



On the association between gender-science stereotypes' endorsement and gender bias attribution

Elena De Gioannis¹

Accepted: 28 October 2023
© The Author(s) 2023

Abstract

The existence and persistence of stereotypes on gender and science, as well as their influence on attitudes and behaviors, have been largely studied worldwide. Current measures of gender-science stereotypes are mainly descriptive and do not ask respondents their opinions about the perceived cause(s) of these gender differences. However, empirical evidence suggests that gender bias attribution, i.e., the difference in the causes to which gender differences are attributed, has heterogeneous consequences. Here, it was exploited the fact that Project Implicit includes both instruments of gender-science stereotypes and gender bias attribution to test whether and to what extent two components of gender bias attribution, i.e., causes attributed to personal characteristics and those attributed to social/contextual factors, were associated with the endorsement of implicit and explicit gender-science stereotypes. Both an exploratory and confirmatory factor analysis tested whether the instrument on gender bias attribution in Project Implicit could be decomposed into two components, while an SEM (Structural Equation Modeling) analysis tested the hypothesized association. The factor analysis confirmed that bias attribution should be decomposed into two distinct components, “internal factors” and “external factors”. Finally, the association between these two components and implicit and explicit gender stereotypes varied depending on the participant’s gender. Explicit gender stereotypes’ endorsement was positively associated with the external component in the case of women and with the internal component in the case of men. Conversely, the association between attribution and implicit gender stereotypes was null.

Keywords Gender stereotypes · Attribution theory · Structural equation modeling · STEM

1 Introduction

The pervasive nature of gender stereotypes is a well-known issue in academic research on gender differences. In the case of the underrepresentation of women in STEM (Science, Technology, Engineering, and Math), stereotypes about the different abilities of women

✉ Elena De Gioannis
elena.degioannis@unimi.it

¹ Department of Social and Political Sciences, University of Milan, Via Conservatorio 7, 20122 Milan, Italy

and men in math-related tasks (B. A. Nosek et al. 2009) are recognized as playing a pivotal role in determining women's abandonment of this sector (Wang and Degol 2017).

While there is a large consensus in academia on the fact that observed gender differences in performance, interests, and attitudes toward STEM are ascribable to social explanations (Eagly and Steffen 1984; Miller et al. 2015), less is known about whether outside academia people still attribute gender differences in STEM domains to other factors, such as biological causes (Kersey et al. 2019; Spelke 2005). However, this distinction may have relevant consequences on attitudes and behaviors.

Studies on the perceived causes underlying gender stereotypes are few and mainly date back to the 90s and the first decade of the twenty-first century. For instance, in a study on beliefs about power-related gender traits of 264 young adults, Neff and Terry-Schmitt (2002) found that those who attributed gender differences in those traits to social causes were linked to egalitarian attitudes toward women, as measured by the Attitudes toward Women Scale¹ (Spence and Helmreich 1972), while those who attributed differences to biological causes were linked to traditional attitudes toward women. However, the latter association was found only among male participants. Further research on this is pivotal, as the attribution of gender differences to, for instance, biological –thus immutable – causes is likely to have significantly different consequences than the attribution of these differences to social causes.

This study aims to verify the association between the endorsement of gender-science stereotypes and the causes to which differences between women and men in STEM are attributed. Usually, the instruments used to test peoples' beliefs about women in STEM do not specify the causes that could explain gender differences. However, Project Implicit (B. A. Nosek et al. 1998), a platform where users can test their implicit gender stereotypes on science, collects data on stereotypes endorsement, attitudes toward science, and gender bias attribution.

While previous studies on the STEM gender gap used the data from Project Implicit (Lewis and Lupyan 2020; Miller et al. 2015; Smyth and Nosek 2015), its dataset is usually valued for the availability of a measure of both implicit and explicit stereotypes for several countries. To our knowledge, this is the first time in which the instrument on gender bias attribution is used.

The paper proceeds as follows. First, it reviews the literature on gender-science stereotypes and attribution theory, showing how the two are related. It, then, describes the objective of the study and the methodology used to test it. Finally, it shows the outcomes of the structural equation modeling and discusses these results.

2 Gender-science stereotypes

Stereotypes can be defined as “general expectations about members of particular social groups [...] that lead people to overemphasize differences between groups and underestimate variations within groups” (Ellemers 2018). As regards STEM, gender stereotypes on science and math are multiple and concern several gender differences (De Gioannis 2022).

¹ The Attitudes toward Women Scale measures attitudes about the rights and roles of women in society in relation to men in different domains, i.e., vocational, educational, and intellectual roles freedom and independence dating, courtship, and etiquette drinking, swearing, and dirty jokes sexual behavior marital relations and obligations.

Men are considered to have higher abilities than women in math- and science-related tasks (Schmader et al. 2004), and to be more suitable (Farrell et al. 2015) and interested (Ertl et al. 2017) in science-related fields. STEM, in turn, is usually perceived as a masculine field (del Río et al. 2019; Liu et al. 2010), both because it is considered to require characteristics that have been traditionally attributed to men, e.g., agentic traits (Abele and Wojciszke 2019; Diekman et al. 2017; Sczesny et al. 2018), and because women are still underrepresented in STEM academic and career paths (Alam and Sanchez Tapia 2020). This, in turn, even reinforces the association of the scientific sector with boys and men.

Research on the STEM gender gap has long emphasized that gender stereotypes play a pivotal role in determining both the existence and persistence of the underrepresentation of women in STEM. Previous studies found that the endorsement of gender stereotypes is associated with lower performance in math (Cvencek et al. 2015; Kiefer and Sekaquaptewa 2007; Ramsey and Sekaquaptewa 2011; Sanchis-Segura et al. 2018) and negative attitudes toward math (B. A. Nosek et al. 2002; B. A. Nosek and Smyth 2011). The influence of gender stereotypes crosses the borders of school and education, as it extends also to career intentions and aspirations (Schuster and Martiny 2017; Steffens et al. 2010).

Gender stereotypes can be endorsed both at the implicit and explicit levels. In the first case, they are automatic associations of men with STEM-related fields, whereas in the second case, they are self-reported, thus conscious, beliefs (Whitley and Kite 2016). There is a long debate on whether and to what extent the two capture similar constructs (Greenwald et al. 1998; Hofmann et al. 2005; Zitelny et al. 2017), which is however beyond the scope of this study. Given the explorative nature of the study, here, both implicit and explicit gender stereotypes were taken into account, as explained in the following sections.

3 Attribution theory and stereotypes

The attribution theory, initially developed by Heider (1944) and then refined by Weiner (1979, 1985), describes how people explain the causes of behaviors and events. In particular, it examines how information is gathered and then used to form a causal judgment (Fiske and Taylor 1991). According to this theory, there are three possible attributional dimensions, i.e., locus, stability, and controllability. The locus refers to the type of causes, i.e., the locus of attribution is internal when the cause of a behavior is assigned to a person's characteristic, while it is external when it is assigned to situations or contextual factors that are, thus, outside a person's control. The attribution can be also characterized as either stable or unstable and as either controllable or uncontrollable.

These attributions are associated with several factors, among which are stereotypes. As argued in the model on stereotypes and attribution developed by Reyna (2000), in the case of stereotypes, their attribution falls into three potential categories. Using the classification of the attribution theory, the first category assigns the cause to internal but controllable factors, the second to internal and uncontrollable factors, while the third assigns the cause to external, thus uncontrollable, factors. The type of attribution endorsed by the individual is relevant as "each attributional signature is associated with specific emotions and behavioral responses following either desirable or undesirable events" (Reyna 2000, p. 91).

Empirical evidence confirms that differences in the causes to which gender differences are attributed have heterogeneous consequences on behavior. For instance, Dar-Nimrod and Heine (2006) tested the effect of sharing attribution information about gender differences in math on 133 female college students. Those who were either informed that gender

differences in math performance are due to experiential – thus malleable – causes or that there are no gender differences outperformed participants who were informed that these differences are due to genetic – thus fixed – causes. Thoman et al. (2008) replicated the same experiment on 66 undergraduate female students. In that case, participants were either informed that men are better than women in math because of natural ability or because of different levels of effort, i.e., higher for men. Results showed that participants exposed to the effort condition completed fewer math problems but correctly solved a higher percentage of them compared to those exposed to the natural ability condition or not explicitly informed about the gender difference.

Furthermore, a few studies tested whether the strength of endorsement of gender stereotypes is related to different beliefs about gender differences' attribution. Results suggest that the endorsement of strong stereotypes tends to be associated with essentialist thinking, i.e., the belief that social categories are fixed and natural (Bastian and Haslam 2006; Pauker et al. 2010). In a study conducted by Cundiff and Vescio (2016), female undergraduate students who endorsed strong explicit gender stereotypes – measured as the difference in the assignment of stereotypic traits and the assignment of counter-stereotypic traits to women and men – were less likely to attribute gender disparities in the workplace to discrimination.

Empirical evidence seems, thus, to confirm that the association between the attribution of gender differences and gender stereotypes' endorsement is relevant. Still, contributions to this theme are scarce. Studies on gender-science stereotypes only rarely investigated participants' beliefs about the causes of STEM gender bias (De Gioannis 2022). Some instruments of gender-science stereotypes' endorsement include items that specify the cause to which gender differences are attributed, e.g., “Compared to girls, boys mostly increase their mathematical achievement, because of the support of their teachers” (Nurlu 2017), “There are innate biological differences in math abilities of women and men” (Carlana 2019). Unfortunately, testing the difference between these different attributions was outside the scope of those studies.

4 Project implicit

Project Implicit is a non-profit organization of researchers investigating implicit social cognition (B. A. Nosek et al. 1998). The website hosts several implicit association tests (IAT), a test designed by Greenwald et al. (1998) to measure individual differences in implicit cognition and, in particular, implicit stereotypes. In the case of gender and science, the test evaluates the automatic association between two dimensions, e.g., gender and science/humanities sectors.

Along with the IAT, Project Implicit asks users to complete a questionnaire that includes sociodemographic questions, a measure of explicit gender stereotypes, and a question on the causal attribution of women's underrepresentation in STEM. Since the sample of this dataset is self-selected, it cannot be considered a representative sample of a definable population other than that of volunteer visitors to the Project Implicit site. However, it collects data across multiple countries and multiple years, and it is also characterized by a higher variation in age and educational levels than more traditional samples of college students (Smyth and Nosek 2015). Furthermore, the quality and validity of this dataset have already been tested and proven (Charlesworth et al. 2022; Hehman et al. 2019).

The Project Implicit dataset has already been used in previous studies, also on the theme of the STEM gender gap. Just to mention a few, the study by Miller et al. (2015) used Project Implicit data to show the association between implicit gender-science stereotypes and women's representation in STEM in 66 countries. Similarly, Smyth and Nosek (2015) tested the difference in the strength of gender-science stereotypes' endorsement based on whether women belonged to a female- vs. a male-dominated sector. However, to our knowledge, this is the first time that the instrument on gender bias attribution has been used.

5 Research questions and hypotheses

The current study exploited the fact that Project Implicit includes both instruments to measure gender-science stereotypes and questions on gender differences' attribution to verify the association between the two.

The objective of the study was to test whether and to what extent the two components of gender differences' attribution proposed by the attribution theory – one related to causes assigned to personal characteristics and the other related to causes assigned to social/contextual factors –were associated with the endorsement of implicit and explicit gender-science stereotypes.

This was tested by taking into account potential differences related to participants' gender. Based on previous studies on essentialist beliefs and stereotypes, the hypothesis was that attribution is associated with gender stereotypes' endorsement.

6 Materials and methods

6.1 The sample

Following previous studies using the data² of Project Implicit (Xu et al. 2020), the sample included data collected from 2007 to 2019 on the gender-science IAT (Charlesworth and Banaji 2019). Only cases from the U.S. were retained, who both performed the test and answered self-reported questions in the survey, and who were at least 18 years old and no more than 30 years old, to be coherent with college samples usually used in studies on gender-science stereotypes (Smyth and Nosek 2015). The final sample consisted of 150,749 individuals ($M_{age}=22$, $SD_{age}=3.36$). The majority of participants were female (70%), identified as White (68%), and attended at least some years of college (42%).

6.2 Measures

Explicit gender-science stereotypes' endorsement was measured by the question “Please rate how much you associate science with males or females”, measured on a 7-point

² Data collection was approved by the Institutional Review Board for Social and Behavioral Sciences at the University of Virginia (protocol number: 2186). All participants provided informed consent before performing the test and answering the questionnaire. The results reported here constitute secondary analyses of de-identified data.

Table 1 Descriptive statistics and correlations between variables

	1	2	3	4	5	6	7	8
1. Ability	–							
2. Power	0.45	–						
3. Interest	0.42	0.49	–					
4. Family	0.31	0.53	0.45	–				
5. Encouragement	–0.04	–0.06	–0.08	0.03	–			
6. Discrimination	–0.00	–0.04	–0.10	0.01	0.57	–		
7. Implicit gender stereotypes	0.04	0.02	0.06	0.04	–0.01	–0.03	–	
8. Explicit gender stereotypes	0.07	0.05	0.09	0.09	0.05	–0.00	0.17	–
<i>M</i>	20.64	20.33	20.21	20.73	30.85	30.82	0.29	40.57
<i>SD</i>	10.31	10.32	10.27	10.28	10.24	10.27	0.43	10.20
Skewness	0.14	0.49	0.62	0.15	–0.91	–0.89	–0.39	–0.33
Kurtosis	10.85	10.97	20.19	10.96	20.87	20.76	30.27	30.95

The correlation score reports the Spearman rank correlation coefficient (Khamis 2008)

Likert scale where 1 corresponded to “Strongly female”, 4 to “Neither female nor male”, and 7 to “Strongly male”.

Implicit gender stereotypes’ endorsement was inferred by the D-score resulting from the IAT (Greenwald et al. 2003). The test evaluates the automatic association between two dimensions, each consisting of two categories, by measuring the difference in the time needed to do an association in the case of compatible constructs and the time needed in the case of incompatible constructs. The version proposed by Project Implicit includes gender (e.g., man, son, woman, daughter) as the target and science and liberal arts majors (e.g., astronomy, math, history, arts) as categories. In this case, the compatible association is that of men with science majors and women with liberal arts majors, while the incompatible association is that of men with science majors and women with liberal arts majors.

As regards *attribution to gender bias in science*, in the questionnaire participants were presented with the following statement “Women hold a smaller portion of the science and engineering faculty positions at top research universities than do men. The following factors were typically included to explore possible explanations of these differences”. They were then asked to rate each of the following six items on a 5-point Likert scale, where 1 corresponded to “Not at all important” and 5 to “Extremely important”.

- Different proportions of men and women are found among people with the very highest levels of math ability (item ability).
- On average, men and women differ in their willingness to devote the time required by such high-powered positions (item power).
- On average, men and women differ naturally in their scientific interest (item interest).
- On average, men and women differ in their willingness to spend time away from their families (item family).
- Directly or indirectly, boys and girls tend to receive different levels of encouragement for developing scientific interest (item encouragement).
- On average, whether consciously or unconsciously, men are favored in hiring and promotion (item discrimination).

Table 1 reports the correlation between those variables, their means, and standard deviations, as well as the kurtosis and skewness scores. By looking at the correlation between the six items of gender bias attribution, there seems to be two distinct components, one grouping *ability*, *interest*, *family*, and *power*, and the other grouping *encouragement* and *discrimination*. As regards gender stereotypes' endorsement, both implicit and explicit gender-science stereotypes are not correlated with the identified causes for gender differences in STEM, while there is a negligible correlation between explicit and implicit stereotypes.

6.3 Analytical methods

The study's objective was tested using structural equation modeling (SEM). SEM is a family of statistical techniques used to estimate the relationships among constructs, as it is a combination of factor analysis and path analysis (Weston and Gore 2006). It consists of two components: a measurement model describing the relationship between observed variables and latent constructs, and a structural model describing the interrelationship among constructs (Weston and Gore 2006, p. 724). Compared to regression analysis, SEM is, thus, suitable in those cases in which the interest is in the interrelationships between both observed and unobserved, latent variables.

In this case, the measurement model captured the relationship between the six attribution items and the hypothesized two latent constructs, i.e., internal and external causes for gender differences in STEM. The structural model tested the relationship between the two latent constructs and the endorsement of implicit and explicit gender stereotypes.

Following Weston and Gore (2006), the measurement model was tested using factor analysis and the structural model using path analysis. More details on the factorability of the data can be found in the Appendix. After randomly splitting the sample into two subsamples, an exploratory factor analysis (EFA) on the first subsample assessed the number of latent components, while a confirmatory factor analysis (CFA) tested the measurement model with the suggested number of latent constructs on the other subsample.

Finally, the fit of the structural model was tested using path analysis. It was also checked whether there were differences in the model related to the respondents' gender, a property known as "measurement invariance" (Van De Schoot et al. 2015). Measurement invariance was not reached, meaning that the association between the items and the latent factors (i.e., factor loadings, item intercepts, and item residual variances) varied depending on the gender of the respondents. The structural model was, thus, estimated on the male and female samples separately.

The analysis was conducted in R (R Core Team 2013), using the package *psych* (Revelle 2020) to perform the EFA, and the package *lavaan* (Rosseel 2012) to perform the CFA and the path analysis. The package *semTools* (Jorgensen et al. 2020) was used to compare the male and female groups. The categorical nature of the data imposed some conditions on the application of SEM to the dataset. In particular, the factor analysis was based on the polychoric correlation matrix rather than the Pearson correlation matrix (Holgado-Tello et al. 2010). In the EFA, the Principal Axis extraction method was used. Furthermore, the use of a diagonally weighted least square (DWLS) estimator, robust standard errors, and mean- and variance-adjusted tests, minimized bias on estimates and standard errors (Finney and DiStefano 2013; Kořar and Yilmaz Kořar 2015; Li 2016b, 2016a).

Recently, some researchers have cast doubt on the computation and validity of fit indices in the case of categorical data. Research on the correction of fit indices and new cut-off values for ordinal data (Savalei 2018, 2020; Shi and Maydeu-Olivares 2020)

Table 2 Exploratory factor analysis—output of the two-factor solution

Item	Factor 1	Factor 2	Communalities	Uniquenesses	Item Complexity
Ability	0.59		0.35	0.65	10.00
Interest	0.74		0.55	0.45	10.03
Family	0.68		0.48	0.52	10.02
Power	0.83		0.69	0.31	10.00
Discrimination		0.81	0.66	0.34	10.00
Encouragement		0.80	0.64	0.36	10.00
Sum of squared loadings	20.05	10.32			
Proportion variance	0.34	0.22			
Proportion explained	0.61	0.39			
Adequacy tests					
Root mean square of residuals (RMSR)=0.03					
Tucker-Levis Index=0.909					
RMSEA index=0.106 and the 90% confidence intervals are 0.103, 0.109					

Extraction method: Principal Axis (PA)

Rotation: orthogonal varimax

Correlation matrix: polychoric

recommends the use of the Standardized Root Mean Square Residual (SRMR) as a fit index, with the traditional thresholds of 0.08 (Hu and Bentler 1999). Indeed, this fit index is the only one insensitive to the choices of estimators (Shi and Maydeu-Olivares 2020) and was thus used to evaluate the model's goodness of fit.

7 Results

7.1 Results from the factor analysis

The exploratory factor analysis compared a one-factor, two-factor, and three-factor solution. Coherently with the initial hypothesis, results suggested that the six items can be grouped into two distinct factors, the former including *discrimination* and *encouragement* – called external factor – the latter including *power*, *family*, *interest*, and *abilities* – called internal factor (more details on the scree test can be found in the Appendix). Table 2 shows the factor loadings for each item in the two-factor solution.

The fit of the two-factor model was tested in confirmatory factor analysis. The CFA confirmed the result of the EFA, as the two-factor model had a better fit compared to the one-factor model ($SRMR_{one-factor} = 0.158$, $SRMR_{two-factor} = 0.017$). The fit of the model as indicated by the SRMR is acceptable (Hu and Bentler 1999). Figure 1 shows the factor loadings and the correlation between the latent constructs from the CFA. Note that the external and internal latent components seemed uncorrelated, while there was a negative and weak correlation between the item interest and the item power. Reliability analysis reported an acceptable Cronbach alpha for both the internal ($\alpha = 0.76$) and external ($\alpha = 0.73$) components.

Fig. 1 Results from the confirmatory factor analysis

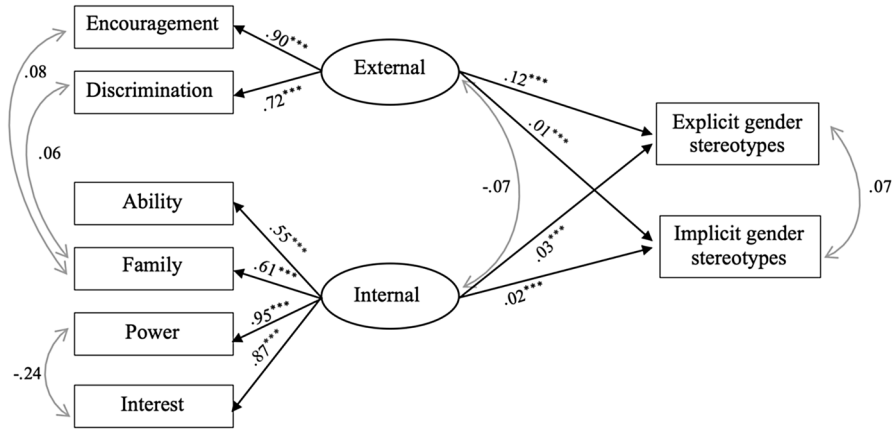
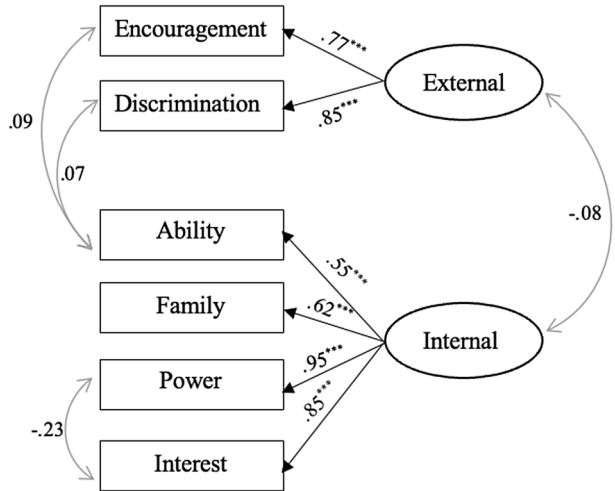


Fig. 2 Results from the structural and measurement model (female group)

7.2 Results from path analysis

The Fig. 2 and 3 show the structural and measurement models for the female group and the male group, respectively. As regards the female sample (SRMR = 0.014), both the external and internal components were positively associated with the endorsement of gender-science stereotypes. However, path coefficients suggest that this association was almost null except for that of the external component with explicit gender stereotypes.

As regards the male sample (SRMR = 0.022), the internal component was positively associated with gender stereotypes' endorsement, while the external component was positively associated with explicit gender stereotypes and negatively with implicit gender stereotypes. However, path coefficients suggest that the association was almost null except for that of the external component with explicit gender stereotypes.

While all coefficients were statistically significant, this could be due to the large size of the sample. To check for the robustness of the results, the same model was estimated on a

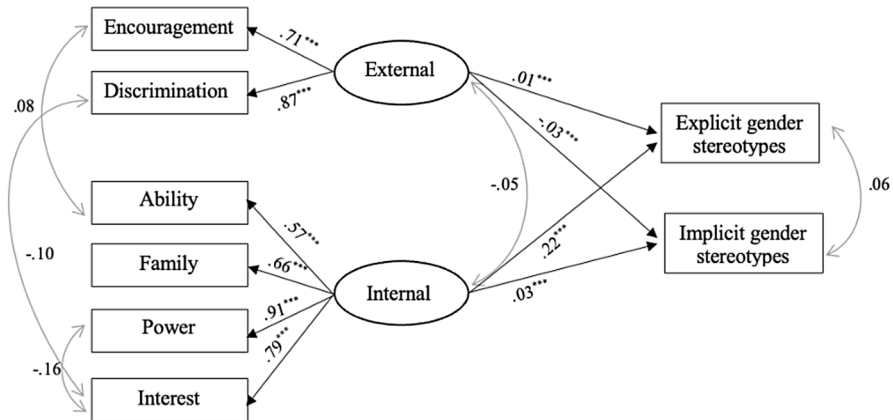


Fig. 3 Results from the structural and measurement model (male group)

smaller sample randomly selected from that used for the CFA. The results are reported in the Appendix. Only the largest coefficients remained statistically significant, i.e., the association between external and internal components and explicit stereotypes for women, and the association between the internal component and explicit stereotypes for men.

8 Discussion and conclusions

The study tested whether the instrument proposed by Project Implicit, assessing the causes attributed to the gender gap in science, can be decomposed into two components, coherent with the attribution theory, i.e., a personal, internal component and a contextual, external component. Furthermore, it also verified whether and how the two components of attribution were associated with the endorsement of both explicit and implicit gender-science stereotypes.

Results confirmed the initial hypothesis that attribution to the gender gap in science is decomposed into two distinct components, one internal and the other external, as suggested by Reyna (2000). The internal component of attribution included causes pertaining to individual characteristics, while the external component included those referring to social or contextual events. It was, then, tested whether and how the two components were associated with the endorsement of gender-science stereotypes.

Results suggested that this configuration was not equivalent for men and women. Indeed, when looking at the association between the two components and explicit and implicit gender stereotypes, there was an important difference. On the one hand, in both groups, neither the external nor the internal component had a strong – although statistically significant – association with implicit gender-science stereotypes. The robustness check found that the statistical significance of these coefficients was likely due to the large size of the sample. On the other hand, in the female group, explicit gender-science stereotypes had a small and significant association with the external component, whereas, in the male group, they had a significant association with the internal component. The results were confirmed in the robustness check.

This suggests that the instrument measuring gender bias attribution proposed by Project Implicit can be decomposed based on the *locus* dimension as proposed by the attribution theory. The first component groups those causes with an internal *locus* as they are characteristics or behaviors of the individual, e.g., interest in science, or performance. Conversely, the second component groups causes that have an external *locus*, e.g., receiving encouragement for pursuing scientific interests. While the model proposed by Reyna (2000) suggests that stereotypes' attribution is classified along a second dimension, i.e., controllability, in this case, the factor analysis suggested retaining only two factors, and the internal component was not further classified into characteristics that are controllable or uncontrollable – i.e., that have biological roots. However, this is likely to be due to the low number of items. In the instrument used by Project Implicit, only one of the potential causes of gender differences explicitly referred to biological reasons, i.e., “On average, men and women differ naturally in their scientific interest”.

Even if the Project Implicit instruments do not allow for distinguishing between controllable and uncontrollable attributions, results from the structural model revealed an interesting pattern. Men and women seem to differ in the type of attribution they associate with the evaluation of science as either a feminine or masculine sector. When associating science with males women tended to refer to external, contextual causes, while men tended to refer to internal, individual causes. This difference may be relevant when assessing the endorsement of gender-science stereotypes.

To assess the level of gender stereotypes' endorsement, researchers tend to use instruments that do not measure why that gender bias is observed. However, as suggested by the application of the attribution theory to stereotypes, this is relevant and should be taken into account. Results from this study seem to suggest that when considering science as a strongly masculine field women have in mind social reasons for this polarization, while men have personal reasons for it. In the first case, it means that while they perceive the sector as masculine, they are also more inclined to believe that it should not be so, as the reason for the overrepresentation of men in this sector is due to contextual factors, such as discrimination. Conversely, men who perceive the sector as masculine are more inclined to believe that this is because of gender differences in attitudes and inclinations toward the scientific sector.

As shown in previous studies, if the cause is perceived as immutable, stereotypes may have a different consequence compared to when it is perceived as not fixed (Dar-Nimrod and Heine 2006; Thoman et al. 2008). In this regard, it is worth mentioning the discussion by Donovan et al. (2019, p. 742). In their field experiment, they found that girls who thought that science ability is innate exhibited also lower interest in STEM. As hypothesized by the authors, “if this fatalistic-essentialist schema is activated within the sociocultural context of the science classroom, then it could be more damaging to girls' future interest in science as messages that increase the belief that science ability is innate and unchangeable could make some girls feel like they do not belong in science”.

The consequences of this association between stereotypes' endorsement and attribution deserve further investigation. Since instruments measuring gender-science stereotypes' endorsement are highly heterogeneous (De Gioannis 2022), it would be necessary to verify whether this association changes when using other explicit gender stereotypes instruments. The one proposed in the Project Implicit survey is quite generic, as it does not specify what is meant by “association of men/women with science”. This could refer to the percentage of women and men working in this sector or the preference of women and men toward scientific topics.

Finally, note that the results suggest that the automatic association between men/women and scientific/humanistic majors was not associated with gender differences' attribution, as the path coefficients were almost zero. This heterogeneity in the results may be related to the fact that the "IAT taps into different constructs than those tapped by the explicit measures used in research on the gender-science stereotype" (Zitely et al. 2017:6). The study conducted by Cundiff and Vescio (2016) tested the association between stereotypes' attribution and explicit but not implicit gender stereotypes. Similarly, an experiment conducted on a mixed-gender sample of 127 undergraduates (Brescoll and LaFrance 2004) found that exposure to biological explanations for sex differences in the ability to identify plants caused participants to endorse more gender stereotypes. Also in that case gender stereotypes were measured at the explicit level as the difference in the attribution of stereotypical traits to men and women. Therefore, there is a lack of evidence on the association between stereotypes' attribution and implicit, rather than explicit stereotypes.

Having noted this, the research does have certain limitations. First, due to the cross-sectional nature of data, it is not possible to determine whether beliefs about stereotypes' attribution influence gender stereotypes' endorsement or whether the latter influences beliefs about stereotypes' attribution. For instance, in the above-mentioned study by Cundiff and Vescio (2016), the experiment tested the effect of stereotypes' endorsement on the attribution of gender disparities in the workplace to discrimination. Here, we are limited to assessing that there is an association and cannot discuss causality.

Furthermore, the sample on which the analysis was performed has some limitations. The Project Implicit website does not impose any restriction on access to the tests, which can be performed on the website for free at any time. This greatly limits the control over participants, who are likely to be already interested in, at least, the notion of implicit bias. This makes the sample likely to be skewed in some ways and not representative of the population. Moreover, note that here it was decided to limit the sample only to U.S. participants, because of the high rate of participation in Project Implicit from this country. This choice allows for a more homogenous sample but limits the possibility of assessing potential heterogeneity. In interpreting the results, it is necessary to keep in mind that this may not hold for individuals coming from other countries and regions of the world. Finally, as mentioned before, the instrument used by Project Implicit to measure the attribution of gender differences in STEM did not allow us to test the difference between internal but controllable causes and internal but uncontrollable ones. Further research is needed, as the dimension of control may be relevant in this context. For instance, while the willingness to spend time away from families is, indeed, a cause related to individual, thus internal, characteristics, it could be, in turn, influenced by social norms and traditional gender roles. However, to test the difference between biological and other controllable internal causes it would be necessary to design a new, ad-hoc instrument as the existing ones cannot serve this purpose.

9 Conclusion

This study suggested that explicit gender-science stereotypes were associated with the attribution of gender differences to external – social, and contextual – factors for women, while they were associated with the attribution of those differences to internal – personal – characteristics for men. This may have relevant implications and further research would be necessary to understand the consequences on choices and behaviors.

As already emerged in previous studies, the consequences of attributing gender differences to biological rather than social factors can be damaging as they, on the one hand, justify the existence of these disparities and, on the other hand, can discourage women from entering the STEM sector. As suggested by Cundiff and Vescio (2016, p. 135), "providing non-essentialist views of gender that emphasize gender differences as malleable and due to social factors, rather than fixed and rooted biology, may offer an avenue for potential attitude change and increased support for strategies designed to address bias".

To do so, it is necessary to conduct further research on the theme. This will require the use of new, more refined, instruments that – contrary to existing ones measuring gender stereotypes – are specially designed to measure beliefs about causal attribution of gender differences. As proven in this study, the instrument included in the Project Implicit dataset may be a good starting point.

Appendix

Check for assumptions' violation

Before proceeding to the analysis, it is necessary to check whether data are suitable to be analyzed using factor analysis. A set of conditions must be satisfied to reach suitability, i.e., (1) an adequate sample size (greater than 200), (2) normality, (3) absence of multicollinearity and (4) factorability of the correlation matrix (Yong and Pearce 2013).

The sample included 150,749 cases and was thus far beyond the limit for an adequate sample size. As stated in the manuscript, the variables used in the analysis are categorical. This requires to use of the polychoric correlation matrix for the exploratory factor analysis. Table 3 reports the kurtosis and skewness values for the six analyzed items.

Note that the third and fourth assumptions were tested using the function `FACTORABILITY` from the package `EFA.dimensions` in R (O'Connor 2022), which allows using a polychoric correlation matrix.

Figure 4 shows the polychoric correlation matrix. As can be seen, seven inter-item correlations are greater than 0.3. More specifically, these correlations suggest that there are two latent components, one grouping 'ability', 'family', 'power' and 'interest', and one grouping 'encouragement' and 'discrimination'. This is coherent with what emerged in the exploratory factor analysis.

Table 3 Kurtosis and skewness values

	Kurtosis	Skewness
Ability	10.85	0.14
Family	10.96	0.15
Power	10.97	0.49
Interest	20.19	0.63
Discrimination	20.76	-0.89
Encouragement	20.87	-0.91

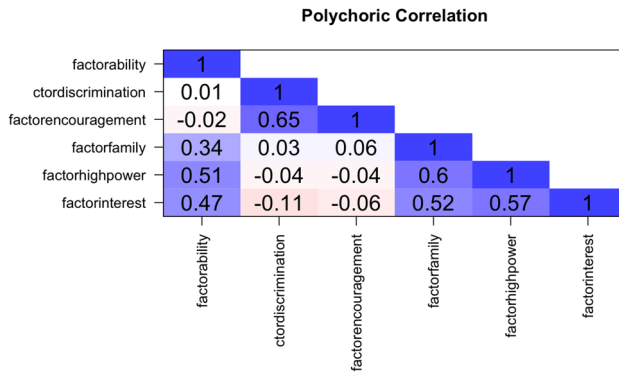


Fig. 4 Polychoric correlation matrix

Table 4 Test of factorability

Test	Result
Determinant of the correlation matrix	0.15
Bartlett test of sphericity	Chi2 = 281,300.66; df = 12; <i>p</i> value = 0.000
Kaiser–Meyer–Olkin measure	Measure of sampling adequacy (MSA) = 0.68

Determinant score rule of thumb > 0.00001;

Bartlett test requires a significant difference for factorability;

The overall MSA should be greater than .50 and items with MSA values < 0.5 should be excluded

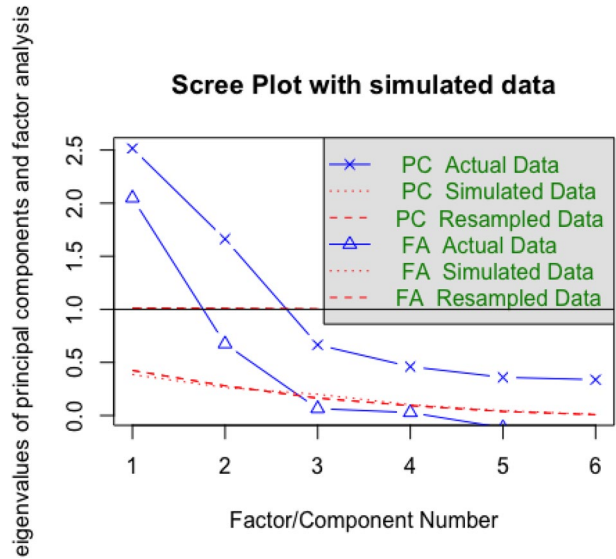
The determinant of the correlation matrix gives information about multicollinearity, as a score greater than 0.00001 is considered an indicator of the absence of multicollinearity. As regards factorability, this can be tested in multiple ways, e.g., by checking that there are at least some small-moderated sized correlations in the correlation matrix and assessing some measures of sampling adequacy. Here I used the Bartlett test of sphericity and the Kaiser–Meyer–Olkin (KMO) Measure of Sampling Adequacy. The first tests whether there is a certain level of redundancy among the items that can be reduced by grouping them into a few factors (Bartlett 1950). The null hypothesis is that variables are orthogonal, i.e., not correlated. The second compares the magnitudes of the observed correlation coefficients to the magnitudes of the partial correlation coefficients so that higher values for this measure indicate that the variables are suitable for factor analysis (Dziuban and Shirkey 1974). However, note that based on the KMO data suitability for factor analysis is quite low.

Table 4 reports the results from the two above-mentioned tests and the determinant score. As it can be seen, all three suggest that the correlation matrix is suitable for factor analysis, i.e., the determinant score is greater than the rule of thumb, Bartlett's test is statistically significant and the KMO measure is greater than 0.50 and none of the items had an MSA lower than 0.50 (Kaiser and Rice 1974).

Parallel analysis

I used parallel analysis and scree test to check the number of factors that should be retained. Figure 5 shows the scree plot. The results suggested retaining two factors.

Fig. 5 Scree plot



Robustness check

Figures 6–9 shows the results of the structural and measurement models estimated on a smaller sample size for both the female sample (Figs. 6, 7) and the male sample (Figs. 8, 9).

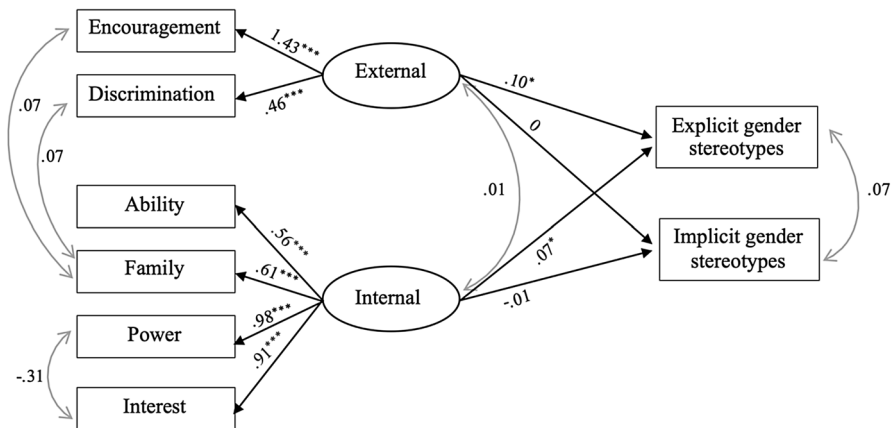


Fig. 6 Structural and measurement models for women (sample size = 1500)

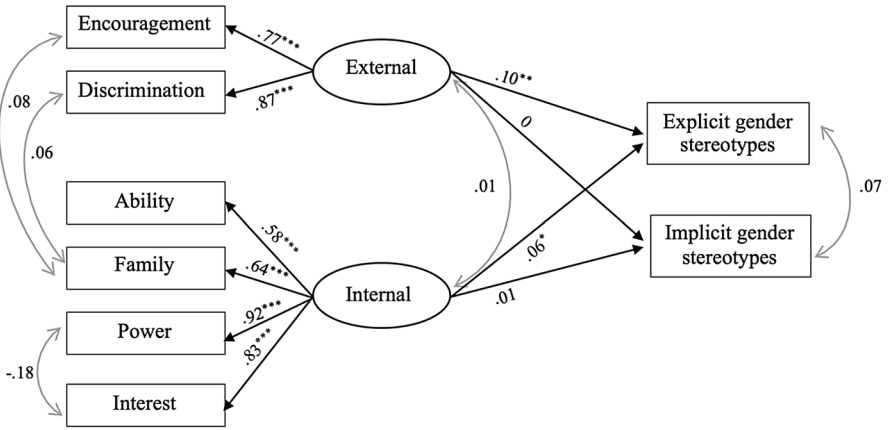


Fig. 7 Structural and measurement models for women (sample size = 2000)

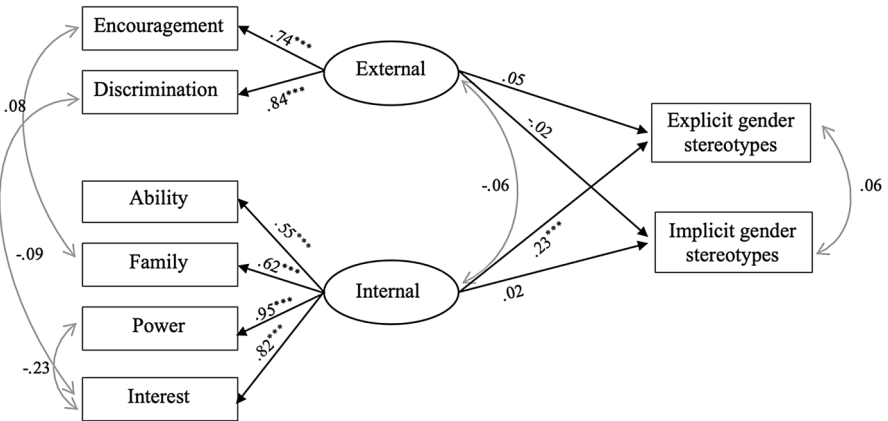


Fig. 8 Structural and measurement models for men (sample size = 1500)

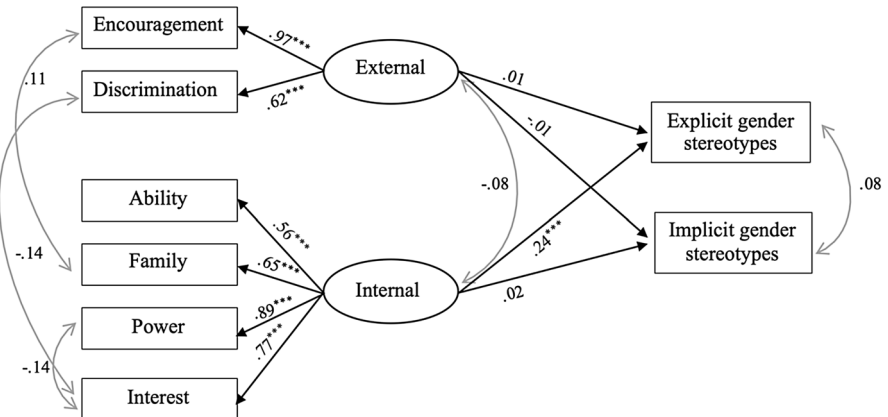


Fig. 9 Structural and measurement models for men (sample size = 2000)

Funding Open access funding provided by Università degli Studi di Milano within the CRUI-CARE Agreement. No funding was received for conducting this study.

Data availability The data that support the findings of this study were derived from the following resources available in the public domain: <https://osf.io/v9dec/>. The dataset is anonymized to protect privacy.

Declarations

Competing Interests The authors have not disclosed any competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abele, A.E., Wojciszke, B. (eds.): Agency and Communion in Social Psychology, 1st edn. Routledge, Taylor & Francis Group (2019)
- Alam, A., Sanchez Tapia, I.: Mapping Gender Equality in STEM from School to Work. UNICEF Office of Global Insight and Policy, New York (2020)
- Bartlett, M.S.: Tests of significance in factor analysis. *Br. J. Stat. Psychol.* **3**(2), 77–85 (1950). <https://doi.org/10.1111/j.2044-8317.1950.tb00285.x>
- Bastian, B., Haslam, N.: Psychological essentialism and stereotype endorsement. *J. Exp. Soc. Psychol.* **42**(2), 228–235 (2006). <https://doi.org/10.1016/j.jesp.2005.03.003>
- Brescoll, V., LaFrance, M.: The Correlates and consequences of newspaper reports of research on sex differences. *Psychol. Sci.* **15**(8), 515–520 (2004). <https://doi.org/10.1111/j.0956-7976.2004.00712.x>
- Carlana, M.: Implicit stereotypes: evidence from teachers' gender bias*. *Q. J. Econ.* **134**(3), 1163–1224 (2019). <https://doi.org/10.1093/qje/qjz008>
- Charlesworth, T.E.S., Banaji, M.R.: Patterns of implicit and explicit attitudes: I. Long-term change and stability from 2007 to 2016. *Psychol. Sci.* **30**(2), 174–192 (2019). <https://doi.org/10.1177/0956797618813087>
- Charlesworth, T.E.S., Navon, M., Rabinovich, Y., Lofaro, N., Kurdi, B.: The project implicit international dataset: Measuring implicit and explicit social group attitudes and stereotypes across 34 countries (2009–2019). *Behav. Res. Methods. Res. Methods* **55**(3), 1413–1440 (2022). <https://doi.org/10.3758/s13428-022-01851-2>
- Cundiff, J.L., Vescio, T.K.: Gender stereotypes influence how people explain gender disparities in the workplace. *Sex Roles* **75**(3–4), 126–138 (2016). <https://doi.org/10.1007/s11199-016-0593-2>
- Cvencek, D., Kapur, M., Meltzoff, A.N.: Math achievement, stereotypes, and math self-concepts among elementary-school students in Singapore. *Learn. Instr.* **39**, 1–10 (2015). <https://doi.org/10.1016/j.learninstruc.2015.04.002>
- Dar-Nimrod, I., Heine, S.J.: Exposure to scientific theories affects women's math performance. *Science* **314**(5798), 435–435 (2006). <https://doi.org/10.1126/science.1131100>
- De Gioannis, E.: The conundrum of gender-science stereotypes: a review and discussion of measurements. *Qual. Quant.* (2022). <https://doi.org/10.1007/s11135-022-01512-8>
- del Río, M.F., Strasser, K., Cvencek, D., Susperreguy, M.I., Meltzoff, A.N.: Chilean kindergarten children's beliefs about mathematics: family matters. *Dev. Psychol.* **55**(4), 687–702 (2019). <https://doi.org/10.1037/dev0000658>
- Diekmann, A.B., Steinberg, M., Brown, E.R., Belanger, A.L., Clark, E.K.: A goal congruity model of role entry, engagement, and exit: understanding communal goal processes in STEM gender gaps. *Pers. Soc. Psychol. Rev.* **21**(2), 142–175 (2017). <https://doi.org/10.1177/1088868316642141>

- Donovan, B.M., Stuhlsatz, M.A.M., Edelson, D.C., Buck Bracey, Z.E.: Gendered genetics: how reading about the genetic basis of sex differences in biology textbooks could affect beliefs associated with science gender disparities. *Sci. Educ.* **103**(4), 719–749 (2019). <https://doi.org/10.1002/sce.21502>
- Dziuban, C.D., Shirkey, E.C.: When is a correlation matrix appropriate for factor analysis? Some Decis. Rules Psychol. Bull. **81**(6), 358–361 (1974). <https://doi.org/10.1037/h0036316>
- Eagly, A.H., Steffen, V.J.: Gender stereotypes stem from the distribution of women and men into social roles. *J. Pers. Soc. Psychol.* **46**(4), 735–754 (1984). <https://doi.org/10.1037/0022-3514.46.4.735>
- Ellemers, N.: Gender stereotypes. *Annu. Rev. Psychol.* **69**(1), 275–298 (2018). <https://doi.org/10.1146/annurev-psych-122216-011719>
- Ertl, B., Luttenberger, S., Paechter, M.: The impact of gender stereotypes on the self-concept of female students in STEM subjects with an under-representation of females. *Front. Psychol.* **8**, 703 (2017). <https://doi.org/10.3389/fpsyg.2017.00703>
- Farrell, L., Cochran, A., McHugh, L.: Exploring attitudes towards gender and science: The advantages of an IRAP approach versus the IAT. *J. Context. Behav. Sci.* **4**(2), 121–128 (2015). <https://doi.org/10.1016/j.jcbs.2015.04.002>
- Finney, S.J., DiStefano, C.: Nonnormal and categorical data in structural equation modeling. In: Finney, S.J. (ed.) *Structural Equation Modeling: A Second Course*, 2nd edn., pp. 439–492. IAP Information Age Publishing, Charlotte (2013)
- Fiske, S.T., Taylor, S.E.: *Social Cognition*. MacGraw-Hill, Noida (1991)
- Greenwald, A.G., McGhee, D.E., Schwartz, J.L.K.: Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* **74**(6), 1464–1480 (1998). <https://doi.org/10.1037/0022-3514.74.6.1464>
- Greenwald, A.G., Nosek, B., Banaji, M.R.: Understanding and using the implicit association test: I. An improved scoring algorithm. *J. Personal. Soc. Psychol.* **85**(2), 197–216 (2003). <https://doi.org/10.1037/0022-3514.85.2.197>
- Helman, E., Calanchini, J., Flake, J.K., Leitner, J.B.: Establishing construct validity evidence for regional measures of explicit and implicit racial bias. *J. Exp. Psychol. Gen.* **148**(6), 1022–1040 (2019). <https://doi.org/10.1037/xge0000623>
- Heider, F.: Social perception and phenomenal causality. *Psychol. Rev.* **51**(6), 358–374 (1944). <https://doi.org/10.1037/h0055425>
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., Schmitt, M.: A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Pers. Soc. Psychol. Bull.* **31**(10), 1369–1385 (2005). <https://doi.org/10.1177/0146167205275613>
- Holgado-Tello, F.P., Chacón-Moscoso, S., Barbero-García, I., Vila-Abad, E.: Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Qual. Quant.* **44**(1), 153–166 (2010). <https://doi.org/10.1007/s11135-008-9190-y>
- Hu, L., Bentler, P.M.: Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Modell.* **6**(1), 1–55 (1999). <https://doi.org/10.1080/10705519909540118>
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2020). *semTools: Useful tools for structural equation modeling* (R package version 0.5–3). <https://CRAN.R-project.org/package=semTools>
- Kaiser, H.F., Rice, J.: Little Jiffy, Mark IV. *Educ. Psychol. Meas.* **34**(1), 111–117 (1974)
- Kersey, A.J., Csumitta, K.D., Cantlon, J.F.: Gender similarities in the brain during mathematics development. *NPJ Sci. Learn.* **4**(1), 19 (2019). <https://doi.org/10.1038/s41539-019-0057-x>
- Khamis, H.: Measures of association: how to choose? *J. Diagn. Med. Sonogr.* **24**(3), 155–162 (2008). <https://doi.org/10.1177/8756479308317006>
- Kiefer, A.K., Sekaquaptewa, D.: Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *J. Exp. Soc. Psychol.* **43**(5), 825–832 (2007). <https://doi.org/10.1016/j.jesp.2006.08.004>
- Koğar, H., Yılmaz Koğar, E.: Comparison of different estimation methods for categorical and ordinal data in confirmatory factor analysis. *Eğitimde Psikolojide Ölçme Değerlendirme Dergisi* **6**(2), 94857 (2015). <https://doi.org/10.21031/epod.94857>
- Lewis, M., Lupyan, G.: Gender stereotypes are reflected in the distributional structure of 25 languages. *Nat. Hum. Behav.* **4**(10), 1021–1028 (2020). <https://doi.org/10.1038/s41562-020-0918-6>
- Li, C.-H.: Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behav. Res. Methods* **48**(3), 936–949 (2016a). <https://doi.org/10.3758/s13428-015-0619-7>
- Li, C.-H.: The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychol. Methods* **21**(3), 369–387 (2016b)

- Liu, M., Hu, W., Jiannong, S., Adey, P.: Gender stereotyping and affective attitudes towards science in Chinese secondary school students. *Int. J. Sci. Educ.* **32**(3), 379–395 (2010). <https://doi.org/10.1080/09500690802595847>
- Miller, D.I., Eagly, A.H., Linn, M.C.: Women's representation in science predicts national gender-science stereotypes: evidence from 66 nations. *J. Educ. Psychol.* **107**(3), 631–644 (2015). <https://doi.org/10.1037/edu0000005>
- Neff, K.D., Terry-Schmitt, L.N.: Youths' attributions for power-related gender differences: nature, nurture, or God? *Cogn. Dev.* **17**(2), 1185–1202 (2002). [https://doi.org/10.1016/S0885-2014\(02\)00094-1](https://doi.org/10.1016/S0885-2014(02)00094-1)
- Nosek, B.A., Smyth, F.L.: Implicit social cognitions predict sex differences in math engagement and achievement. *Am. Educ. Res. J.* **48**(5), 1125–1156 (2011). <https://doi.org/10.3102/0002831211410683>
- Nosek, B.A., Banaji, M.R., Greenwald, A.G.: Math = male, me = female, therefore math \neq me. *J. Pers. Soc. Psychol.* **83**(1), 44–59 (2002). <https://doi.org/10.1037/0022-3514.83.1.44>
- Nosek, B.A., Smyth, F.L., Sriram, N., Lindner, N.M., Devos, T., Ayala, A., Bar-Anan, Y., Bergh, R., Cai, H., Gonsalkorale, K., Kesebir, S., Maliszewski, N., Neto, F., Olli, E., Park, J., Schnabel, K., Shiomura, K., Tulbure, B.T., Wiers, R.W., Greenwald, A.G.: National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proc. Natl. Acad. Sci.* **106**(26), 10593–10597 (2009). <https://doi.org/10.1073/pnas.0809921106>
- Nosek, B. A., Banaji, M. R., and Greenwald, A. G. (1998). *Project Implicit*. <https://implicit.harvard.edu/implicit/index.jsp>
- Nurlu, Ö.: Developing a teachers gender stereotype scale toward mathematics. *Int. Electr. J. Elem. Educ.* **10**(2), 287–299 (2017). <https://doi.org/10.26822/iejee.2017236124>
- O'Connor, B. P. (2022). *EFA. Dimensions: Exploratory Factor Analysis Functions for Assessing Dimensionality* [R Package version 0.1.7.4.]. <https://CRAN.R-project.org/package=EFA.dimensions>
- Pauker, K., Ambady, N., Apfelbaum, E.P.: Race salience and essentialist thinking in racial stereotype development: racial stereotype development. *Child Dev.* **81**(6), 1799–1813 (2010). <https://doi.org/10.1111/j.1467-8624.2010.01511.x>
- R Core Team. (2013). *R: A language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Ramsey, L.R., Sakaquaptewa, D.: Changing stereotypes, changing grades: A longitudinal study of stereotyping during a college math course. *Soc. Psychol. Educ.* **14**(3), 377–387 (2011). <https://doi.org/10.1007/s11218-010-9150-y>
- Revelle, W. (2020). *Psych: Procedures for Psychological, Psychometric, and Personality Research* (R package version 2.0.9). Northwestern University. <https://CRAN.R-project.org/package=psych>
- Reyna, C.: Lazy, dumb, or industrious: When stereotypes convey attribution information in the classroom. *Educ. Psychol. Rev.* **12**(1), 85–110 (2000). <https://doi.org/10.1023/A:1009037101170>
- Rosseel, Y.: lavaan: an R package for structural equation modeling. *J. Stat. Softw.* **48**(2), 1–36 (2012). <https://doi.org/10.18637/jss.v048.i02>
- Sanchis-Segura, C., Aguirre, N., Cruz-Gómez, Á.J., Solozano, N., Forn, C.: Do gender-related stereotypes affect spatial performance? exploring when, how and to whom using a chronometric two-choice mental rotation task. *Front. Psychol.* **9**, 1261 (2018). <https://doi.org/10.3389/fpsyg.2018.01261>
- Savalei, V.: On the computation of the RMSEA and CFI from the mean-and-variance corrected test statistic with nonnormal data in SEM. *Multivar. Behav. Res.* **53**(3), 419–429 (2018). <https://doi.org/10.1080/00273171.2018.1455142>
- Savalei, V.: Improving fit indices in structural equation modeling with categorical data. *Multivar. Behav. Res.* **2**, 1–18 (2020). <https://doi.org/10.1080/00273171.2020.1717922>
- Schmader, T., Johns, M., Barquissau, M.: The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. *Sex Roles* **50**(11), 835–850 (2004). <https://doi.org/10.1023/B:SERS.0000029101.74557.a0>
- Schuster, C., Martiny, S.E.: Not feeling good in STEM: effects of stereotype activation and anticipated affect on women's career aspirations. *Sex Roles* **76**(1–2), 40–55 (2017). <https://doi.org/10.1007/s11199-016-0665-3>
- Szesny, S., Nater, C., Eagly, A.H.: Agency and communion: Their implications for gender stereotypes and gender identities. *Agency Commun. Soc. Psychol.* **15**, 103–116 (2018). <https://doi.org/10.4324/9780203703663>
- Shi, D., Maydeu-Olivares, A.: The effect of estimation methods on SEM fit indices. *Educ. Psychol. Measur.* **80**(3), 421–445 (2020). <https://doi.org/10.1177/0013164419885164>
- Smyth, F.L., Nosek, B.: On the gender–science stereotypes held by scientists: Explicit accord with gender-ratios, implicit accord with scientific identity. *Front. Psychol.* **6**, 415 (2015). <https://doi.org/10.3389/fpsyg.2015.00415>

- Spelke, E.S.: Sex differences in intrinsic aptitude for mathematics and science?: A critical review. *Am. Psychol.* **60**(9), 950–958 (2005). <https://doi.org/10.1037/0003-066X.60.9.950>
- Spence, J.T., Helmreich, R.: The attitudes toward women scale: An objective instrument to measure attitudes toward the rights and roles of women in contemporary society. *Catalog Sel. Doc. Psychol.* **2**(66), 5965236 (1972)
- Steffens, M.C., Jelenec, P., Noack, P.: On the leaky math pipeline: Comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *J. Educ. Psychol.* **102**(4), 947–963 (2010). <https://doi.org/10.1037/a0019920>
- Thoman, D.B., White, P.H., Yamawaki, N., Koishi, H.: Variations of gender–math stereotype content affect women’s vulnerability to stereotype threat. *Sex Roles* **58**(9–10), 702–712 (2008). <https://doi.org/10.1007/s11199-008-9390-x>
- Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., Zondervan-Zwijenburg, M.: Editorial: measurement invariance. *Front. Psychol.* **6**, 1064 (2015). <https://doi.org/10.3389/fpsyg.2015.01064>
- Wang, M.-T., Degol, J.L.: Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educ. Psychol. Rev.* **29**(1), 119–140 (2017). <https://doi.org/10.1007/s10648-015-9355-x>
- Weiner, B.: A theory of motivation for some classroom experiences. *J. Educ. Psychol.* **71**(1), 3–25 (1979). <https://doi.org/10.1037/0022-0663.71.1.3>
- Weiner, B.: An attributional theory of achievement motivation and emotion. *Psychol. Rev.* **92**(4), 548–573 (1985). <https://doi.org/10.1037/0033-295X.92.4.548>
- Weston, R., Gore, P.A.: A brief guide to structural equation modeling. *Couns. Psychol.: Psychol.* **34**(5), 719–751 (2006). <https://doi.org/10.1177/0011000006286345>
- Whitley, B.E., Kite, M.E.: *Psychology of Prejudice and Discrimination*, 3rd edn. Routledge, Routledge/Taylor & Francis Group (2016)
- Xu, K., Nosek, B., Greenwald, A., Ratliff, K., Bar-Anan, Y., Umansky, E., Banaji, M., Lofaro, N., Smith, C., Axt, J., Lai, C.: *Proj. Implicit Demo Web. Datasets* (2020). <https://doi.org/10.17605/OSF.IO/Y9HIQ>
- Yong, A.G., Pearce, S.: A beginner’s guide to factor analysis: focusing on exploratory factor analysis. *Tutor. Quant. Methods Psychol.* **9**(2), 79–94 (2013). <https://doi.org/10.20982/tqmp.09.2.p079>
- Zitelny, H., Shalom, M., Bar-Anan, Y.: What is the implicit gender-science stereotype? Exploring correlations between the gender-science IAT and self-report measures. *Soc. Psychol. Personal. Sci.* **8**(7), 719–735 (2017). <https://doi.org/10.1177/1948550616683017>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.