What is the Implicit Gender-Science Stereotype?

Exploring Correlations between the Gender-Science IAT and Self-Report Measures

Hila Zitelny, Michal Shalom, and Yoav Bar-Anan

Ben-Gurion University of the Negev, Beer-Sheva, Israel

Word-count: 4985

Author Biographies:

Hila Zitelny completed her master's degree in clinical psychology at the Department of Psychology at Ben-Gurion University in the Negev. Her research focuses on gender and stereotypes.

Michal Shalom is a master's degree student in educational psychology at the Department of Psychology at Ben-Gurion University in the Negev. Her research focuses on emotion regulation and embodiment.

Yoav Bar-Anan is a senior lecturer at the Department of Psychology at Ben-Gurion University of the Negev. His research focuses on evaluative processes and self-knowledge.

Author's note: Correspondence should be addressed to: Hila Zitelny, Department of Psychology, Ben-Gurion University in the Negev Be'er Sheva, Israel. E-mail: zitelnyh@post.bgu.ac.il.

This project was supported by grants from the Israeli Science Foundation [779/16], from Project Implicit Inc. and from the United States – Israel Binational Science Foundation [2013214] to Y. B.-A

Abstract

Implicit measures of the gender-science stereotype are often better than explicit measures in predicting relevant outcomes. This finding could reflect a discrepancy between implicit and explicit stereotypes, but an alternative is that the implicit measure is sensitive to constructs other than the stereotype. Analyzing an archival dataset (total *N* = 478,550), we found that self-reported liking of science versus liberal arts was the best predictor of the gender-science Implicit Association Test (IAT). In a re-analysis of a previous study and a replication of another study, we found that evidence for the IAT's advantage over explicit stereotypes in predicting relevant outcomes disappeared when controlling for self-reported liking. Therefore, perhaps the IAT has often outperformed the explicit stereotype because the gender-science IAT captures personal attraction whereas the explicit stereotype does not. It is premature to conclude that implicit constructs are superior to explicit constructs in predicting science-related plans and behavior.

Abstract word-count: 148

Key words: implicit association test; implicit stereotypes; gender-science stereotype; gender; stereotypes

What is the implicit gender-science stereotype? Exploring correlations between the gender-science IAT and self-report measures

According to current theory and research, the gender-science stereotype has a central role in the underrepresentation of women in occupations related to science. The basic premise is that people expect women to be unskillful or uninterested in science, and those expectations influence judgment and behavior toward women (e.g., by teachers and parents, Gunderson, Ramirez, Levine, & Beilok 2012; by potential employers, Reuben, Sapienza, & Zingales, 2014), as well as women's self-concepts (e.g., math ability self-concept, Sáinz & Eccles, 2012) and aspirations (e.g., career intentions, Schmader, Johns, & Barquissau, 2004). Of special interest are implicit stereotypes – "social category associations that become activated without the perceiver's intention or awareness when […] presented with a category cue" (Blair, Ma, & Lenton, 2001, p. 828). Even people who do not endorse the stereotype explicitly might still have those mental associations from exposure to the prevalent beliefs that constitute the stereotype. Those associations might influence behavior and judgment automatically, without intention and awareness (Gawronski, Geschke, & Banse, 2003; Gawronski, Hofmann, & Wilbur, 2006; Greenwald & Banaji, 1995; Jost et al., 2009).

Previous studies have found that the implicit gender-science stereotype predicts judgment and behavior that contribute to women's underrepresentation in science-related activities and occupations. Appendix A includes a summary of all the studies that we found that measured implicit gender-science stereotypes and explicit gender-science stereotypes or other explicit beliefs. Those studies found that implicit stereotypes predicted participants' math engagement (Nosek & Smyth, 2011), performance and achievement (Ramsey & Sekaquaptewa, 2011), intentions to pursue science-related majors, academic programs (Lane, Goh, & Driver-Linn ,2012; Smyth, Greenwald, & Nosek, 2009), and career (Cundiff, Vescio, Loken, & Lo, 2013). These relations were usually moderated by gender. Among women, stronger implicit stereotypes predicted worse math performance and achievement, and weaker identification with math and science. Among men, the implicit stereotypes sometimes had no predictive value, and on other studies, stronger implicit stereotypes predicted better performance, achievements, and identification with math and science.

Seventeen of the studies listed in Appendix A compared the relation of a third outcome measure with implicit versus explicit stereotypes. Fifteen of those studies found that the implicit stereotype had a stronger relation with an outcome measure than the explicit stereotype. The superiority of the implicit stereotypes could reflect a unique role for automatic activation of stereotypes in judgment and behavior. Lane et al. (2012) argued that "sincere and conscious beliefs that men and women are equally well-suited for STEM fields do not preclude internalization of these beliefs at a less conscious level" (p. 222). Likewise, Muzzatti and Agnoli (2007) speculated that the implicit gender-science stereotype is present even when "participants are not aware of (or deny) the stereotype" (p. 758). It was further speculated that implicit stereotypes "shape choices by subtly constraining preferences without the individual's awareness or conscious exertion of choice" (Nosek, Banaji, & Greenwald, 2002, p. 50). Moreover, Nosek and Smyth (2011) argued that implicit stereotypes can shape certain outcomes (e.g., math engagement and achievement) through mechanisms that operate "under the surface".

Thus, it is common to interpret the advantage of implicit measures of the gender-science stereotype over their explicit counterparts as revealing the important role of automatic activation of stereotypes in those areas. In this article, we suggest that an alternative account is as likely. According to that alternative, the measure used so far for measuring the implicit gender-science stereotype is not the implicit counterpart of the common explicit stereotype measures. In addition to stereotypes, the implicit measure taps other constructs linked to science-related behavior and intentions. Those constructs, rather than implicit processes or constructs, might be the reason for the superiority of the implicit measure over the explicit stereotype measures in predicting important outcomes.

Implicit stereotypes are almost exclusively measured with indirect measures that are considered sensitive to mental associations, mainly the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). In all the studies that we found, the implicit gender-science stereotype was measured with the IAT (or an IAT variant) as the association between nouns representing science and non-science (*Math*/*Humanities*, *Science*/*Arts*, *Mathematics*/*Language*, *Math words*/*Reading words*, *Science*/*Liberal Arts*, *Math*/*English*, *Scientific*/*Humanistic*) and nouns representing gender (*Boys*/*Girls*, *Girls names*/*Boys names*, *Female*/*Male*, *Masculine*/*Feminine*, *Men*/*Woman*). It is not obvious that the implicit/explicit distinction is the only difference between such an IAT and a measure of the belief that, in comparison to women, men are better or are more interested in science. Many other beliefs could map into the gender/science associations. Based on that notion, the original goal of the present research was to test a simple hypothesis: self-reported associations would be more strongly related to the IAT than self-reported gender-science beliefs. Such a result would cast doubt on the common interpretation of previous findings that the IAT was better than self-reported beliefs in predicting important outcomes. Perhaps those results reflected a superiority of associations over beliefs, not the superiority of implicit constructs over explicit constructs.

In the present research, we analyzed a large sample of participants (N = 478,550) to test whether self-reported associations are related to the IAT more than self-reported beliefs pertaining to the gender-science stereotype. Although the analyses confirmed our hypothesis, they also found that self-reported liking of science was related to the IAT even stronger than self-reported associations. This finding suggests that the advantage of the gender-science IAT over explicit gender-science beliefs in predicting relevant outcomes might reflect only the advantage of personal attraction over gender-science beliefs in predicting those outcomes, not the advantage of implicit constructs. In the second part of the present investigation, we searched for evidence that the IAT has any advantage over self-reported stereotypes, after controlling for self-reported liking.

Method

Participants

Participants were volunteers who completed the gender-science IAT demonstration task in the Project Implicit website (implicit.harvard.edu; Nosek, 2005) between January 13th, 2003 and December 31st, 2013. We excluded participants who did not indicate their gender. We separated the dataset to 11 studies, one for each year, because the self-report measures changed over time (see Table 1 for details).

Table 1

Sample size, women rate, and mean age, in each year (study)

Year	*N*	% females	Mean age (*SD*)
2003	27,397	66.99	26.04 (10.43)
2004	36,966	69.15	25.69 (10.69)
2005	67,119	67.34	27.93 (11.81)
2006	41,311	64.86	27.54 (11.73)
2007	44,767	68.32	27.28 (11.54)
2008	37,348	67.89	26.76 (11.55)
2009	43,819	67.22	27.21 (11.85)
2010	39,197	69.05	26.08 (11.09)
2011	40,789	69.61	26.02 (10.86)
2012	51,295	67.41	26.70 (11.28)
2013	48,542	67.05	27.60 (11.66)

Note. Because many of the self-report measures changed on December 7th, 2006, we added the sessions that followed that date (until the end of 2006) to the 2007 sample.

Measures

Complete information about the dataset, methods, and measures is available online, at osf.io/f7jzb.

Implicit Association Test. The categories were *Male* (items: *Man*, *Boy*, *Father*, *Male*, *Grandpa, Husband*, *Son, Uncle*), *Female* (*Girl, Female*, *Aunt, Daughter, Wife*, *Woman, Mother*, *Grandma*), *Scienc*e (*Biology*, *Physics, Chemistry*, *Astronomy, Engineering, Neuroscience, Biochemistry;* the last two were replaced with *Math* and *Geology* at 2007), and *Liberal Arts* (*Philosophy*, *Humanities*, *Arts*, *English*, *Music*, *History, Latin; Latin* was replaced with *Literature* at 2007). The IAT consisted of seven trial blocks, and was scored with the D1 algorithm (Greenwald, Nosek, & Banaji, 2003). Positive scores indicated faster performance when words related to males and science shared the same key than when words related to females and science shared the same key.

Self-Report Measures. Participants answered direct questions related to their own attitudes about science and liberal arts, and about their beliefs regarding gender differences in those subjects. We analyzed only questions relevant to the present investigation.

Self-reported associations. Participants reported how much they associated science with males versus with females, and how much they associated liberal arts with males versus females. The response scales changed over the years, but always ranged from *strongly female* to *strongly male*. The self-reported association score was the difference between these two items, larger numbers indicating stronger association of science with males and liberal arts with females.

Beliefs about natural ability. In 2003-2006, participants reported their level of agreement with the statement "Males perform better than females in science because of greater natural ability" on a 7-point scale. In 2007-2013, participants rated factors explaining why "Women hold a smaller portion of the science and engineering faculty positions at top research universities than do men". One factor pertained to ability: "Different proportions of men and women are found among people with the very highest levels of math ability". Participants rated how important that factor was in explaining this frequency difference, on a 5-point scale.

Beliefs about natural interest. In 2007-2013 participants rated, on a 5-point scale, the importance of the factor "On average, men and women differ naturally in their scientific interest" in explaining the abovementioned frequency difference.

Beliefs about prevalence. In years 2007-2013, participants estimated how many out of ten men at U.S. universities graduate with a scientific major, and answered the same question about women. The difference between the two responses was the prevalence score.

Personal Liking. Participants reported, on a 5-point scale, how much they like science and how much they like liberal arts. We computed a preference for the topic stereotypically associated with the participant's gender.

Personal Importance. In years 2007-2013, participants rated, on a 5-point scale, how important it was for them to become knowledgeable in science, math, and liberal arts. We averaged the importance of science and math together, and computed a difference score indicating preference of becoming knowledgeable in the topic stereotypically associated with the participant's gender.

Results

The scores were stable over the years (see Appendix B), with the IAT showing a positive score (M_{min} = 0.34, M_{max} = 0.38). Figure 1 shows highly consistent rank order of the correlations of the IAT with the different self-report measures. In all years, the IAT was more strongly related to self-reported associations (r_{min} = .198, r_{max} = .218, minimum and maximum values are from the eleven correlations computed for the eleven samples) than reported beliefs about natural differences between the genders in math ability (r_{min} = .035, r_{max} = .119), in interest in science (r_{min} = .056, r_{max} = .087), and in estimated prevalence of students who major in science (r_{min} = .137, r_{max} = .161). These results were replicated among women and among men (Figures 2a and 2b).

Unexpectedly, two sets of questions were related to the IAT more strongly than self-reported associations (Figure 1 and Table 2). These were self-reported liking (r_{min} = .217, r_{max} = .290), and self-reported importance (r_{min} = .228, r_{max} = .246). These relations indicated that stronger men/science and women/liberal arts associations predicted stronger preference for science among men and stronger preference for liberal arts among women. Self-reported liking had the strongest relations to the IAT, and as Figures 2a and 2b show, this superiority was more pronounced among women than among men (even among men, personal liking had the strongest correlation with the IAT in 10 of the 11 studies). Self-reported importance and self-reported liking were strongly related (r_{min} = .607, r_{max} = .621). In all the years, self-reported liking and importance were related to the IAT significantly more than to self-reported beliefs about natural differences between the genders in math ability (liking: r_{max} = .135; importance: r_{max} = .053), in interest in science (liking: r_{max} = .111; importance: r_{max} = .080), and in estimated prevalence of students in science majors (liking: r_{max} = .094 ; importance: r_{max} = .080).

We also used multiple regression analyses to predict the IAT score, in each year, from self-reported associations, ability stereotype beliefs, interest stereotype beliefs, prevalence stereotype beliefs, and personal liking. In all years, reported liking shared the largest unique variance with the IAT, and reported associations was always the second-best predictor (Figures 3a and 3b show separate results for men and women). The consistency of the ranking of predictors attests for their statistical reliability. The chances of one predictor being stronger than another predictor in 11 studies, when there is actually no difference between the two, is p = .0009765625 ($2*(1/2^{11})$).

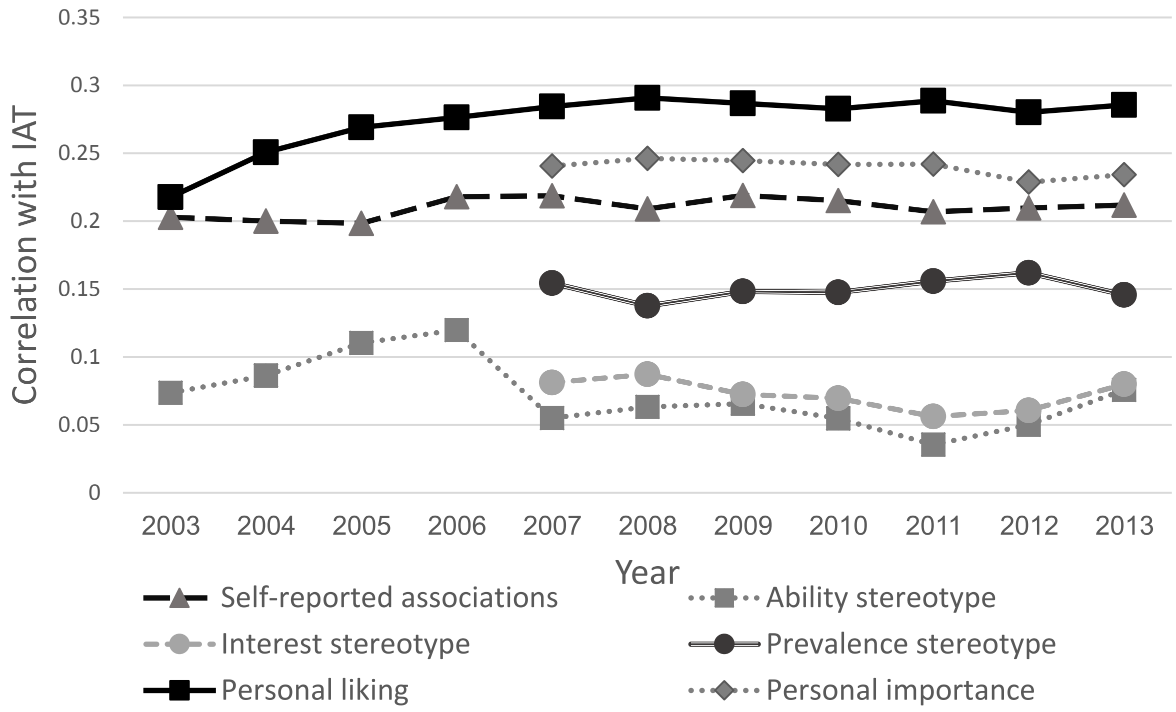

Figure 1. Correlation of the IAT with the six relevant self-report measures, by year.

Table 2

Correlation of the IAT with the six relevant self-report measures, by year

Year	Personal liking	Personal importance	Self-reported associations	Ability stereotype	Interest stereotype	Prevalence stereotype
2003	$.21765_a$		$.20283_a$	$.07354_b$		
2004	$.25073_a$		$.20005_b$	$.08623_c$		
2005	$.26900_a$		$.19832_b$	$.11022_c$		
2006	$.27639_a$		$.21802_b$	$.11975_c$		
2007	$.28430_a$	$.24061_b$	$.21866_c$	$.05501_f$	$.08109_e$	$.15434_d$
2008	$.29079_a$	$.24617_b$	$.20902_c$	$.06306_f$	$.08717_e$	$.13749_d$
2009	$.28667_a$	$.24463_b$	$.21888_c$	$.06548_e$	$.07228_e$	$.14826_d$
2010	$.28277_a$	$.24171_b$	$.21514_c$	$.05463_e$	$.06962_e$	$.14753_d$
2011	$.28857_a$	$.24185_b$	$.20688_c$	$.03530_f$	$.05604_e$	$.15585_d$
2012	$.28001_a$	$.22868_b$	$.20960_c$	$.05009_e$	$.06069_e$	$.16194_d$
2013	$.28556_a$	$.23429_b$	$.21186_c$	$.07563_e$	$.07994_e$	$.14568_d$

Notes. On each row, different subscripts indicate significant difference ($p < .05$).

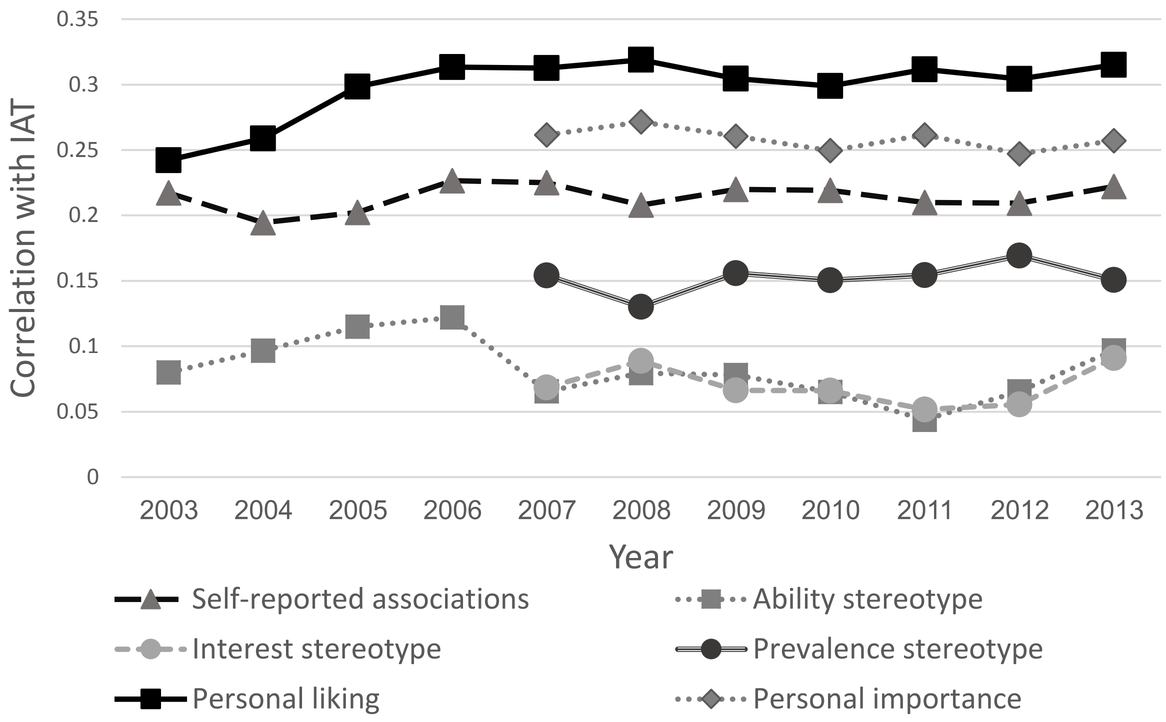

Figure 2a. Female participants: Correlations of the IAT with the six relevant self-report measures, by year.

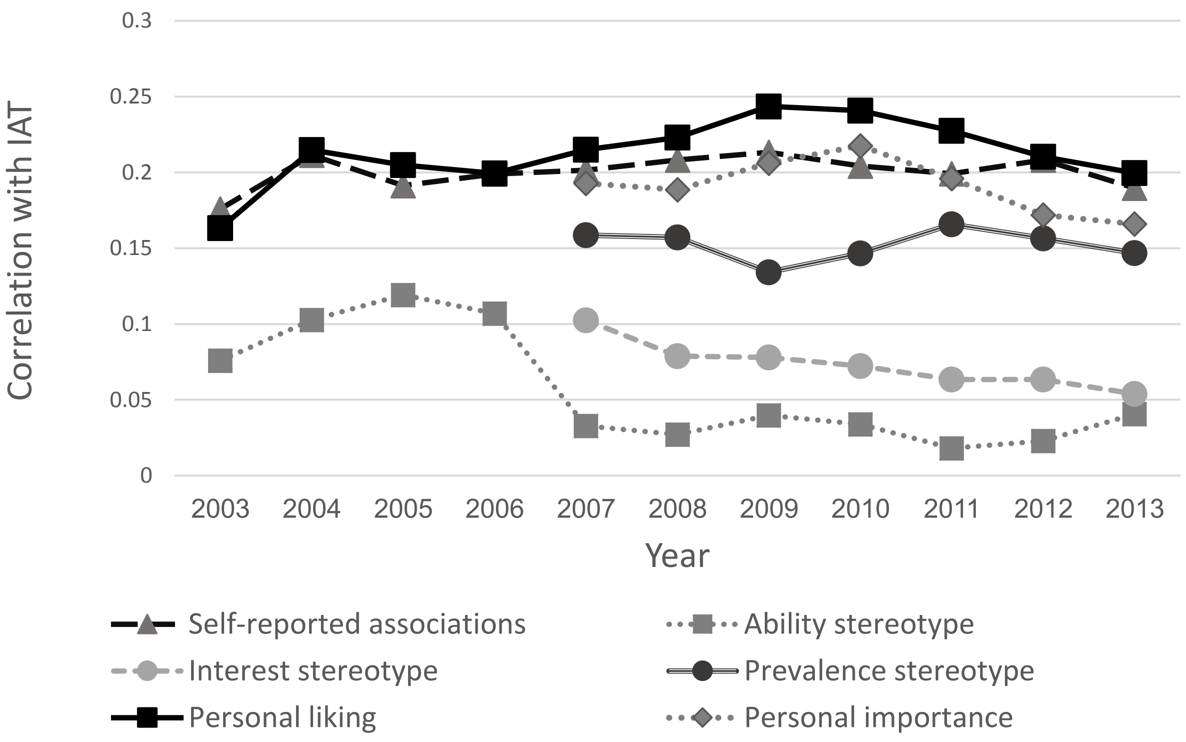

Figure 2b. Male participants: Correlations of the IAT with the six relevant self-report measures, by year.

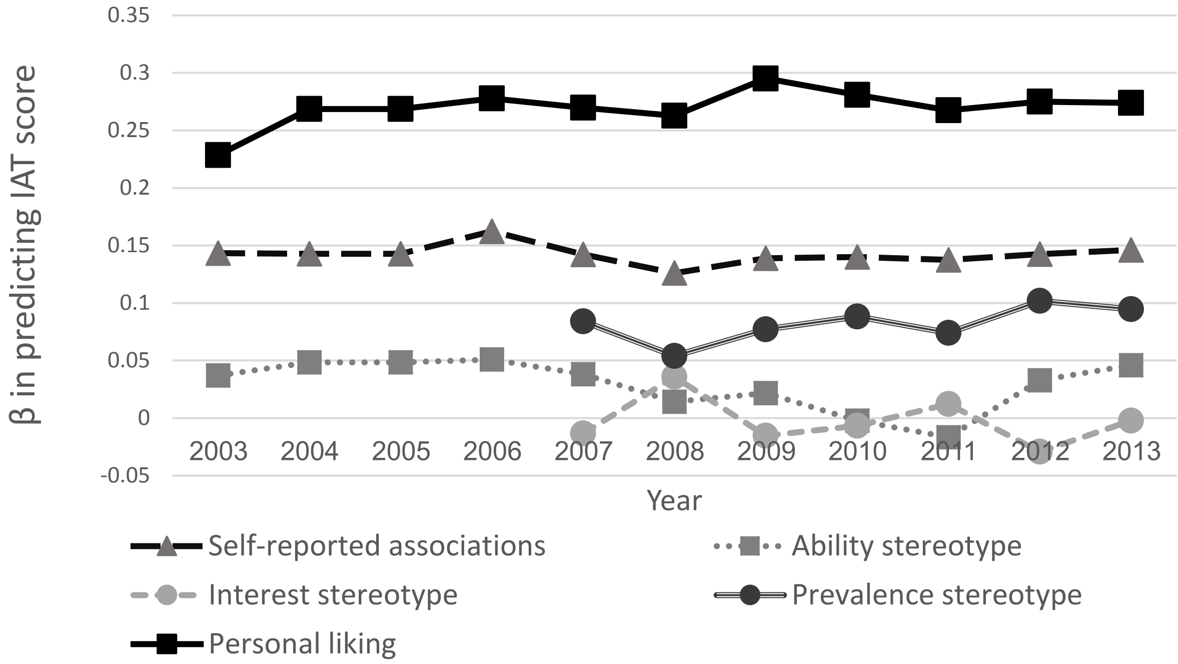

Figure 3a. Female participants: unique variance of the five relevant self-report measures in predicting the IAT score, by year.

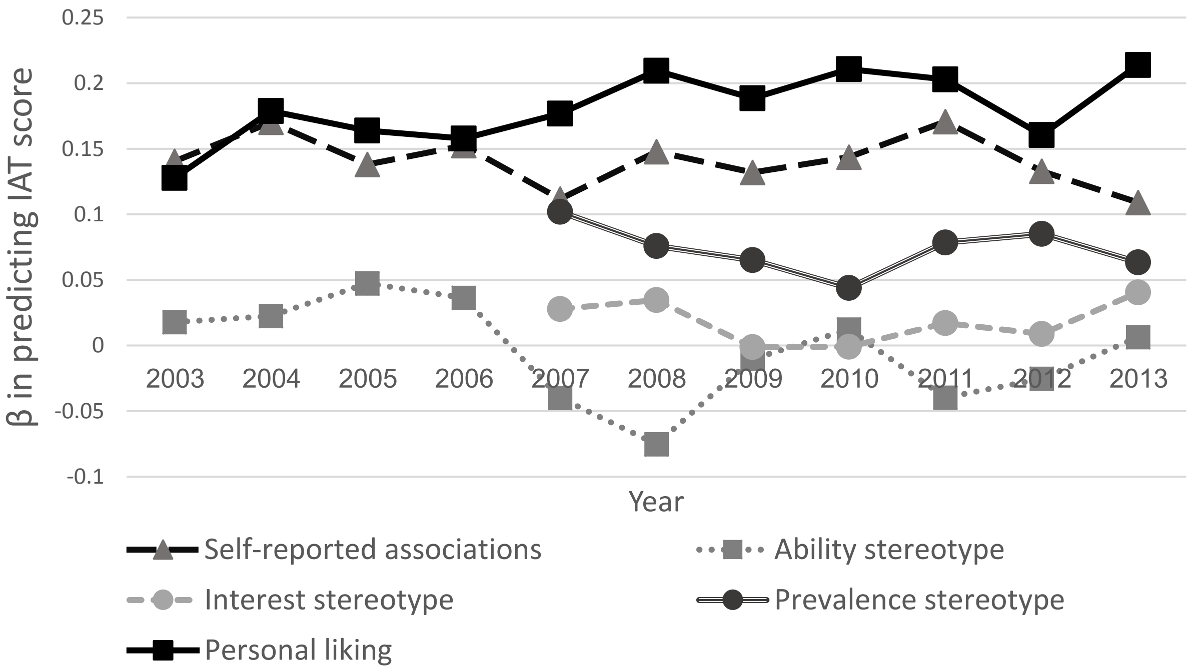

Figure 3b. Male participants: unique variance of the five relevant self-report measures in predicting the IAT score, by year.

Part 2: Can Liking Explain the IAT's Advantage?

The results so far show that direct reports about mental associations were related to the IAT more than self-reported beliefs about stereotypes. Unexpectedly, the IAT and self-reported preference for the topic stereotypically associated with the participant's gender were related to each other more than each of these measures was related to other self-reported beliefs and associations. Therefore, perhaps previous findings that the IAT is better than self-report measures in predicting important outcomes is due to the IAT's relation to personal liking, which, in turn, is a better predictor of science-related outcomes than gender-science beliefs. To refute that possibility, we returned to previous research that found an advantage of the IAT, and tested whether the evidence for this advantage persists even when controlling for self-reported liking.

Predicting Math Performance

Nosek and Smyth (2011) compared the gender-science IAT and the explicit stereotype (measured with two self-reported items: *Men are better at math than women are* and *Women can achieve as much as men in math*) in predicting math-related outcomes. We re-analyzed that data and found only one variable, the difference between math and verbal SAT scores, that had reliably stronger relation with the IAT (r = .193, p < .001) than the explicit stereotype (r = .090, p = .004), Williams' t(1040) = 2.590, p = .010.

In Nosek and Smyth's study, participants rated the warmth of their feelings toward math and a contrast category, and reported a preference between the two. With those measures, we computed a preference for the topic stereotypically associated with the participant's gender over the other topic. Consistent with our findings, the IAT/attitudes relation (r = .281, p < .001) was stronger than the IAT/explicit stereotype relation (r = .136, p < .001), and the explicit stereotype/attitudes relation (r = .145, p < .001), Williams' ts(2918) = 6.199, 5.790, respectively, ps < .001.

We used PROCESS macro for simple mediation (Model 4) for SAS (Hayes, 2013) to find unstandardized estimates with 95% confidence intervals (CI) of the reduction in the effect of each stereotype measure on the SAT difference, due to controlling for attitudes. We entered the IAT as the independent variable, attitudes as a mediator, SAT as the outcome, and explicit stereotypes as a covariate. We replaced the roles of the IAT and the explicit stereotypes when testing the explicit stereotype. In the present context, rather than mediation effects, this analysis tested whether the relation between each stereotype measure and the SAT was significantly reduced when controlling for attitudes. Bootstrap tests with 10,000 resamples showed a significant reduction in the IAT's effect, $b = .137$, $SE = .017$, 95% CI [0.104, 0.171], and in the explicit stereotype's effect, $b = .052$, $SE = .015$, 95% CI [0.023, 0.081], when controlling for attitudes. The regression analysis provided by the PROCESS macro showed that after the reduction due to controlling for attitudes, the IAT's effect, $b = .044$, SE = .028, $t(1040) = 1.55$, $p = .121$, 95% CI [-0.011, 0.100], and the explicit stereotype's effect, $b = .009$, SE = .027, $t(1040) < 1$, $p = .731$, 95% CI [-0.044, 0.063], were no longer significant. Importantly, when we computed partial correlations between each measure and the SAT, partialling out shared variance with attitudes, there was no longer reliable evidence for an IAT advantage: The IAT/SAT relation (r =.049, $p = .115$) was not significantly better than the explicit stereotype/SAT relation ($r = .017$, $p = .584$), Williams' $t(1040) < 1$, $p = .442$.

Predicting Plans to Pursue Science

Lane et al. (2012) found that the gender-science IAT predicted students' plans to pursue science versus humanities ($r = .34$, $p < .0001$) better than the self-reported stereotype ($r = .12$, *ns*), Williams' $t(150) = 2.158$, $p = .03$. We repeated that study with similar materials and procedure, adding attitude and importance measures, measured and scored identically to our main present study (full details about the replication are in Appendix C, and at osf.io/vc68r). We repeated the same analysis strategy as before. Replicating Lane et al., we found significant advantage for the IAT (r = .245, p < .001) over explicit stereotypes (r = .111, p = .012), Williams' t(511) = 2.304, p = .022, in predicting intentions to pursue the topic stereotypically associated with the participant's gender. The bootstrap tests in the mediation-like analyses found that controlling for attitudes significantly reduced the IAT's effect, b = .168, SE = .028, 95% CI [0.114, 0.224], and the explicit stereotype's effect, b = .080, SE = .030, 95% CI [0.019, 0.138]. The regression analyses showed that although the IAT's effect was reduced, it remained significant, b = .067, SE = .032, t(511) = 2.132, p = .034, 95% CI [0.005, 0.130], suggesting that attitudes might not be the only reason for the IAT/pursuit relations. The explicit stereotype's effect was reduced to being non-significant, b = .010, SE = .031, t(511) < 1, p = .755, 95% CI [-0.051, 0.070]. Importantly, we did not find evidence that the IAT maintained its advantage over the explicit stereotypes after controlling for attitudes. When we partialled out shared variance with attitudes, the IAT/pursuit relation (r = .095, p = .032) was no longer reliably stronger than the explicit stereotype/pursuit relation (r = .019, p = .662), Williams' t(511) = 1.251, p = .212.

General Discussion

Research about gender-science stereotypes has often found that implicit measures of the gender-science stereotype are better than explicit measures in predicting performance, motivation, intentions, self-concept, and decision making related to math and science. It is common to interpret such findings as revealing the important role that automatic activation of stereotypes plays in those areas. In this article, we challenge that interpretation. Had previous research used an IAT with the concepts *science/liberal arts*, *pleasant/unpleasant* to measure of implicit gender-science stereotype, many would have doubted a claim that discrepancies between implicit and explicit gender-science stereotype reflect discrepancies between implicit and explicit constructs or processes. It would not seem that the only difference between the implicit measure and self-reported gender-science beliefs is their sensitivity to automatic versus deliberate processes. We suggest that this threat also applies to the actual measure that has been used so far to assess the implicit gender-science stereotype, an IAT with the concepts *science*/*liberal arts*, *male/female.* Perhaps that IAT taps into different constructs than those tapped by the explicit measures used in research on the gender-science stereotype.

We suspected that previous findings about discrepancies between the implicit and the explicit gender-science stereotype might have reflected discrepancies between associations and beliefs, rather than between implicit and explicit constructs. Indeed, we found that people's direct report on their mental associations between gender and science had a stronger correlation with the IAT than any self-reported belief. That finding favors previous research that measured self-reported associations (e.g., Nosek et al., 2009) over research that measured only beliefs (e.g., Ramsey & Sekaquaptewa, 2011) as an investigation of implicit/explicit discrepancies rather than associations/beliefs discrepancies.

Unexpectedly, our research also found that the IAT's strongest relation was not with self-reported associations but with self-reported personal liking of science in comparison to liberal arts. Importantly, that self-report measure was related to the IAT more than to self-reported stereotypic beliefs. Thus, whereas the explicit gender-science stereotype has very little to do with people's self-reported liking of science, the gender-science IAT is related to self-reported liking more than to any other belief.

A cross-study overview of previous research (see Appendix A) finds evidence compatible with our present findings. First, across a variety of direct measures (not including reported associations), previous research found weak relations of explicit gender-science stereotype with the IAT score (r_{min} = .01, r_{max} =.191). Such implicit/explicit correlations are weaker than what is usually found between implicit and explicit measures of attitudes (Bar-

Anan & Nosek, 2014; Nosek, 2005) and stereotypes (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005). Further, across studies, among woman participants, these implicit/explicit relations were weaker than the relations observed between the IAT and self-reported attitudes towards science (r_{min} = .15, r_{max} = .35), and self-reported identification with science (r_{min} = .17, r_{max} = .36). Among men, previous results were less conclusive than our present findings (IAT/liking: r_{min} = .01, r_{max} =.35; IAT/identity: r_{min} = .01, r_{max} = .24).

Our results, mostly supported by a cross-study overview of previous research, are compatible with the possibility that the gender-science IAT and self-reported gender-science stereotypes (or associations) are different not only in automaticity/controllability of the processes that influence them or the implicitness/explicitness of the constructs that they reflect. These measures are also different in the specific beliefs or attitudes that they capture. Whereas the explicit measure captures people's beliefs about gender and science, the IAT is also related to personal attraction to science (versus non-science topics). Therefore, a difference between specific beliefs and attitudes, rather than a difference in controllability or implicitness, might explain previous findings that implicit gender-science stereotypes are better than explicit gender-science stereotypes in predicting important outcome variables.

An overview of previous research finds that, indeed, most of the outcomes that were predicted better by the gender-science IAT than by the explicit gender-science measures are linked to liking math and science. Among those outcomes were plans and intentions to pursue science (Cundiff et al., 2013; Lane et al., 2012), science aspirations (Lane et al., 2012; Phelan, 2010), choice of major (Smyth et al., 2009), math engagement and achievements (Nosek et el., 2002; Nosek & Smyth, 2011), math performance, and the desire to pursue math-related careers (Kiefer & Sekaquaptewa, 2007b; Ramsey & Sekaquaptewa, 2011), math self-perceived ability and math participation (Nosek & Smyth, 2011), and sensitivity to stereotype threat (Galdi, Cadinu, & Tomasetto, 2013; Kiefer & Sekaquaptewa, 2007a).

People who like science are more likely to perceive science abilities as important, plan to pursue science, choose a science major, engage in a related activity, and reach more successful achievements in that activity. Regarding sensitivity to stereotype threat, women with lower grades in math and those who perceive math-related abilities as relatively unimportant are affected to a lesser extent by stereotype threat (Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003; Steinberg, Okun, & Aiken, 2012). Therefore, liking math should predict sensitivity to stereotype threat.

The alternative account that we consider here does not argue that the gender-science IAT is not a measure of automatic processes or implicit constructs. Our argument pertains only to the reason for the IAT's superiority over explicit stereotype measures in predicting science-related outcomes. We argue that the predictive advantage that previous research found for the IAT might reflect a stronger relation of the outcome variable with personal attraction to science than with gender-science beliefs, rather than a stronger relation of the outcome variable with implicit than with explicit constructs or processes.

One course of action to refute the argument proposed in the present article is to show that the IAT is a superior predictor of science-related behavior and cognition even when controlling for self-reported liking of science. Following that logic, we re-analyzed data from one previous study (Nosek & Smyth, 2011) and replicated another (Lane et al., 2012) to examine what happens to the advantage of the gender-science IAT over self-report measures in predicting a science-relevant outcome, when attitudes are added to the model. We found that shared variance with attitudes explains much of the variance the IAT shared with the outcome measure (when we controlled for attitudes, the IAT's effect decreased significantly). We also found that the IAT's advantage over explicit stereotypes was no longer significant when controlling for attitudes. Unfortunately, we did not find a statistical method to test whether the IAT's advantage over the explicit measure was significantly reduced when attitudes were added to the model[1]. Therefore, our findings only failed to refute the alternative account we proposed here, rather than provide more empirical support for that account.

Why would an IAT with the nouns *male/female* and *science/liberal arts* as category names capture one's attitudes toward science and liberal arts? Perhaps women tend to map male/female to not-me/me (and men show the opposite mapping). The self-concept IAT and the attitude IAT are strongly related (e.g., Nosek et al., 2002, r = .58; Nosek & Smyth, 2011: r among women = .53, r among men = .39). Therefore, perhaps IATs with gender categories are related to people's self-concepts more than to people's beliefs about the genders. In turn, self-concepts are strongly related to attitudes (e.g., Nosek & Smyth, 2011: r among women = .84, r among men = .88; Young, Rudman, & Buettner, 2013: r among women = .53, r among men = .54). For that reason, the IAT used so far to measure the gender-science stereotype was sensitive to people's attitudes toward science more than to beliefs about gender differences. Compatible with that hypothesis are previous findings that on the IAT, people tend to show an association between their gender and favorable concepts (Rudman, Greenwald, & McGhee, 2001).

Limitations and Future Research

The most obvious challenge to our alternative account is the possibility that our findings are just another example for the superiority of the implicit over the explicit gender-science stereotypes in predicting important psychological variables related to math and science (in this case, science-related attitudes). Perhaps the predictive power of the IAT diminishes when controlling for self-reported liking due to shared variance between three *distinct* constructs: implicit stereotypes, self-reported liking, and the predicted outcome. That shared variance could reflect various causal relations, and some of them would suggest an

[1] See our exploration of this statistical challenge at osf.io/sazk4.

important role for implicit constructs and processes. For instance, perhaps the automatic activation of the gender-science stereotype affects attraction to science, which further influences aspirations and skills in science. Indeed, we have not ruled out the possibility that an implicit construct is responsible for the IAT's superiority documented in previous studies, and for our present findings. What we have done is to propose an alternative account for the IAT's superiority in the gender-science domain that is as likely as the common account. The only argument in support of the common account is that in other domains, there is good evidence that the IAT reflects implicit constructs. That is not sufficient evidence that implicit constructs are responsible for the IAT's superiority over explicit stereotypes in predicting important science-related outcomes.

To investigate what contributes to the IAT's superiority in the gender-science domain, we recommend three future directions. First, as we have done in the present re-analysis and replication, future research on the relation between the implicit gender-science stereotype and relevant outcomes should control for participants' attitudes toward science (versus a non-science concept). Unique variance between the IAT and the target outcome measure, not shared with any of the self-report measures, could help establish the gender-science IAT as a measure of a psychological construct that has an important role in judgment and behavior related to people's pursuit of math and science.

Second, it is necessary to test whether the gender-science IAT predicts automatic behavior and judgment related to gender and science. For instance, research should test whether the gender-science IAT is a better predictor of the choice to pursue science when people choose under conditions that reduce controllability (e.g., time pressure and cognitive load) than under conditions that allow control. It is also important to test whether the gender-science IAT is related to people's feelings about stereotypical gender-science beliefs more than to their cognitions on those beliefs. Such evidence has helped to establish the IAT as a measure of automatic evaluation and to document the unique role that automatic evaluation plays (e.g., Friese, Hofmann, & Wänke, 2008; Gawronski & LeBel, 2008; Hofmann, Rauch, & Gawronski, 2007).

Third, and perhaps most importantly, so far, the research we reviewed and our present research all used correlational designs. Research on the role of implicit processes in science-related behaviors and goals is doomed to remain limited without experimental studies. It would be important to test whether a direct manipulation of mental associations between gender and science affects important outcomes related to math and science. Such an effect could increase the confidence that the predictive advantage of implicit over explicit measures reflects a causal link, and is not only due to the fact that the outcome variables and the IAT are both sensitive to variance in personal attraction to science. Similarly, research that would manipulate personal attraction to science and find changes in the gender-science IAT might support the alternative account proposed here.

Summary

The present research found that whereas the gender-science IAT is hardly related to explicit beliefs about gender and science, it is related to personal attitudes and goals pertaining to science. This finding points to the possibility that the IAT's advantage over explicit measures of the gender-science stereotype is not only due to the automaticity versus controllability of the processes that influence each measure or to the implicitness versus explicitness of the constructs captured by each measure. Rather, perhaps it is due to discrepancy in the explicit beliefs and attitudes captured by each measure. The present findings emphasize that much evidence is still missing for understanding the theoretical implications of previous findings about implicit gender-science stereotypes. In order to examine the unique role of implicit gender-science stereotypes, one must measure not only explicit stereotypes, but also self-reported associations and self-reported attraction to math and science. Further, research must examine whether the gender-science IAT predicts *automatic* processes that influence science-related behavior and judgment. Finally, it is essential to conduct experiments that directly manipulate the automatic gender-science stereotype and examine its effect on relevant behavior and judgment.

References

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior research methods*, *46*, 668-688. doi: 10.3758/s13428-013-0410-6

Betz, D. E. (2013). *Feminine stem role models: Attempts to improve women's motivation in science, technology engineering, and mathematics fields by countering the unfeminine-stem stereotype* (Doctoral dissertation). Retrieved from hdl.handle.net/2027.42/99856

Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: the moderation of implicit stereotypes through mental imagery. *Journal of personality and social psychology*, *81*, 828-841. doi: I0.1037//0022-3514.81.5.828

Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology*, *33*, 267-285. doi: 10.1002/ejsp.145

Cundiff, J. L., Vescio, T. K., Loken, E., & Lo, L. (2013). Do gender–science stereotypes predict science identification and science career aspirations among undergraduate science majors?. *Social Psychology of Education*, *16*, 541-554. doi: 10.1007/s11218-013-9232-8

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child development*, *82*, 766-779. doi: 10.1111/j.1467-8624.2010.01529.x

Cvencek, D., Meltzoff, A. N., & Kapur, M. (2014). Cognitive consistency and math–gender stereotypes in Singaporean children. *Journal of experimental child psychology*, *117*, 73-91. doi: 10.1016/j.jecp.2013.07.018

Friese, M., Hofmann, W., & Wänke, M. (2008). When impulses take over: Moderated predictive validity of explicit and implicit attitude measures in predicting food choice and consumption behaviour. *British Journal of Social Psychology*, *47*, 397-419. doi: 10.1348/014466607X241540

Galdi, S., Cadinu, M., & Tomasetto, C. (2014). The roots of stereotype threat: when automatic associations disrupt girls' math performance. *Child development*, *85*, 250-263. doi: 10.1111/cdev.12128

Gawronski, B., Geschke, D., & Banse, R. (2003). Implicit bias in impression formation: Associations influence the construal of individuating information. *European Journal of Social Psychology*, *33*, 573-589. doi: 10.1002/ejsp.166

Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are “implicit” attitudes unconscious?. *Consciousness and cognition*, *15*, 485-499. doi: 10.1016/j.concog.2005.11.007

Gawronski, B., & LeBel, E. P. (2008). Understanding patterns of attitude change: When implicit measures show change, but explicit measures do not. *Journal of Experimental Social Psychology*, *44*, 1355-1361. doi: 10.1016/j.jesp.2008.04.005

Gilbert, P. N., O’Brien, L. T., Garcia, D. M., & Marx, D. M. (2015). Not the Sum of Its Parts: Decomposing Implicit Academic Stereotypes to Understand Sense of Fit in Math and English. *Sex Roles*, *72*, 25-39. doi: 10.1007/s11199-014-0428-y

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological review*, *102*, 4-27. doi: 10.1037//0033-295X.102.1.4

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, *74*, 1464-1480. doi: 10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of personality and social psychology*, *85*, 197-216. doi: 10.1037/0022-3514.85.2.197

Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles*, *66*, 153-166. doi: 10.1007/s11199-011-9996-2

Hayes, A. F. (2013*). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*, 1369-1385.
doi: 10.1177/0146167205275613

Hofmann, W., Rauch, W., & Gawronski, B. (2007). And deplete us not into temptation: Automatic attitudes, dietary restraint, and self-regulatory resources as determinants of eating behavior. *Journal of Experimental Social Psychology*, *43*, 497-504.
doi: 10.1016/j.jesp.2006.05.004

Jost, J. T., Rudman, L. A., Blair, I. V., Carney, D. R., Dasgupta, N., Glaser, J., & Hardin, C. D. (2009). The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Research in organizational behavior*, *29*, 39-69.
doi: 10.1016/j.riob.2009.10.001

Kiefer, A. K., & Sekaquaptewa, D. (2007a). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology*, *43*, 825-832. doi: 10.1016/j.jesp.2006.08.004

Kiefer, A. K., & Sekaquaptewa, D. (2007b). Implicit stereotypes, gender identification, and math-related outcomes a prospective study of female college students. *Psychological Science*, *18*, 13-18. doi: 10.1111/j.1467-9280.2007.01841.x

Lane, K. A., Goh, J. X., & Driver-Linn, E. (2012). Implicit science stereotypes mediate the relationship between gender and academic participation. *Sex Roles*, *66*, 220-234. doi: 10.1007/s11199-011-0036-z

Liu, M., Hu, W., Jiannong, S., & Adey, P. (2010). Gender stereotyping and affective attitudes towards science in Chinese secondary school students. *International Journal of Science Education*, *32*, 379-395. doi: 10.1080/09500690802595847

Miller, D. I., Eagly, A. H., & Linn, M. C. (2015). Women's representation in science predicts national gender-science stereotypes: Evidence from 66 nations. *Journal of Educational Psychology*, *107*, 631-644. doi: 10.1037/edu0000005

Muzzatti, B., & Agnoli, F. (2007). Gender and mathematics: Attitudes and stereotype threat susceptibility in Italian children. *Developmental Psychology*, *43*, 747-759. doi: 10.1037/0012-1649.43.3.747

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, *134*, 565-584. doi: 10.1037/0096-3445.134.4.565

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social cognition*, *19*, 625-666. doi: 10.1521/soco.19.6.625.20886

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math= male, me= female, therefore math≠ me. *Journal of personality and social psychology*, *83*, 44-59. doi: 10.1037//0022-3514.83.1.44

Nosek, B. A., & Smyth, F. L. (2011). Implicit social cognitions predict sex differences in math engagement and achievement. *American Educational Research Journal*, *48*, 1125-1156. doi: 10.3102/0002831211410683

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., ... & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18*, 36-88. doi: 10.1080/10463280701489053

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., ... & Greenwald, A. G. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, *106*, 10593-10597. doi: 10.1073/pnas.0809921106

Park, L. E., Cook, K. E., & Greenwald, A. G. (2001). Implicit indicators of women's persistence in math, science, and engineering. *Psi Chi Journal of Undergraduate Research*, *6*, 145-152.

Passolunghi, M. C., Ferreira, T. I. R., & Tomasetto, C. (2014). Math–gender stereotypes and math-related beliefs in childhood and early adolescence. *Learning and Individual Differences*, *34*, 70-76. doi: 10.1016/j.lindif.2014.05.005

Phelan, J. E. (2010). *Increasing women's aspirations and achievement in science: the effect of role models on implicit cognitions* (Doctoral dissertation). Retrieved from rucore.libraries.rutgers.edu/rutgers-lib/27366

Ramsey, L. R., & Sekaquaptewa, D. (2011). Changing stereotypes, changing grades: A longitudinal study of stereotyping during a college math course. *Social Psychology of Education, 14,* 377-387. doi: 10.1007/s11218-010-9150-y

Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences, 111*, 4403-4408. doi: 10.1073/pnas.1314788111

Rudman, L. A., Greenwald, A. G., & McGhee, D. E. (2001). Implicit self-concept and evaluative implicit gender stereotypes: Self and ingroup share desirable traits. *Personality and Social Psychology Bulletin, 27,* 1164-1178. doi: 10.1177/0146167201279009

Sáinz, M., & Eccles, J. (2012). Self-concept of computer and math ability: Gender implications across time and within ICT studies. *Journal of Vocational Behavior, 80*, 486-499. doi: 10.1016/j.jvb.2011.08.005

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. doi: 10.1023/b:sers.0000029101.74557.a0

Smyth, F. L., Greenwald, A. G., & Nosek, B. A. (2009). *Implicit gender-science stereotype outperforms math scholastic aptitude in identifying science majors*. Unpublished manuscript.

Smyth, F. L., & Nosek, B. A. (2015). On the gender–science stereotypes held by scientists: explicit accord with gender-ratios, implicit accord with scientific identity. *Frontiers in psychology*, *6,* 1-19. doi: 10.3389/fpsyg.2015.00415

Steffens, M. C., Jelenec, P., & Noack, P. (2010). On the leaky math pipeline: Comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *Journal of Educational Psychology*, *102*, 947-963. doi: 10.1037/a0019920

Steinberg, J. R., Okun, M. A., & Aiken, L. S. (2012). Calculus GPA and math identification as moderators of stereotype threat in highly persistent women. *Basic and Applied Social Psychology*, *34,* 534-543. doi: 10.1080/01973533.2012.727319

Young, D. M., Rudman, L. A., Buettner, H. M., & McLean, M. C. (2013). The influence of female role models on women's implicit science cognitions. *Psychology of Women Quarterly*, *37*, 283-292. doi: 10.1177/0361684313482109

Appendix A

Overview of Previous Research

Table A1 includes studies that used an implicit measure of the gender-science stereotype, in addition to an explicit measure of that stereotype or other self-report measures. The table details the wording that every study used in order to measure the explicit stereotype, and the correlations between the implicit and explicit measures of the stereotype. In addition, most studies investigated the predictive validity of the two stereotype measures in predicting various outcome measures. When reported, we noted the discrepancies between those two measures in predicting outcomes related to math and science (e.g., performance, achievements, self-concept, plans and intentions). Finally, the table details correlations of the implicit stereotype and various self-report measures. When information was available, we separated the correlations by gender.

Table A1

Previous studies on implicit and explicit gender-science stereotypes

Study	**Measure of Explicit Stereotype**	**I/E Correlations**	**Predictive Discrepancy**	**Correlations of Implicit Stereotype with Other Self-Report Measures**
Betz, 2013	Agreement with the statement: “I think that women in STEM really do look less feminine than women in more traditional fields”.	-.01	The effect of feminine STEM role-models on math plans (the likelihood of taking math in high school and in college) and math attempts (in a math persistence task) was moderated by the implicit stereotype (β =.85**, β = -1.97*	Feminine Appearance Endorsement: .08

			respectively) but not by the explicit stereotype (all *p*s>.29).	
Cundiff, Vescio, Loken, & Lo, 2013	An average of agreement with the following statements: "It is possible that men have more ability in science than do women" and "In general, men may be better than women at science".	F: .07* M: .04	Among women, intentions to persist in science was negatively related to implicit stereotypes (r = -.09**) but not to explicit stereotype (r = -.06).	F: Explicit science identification: -.17** Explicit gender identity: .01 Intentions to persist in science: -.09** M: Explicit science identification: .11** Explicit gender identity: .04 Intentions to persist in science: .05
Cvencek, Meltzoff, & Greenwald, 2011	Which character (boy or girl) likes to do math more, and to what extent.	.14*	Not reported	Not reported
Cvencek, Meltzoff, & Kapur, 2014	Which character (boy or girl) likes to do math more, and to what extent.	-.02	Not reported	Explicit gender identity: .22** Explicit math self-concept: .15
Galdi, Cadinu, & Tomasetto, 2013	"Who is better in math?" A six-year old boy, a six-year-old girl, or both equally.	F: -.115 M: .000	Implicit stereotypes mediated the effect of exposure to stereotype consistent (versus stereotype inconsistent) painting task on girls' math performance (r = -.267*), whereas explicit stereotypes did not (r = .142).	Math performance: F: -.267** M: -.045
Gilbert, O'Brien, Garcia, & Marx, 2015	N/A	N/A	N/A	An explicit 'math sense of fit' (a variable that integrates scales from three domains: identification, sense of belonging, and enjoyment) was

				correlated with four mental associations measured with the Go/No-go Association Task (Nosek & Banaji, 2001): F: 'Women-Math': .22*** 'Men-Math': .1 Women-English': -.06 'Men- English': .1 M: 'Women-Math': -.12 'Men-Math': -.2 'Women-English': .4 'Men- English': -.18*
Kiefer & Sekaquaptewa, 2007a	Percentage of men that are believed to be good at math minus percentage of women that are believed to be good at math.	.07	Implicit stereotypes moderated the effect of stereotype threat on performance (β = -.19*) whereas explicit stereotypes did not (*p*s >.1).	F: Math test score: .04 Math class difficulty: .36***
Kiefer & Sekaquaptewa, 2007b	An average of agreement with the following statements: "It is possible that men have more math ability than do women", "In general, men may be better than women at math", and "I don't think that there are any real	.191	The interaction term between gender identity and implicit stereotypes predicted women's performance in their calculus final exams (β =.27*) and their desire to pursue math-related careers (β = -.32*), whereas explicit stereotype did not (β =.15, β = -.19 respectively).	F: Career goals: -.255* Math test score: -.098 Explicit gender identification: .083 Math sat score: -.118

	gender differences in math ability".			
Lane, Goh, & Driver-Linn, 2012	An average of agreement with the following statements: "Men are just better at science than women", "If I were having trouble with a math problem, I would go to a man instead of a woman for help", and an indication of the extent to which the genders differed in skill at sciences and humanities (from *men are much better* to *women are much better*).	F: -.19[#] M: .1	Implicit stereotypes predicted plans to pursue science versus humanities (*r* = .34***), whereas explicit stereotypes did not (*r* = .12).	F: Plans to pursue science versus humanities: .22* Explicit gender identity: .26* M: Plans to pursue science versus humanities: .32** Explicit gender identity: -.07
Liu, Hu, Jiannong, & Adey, 2010	The extent to which participants associate science/humanities with gender attributes (from *strongly male* to *strongly female*).	7th grade: F: .057 M: .231 8th grade: F: .013 M: .184 9th grade: F: -.239 M: -.001 10th grade: F: .057	Not reported	Attitude towards science: F: .445*** M: .000

		M: .217 11th grade: F: .132 M: .209		
Miller, Eagly, & Linn, 2015	"Please rate how much you associate the following domains [liberal arts\science] with males or females" (from *strongly male* to *strongly female*).	.19***	Female employment rates in the science-related workforce predicted explicit stereotypes but not implicit stereotypes. The difference between women's representation in science education versus researcher's workforce predicted implicit stereotypes but not explicit stereotypes (exact statistics were not reported; only p-values).	Not reported
Nosek et al., 2007	"Which statement best describes you?" From *I strongly associate liberal arts with females and science with males* to *I strongly associate liberal arts with males and science with females.*	F: .22¬ M: .22¬	Not reported	Not reported
Nosek et al., 2009	"Please rate how much you associate the following domains [liberal arts\science] with males or	.13¬	Implicit stereotypes were uniquely related to gender inequality in science and math achievement, accounting for 19% and 24% of variance (respectively), while	Not reported

	females" (from *strongly male* to *strongly female*).		explicit stereotypes accounted only for 2% and 1% of the variance.	
Nosek, Banaji, & Greenwald, 2002	The degree to which math and arts are associated with males and females, using semantic differential item (*males–female* as anchor points).	Not reported	The interaction term between gender and implicit stereotypes predicted several variables: implicit (β = -.32**) and explicit (β = -.44***) math attitudes, implicit (β = -.28*) and explicit (β = -.54***) identification with math, and SAT performance (β = -.31**). The interaction term between gender and explicit stereotypes did not show any significant relations.	F: SAT performance: -.16 Implicit and explicit math attitude, math identification and math SAT performance: -.25 (average) M: SAT performance: .51*** Implicit and explicit math attitude, math identification and math SAT performance: .50 (average)
Nosek & Smyth, 2011	Math stereotyping: An average of agreement with the following statements: "Men are better at math than women are", "Women can achieve as much as men in math". Gender stereotyping: Which statement best describes your belief? From "I strongly associate math with females and arts with males" to "I strongly	Math stereotyping: F: .13* M: .14* Gender stereotyping: F: .18* M: .21*	The interaction term between math stereotypes and gender was a better predictor of math-verbal difference SAT scores (β = -.23***) than the interaction term of gender and math stereotypes (β = -.08**). Similar results were found when predicting math engagement: math attitude (I: β = -.25***, E: β = -.11***), math identity (I: β = -.24***, E: β = -.10***), math anxiety (I: β = -.16***, E: β = -.13***), self-perceived ability (I: β = -.20***, E: β = -.14***),	F: Explicit math identification: -.36* Explicit math attitude: -.35* Math - Verbal SAT performance: -.19* M: Explicit math identification: .08* Explicit math attitude: .09* Math - Verbal SAT performance: .19*

	associate arts with females and math with males".		and math participation (I: β = -.17***, E: β = -.10***).	
Park, Cook, & Greenwald, 2001	An average of agreement with statements such as "I think that in general, men are better at math, science and engineering than women", and placing an X on a semantic differential scale closer to the word (*male* or *female*) in reference to the fields of math, science, engineering, art and English.	Agreements: F: -.20 M: .11 Semantic differential scale: F: -.01 M: .12	Not reported	Not reported
Passolungh, Ferreir, & Tomasetto, 2014	An average of agreement with the following: "According to your teachers, who is better at math between girls and boys?", "According to your classmates, who is better at math between girls and boys?", and "In your opinion, who is better at math between girls and boys?"	.01	The interaction term between explicit stereotypes and gender predicted self-perception of math ability (β =.19*) while the interaction term of implicit stereotypes and gender did not (β = -.02).	Perception of math ability: -.01 Belief in math's value: -.02

Phelan, 2010	Indicate what percentage of U.S. biologists, chemists or physicists (depending on the class) are female as well as what percentage of scientists, writers, and poets are females.	Time 1 (beginning of school year): F: .048 M: .211** Time 2 (end of school year): F: .047 M: .182**	Time 2 implicit stereotypes predicted explicit science attitudes (F: *r* = -.287***, M: *r* = .163*), explicit identification with science (F: *r* = -.305***, M: *r* = .231**), and explicit science aspirations (F: *r* = -.348***, M: *r* = .251***), while explicit stereotypes did not (F: *r* = .102, *r* = .090, M: *r* =.011, *r* = .021, respectively), or did, but to a lesser extent (explicit science aspirations: F: *r* = .128*, M: *r* = .175*). Among males, final grades were correlated with explicit (*r* = .179*) but not with implicit stereotypes (*r* = .070).	Time 1: F: Explicit science attitudes: -.223*** Explicit science identification: -.182** M: Explicit science attitudes: .127* Explicit science identification: .119 Time 2: F: Explicit science attitudes: -.287*** Explicit science identification: -.305*** M: Explicit science attitudes: .163* Explicit science identification: .231**
Ramsey & Sekaquaptewa, 2011	An average of agreement with four statements including "It is possible that men have more math ability than do women", "In general, men may be better than women at math", and "I don’t think that there are any real gender differences in math ability".	Time 1 (midterm exam): .06 Time 2 (final exam): .01	The interaction term between changes in implicit stereotypes and gender predicted math course performance (β= .21*), whereas the interaction term between gender and changes in explicit stereotypes was not significant (β= .04).	Final grade in a math course: Time 1: .32**, Time 2: .19[#]

Reuben, Sapienza, & Zingales, 2014	N/A	N/A	N/A	Participants' performance in an arithmetic task: F: 0.166[#] M: 0.190[#] Participants' expected performance by Employers: for male candidates: 0.177[#], for female candidates: 0.170[#], and the difference between male and female candidates: 0.265**
Smyth, Greenwald, & Nosek, 2009	"Please rate how much you associate the following domains [liberal arts\science] with males or females" (from *strongly male* to *strongly female*).	Study 1: F: .18*** M: .18*** Study 2: F: .21*** M: .22***	While Implicit stereotypes uniquely predicted STEM majors (positively for men and negatively for women), accounting for 5.9%-7.4% of the variance, explicit stereotypes accounted only for less than 0.5% of the variance.	Study 1: Math SAT: F: -.11*** M: .13*** Study 2: Math SAT: F: -.17*** M: .09***
Smyth & Nosek, 2015	"Please rate how much you associate the following domains [liberal arts\science] with males or females" (from *strongly male* to *strongly female*).	Not reported	Among participants who reported an occupation related to science, the men/women rate in that occupation was related to the participant's explicit stereotypes but not to the implicit stereotype (exact statistics were not reported).	Not reported

Steffens, Jelenec, & Noack, 2010	An average of the differences between ratings of math and German giftedness among girls and boys (e.g., "Boys are often talented for doing German"), and the statement: "Math/German is rather a typical subject for…" (*boys* and *girls* as anchor points).	F: .14 M: .24	Among females, implicit stereotypes were slightly a better predictor for implicit math self-concept (I: β = -.22*, E: β = -.18*), whereas explicit stereotypes were better than implicit stereotypes in predicting explicit math self-concept (I: β = -.11*, E: β = -.26*), school grades (I: β = -.19*, E: β = -.22*), and enrollment preferences (I: β = -.15*, E: β = -.25*). Among males, explicit stereotypes were better than implicit stereotypes in predicting school grades (I: β = .13*, E: β = .16*), and only explicit stereotypes predicted explicit math self-concept (I: β = .04, E: β = .28*) and enrollment preferences (I: β = .06, E: β = .32*).	F: Explicit math self-concept: -.15* Enrollment preferences: -.19* M: Explicit math self-concept: .11 Enrollment preferences: .13*
Young, Rudman, Buettner, & McLean, 2013	Rating science, chemistry, engineering, and physics on a scale ranging from 1 (*masculine*) to 7 (*feminine*).	F: .16* M: -.13	Among females, identification with female role models (science professors) was negatively correlated with implicit stereotypes (β = -.25***), but not explicit stereotypes (β value is not reported)). In addition, among females, attitude toward science	F: Explicit science identification: -.21** Explicit science attitude: -.15* Role model and career Aspirations: -.06 M: Explicit science identification: .01 Explicit science attitude: .01 Role model and career Aspirations: -.03

	and identification with science were negatively correlated with implicit stereotypes (*r* = -.15*, *r* = -.21**, respectively) but not with explicit stereotypes (*r* = .05, *r* = .13, respectively).	

Notes. All the studies used the IAT to measure implicit stereotype, excluding Gilbert, O'Brien, Garcia, & Marx, 2014 who used the Go/No-go Association Test (Nosek & Banaji, 2001). Under *Predictive Discrepancy* are differences found between the implicit and explicit stereotypes in predicting outcomes related to math and science; I = implicit; E = explicit; M = males; F = Females; $^{\#}p < .1$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$, ¬ statistical probability not reported.

Appendix B

Scores of the Relevant Measures Over the Years

Table B1

Means (SD) of the relevant measures, in each year

Year	IAT	Self-reported associations	Ability stereotype	Interest stereotype	Prevalence stereotype	Personal liking	Personal importance
2003	0.38 (0.39)	1.19 (1.16)	-1.54 (1.68)			0.35 (1.41)	
2004	0.36 (0.38)	1.21 (1.17)	-1.23 (1.73)			0.40 (1.45)	
2005	0.34 (0.39)	1.14 (1.14)	-1.28 (1.74)			0.34 (1.41)	
2006	0.34 (0.40)	1.10 (1.14)	-1.27 (1.75)			0.30 (1.40)	
2007	0.37 (0.40)	1.71 (1.73)	2.67 (1.23)	2.45 (1.21)	1.97 (1.84)	0.25 (1.45)	0.05 (1.35)
2008	0.36 (0.40)	1.69 (1.70)	2.65 (1.22)	2.43 (1.20)	1.95 (1.85)	0.24 (1.45)	0.05 (1.36)
2009	0.37 (0.39)	1.68 (1.72)	2.67 (1.22)	2.46 (1.21)	1.91 (1.84)	0.25 (1.47)	0.05 (1.38)
2010	0.36 (0.40)	1.69 (1.75)	2.70 (1.21)	2.48 (1.21)	1.97 (1.86)	0.25 (1.48)	0.05 (1.38)
2011	0.35 (0.40)	1.71 (1.77)	2.70 (1.23)	2.49 (1.22)	1.91 (1.89)	0.22 (1.48)	0.001 (1.38)
2012	0.35 (0.41)	1.68 (1.75)	2.63 (1.25)	2.41 (1.23)	1.85 (1.84)	0.20 (1.46)	0.01 (1.40)
2013	0.34 (0.41)	1.62 (1.74)	2.54 (1.26)	2.31 (1.22)	1.86 (1.84)	0.16 (1.45)	-0.02 (1.40)
Females	0.35 (0.40)	1.46 (1.57)	0.66 (2.55)	2.39 (1.20)	1.96 (1.85)	0.24 (1.48)	-0.17 (1.33)
Males	0.37 (0.40)	1.53 (1.54)	0.93 (2.32)	2.51 (1.25)	1.80 (1.86)	0.35 (1.37)	0.44 (1.40)

Notes. Larger positive scores reflect stronger gender-science stereotype; *Personal liking* and *personal importance* reflect stronger preference for the topic stereotypically associated with the participant's gender over the topic stereotypically associated with the other gender; Until 2007, minimum and maximum values are as follows: *Self-reported associations*: -4, 4; *Ability stereotype*: -3, 3. In subsequent years: *Self-reported associations*: -6, 6; *Ability stereotype*: 1, 5; *Interest stereotype*: 1, 5; *Prevalence stereotype*: -10, 10; *Personal importance*: -4, 4; Minimum and maximum values for *Personal liking* remain the same across the years: -4, 4.

Appendix C

A Replication of Lane, Goh, and Driver-Linn ,2012: Predicting Plans to Pursue Science

Method

Participants. Participants were volunteers at *Project Implicit* website (Nosek, 2005) which were from the United States and born between 1994 and 1998. 514 participants completed the experiment (68.48% women, M_{age} = 20.11, SD_{age} = 1.28, 83.46% students). Like Lane et al., we restricted the age of the participants to 18-22. Unlike in Lane et al.'s study, our participants were not all from the same university, they were not compensated for their participation, and they participated in the study during the spring semester instead of the fall semester. To obtain more statistical power, unlike Lane et al., we did not restrict our sample to 1st year undergraduate students. However, we analyzed also that subset in our sample (n = 164, 74.39% women, M_{age} = 19.38, SD_{age} = 0.78).

Measures. All measures beside those that measured attitudes and personal importance were taken from Lane, Goh, and Driver-Linn (2012).

Implicit Association Test. The categories and stimuli were: *Male* (items*: Father, Brother*, *Son*, *Uncle*, *Boy*), *Female* (*Mother*, *Sister*, *Daughter*, *Aunt*, *Girl*), *Science* (*Chemistry, Biology*, *Engineering*, *Physics, Math*), and *Humanities* (*Classics, Literature, History, Music*, *Philosophy*). The IAT consisted of seven trial blocks, and was scored with the D1 algorithm (Greenwald, Nosek, & Banaji, 2003). We computed the score such that a positive score indicated stronger associations of *science* with *male* and *humanities* with *female*.

Explicit Stereotype. Participants rated their agreement with the statements: "Men are just better at science than women", and "If I were having trouble with a math problem, I would go to a man instead of a woman for help" (scale: from 1, *strongly disagree*, to 7, *strongly agree*), and answered the questions "To what extent do you think that men and women are different in terms of science skills?", and "To what extent do you think that men and women are different in terms of humanities skills?" (scale: from 1, *men are much* better, to 5, *women are much better*). We averaged the standardized values of the first two statements and the standardized values of the difference between the last two statements to compute the explicit stereotype score.

Behavioral intentions. Participants reported plans to pursue academic subjects by rating their agreement (same 7-pointy scale as before) with three items for science: "I will concentrate in math or a science related subject", "I enjoy reading science literature or watching science programs even if they're not required", and "I doubt I will attend many science lectures out of those required for my courses" (reverse-coded); and two items for humanities: "I will concentrate in a humanities subject" and "I enjoy reading literature or watching programs related to the humanities even if they're not required". We computed a score that reflected stronger plans to pursue the topic stereotypically associated with the participant's gender than to pursue the other topic as the difference between the average of the science items and the humanities items (the direction depended on the participant's gender).

Personal Liking. Similarly to our main study, participants reported how much they like science and how much they like humanities on a scale ranging from 1 (*strongly dislike*) to 5 (*strongly like*). The difference between the responses was the attitude score, coded to reflect a preference for the topic associated with the participant's gender over the topic associated with the other gender.

Personal Importance. Similarly to our main study, participants reported how important it was for them to become knowledgeable in science, math, and humanities on a scale ranging from 1 (*not at all important*) to 5 (*extremely important*). The score was the difference between the average rating of science and math and the humanities rating, coded to reflect an importance of the topic associated with the participant's gender over the topic associated with the other gender.

Procedure. Participants completed the IAT task, followed by a self-report questionnaire of the explicit measures: explicit stereotype, behavioral intentions, self-reported liking of science and humanities, self-reported importance of science, math and humanities, and a question about student status (all presented in a random order). Because they were not relevant for the present research, we excluded measures of gender identity that Lane et al. used.

Results

Table C1 presents descriptive statistics for the measures. As Table C2 shows, in line with our findings in the main study, the IAT/attitudes relation (r = .252, p < .001) was stronger than the IAT/explicit stereotype relation (r = .092, p = .038), Williams' $t(511)$ = 2.845, p = .005, as well as than the explicit stereotype/attitudes relation (r = .137, p = .002), Williams' $t(511)$ = 2.001, p = .046. Similarly, the IAT/importance relation (r = .250, p < .001) was stronger than the IAT/explicit stereotype relation, Williams' $t(511)$ = 2.746, p = .006, and stronger than the explicit stereotype/importance relation (r = .097, p = .027), Williams' $t(511)$ = 2.638 , p = .009.

Replicating Lane et al.'s results, the IAT was a better predictor of pursuit of the topic stereotypically associated with the participant's gender (r = .245, p < .001), than the explicit stereotype (r = .111, p = .012), Williams' $t(511)$ = 2.304, p = .022. We used PROCESS macro for simple mediation (Model 4) for SAS (Hayes, 2013), to find unstandardized estimates with 95% confidence intervals (CI) of the reduction in the effect of each stereotype measure on pursuit, due to controlling for attitudes/importance. Rather than mediation effects, this strategy was used to test whether the relation between each stereotype measure and pursuit was significantly reduced when controlling (separately) for attitudes and importance. We repeated the same analyses to test the reduction in the stereotype measures' effects due to controlling for attitudes, and due to controlling for importance. These were four tests: for each stereotype measure and for each variable we controlled for (attitude or importance). When we tested the IAT we controlled for explicit stereotypes, and when we tested the explicit stereotypes we controlled for the IAT.

Bootstrap tests with 10,000 resamples showed that controlling for attitudes significantly reduced the IAT's effect, $b = .168$, $SE = .028$, 95% CI [0.114, 0.224], and the explicit stereotype's effect, $b = .080$, $SE = .030$, 95% CI [0.019, 0.138]. The regression analysis provided by the PROCESS macro showed that although the IAT's effect was reduced, it remained significant, $b = .067$, SE = .032, $t(511) = 2.132$, $p = .034$, 95% CI [0.005, 0.130]. The explicit stereotype's effect was reduced to being non-significant, $b = .010$, SE = .031, $t(511) < 1$, $p = .755$, 95% CI [-0.051, 0.070]. When we computed partial correlations to partial out shared variance with attitudes, the IAT/pursuit relation ($r = .095$, $p = .032$) was no longer reliably stronger than the explicit stereotype/pursuit relation ($r = .019$, $p = .662$), Williams' $t(511) = 1.251$, $p = .212$.

Conducting the same analyses when controlling for importance showed the same results for the IAT and slightly different results for explicit stereotypes. The bootstrap tests showed a significant reduction in the IAT's effect, $b = .147$, $SE = .028$, 95% CI [0.093, 0.203], but not in the explicit stereotype's effect, $b = .045$, $SE = .026$, 95% CI [-0.006, 0.093]. The regression analysis showed that despite the reduction in IAT's effect, it remained significant, $b = .089$, SE = .035, $t(511) = 2.530$, $p = .012$, 95% CI [0.020, 0.157], and although the explicit stereotype's effect was not significantly reduced, it was no longer significant, $b = .044$, SE = .034, $t(511) = 1.289$, $p = .198$, 95% CI [-0.023, 0.111]. When shared variance with personal importance was partialled out, the IAT/pursuit relation (r =

.115, $p = .009$) was no longer reliably stronger than the explicit stereotype/pursuit relation ($r = .064$, $p = .145$), Williams' $t(511) < 1$, $p = .397$.

When we explored the results in the freshmen sample (the same restrictions used by Lane et al.), the findings did not replicate the results of our entire sample or Lane et al.'s results. Specifically, the explicit stereotype/attitudes relation ($r = .157$, $p = .045$) was not significantly different from the IAT/explicit stereotype relation ($r = .047$, $p = .550$), Williams' $t(161) = 1.039$, $p = .300$, and from the IAT/attitudes relation ($r = .078$, $p = .322$), Williams' $t(161) < 1$, $p = .463$. Similarly, the IAT/importance relation ($r = .143$, $p = .067$) was not significantly different from the explicit stereotype/importance relation ($r = .113$, $p = .148$), Williams' $t(161) < 1$, $p = .781$, or the IAT/explicit stereotype relation, Williams' $t(161) < 1$, $p = .356$. In the freshman sample, the IAT was not a better predictor of self-reported pursuit of the topic stereotypically associated with the participant's gender ($r = .111$, $p = .157$) than the explicit stereotype ($r = .157$, $p = .045$), Williams' $t(161) < 1$, $p = .669$. Because we replicated Lane et al.'s results with the entire sample, we think that it is likely that the failure to replicate the previous findings with a sample more similar to Lane et al.'s original sample reflects random error, due to small sample size, rather than an actual limitation in the generalizability of the previous findings.

Table C1

Means and SD of all the measures

	Men		Women	
	M	*SD*	*M*	*SD*
IAT	0.21 (0.24)	0.43 (0.45)	0.24 (0.25)	0.41 (0.41)
Explicit Stereotype	0.08 (0.17)	0.79 (0.99)	-0.04 (-0.06)	0.72 (0.67)
Plans to Pursue Humanities	4.52 (4.08)	1.64 (1.69)	4.80 (4.71)	1.54 (1.55)
Plans to Pursue Science	4.64 (4.90)	1.46 (1.37)	4.35 (4.40)	1.45 (1.50)
Plans to Pursue (difference score)	0.13 (0.82)	2.31 (2.27)	0.45 (0.31)	2.22 (2.32)
Personal Liking	0.10 (0.52)	1.47 (1.58)	0.43 (0.41)	1.58 (1.67)
Personal Importance	0.08 (0.39)	1.30 (1.34)	0.15 (0.06)	1.10 (1.05)

Notes. In parentheses: the freshmen restricted sample; More positive values on the stereotype measures reflect stronger associations of science with men, and humanities with women; More positive values on *plans to pursue (difference score)*, *personal liking*, and *personal importance* reflect stronger preference for the topic stereotypically associated with the participant's gender over the other topic.

Table C2

Correlations of all variables

	IAT	Explicit Stereotype	Plans to Pursue Humanities	Plans to Pursue Science	Plans to Pursue (diff.)	Personal Liking	Personal Importance
IAT		.092*	-.211***	.151**	.245***	.252***	.250***
Explicit Stereotype	.047		-.095*	.070	.111*	.137**	.097*
Plans to Pursue Humanities	-.194*	-.115		-.106*	-.767***	-.520***	-.553***
Plans to Pursue Science	-.035	.122	-.154*		.719***	.548***	.382***
Plans to Pursue (diff.)	.111	.157*	-.782***	.736***		.717***	.633***
Personal Liking	.078	.157*	-.580***	.516***	.723***		.600***
Personal Importance	.143	.113	-.560***	.415***	.645***	.594***	

Notes. Above the diagonal: correlations in the entire sample; Below the diagonal: correlations in the freshmen restricted sample; Stereotype measures reflect stronger associations of science with men, and humanities with women*;* More positive value on *plans to pursue (difference score)*, *personal liking*, and *personal importance* reflect stronger preference for the topic stereotypically associated with the participant's gender over the other topic; *p < .05, **p < .01, ***p < .001