin to the existing reality in many scientific fields) felt less belonging, less desire to participate in the conference, and even more physiological markers related to threat than female science majors who saw a gender-balanced conference video (34). Social reinforcements that support the “boys are better” stereotype only add to the blatant fact of visible covariation.

Applying these bi-directional relations to a cultural level of analysis, national sex differences in science participation or performance may create “science = male” stereotypes, and science = male stereotypes may create national sex differences in science participation or performance. Mutually reinforcing mechanisms could lead some cultures to maintain larger sex gaps in science participation and performance than others. It is difficult to establish causal relations across cultures because cultures are not amenable to random assignment of treatments.

Author contributions: B.A.N., N.S., M.R.B., and A.G.G. designed research; B.A.N., F.L.S., N.S., T.D., A.A., Y.B.-A., R.B., H.C., K.G., S.K., N.M., F.N., E.O., J.P., K. Schnabel, K. Shiomura, B.T.T., R.W.W., M.S., N.A., B.E., and M.V. performed research; B.A.N., F.L.S., N.M.L., and A.G.G. analyzed data; and B.A.N., F.L.S., M.R.B., and A.G.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed at: Department of Psychology, University of Virginia, Box 400400, Charlottesville, VA 22904-4400. E-mail: nosek@virginia.edu.

²Present address: Department of Psychology, Ben-Gurion University of the Negev, Beer-Sheva, 84105, Israel.

This article contains supporting information online at www.pnas.org/cgi/content/full/0809921106/DCSupplemental.

Even so, if a standardized measure of the gender–science stereotypes were available across multiple countries, then we could investigate whether there is a performance–stereotype relationship across nations. The present investigation does exactly that because, for the first time, such a measure was available from a very large international sample.

Given the foregoing review, it is surprising to note that relatively few people explicitly endorse gender–science stereotypes (17, 18). With weak endorsement, it is not obvious how stereotypes can be importantly related to achievement in math and science. Based on many decades of research on the limits of self-report (35, 36), we know that the lack of stereotype endorsement does not therefore imply a lack of its influence on choices and behavior. For example, the research on social identity threat, noted earlier, suggests that stereotypes need not be explicitly endorsed to influence individual behavior. Likewise, research relying on the Implicit Association Test (IAT) (37) shows that most men and women associate male with science and female with liberal arts more easily than the reverse (38). The IAT is a behavioral measure in which participants categorize words into their superordinate categories in 2 different sorting conditions. In one condition, participants categorize items representing male (e.g., he, boy) and science (e.g., physics, chemistry) with one response key, while categorizing items representing female (e.g., she, girl) and liberal arts (e.g., arts, history) by using another response key. In the other opposing condition (randomly completed before or after the first condition) (39), participants categorize the same words but they are paired differently: This time male and liberal arts items are categorized with one key whereas female and science items are categorized with the other. Most people are able to categorize the words faster and more accurately in the former condition (male = science) compared with the latter (female = science). This differential ease is taken to reflect stronger associations of science with male than female. We have interpreted this result as reflecting an implicit gender–science stereotype because participants do not introspect or express their conscious beliefs about gender and science. The implicit stereotype may differ from self-reported stereotype because people are unaware of it, do not endorse it, or do not wish to reveal that they endorse it.

Individual differences in the tendency to associate male with science (or math) on the IAT predicts interest, participation, and performance in scientific domains (14, 40, 41). For example, women who find it easier to associate men with science (and women with liberal arts) report less liking for math and science domains, less interest in pursuing science in the future, perform worse on standardized math exams like the SAT and ACT, and are less likely to be a math or science majors compared with women who do not have that association (41). Also, a prospective study of women taking college calculus found that, for weakly gender-identified women, stronger implicit, but not explicit, math = male stereotyping at the start of a semester predicted worse final examination performance (14).

We operated a virtual laboratory at which participants could complete the gender–science IAT described above (<https://implicit.harvard.edu/>). The site was available in 17 languages and attracted a very large and diverse sample. The accumulated dataset is wholly unique in being a large-scale assessment of implicit gender–science stereotypes. Across >500,000 completed tests from around the world, >70% of men and women showed a tendency to associate male with science and female with liberal arts more easily than the reverse on the IAT (38), and the implicit stereotype was relatively weakly related with self-reported stereotyping ($r = 0.22$). Even so, there was substantial variability in implicit stereotyping across individuals and across cultures. Because of the very large sample size, we could compute national estimates of implicit gender–science stereotypes for dozens of countries.

The present research investigated whether implicit gender–science stereotypes could account for sex differences in science performance across nations. As a national indicator, implicit stereotypes may index the extent to which associations of male with science are manifest in the national culture, even if people in that society are relatively unwilling to endorse such stereotypes. Culture is a powerful force for shaping the beliefs and behavior of its members (42). As such, we hypothesized that the strength of implicit stereotyping at a national level would be positively related with the extent to which sex differences in science performance are observed in that culture. This effect would suggest that implicit gender–science stereotyping is a national indicator of gender (in)equality in science achievement.

Results

In 2003, the TIMSS conducted standardized exams of math and science achievement among representative samples of 8th graders (43, 44). We used the 34 countries that followed the TIMSS sampling guidelines as our sample of nations to ensure comparable results (1). Because our implicit stereotype assessment was specific to science, we began with analysis of the mean science scores reported for boys and girls in each country (median country boys' science score is 516, girls' is 506, and median country overall SD = 75). We created an index of national sex differences in science achievement by subtracting the mean for girls from that for boys. With a median advantage of 9.5 points (mean = 8.6), boys averaged significantly higher science achievement in 65% of the countries. However, cross-country variation was substantial, ranging from raw score gaps favoring boys of -27 to 29 (SD = 11.3), and, in terms of Cohen's d effect sizes, from -0.31 to 0.40 (SD = 0.15).

For assessment of implicit gender–science stereotypes, we used IAT (39, 45) data collected at the Project Implicit website (<https://implicit.harvard.edu/>) (38). Over a half million gender–science IATs were completed between May 2000 and July 2008. We focus on the $n = 298,846$ from citizens of the 34 TIMSS nations (mean participant age = 27, SD = 11; 65% female). National indicators of implicit gender–science stereotypes were operationalized as the mean score of all valid IAT scores of citizens of each nation (see *SI Appendix* for details). IAT participation by nation within the subsample of 34 TIMSS nations varied widely, with the largest sample coming from the United States ($n = 248,306$) and the smallest from Moldova ($n = 15$). The median sample size was 473.

We tested our hypothesis by regressing the TIMSS 2003 national sex differences in 8th-grade science performance on the national estimates of implicit stereotyping, first alone and then including a variety of covariates. Given their variability, the inverses of the country variances for the implicit stereotype and TIMSS estimates were averaged and then used as weighting for the regressions (weighting details are provided in the *SI Appendix*).

Fig. 1 presents the nation-level scatter plot of the key relationship: National implicit stereotyping of science as male was strongly related to national sex differences in 8th-grade science performance ($r = 0.60$, 95% confidence interval: 0.31, 0.77). In terms of regression, the estimated effect of a 1-standard-deviation increase in implicit stereotyping on the male science advantage was 6.3 points, $P < 0.001$ (SD of the sex difference in achievement scores is 11.3), or a standardized effect (β) of 0.56. This is the key relationship between implicit stereotypes and sex differences in science performance. To test the robustness of this relationship, we subjected the data to an increasingly stringent series of tests. Given the small sample of nations, this sequence of regressions provides a very conservative estimate of the reliability of the relationship. Regression diagnostics identified 1 high leverage outlier. The effect persisted after removing that outlier, $\beta = 0.66$ ($P < 0.001$). Next, we included in the model 7 country-level covariates derived from TIMSS and Project Im-

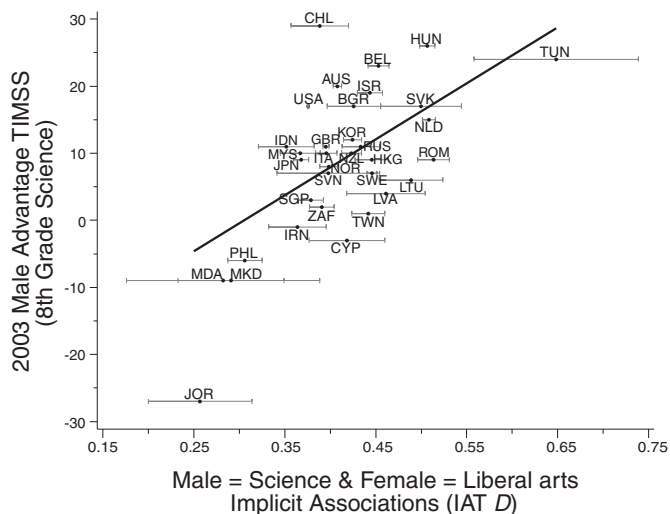


Fig. 1. The relationship between implicit gender–science stereotyping and national sex differences in science performance for 2003 TIMSS data. Horizontal error bars represent the standard error of the mean for implicit stereotype data. Regression estimated with covariates and reflects weighting that is detailed in the *SI Appendix*. Country codes: AUS, Australia; BEL, Belgium; BGR, Bulgaria; CHL, Chile; CYP, Cyprus; GBR, United Kingdom; HKG, Hong Kong—China; HUN, Hungary; IDN, Indonesia; IRN, Iran; ISR, Israel; ITA, Italy; JOR, Jordan; JPN, Japan; KOR, South Korea; LTU, Lithuania; LVA, Latvia; MDA, Moldova; MKD, Macedonia; MYS, Malaysia; NLD, The Netherlands; NOR, Norway; NZL, New Zealand; PHL, Philippines; ROM, Romania; RUS, Russia; SGP, Singapore; SVK, Slovakia; SVN, Slovenia; SWE, Sweden; TUN, Tunisia; TWN, Taiwan; USA, United States; ZAF, South Africa.

implicit data: TIMSS 2003 8th-grade science mean, mean explicit (self-reported) science and arts stereotypes, implicit–explicit science stereotype correlation, mean IAT trial latency across the task’s response conditions, percentage of IAT sample that was male, and average participant age of the IAT sample. Implicit stereotyping remained a uniquely significant predictor ($\beta = 0.55$, $P = 0.01$) and none of the covariates were significant predictors (see *SI Appendix* for details of covariate analyses). Finally, following Guiso and colleagues (5), we added 2 additional covariates: national gross domestic product (GDP) and an indicator of gender equality, the Gender Gap Index (GGI; see *SI Appendix* for variable details), the latter of which was not available for 2 of the TIMSS nations. Despite the reduced sample size ($n = 31$) and power (9 covariates), the effect of implicit stereotyping remained significant ($\beta = 0.71$, $P < 0.01$), with an estimated effect of 1 standard deviation increase in stereotyping predicting a 0.7 standard deviation increase in the male 8th-grade science advantage (see Table 1 for summary of model tests).

Math achievement was also measured by TIMSS and, although distinct from science, it is a closely associated discipline. Math is assumed to be an important component of most science fields, and a key skill for scientific excellence. We tested whether implicit gender–science stereotypes would also predict these national sex differences in math performance. Replicating the analytic approach as described for the science outcome revealed the same pattern of results for the math sex gaps: National implicit science stereotyping was significantly positively related ($\beta = 0.63$, $P < 0.0001$) and persisted after removing the same high-leverage outlier ($\beta = 0.52$, $P < 0.05$). In the full multiple regression model with 9 covariates (except the country 8th-grade math mean from TIMSS was used instead of the science mean) and with 2 nations removed that did not have gender gap index (GGI) data, implicit gender–science stereotyping remained a significant unique predictor of the math gap ($\beta = 0.67$, $P < 0.05$).

Again, with covariates accounted for, the estimated effect of a SD increase in implicit science stereotype was a roughly 0.7-SD increase in the male advantage in 8th-grade math performance. In summary, variation in science = male IAT scores across 34 nations predicted variation in sex differences in both 2003 science and math performance even after removing 12 degrees of freedom (1 outlier, 2 with missing data, 9 covariates).

TIMSS conducted another international data collection in 1999, before any of the IAT data had been collected. These data offered an opportunity for replication when performance had temporal precedence to stereotyping (see *SI Appendix* for notes on 1995 and 4th-grade TIMSS). Only 29 nations participated in the 1999 TIMSS, reducing statistical power to detect relationships still further. Even so, the implicit science stereotyping and the earlier science–gender gap were significantly positively related ($\beta = 0.46$, $P < 0.01$). And, that significant effect persisted after removing 2 influential outliers ($\beta = 0.43$, $P < 0.05$). With n now at just 27, the significant contribution of implicit stereotyping was lost when the 7 covariates of model M4 were introduced (see Table 1). However, none of these covariates alone was significantly related to the TIMSS99 science difference (see *SI Appendix* for effects of each covariate alone and in combination with implicit stereotyping). GGI, added with GDP in model M5, was the only covariate by itself significantly related to the TIMSS99 science outcome (and its inclusion further reduced the sample size to 25). Of note, when implicit stereotyping was included in a model with GGI as the lone covariate, both remained independently predictive of the TIMSS99 science-achievement gap, with the estimated effect of stereotyping at $\beta = 0.48$ ($P = 0.01$). The 1999 math gap was positively related to implicit science stereotyping ($\beta = 0.37$, $P < 0.05$). However, that relationship disappeared after removing one influential outlier ($\beta = 0.06$, $P = 0.79$). The effect did not return to statistical significance after including the 9 covariates, even though the effect size estimate was larger than the initial one ($\beta = 0.41$, $P = 0.31$).

In summary, we observed a positive relationship between implicit gender–science stereotyping in all 4 comparisons (2003 and 1999 science and math performances). In 3 cases, the effect was still reliable after removing high-leverage outliers. In 2 cases, the effect was still reliable even after adding 9 covariates (and losing an additional 2 countries with missing data on one covariate). There are multiple possible explanations for the variation in robustness of the effect. The most likely is statistical power (see *SI Appendix*). By using the initial 2003 IAT–science relationship as a baseline ($R^2 = 0.35$), the power to detect that effect with $\alpha = .05$ and 14 degrees of freedom (the final 1999 science df) was 0.52. To achieve 80% power to detect the original effect size in the covariate analysis, we would have needed 57 nations in the sample (13).

There may also be substantive reasons for the less robust effect in the 1999 data compared with the 2003 data. For example, socio-cultural stereotyping across nations is likely to be shifting over time. The stereotyping data were collected over an 8-year span from 2000 to 2008. The 1999 TIMSS data may have a weaker relationship with the national indicators of implicit stereotyping because it was temporally before the entire stereotype data collection (see *SI Appendix*). There is not enough data within each nation to test temporal hypotheses with confidence, but future investigations may be able to shed light on these and other possible explanations.

Self-report measures of gender–science stereotyping were also included at the Project Implicit websites offering an opportunity to test whether both implicit and explicit stereotyping contributed to predicting the sex gap in performance. Explicit science = male stereotyping was significantly correlated with the 2003 TIMSS sex gaps in both science and math (science-weighted: $r = 0.39$, 95% CI: 0.05, 0.64; math: $r = 0.34$, 95% CI: 0.00, 0.61), but

Table 1. Estimated effects of country-level implicit gender–science stereotype on country male–female score differences in 8th-grade TIMSS science and math in 2003, 1999, and in aggregate

Regression models	DVs are 2003 male–female test score differences						DVs are 1999 male–female test score differences						Aggregate IAT effect, std
	Science			Math			Science			Math			
	<i>df</i>	<i>SD</i>	IAT effect, raw (std)	<i>df</i>	<i>SD</i>	IAT effect, raw (std)	<i>df</i>	<i>SD</i>	IAT effect, raw (std)	<i>df</i>	<i>SD</i>	IAT effect, raw (std)	
Unweighted													
M1 (all countries) IAT only	32	11.3	7.3 (0.65)	32	9.5	5.9 (0.62)	27	10.4	5.3 (0.51)	27	8.7	2.6 (0.30)	0.53
Weighted*													
M2 (all countries) IAT only	32	11.3	6.3 (0.56)	32	9.5	6.0 (0.63)	27	10.4	4.8 (0.46)	27	8.7	3.2 (0.37)	0.51
M3 (outliers dropped) IAT only	31	11.1	7.3 (0.66)	31	8.5	4.4 (0.52)	25	8.4	3.6 (0.43)	26	7.9	0.5 (0.06)	0.43
M3a (outliers dropped) mid-90% IAT only	31	11.1	7.5 (0.68)	31	8.5	4.7 (0.55)	25	8.4	3.7 (0.44)	26	7.9	0.6 (0.08)	0.45
M4 = M3 + 7 covariates	24	11.1	6.1 (0.55)	24	8.5	4.6 (0.54)	18	8.4	−0.1 (−0.01)	19	7.9	0.6 (0.08)	0.32
M5 = M4 + 2 covariates, GDP/GGI†	20	11.4	8.1 (0.71)	20	8.8	5.9 (0.67)	14	8.7	1.8 (0.21)	15	8.1	3.3 (0.41)	0.53

Bold font indicates significant IAT parameter estimate, $P < 0.05$. The 7 covariates of M4 are from either Project Implicit or TIMSS data: (i) explicit “science = male” and (ii) “liberal arts = female” stereotypes; (iii) correlation of IAT and explicit “science = male” scores; (iv) overall country mean TIMSS 8th-grade score for given year and test; (v) mean IAT trial latency collapsed across experimental conditions; (vi) percentage of men taking the IAT; and (vii) mean age of Project Implicit participants. *df* is the degrees of freedom for the model error term. *SD* is the standard deviation of country-mean TIMSS male–female score differences for the given year and test. Raw IAT effect is the estimated change in the given DV for a 1-standard-deviation change in country IAT score; standard (std) IAT effect is scaled by the *SD* of the given TIMSS sex difference. Aggregate IAT effect is the mean of the 4 standardized estimates, each weighted by the respective *df*. Effect-raw is the estimated effect.

*Regression weighting details are in the *SI Appendix*.

†GGI was not available for Chinese Taipei and Hong Kong.

not with the 1999 TIMSS differences ($r = 0.27$ and 0.25 , respectively). However, when both implicit and explicit stereotypes were included in regression models, explicit stereotyping did not contribute uniquely to the prediction of 2003 science or math gender differences, but implicit stereotyping continued to be significantly predictive for both. In other words, explicit stereotypes uniquely accounted for 2% of variance in the science sex gap and 1% of the math sex gap, whereas implicit stereotypes uniquely accounted for 19% and 24%, respectively.

Self-selection of the IAT respondents is a potential threat to the validity of inference. However, if the same selection pressures are operating across nations, comparisons within the overall sample would not be undermined. Also, nations self-selected for participation in the TIMSS data collection. It is possible that national and individual self-selection factors could be artificially inflating the observed correlations if those factors varied systematically with both stereotyping magnitude and performance differences. Other tasks, beside implicit gender stereotyping, appeared at the Project Implicit website and were subject to similar selection influences, and offer an opportunity to test discriminant validity. We calculated national implicit race and age bias estimates with tasks that measured associations between black and white faces (or young and old faces) and good and bad words (38). Repeating the regressions with covariates, national implicit race and age bias did not reliably predict sex differences in TIMSS science or math performance in 2003 or 1999 (all 8 P values > 0.28). Thus, the prediction of TIMSS science performance was specific to implicit gender–science stereotypes.

Discussion

We found that a national indicator of implicit gender–science stereotyping was related to nations’ sex differences in science and math achievement. National sex differences in science and

math achievement were based on the international TIMSS standardized examination of 8th graders, whereas estimates of national implicit gender–science stereotyping were calculated from IATs completed by a large volunteer sample at Project Implicit (<https://implicit.harvard.edu/>). The mean level of implicit stereotyping among national citizens, regardless of age or gender, predicted the sex differences in TIMSS performance among the 8th graders of that nation from 2003 and 1999.

The finding is especially compelling given that 2 distinct samples provided (i) the societal indicators of implicit stereotyping and (ii) science and math performance estimates. There is no reason to expect that members of the IAT sample had any particular interaction with or specific exposure to 8th graders in 1999 and 2003. Rather, a more likely cause of the relation is that both the 8th grade test takers and the diverse IAT participants of a given country are influenced by the same socio-cultural context. That social context embodies the reciprocal influence of stereotyped science = male associations and sex differences in engagement in science and mathematics. This significant relationship persisted even after accounting for a general indicator of societal gender inequality, the GGI. Thus, the relation between implicit gender–science stereotypes and science and math achievement gaps is specific to science and math domains, and not simply a consequence of generalized national gender inequality.

If implicit gender stereotypes and sex gaps in scientific engagement are mutually reinforcing, then national policy initiatives addressing both factors simultaneously stand the best chance to maximize national scientific achievement. Education campaigns attempting to bolster women’s participation and performance must overcome the pervasive implicit stereotypes that are already embodied in individual minds. Likewise, interventions aimed at altering implicit stereotypes must contend with the influence of persisting cultural realities that fewer

women pursue scientific careers and are in positions of scientific leadership. Even so, whereas mutually reinforcing influences make it more difficult to jerk the system out of homeostasis, an effective intervention that changes implicit stereotypes or the performance and participation gaps can have cascading influences. Change on one factor can produce change on the other and move the system toward a new homeostasis point.

Our findings suggest that a nation's average implicit stereotyping (and not explicit) is uniquely related to gender inequality in science and math achievement and, by extension, to other markers of a diverse scientific workforce such as interest, participation, and presence in scientific leadership. Experimental research has frequently demonstrated causal effects of implicit stereotypes on such inequalities, and suggests that observation of inequalities can influence stereotypes. Changing implicit stereotypes is not just a matter of influencing intentions; it also requires consideration of the social realities that shape minds without intention.

- Gonzales P, et al (2004) *Highlights from the Trends in International Mathematics and Science Study (TIMSS) 2003* (US Dept of Ed, Washington, DC) (NCES no. 2005-005).
- Martin MO, Mullis IVS, Chrostowski SJ (2004) *TIMSS 2003 Technical Report* (TIMSS and PIRLS Int Study Ctr, Boston College, Chestnut Hill, MA).
- Jacobs JE, Eccles JS (1985) Gender differences in math ability: The impact of media reports on parents. *Ed Reseachr* 14:20–25.
- Hyde JS, Lindberg SM, Linn MC, Ellis AB, Williams CC (2008) Gender similarities characterize math performance. *Science* 321:494–495.
- Guiso L, Monte F, Sapienza P, Zingales L (2008) Culture, gender and math. *Science* 320:1164–1165.
- Hyde JS, Linn MC (2006) Gender similarities in mathematics and science. *Science* 314:599–600.
- Lummis M, Stevenson HW (1990) Gender differences in beliefs and achievement: A cross-cultural study. *Dev Psychol* 26:254–263.
- Baker DP, Jones DP (1993) Creating gender equality: Cross-national gender stratification and mathematical performance. *Sociol Ed* 66:91–103.
- Spelke ES (2005) Sex differences in intrinsic aptitude for mathematics and science? *Am Psychol* 60:950–958.
- Halpern DF, et al (2007) The science of sex differences in science and mathematics. *Psychol Sci Pub Int* 8:1–51.
- Frome PM, Eccles JS (1998) Parents' influence on children's achievement-related perceptions. *J Pers Soc Psychol* 74:435–452.
- Furnham A, Reeves E, Budhani S (2002) Parents think their sons are brighter than their daughters: Sex differences in parental self-estimations and estimations of their children's multiple intelligences. *J Gen Psychol* 163:24–39.
- Faul F, Erdfelder E, Lang A-G, Buchner A (2007) G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Resch Methods* 39:175–191.
- Kiefer AK, Sekaquaptewa D (2007) Implicit stereotypes, gender identification, and math-related outcomes: A prospective study of female college students. *Psychol Sci* 18:13–18.
- Li Q (1999) Teachers' beliefs and gender differences in mathematics: A review. *Ed Rsch* 41:63–76.
- Meece J, Eccles-Parsons J, Kaczala C, Goff S, Futterman R (1982) Sex differences in math achievement: Toward a model of academic choice. *Psychol Bull* 91:324–348.
- Schmader T, Johns M, Barquissau M (2004) The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. *Sex Roles* 50:835–850.
- Hyde JS, Fennema E, Ryan M, Frost LA, Hopp C (1990) Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychol Women Quart* 14:299–324.
- Bell AE, Spencer SJ, Iserman E, Logel CER (2003) Stereotype threat and women's performance in engineering. *J Engin Ed* 92:307–312.
- Inzlicht M, Ben-Zeev T (2000) A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychol Sci* 11:365–371.
- Shih M, Pittinsky TL, Ambady N (1999) Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychol Sci* 10:80–83.
- Spencer SJ, Steele CM, Quinn DM (1999) Stereotype threat and women's math performance. *J Exp Soc Psychol* 35:4–28.
- Steele CM (1997) A threat in the air: How stereotypes shape intellectual identity and performance. *Am Psychol* 52:613–629.
- Steele CM, Spencer SJ, Aronson J (2002) in *Advances in Experimental Social Psychology*, ed Zanna M (Academic, San Diego), pp 379–440.
- Schmader T, Johns M (2003) Converging evidence that stereotype threat reduces working memory capacity. *J Pers Soc Psychol* 85:440–452.
- Ceci SJ, Williams MW (2007) *Why aren't more women in science? Top researchers debate the evidence*, (Amer Psych Assoc, Washington, DC).
- National Academy of Sciences (2007) *Beyond bias and barriers: Fulfilling the potential of women in academic science and engineering* (Nat'l Acad Press, Washington, DC).
- Shanks DR (1995) *The psychology of associative learning* (Cambridge Univ Press, Cambridge).
- De Houwer J, Beckers T (2002) A review of recent developments in research and theories on human contingency learning. *Quart J Exp Psychol* 55B:289–310.
- Eagly AH, Steffen VJ (1984) Gender stereotypes stem from the distribution of women and men into social roles. *J Pers Soc Psychol* 15:543–558.
- Hoffman C, Hurst N (1990) Gender stereotypes: Perception or rationalization? *J Pers Soc Psychol* 58:197–208.
- Jussim L (1991) Social perception and social reality: A reflection–construction model. *Psychol Rev* 98:54–73.
- Jussim L, Harber KD, Crawford JT, Cain TR, Cohen F (2005) Social reality makes the social mind: Self-fulfilling prophecy, stereotypes, bias, and accuracy. *Int Stud: Soc Behav Comm Biol Art Sys* 6:85–102.
- Murphy MC, Steele CM, Gross JJ (2007) Signaling threat: How situational cues affect women in math, science and engineering settings. *Psychol Sci* 18:879–885.
- Nisbett RE, Wilson TD (1977) Telling more than we can know: Verbal reports on mental processes. *Psychol Rev* 84:231–259.
- Greenwald AG, Banaji MR (1995) Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychol Rev* 102:4–27.
- Nosek BA, Greenwald AG, Banaji MR (2006) in *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes*, ed Bargh JA (Psychology Press, New York), pp 265–292.
- Nosek BA, et al (2007) Pervasiveness and correlates of implicit attitudes and stereotypes. *Euro Rev Soc Psychol* 18:36–88.
- Nosek BA, Greenwald AG, Banaji MR (2005) Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Pers Soc Psychol Bul* 31:166–180.
- Kiefer AK, Sekaquaptewa D (2007) Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *J Exp Soc Psychol* 43:825–832.
- Nosek BA, Banaji MR, Greenwald AG (2002) Math = male, me = female, therefore math ≠ me. *J Pers Soc Psychol* 83:44–59.
- Fiske AP, Kitayama S, Markus HR, Nisbett RE (1998) in *The Handbook of Social Psychology*, eds Gilbert DT, Fiske ST, Lindzey G (McGraw-Hill, Boston), pp 915–981.
- Martin MO, Mullis IVS, Gonzalez EJ, Chrostowski SJ (2004) *TIMSS 2003 International Science Report* (TIMSS and PIRLS Int Study Ctr, Boston College, Chestnut Hill, MA).
- Mullis IVS, Martin MO, Gonzalez EJ, Chrostowski SJ (2004) *TIMSS 2003 International Mathematics Report* (TIMSS and PIRLS Int Study Ctr, Boston College, Chestnut Hill, MA).
- Greenwald AG, McGhee DE, Schwarz JLK (1998) Measuring individual differences in implicit cognition: The Implicit Association Test. *J Pers Soc Psychol* 74:1464–1480.

Materials and Methods

Visitors to the Project Implicit website (<https://implicit.harvard.edu/>) could complete IATs about a variety of topics including measuring association strengths between gender (male, female) and academics (science, liberal arts). The participant sample at Project Implicit consists of unselected volunteers. IATs, and accompanying materials, were available in 17 languages during the time frame of data collection. Participants who selected the gender–science IAT completed the IAT, a short questionnaire measuring beliefs and attitudes about math and science, and a demographics questionnaire in a randomized order. The study required ≈10 min to complete. At the end, participants received a debriefing and information about their IAT performance and comparison data with other participants. Additional detail on materials, methods, and analysis are available in the [SI Appendix](#).

ACKNOWLEDGMENTS. This work was supported by the National Institute of Mental Health Grant R01 MH-68447 and the National Science Foundation grant REC-0634041.