

A Replication Attempt of Stereotype Susceptibility (Shih, Pittinsky, & Ambady, 1999)

Identity Salience and Shifts in Quantitative Performance

Carolyn E. Gibson, Joy Losee, and Christine Vitiello

Georgia Southern University, Statesboro, GA, USA

Abstract. Awareness of stereotypes about a person's in-group can affect a person's behavior and performance when they complete a stereotype-relevant task, a phenomenon called stereotype susceptibility. Shih, Pittinsky, and Ambady (1999) primed Asian American women with either their Asian identity (stereotyped with high math ability) or female identity (stereotyped with low math ability) or no priming before administering a math test. Of the three groups, Asian-primed participants performed best on the math test, female-primed participants performed worst. The article is a citation classic, but the original studies and conceptual replications have low sample sizes and wide confidence intervals. We conducted a replication of Shih et al. (1999) with a large sample and found a significant effect with the same pattern of means after removing participants that did not know the race or gender stereotypes, but not when those participants were retained. Math identification did not moderate the observed effects.

Keywords: stereotype susceptibility, priming, group identity, Asian American women, replication, stereotype threat

Stereotype susceptibility is a phenomenon in which awareness of stereotypes about a person's in-group and other out-groups affects a person's behavior and performance on tasks related to the stereotype (Shih, Pittinsky, & Ambady, 1999). Negative stereotypes about a person's in-group can hinder performance (stereotype threat; Steele & Aronson, 1995) and positive stereotypes about a person's in-group can facilitate performance (stereotype boost; Shih, Ambady, Richeson, Fujita, & Gray, 2002). Shih et al. (1999) found that Asian American women performed better on a mathematics test when their ethnic identity was made salient compared to a control group. In contrast, Asian American women performed worse on a mathematics test when their gender identity was made salient compared to a control group.

Studies have shown that stereotype susceptibility can affect other behavior besides performance on a test, such as learning, self-control, aggressive behavior, decision-making (Inzlicht & Kang, 2010), and gender diversity in STEM fields (Shapiro & Williams, 2012; Tsui, 2007). Stereotype susceptibility could have broad implications, but some studies have shown mixed results in real-world applications (Stricker & Ward, 2004; Wei, 2012). Stoet and Geary (2012) argued that there is a mismatch between the empirical evidence for the effect and the certainty reported in academic citations of that evidence. For example,

studies of stereotype susceptibility tend to have small sample sizes and wide confidence intervals (Stricker & Ward, 2004; see supplements for summary table). In novel areas of research, low power ironically can increase *both* Type I and Type II errors (Button et al., 2013) even though power is traditionally associated with Type II error risk alone. Shih and colleagues' (1999) seminal paper has had substantial impact (760 citations; Google Scholar, January 12, 2014), but used small samples. Thus, given its importance and potential applicability, we sought to replicate the original Shih et al. result to improve confidence in the effect.

Constraints on Stereotype Susceptibility Effects

Stereotype Awareness

Explicit awareness of a stereotype is an important factor for stereotype susceptibility effects (Appel, Kronberger, & Aronson, 2011; Inzlicht & Kang, 2010). Based on the suggestion by Shih et al. (1999) that their second study on a Canadian sample of recently immigrated Asian women

did not find significant results with the hypothesis that participants were not aware of gender and racial stereotypes about math, we included a measure of stereotype awareness as a basis for exclusion. This measure was adapted from Inzlicht and Kang (2010) who also measured stereotype awareness as a basis for exclusion.

Math Identification

Though it was not part of the original demonstration, subsequent research suggests that domain identification may moderate stereotype susceptibility effects (Shih, Pittinsky, & Ho, 2012; Shih, Pittinsky, & Trahan, 2006). However, Smith and Johnson (2006) observed evidence that strong domain identification may not lead to stereotype boost in response to positive stereotypes. Shih et al. (2012) suggested that further research with domain identification is needed in order to understand how this moderator could affect stereotype susceptibility. As such, in addition to a faithful replication of the original procedure, we added a measure of math identification to test whether the effect is more likely to be observed among the highly identified.

Method

Participants

A total of 164 Asian Female college students participated in this study, with approximately 52 in each condition so as to detect a medium effect size ($r = .35$; Shih et al., 1999, p. 81) with 80% power (Cohen, 1992). Participants were recruited from six universities in the southeast United States. Participants completed the study either individually in a laboratory setting ($n = 19$) or at student centers ($n = 147$). The student centers were relatively quiet and only people sitting alone were invited to participate. So that the recruited Asian women would not be sensitized to having been selected on the basis of that identity, experimenters also asked others nearby to participate. Data collected from non-Asian women were discarded. Experimenters were all White women.

Our registered sampling plan was to preselect Asian American women who were aware of Asian and female stereotypes about math. However, we could not access the necessary sample size through the original data collection plan. So, we altered the sampling plan by adding schools and eliminating the preselection process. Instead, demographic and stereotype awareness measures were included as a basis for exclusion post data collection. Additionally, to maintain consistency with the in-lab setting, distraction during the course of the math test was added as basis for exclusion prior to observing the results. We collected data from 168 participants. Ten participants were excluded: six for distraction, two because they had completed the test at a previous time, and two because they did not speak enough English to complete the survey. Thus, 158 people comprised the sample, consistent with the original sampling plan.

Of the 158 participants, 32 participants were not aware of either or both of the stereotypes of Asians being better than Caucasians and men being better than women at math. Thus, for tests limited to participants aware of the stereotypes, only 127 participants were included.

Design and Materials

Design and materials are the same as the materials used in Shih et al. (1999). This included an identity prime, a 12-item math test with questions from the Canadian Math Competition, and a manipulation check with questions about enjoyment, the perceived difficulty of the test, and hypotheses of what the experiment was assessing. We added a math identification questionnaire with 16 questions (Smith & White, 2001; e.g., "Math is one of my best subjects") with a response scale from 1 = Strongly Disagree to 5 = Strong Agree. All participants also completed a 7-item stereotype awareness survey assessing awareness of cultural stereotypes including men are better than women at math and Asians are better than Caucasians at math (adapted from Inzlicht & Kang, 2010). All materials and data are available on the Open Science Framework (<http://osf.io/8q769>).

Procedure

For in-lab recruitment, participants received the stereotype awareness survey via e-mail. Those reporting awareness of both stereotypes received a second e-mail inviting them to the laboratory. For out-of-lab recruitment, participants were approached in common areas and asked if they would like to take a math test for \$5 and a candy bar. Participants were also offered a chance to enter a drawing to win \$500. After agreeing to participate, an experimenter provided the informed consent. Study packets were randomized prior to data collection and began with a blank paper on top that assured the experimenter's blindness to condition. Participants first completed the manipulation making their Asian or female identity salient (or control). In the female-priming condition, participants ($n = 54$) answered questions concerning coed versus single sex living arrangements. In the Asian-priming condition, participants ($n = 52$) responded about their family and ethnic involvement. Those in the control condition ($n = 52$) answered four questions about their lives unrelated to gender or ethnicity (e.g., how often they eat out).

Participants then completed the 12-question math test from the original study and were told they would have 20 min to finish it. After the math test, participants answered a math identification questionnaire, a stereotype awareness survey (except for in-lab participants), and a final set of questions including a manipulation check, participants' self-reported math skill, enjoyment, and difficulty of the test. These latter items were not analyzed for the confirmatory report, but summary analyses are available in supplementary materials.

The original study had not included a manipulation check, but we added one to see if participants had knowledge

Table 1. Means and standard deviations for the linear comparison between the replication and original. Standard deviations are in parentheses. Means for accuracy and number of correct responses are included

	Asian		Control		Female	
	<i>N</i>	Mean	<i>N</i>	Mean	<i>N</i>	Mean
Current <i>N</i> = 158 accuracy	52	0.59 (0.23)	52	0.55 (0.26)	54	0.53 (0.23)
Current <i>N</i> = 127 aware only, accuracy	40	0.63 (0.22)	44	0.55 (0.25)	43	0.51 (0.23)
Moon and Roeder <i>N</i> = 139 accuracy	53	0.46 (0.17)	48	0.50 (0.18)	38	0.43 (0.16)
Moon and Roeder <i>N</i> = 106 aware only, accuracy	42	0.47 (0.18)	37	0.50 (0.17)	27	0.43 (0.16)
Shih et al. (1999) accuracy	16	0.54 (0.17)	16	0.49 (0.20)	14	0.43 (0.16)
Current <i>N</i> = 158 correct responses	52	6.52 (2.73)	52	5.75 (2.80)	54	5.72 (2.57)
Current <i>N</i> = 127 aware only, correct responses	40	6.93 (2.68)	44	5.73 (2.60)	43	5.60 (2.66)
Moon and Roeder <i>N</i> = 139 correct responses	53	4.75 (2.16)	48	5.21 (1.91)	38	4.50 (1.96)
Moon and Roeder <i>N</i> = 106 aware only, correct responses	42	4.83 (NR)	37	5.19 (NR)	27	4.30 (NR)
Shih et al. (1999) correct responses	16	5.37 (NR)	16	5.31 (NR)	14	4.71 (NR)

of the theme of the questions in the manipulation that preceded the math test. Many of the participants did not pass the manipulation check (34 people passed the manipulation check, 14 people did not answer, 108 did not pass). The theme of the manipulation questions was a subtle prime, so it is conceivable that people would not realize that there was a theme, but still be influenced by the prime.

Experimenters discreetly observed out-of-lab participants during the math test so as to ensure the conditions were similar to an in-lab session. Participants were excluded if they became distracted during the 20 min of the math test. Distraction criteria were defined by circumstances that would not normally happen in the laboratory. This included talking to another person and leaving the seat during the test. Experimenters were blind to condition and could not assess performance, thus these issues could not bias the decision to exclude participants.

Known Differences From Original Study

Above, we described exclusion criteria for stereotype awareness, and measurement of math identity. These were not included in the original research, but were identified as possible constraints on observing stereotype susceptibility effects in the original article or subsequent reports.

One of the differences between this replication and the original was the region in which it took place (Northeast/Harvard vs. Southeast/Emory, Georgia Tech, UGA, Armstrong-Atlantic, Georgia Southern, and UAB). There could be differences between these samples on average SAT scores and average GPAs that may be important for observing the effect.

Another difference is that the original study used an Asian female experimenter and this study used Caucasian female experimenters. However, Shih et al. (2002) used a Caucasian female experimenter and obtained effects similar to Shih et al. (1999). Even so, we minimized contact between participant and experimenter in case this could have an effect.

Like the original study, some of the participants were run in individual laboratory-based sessions, but most participants completed the study in quiet common areas around campus.

Results

Two separate analyses were conducted. First, per the registered sampling plan, we analyzed the data of 158 participants excluding only those unable to properly complete the experiment. Second, per the registered sampling plan, we analyzed the data of only the 127 participants that were aware of both stereotypes.

Primary Comparisons

Following the original study, data were submitted to a linear contrast analysis. This analysis tested the hypothesis that Asian-primed participants would have the best math performance, control participants would perform in the middle, and the female-primed participants would perform worst.

First, the analysis tested the effects of condition on accuracy. As in Shih et al. (1999), accuracy was calculated by dividing questions answered correctly by questions attempted (see Table 1). There was no significant difference between groups on accuracy, $t(155) = 1.31$, $p = .19$, $\eta^2 = .01$, 95% CI [.00, .06] (Note: contrast analysis effect sizes and confidence intervals are reported as η^2 , and η^2 cannot be less than zero). Next, as in the original study, a two-tailed independent-samples *t*-test analyzed differences in accuracy between female-primed and Asian-primed conditions, but did not show a significant difference, $t(104) = 1.37$, $p = .18$, $d = .27$, 95% CI [−.11, .65].

A second analysis tested the effects of identity salience on accuracy, but included only participants who were aware of the race and gender stereotypes ($N = 127$). With only this subset, we observed a significant difference between groups on accuracy, $t(124) = 2.30$, $p = .02$, $\eta^2 = .04$, 95% CI [.01, .17], and the means followed the predicted pattern, Asian ($M = .63$), Control ($M = .55$), and Female ($M = .51$). Likewise, an independent samples *t*-test showed that female-primed participants performed significantly worse than Asian-primed participants, $t(81) = 2.40$, $p = .02$, $d = .53$, 95% CI [.09, .97]. There were no significant differences between the Asian-primed group and the

control group, $t(82) = 1.51$, $p = .13$, $d = .33$, 95% CI $[-.10, .76]$, or the female-primed group and the control group, $t(85) = .77$, $p = .44$, $d = .17$, 95% CI $[-.25, .59]$.

Shih et al. (1999) also used total number of questions answered correctly as a dependent variable, but did not observe significant differences. In a linear contrast with the full sample, there was no significant difference between groups using this dependent variable, $t(155) = 1.52$, $p = .13$, $\eta^2 = .02$, 95% CI $[.00, .07]$ (see Table 1). An independent-samples t -test showed no difference between Asian-primed and female-primed group on correct responses, $t(104) = 1.55$, $p = .13$, $d = .30$, 95% CI $[-.08, .68]$.

Including only those aware of both stereotypes, there was a significant difference between groups on number of questions answered correctly, $t(124) = 2.27$, $p = .03$, $\eta^2 = .05$, 95% CI $[.01, .17]$, and the means followed the expected pattern: Asian-primed ($M = 6.93$), Control ($M = 5.73$), and female-primed ($M = 5.60$). Asian-primed participants had significantly more correct responses than female-primed participants, $t(81) = 2.25$, $p = .03$, $d = .50$, 95% CI $[.06, .94]$. There was also a significant difference between the Asian-primed group and the control, $t(82) = 2.08$, $p = .04$, $d = .46$, 95% CI $[.02, .89]$, but not between the female-primed group and the control, $t(85) = .22$, $p = .83$, $d = .05$, 95% CI $[-.37, .47]$.

Math Identification as a Moderator

To analyze math identification as a moderator, an ANCOVA analyzed math identification as a covariate. Math identification did not interact with condition and is therefore not considered a moderator, $F(2, 151) = 1.63$, $p = .20$. Math identification was also not a moderator when analyzing the data only with participants aware of both stereotypes, $F(2, 120) = 2.06$, $p = .13$.

Discussion

Shih et al. (1999) found that priming Asian women's racial identity or gender identity led to somewhat better and worse math performance respectively, with a control condition in between. Their linear contrast ($N = 46$) elicited an effect size of $\eta^2 = .07$. We conducted a high-powered replication ($N = 156$) and found the same ordinal relations among the conditions, but a much smaller effect size of $\eta^2 = .01$ and a nonsignificant effect ($p = .18$). However, after excluding participants that were unaware of either the gender or racial stereotypes ($N = 127$) the effect was significant ($p = .02$) and the effect size was larger $\eta^2 = .04$. Beilock, Rydell, and McConnell (2007) determined that a causal mechanism of stereotype threat occurs when one is aware of the negative stereotype, such that it causes an explicit expectation of performance. We used this finding and the precedent set by Inzlicht and Kang (2010) to select participants based on their awareness of the stereotypes in question.

When the original and current study's effect sizes are combined, we estimate an effect size of $\eta^2 = .03$ combined with the full replication sample and $\eta^2 = .05$ combined with the reduced replication sample. Shih et al. (1999) included a second study, but they argued that the study ($N = 19$) was a comparison (Canadian sample) in which the effect was not expected to occur, so that study is not included in the aggregate result.

Caution

We replicated the Shih et al. (1999) effect after excluding participants who did not report awareness of the Asian or female stereotypes about math ability. However, most of the participants completed this stereotype measure at the end of the study. As a consequence, it is conceivable that the manipulation differentially affected responses to the stereotype awareness measure among low and high performers. While we perceive this as unlikely, additional research is needed to rule it out conclusively.

There are some curiosities in the present results that deserve further attention. For one, most participants failed the manipulation check suggesting that they were not aware of the priming theme. This could be news that the prime is influential outside of awareness, or it could be a cause of concern for the interpretability of the results. Had the results shown no evidence for an effect, the lack of effect on the manipulation check might have been taken as cautionary evidence against interpreting the results. We did not consider the manipulation check to be critical in our preregistered plan, but that does not prevent opportunistic interpretation and inflation of alpha levels once the results are known. This suggests that additional evidence for this effect in other contexts will be useful. Also notable is the fact that math identification did not moderate the result despite some prior suggestion that this could be a moderator. These observations suggest that the conditions necessary to obtain the present results are not yet well understood.

Finally, a second replication attempt using our registered protocol was conducted at University of California, Berkeley with a total of 139 participants (Moon & Roeder, 2014). In their study, the three priming groups did not significantly differ on accuracy or number of correct answers, and the pattern of means did not fit the hypothesis. Furthermore, analysis with only participants aware of the stereotypes showed no significant differences between groups. These observations suggest that the conditions necessary to obtain the stereotype susceptibility effect are not yet completely understood.

Conclusion

The present research provided some additional support for the hypothesis that priming Asian women with social identities associated with math stereotypes can influence their performance in mathematics. The results additionally suggest

that this effect is contingent on awareness of these stereotypes, but not on identification with mathematics. Conducted with a preregistered confirmatory design, these results provide additional empirical value for this finding, and research on stereotype threat and stereotype boost more generally. Studies examining stereotype threat in standardized testing have inconsistent results with some showing an effect of stereotype susceptibility with field methods (Cohen, Garcia, Apfel, & Master, 2006; Good, Aronson, & Inzlicht, 2003; Wei, 2012) and others not (Ganley, Mingle, Ryan, Ryan, & Vasilyeva, 2013; Stricker & Ward, 2004). Given the additional open questions about this effect, future studies may benefit from strong confirmatory designs to help clarify the conditions under which these effects can be observed.

Acknowledgments

The authors contributed equally to this article. Thus, authors are listed alphabetically. Designed research: C. G., J. L., C. V.; Performed Research: C. G., J. L., C. V.; Analyzed Data: C. G., J. L., C. V.; Wrote paper: C. G., J. L., C. V. The authors would like to thank Katharine Gaskin for her help collecting data, and Sean Mackinnon and Daniël Lakens for contributing expertise to analyses. The authors would also like to thank Amy Hackney and Brian Nosek for their helpful comments and support throughout the process. We report all data exclusions, manipulations, and measures, and how we determined our sample sizes. The authors declare no conflict-of-interest with the content of this article. The project was funded by a grant from the Center for Open Science. All materials, data, and the preregistered design are available at: <http://osf.io/5itqa>.



References

- Appel, M., Kronberger, N., & Aronson, J. (2011). Stereotype threat impairs ability building: Effects on test preparation among women in science and technology. *European Journal of Social Psychology, 41*, 904–913.
- Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviations, spillover. *Journal of Experimental Psychology: General, 136*, 256–276.
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*, 1–12.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.
- Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science, 313*(5791), 1307–1310.
- Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., & Vasilyeva, M. (2013). An examination of stereotype threat effects girls' mathematics performance. *Developmental Psychology, 49*, 1886–1897.
- Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Applied Developmental Psychology, 24*, 645–662.
- Inzlicht, M., & Kang, S. K. (2010). Stereotype threat spillover: How coping with threats to social identity affects aggression, eating, decision making, and attention. *Journal of Personality and Social Psychology, 99*, 467–481.
- Moon, A., & Roeder, S. S. (2014). A secondary replication attempt of stereotype susceptibility (Shih, Pittinsky, & Ambady, 1999). *Social Psychology, 45*.
- Shapiro, J. R., & Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields. *Sex Roles, 66*, 175–183.
- Shih, M., Ambady, N., Richeson, J. A., Fujita, K., & Gray, H. M. (2002). Stereotype performance boosts: The impact of self-relevance and the manner of stereotype activation. *Journal of Personality and Social Psychology, 83*, 638–647.
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80–83.
- Shih, M., Pittinsky, T. L., & Ho, G. C. (2012). Stereotype boost: Positive outcomes from the activation of positive stereotypes. In M. Inzlicht & T. Schmader (Eds.), *Stereotype threat: Theory, process, and application* (pp. 141–158). New York, NY: Oxford University Press.
- Shih, M., Pittinsky, T. L., & Trahan, A. (2006). Domain specific effects of stereotypes on performance. *Self and Identity, 5*, 1–14.
- Smith, J. L., & Johnson, C. S. (2006). A stereotype boost or choking under pressure? Positive gender stereotypes and men who are low in domain identification. *Basic and Applied Social Psychology, 28*, 51–63.
- Smith, J. L., & White, P. H. (2001). Development of the domain identification measure: A tool for investigating stereotype threat effects. *Educational and Psychological Measurement, 61*, 1040–1057.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797–811.
- Stoet, G., & Geary, D. C. (2012). Can stereotype threat explain the gender gap in mathematics performance and achievement? *Review of General Psychology, 16*, 93–102.
- Stricker, L. J., & Ward, W. C. (2004). Stereotype threat, inquiring about test takers' ethnicity and gender, and standardized test performance. *Journal of Applied Social Psychology, 34*, 665–693.
- Tsui, M. (2007). Gender and mathematics achievement in China and the United States. *Gender Issues, 24*, 1–11.
- Wei, T. E. (2012). Sticks, stones, words, and broken bones: New field and lab evidence on stereotype threat. *Educational Evaluation and Policy Analysis, 34*, 465–488.

Received February 28, 2013

Accepted January 22, 2014

Published online May 19, 2014

Carolyn E. Gibson

Department of Psychology
Georgia Southern University
PO Box 8041
USA
Statesboro, GA 30460
E-mail cegibson504@gmail.com