AMERICAN PSYCHOLOGICAL ASSOCIATION

Developmental Psychology

ISSN: 0012-1649

2025, Vol. 61, No. 3, 579–593
https://doi.org/10.1037/dev0001797

National Disparities Favoring Males Are Reflected in Girls' Implicit Associations About Gender and Academic Subjects

Dario Cvencek[1], Elizabeth A. Sanders[2], M. Francisca del Río[3, 4], María Inés Susperreguy[4, 5], Katherine Strasser[4, 6], Ružica Brečić[7], Dora Gaćeša[7], David Skala[7], Carlo Tomasetto[8], Silvia Galdi[9], Mara Cadinu[10], Manu Kapur[11], Maria Chiara Passolunghi[12], Tania I. Rueda Ferreira[13], Alberto Mirisola[14], Beatrice Mariani[15], and Andrew N. Meltzoff[1, 16]

[1] Institute for Learning and Brain Sciences, University of Washington
[2] Department of Measurement and Statistics, College of Education, University of Washington
[3] Faculty of Education, Universidad Diego Portales
[4] Millennium Nucleus for the Study of the Development of Early Math Skills, Santiago, Chile
[5] Faculty of Education, Pontificia Universidad Católica de Chile
[6] School of Psychology, Pontificia Universidad Católica de Chile
[7] Faculty of Economics and Business, University of Zagreb
[8] Department of Psychology, University of Bologna
[9] Department of Psychology, University of Campania Luigi Vanvitelli
[10] Department of Developmental and Socialization Psychology, University of Padua
[11] Department of Humanities, Social and Political Sciences, Swiss Federal Institute of Technology Zurich
[12] Department of Life Sciences, University of Trieste
[13] Hattiva Lab Cooperativa Sociale, Udine, Italy
[14] Department of Psychology, University of Palermo
[15] Cambiavento Associazioni di Promozione Sociale, Cesena, Italy
[16] Department of Psychology, University of Washington

Based on data for $N = 2{,}756$ children (1,410 girls; $M_{\text{age}} = 8.10$ years) from 16 data sets spanning five nations, this study investigated relations between national gender disparities and children's beliefs about gender and academic subjects. One national-level gender disparity involved inequalities in socioeconomic standing favoring adult males over females (U.N. Human Development Index). The other involved national-level gaps in standardized math achievement, favoring boys over girls (Trends in International Mathematics and Science Study Grade 4). Three novel findings emerged. First, girls' results from a Child Implicit Association Test showed that implicit associations linking *boys* with *math* and *girls* with *reading* were positively related to both national male advantages in socioeconomic standing and national boy advantages in Trends in International Mathematics and Science Study. Second, these relations were obtained for implicit but not explicit measures of children's beliefs linking gender and academic subjects. Third, implicit associations linking gender to academic subjects increased significantly as a function of children's age. We propose a psychological account of why national gender disparities are likely to influence children's developing implicit associations about gender and academic subjects, especially for girls.

Public Significance Statement

In an international study, we examined how national patterns of gender disparities relate to elementary school children's implicit associations between gender and academic subjects. The study involved 2,756 children from five countries. We found that, for girls, national variations in gender inequalities in socioeconomic status and academic achievement significantly predicted stronger implicit associations linking *boys* with *math* and *girls* with *reading*. Moreover, children's implicit associations linking gender and academic subjects significantly increased with age. The findings have implications for psychology, educational equity, and public policy.

Keywords: societal gender inequalities, gender stereotypes, Child Implicit Association Test, age differences, implicit social cognition

Supplemental materials: https://doi.org/10.1037/dev0001797.supp

This article was published Online First August 22, 2024.
Dario Cvencek https://orcid.org/0000-0002-0073-5862
Elizabeth A. Sanders https://orcid.org/0000-0002-1008-8435
M. Francisca del Río https://orcid.org/0000-0002-2050-7963
María Inés Susperreguy https://orcid.org/0000-0001-5584-2692
Katherine Strasser https://orcid.org/0000-0003-2364-6798
Ružica Brečić https://orcid.org/0000-0003-3327-2187
Dora Gaćeša https://orcid.org/0000-0002-0975-5069

continued

Children and adults tend to link males and females to different academic subjects, such as the widespread belief that boys go with computer science and engineering more than girls do. Such beliefs about gender and academic disciplines can be assessed using both implicit and explicit measures. The explicit measures usually involve some form of verbal self-report. The most prominent implicit measure is the Implicit Association Test (IAT; Greenwald et al., 1998). The IAT taps rapid associations that people have between social categories (e.g., gender, age, race) and other attributes (e.g., academic subjects, careers, personal traits). These associations do not require introspection, deliberation, or verbal expression (Greenwald & Lai, 2020; Schmader et al., 2022) and are often referred to as uncontrolled or "automatic" (De Houwer & Boddez, 2022; Ratliff & Smith, 2022). Implicit associations are theorized to be based on statistical patterns in the environment that are often picked up by people without ready introspective access or conscious awareness (Gawronski et al., 2022; Payne et al., 2019) and yet contribute to a person's internal working model of the social world. The original IAT was developed for adults. It has now been modified and adapted for use with children (Baron & Banaji, 2006; Cvencek et al., 2011), including even preschoolers (Cvencek, Brečić et al., 2021; Cvencek et al., 2016).

Understanding the origins and existence of associations between gender and academic subjects is important because they predict a variety of negative outcomes. When held by college-aged women, strong implicit associations of *math* with *men* (and *humanities* with *women*) predict reduced interest in pursuing graduate studies in math-related fields (Kiefer & Sekaquaptewa, 2007). When held by men, implicit associations linking *math* with *men* and *liberal arts* with *women* predict increased biased behavior, such as denial of promotions to women in STEM fields (Régner et al., 2019). In elementary school children, implicit associations about gender and academic subjects predict stronger math self-concepts in boys (math = me) and weaker math self-concepts in girls (math = not-me), which are, in turn, predictive of children's math achievement on standardized tests (Cvencek et al., 2015). Given these and other negative consequences of gender-linked associations between math and reading, it is useful to investigate contributors to these implicit associations during childhood before they begin to impact career pursuits (Early Childhood STEM Working Group, 2017).

Although implicit associations linking *boys* with *math* and *girls* with *reading* have been detected in children during elementary school (Cvencek, Brečić et al., 2021; Cvencek et al., 2011; Galdi et al., 2014; Levine & Pantoja, 2021), little is known about the sources of these associations. To date, three studies have tested how children's implicit associations between gender and academic subjects relate to those held by their parents. The findings indicate that the correlations are either not significant (del Río et al., 2019, 2021) or weak (Galdi et al., 2017). This suggests that children's implicit associations between gender and academic subjects may also have roots in societal sources that lay beyond the family environment itself.

Candidate sources beyond the family are societal-level patterns of disparities, often referred to as structural or systemic biases. Across many cultures (but certainly not all), gender disparities favoring men are evident in terms of standardized math and science achievement tests, particularly in the higher grades (e.g., Breda et al., 2020; Nosek et al., 2009). Math achievement gaps in standardized tests favoring boys can be considered a form of societal gender disparity because there is evidence that such advantages are closely related to—and possibly driven by—societal-level variations in opportunity structures for girls and women (Else-Quest et al., 2010). To quantify "national math gender gaps," we used the Trends in International Mathematics and Science Study (TIMSS). TIMSS is a widely used, standardized international assessment that is designed to rank and compare education systems worldwide using large, representative student samples (Mullis et al., 2020). In the 2019 cycle, 58 nationally representative samples totaling more than 330,000 students and 11,000 schools were involved (Mullis et al., 2020). At each cycle, the international rankings are prominently publicized and discussed; policymakers and educators often strive to improve the test scores of their students and achieve higher international rankings.

Another index of gender disparity at the national level derives from an indicator of systemic gender inequalities in the domains of health, education, and economic standing. The Human Development Index (HDI) is the result of joint efforts of several U.N. agencies,

David Skala https://orcid.org/0000-0001-5468-3677
Carlo Tomasetto https://orcid.org/0000-0002-1350-1387
Silvia Galdi https://orcid.org/0000-0002-1343-9245
Mara Cadinu https://orcid.org/0000-0001-9987-4442
Manu Kapur https://orcid.org/0000-0002-2232-6111
Maria Chiara Passolunghi https://orcid.org/0000-0001-6713-866X
Alberto Mirisola https://orcid.org/0000-0002-7591-1058
Andrew N. Meltzoff https://orcid.org/0000-0001-8683-0547

Elizabeth A. Gunderson served as action editor.

This research was supported by the National Science Foundation under Grant SMA-1640889, awarded to Dario Cvencek, and Grant HRD-1661285, awarded to Andrew N. Meltzoff. The Bezos Family Foundation supported the research under Grant awarded to Andrew N. Meltzoff. The authors thank Tony Greenwald for conversation on the matters discussed here, and Craig Maddox for help with data and article preparation. The authors have no conflicts of interest to disclose. All data and analysis code are available on the Open Science Framework (https://osf.io/ksujx/).

Dario Cvencek played a lead role in conceptualization, methodology, writing–original draft, and writing–review and editing and an equal role in data curation, formal analysis and data contribution. Elizabeth A. Sanders played a lead role in formal analysis, a supporting role in writing–original draft, and writing–review and editing, and an equal role in data curation and methodology. M. Francisca del Río played an equal role in data contribution. María Inés Susperreguy played an equal role in data contribution. Katherine Strasser played an equal role in data contribution. Ružica Brečić played an equal role in data contribution. Dora Gaćeša played an equal role in data contribution. David Skala played an equal role in data contribution. Carlo Tomasetto played an equal role in data contribution. Silvia Galdi played an equal role in data contribution. Mara Cadinu played an equal role in data contribution. Manu Kapur played an equal role in data contribution. Maria Chiara Passolunghi played an equal role in data contribution. Tania I. Rueda Ferreira played an equal role in data contribution. Alberto Mirisola played an equal role in data contribution. Beatrice Mariani played an equal role in data contribution. Andrew N. Meltzoff played a lead role in funding acquisition, a supporting role in methodology, data curation, formal analysis, data contribution, and writing–original draft, and an equal role in conceptualization and writing–review and editing.

Correspondence concerning this article should be addressed to Dario Cvencek, Institute for Learning and Brain Sciences, University of Washington, Box 357988, Seattle, WA 98195, United States. Email: dario1@uw.edu

the World Bank, and multiple national agencies to obtain internationally comparable indicators of socioeconomic standing. The HDI is an annual statistic that is reported in 189 nations and measures each nation's overall progress with respect to social and economic dimensions (United Nations Development Programme, 2019). We hypothesize that children could detect gender inequalities in socioeconomic standing by noticing that men in society are more likely to work (or hold higher prestige or more powerful jobs) than women. To quantify "national socioeconomic gender inequalities" favoring men, we computed an HDI male-to-female ratio, in line with others' use of HDI to quantify gender inequalities. This HDI gender ratio has been shown to yield insights about gender inequalities favoring men over women in socioeconomic development (Klasen & Schüler, 2011), with calls for its regular use in scientific literature. The HDI is widely regarded as a key indicator of socioeconomic standing at a national level because it is derived from representative samples and uses a standardized index that facilitates comparisons across nations (Klasen, 2017; Marsh et al., 2021).

We believe that these two types of national gender disparities may play a role in the development of children's implicit associations linking gender with academic subjects. More specifically, we think that societal-level inequalities can be picked up by children as patterns and structures in the environment, and these perceived patterns can, in turn, influence children's developing representations of the social world. If children are exposed to persistent patterns of gender inequalities, this may be a key input for forging implicit associations.

At the same time, we acknowledge that there is a debate in the literature about what implicit associations mean conceptually (De Houwer & Boddez, 2022; Dovidio & Kunst, 2022; Eagly & Chaiken, 2005; Greenwald & Lai, 2020; Schmader et al., 2022). Recognizing this ongoing discussion within contemporary social psychology, several theorists have urged for more empirical evidence regarding the relation between individual-level psychological measures and macrolevel societal measures (Gawronski et al., 2022; Payne et al., 2017). We designed this study to begin to address this point from a developmental psychology perspective—that is, to add to the empirical literature using both implicit and explicit measures in the *same* children and to connect children's developing implicit cognition to larger patterns of societal inequalities that are evident in the cultures in which the children are reared. This approach takes advantage of cultural variations to inform us about the role of environmental context in the development of children's social cognition.

Study Aims

Using 16 international data sets, we investigated whether national adult socioeconomic gender inequality and national-level math gender gaps predict the magnitude of children's implicit associations between gender and academic subjects ("math/reading–gender associations") and whether these relations differ for girls and boys.

Our first hypothesis concerned the relation between children's implicit associations linking *math* with *boys* and *reading* with *girls* and two national disparities—national math gender gaps favoring boys on standardized tests (TIMSS Grade 4; hereafter "TIMSS–4") and national socioeconomic gender inequality favoring men (HDI male-to-female ratio; hereafter "HDI M/F ratio"). We note that Nosek et al. (2009) reported IAT findings in *adults* linking IAT gender–science scores to variations in national gender inequalities. Here, we hypothesized that *children's* gender-linked implicit associations about math and reading would be stronger in nations with larger math gender gaps favoring boys and larger socioeconomic gender inequalities favoring men. In addition to testing this hypothesis, we also examined, in an exploratory fashion, whether the magnitude of such expected positive relations would vary by gender, in part because significant gender differences are usually found with TIMSS (favoring boys) and HDI scores (favoring adult men).

Our second hypothesis concerned the degree to which societal gender disparities may be reflected in children's implicit associations versus their explicit stereotypes. Implicit, gender-linked associations are theorized to be fast, overlearned, "automatic" reflections of patterns found in the social environments within which individuals are immersed (e.g., Cvencek, Brečić et al., 2021; Dasgupta, 2013; del Río et al., 2019; Greenwald & Banaji, 2017; Payne et al., 2019). In contrast, explicitly measured stereotypes typically involve deliberation, and their measurement assumes that respondents (a) can introspectively access relevant information from memory and (b) are motivated to report it verbally in a veridical fashion. These assumptions are not always warranted (Fazio & Olson, 2003; Greenwald et al., 2009). We hypothesized that societal disparities would be reflected more strongly in children's implicit associations than in their explicitly measured stereotypes.

Finally, this study also offers an opportunity to examine age-related variations in children's implicit associations that may not have been detected in previous studies due to the lack of statistical power. The majority of published studies have not reported significant age differences in children's implicit associations linking gender to math and reading, but in the few instances when age differences are found, the effects showed slightly stronger implicit associations in older children (e.g., Cvencek et al., 2015; Passolunghi et al., 2014). Consequently, there is a need for a more detailed investigation of the effects of age on children's implicit associations about gender and academic subjects with larger samples, which the present study with provides.

This study was designed to potentially make three contributions to the literature. First, to our knowledge, no previous study has examined the relation between national indices of gender disparities and children's implicit and explicit beliefs about academic disciplines. Second, the large-scale nature of the study ($N = 2,756$ children, approximately evenly split by gender) allowed us to investigate, in an exploratory fashion, potential gender differences in these relations. Third, this study utilized multilevel linear regression modeling to predict the magnitude of child-level associations about gender and academic subjects from national disparities, which allows for child-level inferences about the effects of societal factors such as the nation-level socioeconomic gender inequalities and nation-level math gender gaps. The use of multilevel modeling is a powerful statistical approach that has gained traction in developmental science (e.g., Muradoglu et al., 2023), because it addresses clustering dependencies and also appropriately tests predictors measured at different levels of child data.

Method

All procedures were approved by the contributing authors' respective Institutional Review Boards. All children provided informed assent before participating, and their parents provided written or verbal consent.

Collection of Data Sets

We started with a literature search to identify and request data for inclusion in the present study. We looked for published studies with the following child-level data available: (a) IAT measures of math/reading–gender associations, (b) explicit gender stereotypes, (c) age of the participant, and (d) gender of the participant. Additionally, the nation in which the data were collected must have (a) participated nationally in the TIMSS-4 math achievement tests within 3 years of the data collection and (b) have U.N.-published HDI data from the year of the data collection period.

Searches were conducted in both APA PsycInfo and Google Scholar using the search terms *implicit*, *explicit*, *math*, *stereotype*, and *children*. These searches yielded eight published studies that met all criteria (three additional published studies were found, but those nations did not use a developmentally appropriate Child IAT to measure children's implicit associations or did not participate in TIMSS within 3 years of the data collection period). The first authors of these eight published studies were asked to share their published data and also asked for unpublished data they were working on. We received nine data sets from the eight published studies that met our criteria, as well as seven more data sets from five unpublished studies, resulting in a total of 16 data sets that were collected between 2006 and 2020. For all of the 16 data sets, the TIMSS and HDI data were available only at the level of national aggregates and not at the level of individual children (see Supplemental Table S1 for additional details about the data sets).

Participants

Across the 16 data sets, we achieved the collection of a large amount of child data, spanning five nations, including Chile (N = 548), Croatia (N = 431), Italy (N = 606), Singapore (N = 267), and the United States (N = 1,073). For three of these five nations, child-level data on race and ethnicity were not available because parents were not asked by researchers to provide race and ethnicity information about their children (i.e., Chile, Croatia, and Italy). Children who were missing age information (5.2%, n = 151) or gender information (0.6%, n = 18) were excluded from analyses (see Analytic Plan for details regarding missing data handling). This resulted in a final analytic sample of N = 2,756 children (1,410 girls) ranging in age from 3 to 15 years old, with M = 8.10 years (SD = 1.98); see Supplemental Table S1 for each data set's sample size and ages.

Implicit and Explicit Measures

Measures of both implicit math/reading–gender associations and explicit stereotypes were obtained.

Children's Implicit Math/Reading–Gender Associations

There are several variants of Child IATs (Baron & Banaji, 2006; Cvencek et al., 2011), as well as detailed validation studies of Child IAT procedures (Cvencek et al., 2016). Previous research on the Child IAT assessing math/reading–gender associations have shown acceptable reliability, Cronbach's α = .74 (Cvencek et al., 2011).

The Child IAT is a computer sorting task in which children categorize pictures or words into four categories as quickly as possible using two response buttons. The principle behind the Child IAT is that the ease and speed with which one can sort the stimuli reflects the associations that the participant finds more natural or "congruent." For example, it would likely be easier for a child to respond quickly to the pairings of *big* with *dinosaurs* and *small* with *birds* ("congruent pairing") than the pairings of *big* with *birds* and *small* with *dinosaurs* ("incongruent pairing"). The math/reading–gender Child IAT included the categories of *math*, *reading*, *boy*, and *girl*. In the congruent pairing, children responded to *math* and *boy* stimuli with one response button and *reading* and *girl* stimuli with the other. In the incongruent pairing, children responded to *math* and *girl* stimuli with one response button and *reading* and *boy* stimuli with the other (the Italian samples used *language* or *arts* instead of *reading* as a contrast to *math*). Children who respond faster when *math* and *boys* share a response button, compared to when *math* and *girls* share a button, are presumed to hold implicit associations linking *boys* with *math* and *girls* with *reading*. All children completed both pairings (order counterbalanced).

The D score algorithm was used to compute the Child IAT score for each individual participant as the difference of mean response times of the *math* = *boy* and *math* = *girl* tasks divided by the pooled standard deviation (Greenwald et al., 2003). This procedure yields a D score ranging from −2 (indicating a stronger association of *math* = *girl* and *reading* = *boy*) to +2 (indicating a stronger association of *math* = *boy* and *reading* = *girl*). The Child IAT (like the adult IAT) has a rational zero value, indicating an equally strong association of *math* with *boys* and *girls* (Cvencek, Meltzoff, et al., 2021). (One study administered a paper-and-pencil version of the Child IAT, which followed the same principle as the computer version, Passolunghi et al., 2014, and those scores were converted to the same scale as the computerized Child IATs.)

Children's Explicit Stereotypes

Each data set included a measure of explicit (self-reported) gender-linked stereotypes about math and reading. For data collected in all nations except Italy (see Supplemental Material Section S5.1 for details), the explicit measure was administered as two separate Likert-scale items (math stereotype and reading stereotype) based on pictures used in Harter and Pike's (1984) Pictorial Scale. For each, participants were shown the pictures of two children (a boy and a girl) from the Harter and Pike picture set and responded by reporting (a) which child they believed possessed an attribute to a greater degree and (b) whether they believed the character possessed the attribute "a little" or "a lot" (the latter was done by the children pointing to one of two differently sized circles). One item requested selecting whether the boy or girl "liked to do math more," and the other item requested selecting whether the boy or girl "liked to read more." The scores on these two items were averaged to arrive at the explicit stereotype score, which ranged from −2 (indicating that girls like math more than boys and boys like reading more than girls) to +2 (indicating that boys like math more than girls and girls like reading more than boys). A score of zero indicated a belief that girls and boys like math and reading equally. All scales used with Italian data sets were scored in the same fashion, ranging from −2 to +2, with a score of zero indicating a belief that girls and boys are equally good at math and reading (see Supplemental Material Section S5.1 for more information).

National Gender Disparity Measures

National data on student standardized achievement tests and adult socioeconomic inequalities were integrated with the child-level data in our study. Specifically, we used national disparity data (TIMSS and HDI) from the same year as each data set's IAT data collection year. If data were unavailable for the exact year that the IAT was administered, the closest year prior was substituted (all national data were collected within 3 years of each data set's child-level IAT data collection year).

National Gender Gap in Math Achievement: TIMSS

Information about national gender gaps in math was accessed by using results from the standardized Grade 4 Trends in International Mathematics and Science Study (TIMSS; see Supplemental Material Section S1.2 for details). The TIMSS is administered internationally every 4 years. TIMSS data were downloaded from the TIMSS and Progress in International Reading Literacy Study International Study Center on April 23, 2020.

The national math gender gap was measured by first subtracting a given nation's mean math score for girls from its mean math score for boys within one TIMSS cycle (i.e., boy–girl difference). This difference score was then assigned to each of the participants in our data set as a gender gap score for that country and year following the rule described above. For example, in 2011, the average math score for Italian *boys* was 512, and the average math score for Italian *girls* was 503; therefore, the national math gender gap for Italy in 2011 was 9 (512–503), indicating a gender gap "favoring boys" (greater than 0). That math gender gap score was then assigned to all Italian participants in our data set who were tested in 2011, as well as the Italian participants tested in 2012 (because there was no TIMSS testing in 2012). The same procedure for computing and matching national math achievement gender gaps was followed for all other participants based on their nation and the year in which they were tested on the math/reading–gender association measures.

The national math gender gap was scored so that positive values indicate a national boy advantage (i.e., boys scoring higher than girls) and negative values indicate a national girl advantage (i.e., girls scoring higher than boys); a value of 0 indicates that, nationally, girls and boys have equal math scores. In the present study, we focused on national Grade 4 gender gaps, rather than Grade 8 gender gaps, because (a) TIMSS-4 was the closest to the mean age of the sample (i.e., 8 years old), and (b) two of our 16 data sets were from Croatia, which did not participate in the Grade 8 TIMSS within a 3-year timeframe of when the child-level math/reading–gender association data were collected.

National Socioeconomic Gender Inequality: HDI

Data on national gender inequalities were accessed by using the U.N. Human Development Index (HDI). HDI scores were also integrated with each of the 16 data sets, again matched by year of math/reading–gender association data collection. The HDI is an annually reported summary measure of each nation's average adult socioeconomic standing across three key dimensions of human development: (a) living a long and healthy life (life expectancy at birth), (b) having access to knowledge (expected years of schooling, mean years of schooling), and (c) having a sufficient standard of living (gross national income per capita). More specifically, the HDI is the geometric mean of normalized indices for each of the dimensions, with scores ranging from 0 to 1 (United Nations Development Programme, 2019). The HDI was originally constructed to measure the "gender gap in human development" (Klasen, 2017), and we followed Dehdarirad et al. (2019), Gozzi et al. (2021), and Troumbis's (2020) description of the HDI as an index of "socioeconomic standing." This characterization makes sense because socioeconomic standing is itself a combination of education, income, and occupation—three dimensions that largely overlap with the dimensions comprising the HDI. Marsh et al. (2021) documented that the HDI correlates at $r = .86$ with a standard socioeconomic status index derived from the Program for International Student Assessment. HDI data were downloaded from the United Nations Development Programme Download Center on May 4, 2021.

For consistency with the TIMSS math gender gap and ease of results interpretation, we computed national gender inequality on the HDI as the ratio of male-to-female HDI scores (HDI M/F ratio) for the respective year of data collection. The HDI M/F ratio is a measure of the socioeconomic gender inequalities favoring men over women.[1] For example, in 2011, the HDI score for Italian *males* was 0.894, and the HDI score for Italian *females* was 0.869; therefore, the socioeconomic gender inequality for Italy in 2011 was computed to be 1.03 (0.894/0.869). That HDI M/F ratio was then assigned to all Italian participants in our data set that were tested in 2011.

The national socioeconomic gender inequality was scored so that HDI M/F ratios greater than 1.00 indicate inequality favoring men, a ratio of 1.00 indicates equality (i.e., women and men having equal outcomes), and ratios less than 1.00 indicate inequality favoring women. It is also worth noting that the HDI male-to-female ratios focus on *within*-nation inequalities (Marsh et al., 2021). In other words, rather than comparing the HDI of women from Italy to the HDI of men from the United States, we are examining the within-nation gender differences by comparing, for example, the HDI of men from Italy to the HDI of women from Italy.

Covariates

To better isolate the unique and interactive effects of national gender disparities and child gender on children's implicit associations, we incorporated two covariates that would be likely to have their own effects on implicit associations: the time period of child-level data collection and child age. With respect to the *time of data collection*, the 16 data sets included in the present study spanned a period ranging from 2006 to 2020. This time period was marked in some respects by a heightened awareness about shifts in real-world gender roles and gender inequalities in the workplace (e.g., the #MeToo movement).

[1] We also reestimated our multilevel statistical models using a male–female HDI difference score (i.e., HDI for males *minus* HDI for females) instead of the male-to-female ratio, and found the same substantive results (i.e., what is and is not significant, as well as the pattern of signs in the coefficients, remained the same).

Such heightened awareness has been implicated in a recent finding using adult participants showing that societal implicit and explicit gender-liked beliefs about academic disciplines are malleable and shifting towards neutrality (Charlesworth & Banaji, 2022). This said, because time periods (years) for our 16 data sets were left-skewed, we used a median split to create a dichotomous predictor at the data set-level for data collected before 2014 (n = 8 data sets) versus after 2014 (n = 8 data sets).[2] By utilizing this time period predictor in our models (see below), we can better evaluate the effects of national gender disparities on math/reading–gender associations independent of variation in data collection year.

Our 16 data sets also varied in the mean *ages* of children who participated (some data sets included very young children, while others tended to sample older children). In keeping with the two-level nature of the data, and because child age was collected at the individual level, we employed two predictors to represent these two levels of child age: (a) the child's age relative to their data set's mean age ("individual-level relative age"), and (b) the mean child age of the data sets ("dataset-level aggregate age"). Decomposing lower-level predictors in a multilevel data structure into orthogonal within-cluster scores (in this case, mean-centering scores within data sets) and between-cluster scores (in this case, creating aggregate scores for each data set) is considered best practice for multilevel analyses (e.g., Enders & Tofighi, 2007; Hamaker & Muthén, 2020). When lower-level predictors are *not* decomposed properly, predictor slope values can be biased, and further, omission of the cluster aggregate can lead to omitted variable bias (e.g., Bell & Jones, 2015). Specifically, in this data report, if an individual child was 6.38 years old, and the data set they participated in had a mean age of 5.61 years, then the *individual-level relative age* would be 0.77 years (6.38–5.61). Use of both "individual-level relative age" (in this case, 0.77 years) and "dataset-level aggregate age" (in this case, 5.61 years) enabled us to control for both: (a) the average effect of child age on the magnitude of implicit associations *within any given data set*, and (b) the average effect, if any, of the mean data set age on the magnitude of implicit associations. Controlling for both allowed us to more precisely isolate national gender disparity effects on implicit associations, as well as to report on age effects, if any, on implicit associations at both individual and sample levels.

Analysis Plan

Models

We analyzed implicit and explicit measures for N = 2,756 children (L1), nested within 16 data sets (L2) from five nations (L3) using three-level, random intercept multilevel linear regression. Although only five nations are represented at L3, we include nations as a random effect to avoid nonindependence errors due to nation membership. When we estimate models as two-level (ignoring the nation level), the model results are essentially the same as the three-level model.

Models were estimated in *R* lme4 (Bates et al., 2015) using full information maximum likelihood, which estimates model fixed effects parameters using the variance–covariance matrix for all variables used in the analysis (and therefore children with some missing data were included in estimates, see Supplemental Material Section S2.4 for details and further justification). Coefficient significance tests employed Satterthwaite degrees of freedom via the *R* lmerTest package (Kuznetsova et al., 2017). Focal predictors included gender (effect coded: 1 = girls, −1 = boys), national gender gaps in TIMSS-4 math scores (with higher scores favoring boys over girls), and national socioeconomic gender inequalities in HDI M/F ratio (scores greater than 1 favoring males over females);[3] covariates included time period of data collection (effect coded: 1 = after 2014, −1 = before 2014) and child age (both relative mean-centered child age as well as data set aggregate mean age). In addition to testing predictor main effects, we tested two-way interactions between child gender and each of the two national gender disparity predictors to evaluate, in an exploratory fashion, whether national gender disparity effects on implicit associations or explicit stereotypes differed for girls and boys. For ease of results interpretation, all continuous predictors at both child and data set levels were standardized as z scores. Thus, our general mixed model for the implicit math/reading–gender associations was as follows:

$$\begin{aligned} Y_{ijk} = {} & \gamma_{000} + \gamma_{010}\text{TimePeriod}_{jk} + \gamma_{100}\text{ZChildAge}_{ijk} \\ & + \gamma_{020}\text{ZChildAgeAgg}_{jk} + \gamma_{200}\text{Gender}_{ijk} \\ & + \gamma_{020}\text{ZTIMSSgap}_{jk} + \gamma_{030}\text{ZHDIratio}_{jk} \\ & + \gamma_{220}(\text{Gender} \times \text{ZTIMSSgap})_{ijk} \\ & + \gamma_{230}(\text{Gender} \times \text{ZHDIratio})_{ijk} + U_{00k} + U_{0jk} + r_{ijk.} \end{aligned} \tag{1}$$

In this model (Equation 1), the ith child score from the jth data set in the kth nation was modeled as a function of the model intercept (γ_{000}, the estimated mean association level) plus predictor fixed effects on association levels ($\gamma_{010} - \gamma_{230}$), variation in association levels among nations (U_{00k}), variation in association levels among data sets within nations (U_{0jk}), and residual variation among children within data sets (r_{ijk}).

Power for Cross-Level Interaction Tests

Given our small nation-level sample size, we evaluated the minimum detectable effect size (MDES) associated with 80% power for both our cross-level interaction tests. In so doing, we used the *simr* package in *R* (Green et al., 2023), which generates simulation-based power estimates for fixed effects tests in multilevel models across a variety of random effects structures as well as different types of dependent variable distributions (Brysbaert & Stevens, 2018; Green & MacLeod, 2016). (Code in Mathieu et al., 2012, is also useful for evaluating power for cross-level interaction tests but is limited to two-level linear models.) In addition to assessing the MDES, we simulated post hoc power for our model-based observed effects. All things being equal, these analyses revealed that our model of the implicit dependent variable had good power for

[2] We note that when we substituted this dichotomized year of data collection with a continuous z scored predictor, our analysis models yielded substantively the same results for the implicit measure, and nearly the same results for the explicit measure (the HDI effect was slightly reduced). We kept the dichotomized version in our models to avoid distortion in the coefficient estimation given the predictor's left skew.

[3] Because these national-level predictors were right-skewed, especially when disaggregated at the child level, we conducted a robustness check on results using dichotomized versions of these predictors and found that results were substantively the same as using continuous versions. For clarity, we retain the continuous versions of these variables in the forthcoming results. More details on robustness checks can be found in the Supplemental Material Section S3.1.

detecting small cross-level interaction effects, but the model for the explicit dependent variable had less power, likely due to the relatively higher variance we observed for that measure, coupled with lower marginal relations among the interaction factors and that dependent variable (see descriptive statistics in Table 1). Additional details about MDES and post hoc power are provided at the end of the Results section.

Transparency and Openness

In the "Implicit and Explicit Measures," "National Gender Disparities Measures," "Covariates," "Models," and "Power for Cross-Level Interaction Tests" sections, we have reported all measures in the study and how we determined our sample size and data exclusions. We also followed the Journal Article Reporting Standards (Kazak, 2018). All data and analysis code are available on the Open Science Framework (https://osf.io/ksujx/). Data were analyzed using R lme4.

Results

Zero-order correlations among variables used in analyses are provided in Table 1 (see Supplemental Table S1 for descriptive statistics of each of the 16 data sets).

Implicit Math/Reading–Gender Associations

Model results and effect sizes for implicit math/reading–gender associations are displayed in the left set of four columns in Table 2 (see Supplemental Material Section S2.3 for effect size computations). The intercept was significantly greater than zero, indicating substantial mean implicit association linking *boys* with *math* and *girls* with *reading*, Coeff = 0.099 (SE = 0.009), $p < .001$, $d = 0.25$, controlling for year of data collection, child age, national math gender gap, and national socioeconomic gender inequality.

Model results for implicit math/reading–gender associations further showed that children who were relatively older than other children in their own sample (data set) and children from data sets that were relatively older (aggregate age) were predicted to have stronger implicit math/reading–gender associations, $p < .001$ and $p = .006$, respectively, holding all else constant (see Table 2). Child gender was also predictive of the magnitude of implicit math/reading–gender associations, $p = .001$, with girls holding stronger implicit associations, all else held constant. Further, national math gender gaps favoring boys in TIMSS–4 scores were uniquely related to the magnitude of implicit math/reading–gender associations (for every standard deviation increase in math gender gaps, there was a predicted increase of 0.027 points [SE = 0.008] in the magnitude of association, $p = .021$).

Interestingly, we found significant two-way interactions among child gender and national math gender gaps favoring boys, $p < .001$, and child gender and national socioeconomic gender inequality favoring men, $p = .002$. To understand the nature of these interactions, we separated analytic models for girls and boys and found that (a) the gender disparity effects on the magnitude of association were only significant for girls, and (b) the national math gender gap effect was more than twice as high as the national socioeconomic gender inequality effect. Specifically, girls from nations with higher levels of gender gap favoring boys in TIMSS-4 math scores exhibited a significant increase in math/reading–gender associations, Coeff = 0.075 (SE = 0.016), $p < .001$, $d = 0.19$ (Figure 1), and a significant increase in math/reading–gender associations was predicted for girls from nations with higher levels of socioeconomic gender inequalities favoring men, Coeff = 0.032 (SE = 0.016), $p = .049$, $d = 0.08$ (Figure 2). For boys, there were no significant relations found for the math gender gap, Coeff = −0.021 (SE = 0.017), $p = .211$, $d = -0.05$ (Figure 1) or socioeconomic gender inequality, Coeff = −0.010 (SE = 0.017), $p = .557$, $d = -0.02$ (Figure 2).

Robustness Checks

Global Gender Gap Index. The Global Gender Gap Index (GGI) was used as an alternative measure of national socioeconomic gender inequality to provide a robustness check for the findings obtained with the HDI M/F ratio measure (see Supplemental Material Section S1.4 for the description of the GGI measure). Reestimating the models using the GGI revealed the same substantive model results as with the HDI M/F ratio (see Supplemental Material Section S3.2 for full statistical details). This robustness check helped ensure that the socioeconomic gender disparity effect (and interaction with

Table 1
Descriptive Statistics and Zero-Order Correlations for Variables Used in Statistical Models

Variable	*M*	*SD*	*N*	1	2	3	4	5	6	7	8
Dependent variables											
1. Implicit math/reading–gender association (child level)	0.10	0.409	2,691	—							
2. Explicit gender stereotype (child level)	0.07	1.212	2,489	.02	—						
Predictors											
3. Time period (1 = after 2014, data set level)	0.62	0.485	2,756	**−.06**	**−.06**	—					
4. Aggregate child age (data set level)	8.10	1.296	2,756	**.07**	.02	**−.32**	—				
5. Mean-centered child age (child level)	0.00	1.491	2,756	**.08**	−.01	.00	.00	—			
6. TIMSS-4 math boy–girl gap (data set level)	7.40	5.584	2,756	**.04**	.01	**.14**	.00	.00	—		
7. HDI M/F ratio (data set level)	1.02	0.012	2,756	−.02	−.04	**−.26**	**−.33**	.00	**−.64**	—	
8. Gender (1 = girl, child level)	0.51	0.500	2,756	**.05**	−.03	−.02	.01	−.04	.01	−.01	—

Note. N = 2,756 children from 16 data sets across five nations. For these descriptives, time period (year of data collection) and child gender were dummy-coded and represent the percentage of child data collected in the later time period (after 2014 = 1, before 2014 = 0) and the percentage of children who are girls (girls = 1, boys = 0). Pearson's r reported for disaggregated data; significant correlations at the .05 level are boldfaced. TIMSS- 4 = Trends in International Math and Science Study, Grade 4; HDI = Human Development Index. M/F ratio = male-to-female HDI ratio.

Table 2
Multilevel Model Fixed Effect Results Predicting Children's Implicit Associations and Explicit Stereotypes

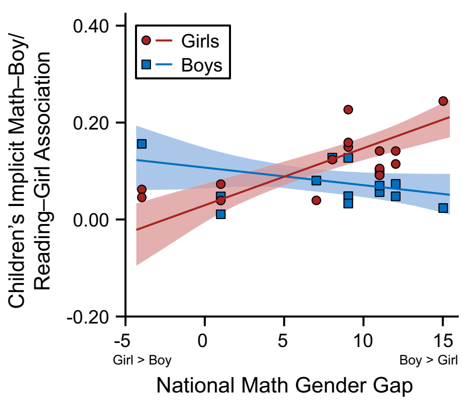
Fixed effect	Implicit math/reading–gender association				Explicit gender stereotype			
	Coeff	*SE*	*p*	*d*	Coeff	*SE*	*p*	*d*
Intercept (mean association/stereotype)	**0.099**	**0.009**	**<.001**	**0.25**	**0.099**	**0.029**	**.003**	**0.08**
Time period (1 = after 2014, data set level)	−0.015	0.009	.106	−0.04	**−0.121**	**0.034**	**.003**	**−0.10**
Aggregate child age (data set level; *z*)	**0.028**	**0.010**	**.006**	**0.07**	−0.049	0.037	.203	−0.04
Mean-centered child age (child level; *z*)	**0.034**	**0.008**	**<.001**	**0.08**	−0.017	0.027	.527	−0.01
TIMSS-4 math boy–girl gap (data set level; *z*)	**0.027**	**0.012**	**.021**	**0.07**	−0.039	0.039	.329	−0.03
HDI M/F ratio (data set level; *z*)	0.011	0.012	.366	0.03	**−0.098**	**0.040**	**.030**	**−0.08**
Gender (1 = girl, child level)	**0.027**	**0.008**	**.001**	**0.07**	−0.040	0.024	.105	−0.03
Child Gender × TIMSS-4 Boy–Girl Gap	**0.050**	**0.011**	**<.001**	**0.12**	−0.027	0.032	.405	−0.02
Child Gender × HDI M/F Ratio	**0.030**	**0.010**	**.002**	**0.07**	−0.004	0.030	.895	0.00

Note. N = 2,756 children from 16 data sets across five nations. All predictors were standardized into z scores (z) except for the two binary predictors, Time period and Gender, which were effect-coded (see the Method section). Parameter estimates derived from 3-level random intercept models estimated with full information maximum likelihood. TIMSS-4 = Trends in International Math and Science Study, Grade 4; HDI = Human Development Index; M/F ratio = male-to-female HDI ratio.

child gender) was not solely due to the unique properties of the HDI M/F measure.

Cross-Validation of Results Using Leave-One-Out Method. As an additional robustness check on our model results, we used a leave-one-out cross-validation approach (e.g., Darlington & Hayes, 2017, pp. 184–185) to ensure results were not due to a particular data set. Specifically, we omitted one data set at a time and re-ran the models for each dependent variable with the remaining data (i.e., we excluded Data set 1, ran the model, then reincluded Data set 1 and excluded Data set 2 and re-ran the models, then reincluded Data set 2 and excluded Data set 3 and re-ran the models, and so forth). Across these analyses, the omission of a given sample still yielded the same pattern of results, as shown in Table 1. More specifically, there were: (a) significant, positive main effects at the child- and sample-levels for age, sample-level TIMSS–4 math boy–girl gap, and child gender; (b) two-way interactions between gender and TIMSS–4 boy–girl gap and HDI M/F ratio on implicit math/reading–gender association, and (c) significant, negative main effects of the time period and HDI M/F ratio on explicit stereotypes.[4]

Figure 1
Children's Implicit Math/Reading–Gender Associations as a Function of National Math Gender Gap



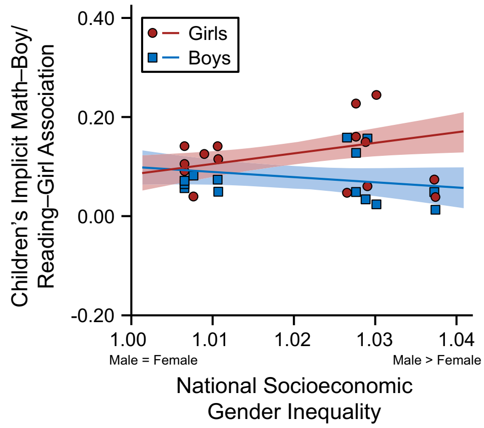
Note. Model-predicted estimates of implicit math/reading–gender associations as a function of national math gender gaps and child gender. The y axis represents model-predicted implicit association (in D scores, see the Method section); positive D scores indicate stronger associations of *math* = *boy* and *reading* = *girl*, negative D scores indicate stronger associations of *math* = *girl* and *reading* = *boy*, with 0 indicating an equally strong association of *math* with *boys* and *girls*. The x axis represents national math gender gaps (TIMSS-4 boy–girl difference; zero indicates no difference, see the Method section). Regression lines represent predicted values derived from model interaction results (shaded regions are 95% CIs), taking into account the intercorrelations among variables, the nested data structure, and the differential sizes of the 16 data sets. Points represent aggregate mean predicted values for each sample by gender. TIMSS-4 = Trends in International Math and Science Study, Grade 4; CIs = confidence intervals. See the online article for the color version of this figure.

Age Effect for Implicit Math/Reading–Gender Associations

Of interest to developmental theories, we also found that the magnitude of implicit math/reading–gender associations increased with age: As shown in Figure 3, data sets with comparatively older children were predicted to have stronger implicit associations (0.028 points more per standard deviation increase in mean data set aggregated child age, p = .006; Table 2). Similarly, older children were predicted to have stronger implicit math/reading–gender associations (0.034 points more per standard deviation increase in child age, all else held constant, p < .001; Table 2).

[4] The exceptions were that: (a) there were three samples (from three different countries) that, when omitted, resulted in a decreased magnitude (but not positive sign) of the TIMSS main effect on implicit math/reading–gender association such that the main effect was no longer statistically significant (but again, the gender interaction with TIMSS was still significant), and (b) there were two samples (from two different countries) that, when omitted, resulted in a decreased magnitude (but not negative sign) of the HDI main effect on explicit stereotype such that the effect was no longer statistically significant. Again, for all of these exceptions, the pattern did not change, just the significance levels of these (data set-level) main effects.

Figure 2
Children's Implicit Math/Reading–Gender Associations as a Function of National Socioeconomic Gender Inequality



Note. Model-predicted estimates of implicit math/reading–gender associations as a function of national socioeconomic gender inequality and child gender. The *y* axis represents model-predicted implicit association (in *D* scores); positive *D* scores indicate stronger associations of *math* = *boy* and *reading* = *girl*, negative *D* scores indicate stronger associations of *math* = *girl* and *reading* = *boy*, with 0 indicating an equally strong association of *math* with *boys* and *girls*. The *x* axis represents national socioeconomic gender inequality (HDI M/F ratio; 1 indicates no difference, see the Method section). Regression lines represent predicted values derived from model interaction results (shaded regions are 95% CIs), taking into account the intercorrelations among variables, the nested data structure, and the differential sizes of the 16 data sets. Points represent aggregate mean predicted values for each sample by gender. HDI = Human Development Index; M/F ratio = male-to-female HDI ratio; CIs = confidence intervals. See the online article for the color version of this figure.

Explicit Gender Stereotypes About Math and Reading

Model results for the explicit measure showed a statistically significant explicit gender stereotype about math and reading, Coeff = 0.099 (SE = 0.029), $p < .003$, $d = 0.08$ (see Table 2, the right set of four columns). In contrast with the results for implicit associations, however, only two variables uniquely predicted explicit stereotypes: data collection time period was negatively related to explicit stereotypes, Coeff = −0.121 (SE = 0.034), $p <$ .003, $d = -0.10$, and higher socioeconomic gender inequality favoring males was negatively related to explicit stereotypes, Coeff = −0.098 (SE = 0.040), $p = .030$. For a full discussion of these effects, see Supplemental Material Section S5.3. Age was not a significant predictor of explicit stereotypes.

Cross-Level Interaction Tests of MDES and Post Hoc Power

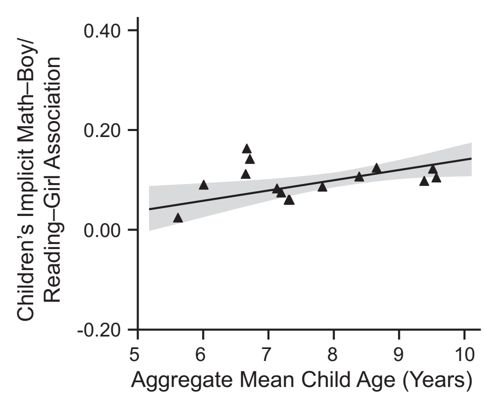
For *implicit* math/reading–gender association, the MDES for the Child Gender × TIMSS-4 boy–girl gap interaction was .031 with 79.90% power (95% CI [77.28%, 82.34%]), and the MDES for the Child Gender × HDI M/F ratio interaction was .027 with 79.90% power (95% CI [77.28%, 82.34%]). As shown in Table 2 and discussed above, our interaction term effect estimates for this outcome were .050 and .030, respectively (ds = 0.12 and 0.07), which exceeded the MDES for each test. Using the observed effect values as true values, our post hoc power was estimated at 99.60% and 86.50% for each test, respectively.

For the *explicit* gender stereotype dependent variable, the MDES for the Child Gender × TIMSS-4 boy–girl gap interaction was −.089 with 80.80% power (95% CI: 78.22%, 83.20%), and the MDES for the Child Gender × HDI M/F ratio interaction was −.079 with 80.80% power (95% CI [78.22%, 83.20%]). As shown in Table 2, our interaction term effect estimates for this outcome were −.027 and −.004, respectively (i.e., smaller in magnitude than the MDESs). Using the observed effect values as true values, our post hoc power was estimated at 14.50% and 5.30% for each test, respectively. As always, increasing the number of data sets and/or size of each data set would improve power, but we note that the interaction effects we observed for this outcome were extremely close to zero (ds = −0.02 and 0.00).

Discussion

The large-scale (N = 2,756) multinational study reported here used child-level data to investigate potential societal sources of children's implicit associations linking *boys* with *math* and *girls* with *reading*. The findings inform our understanding of differential patterns of such implicit associations for girls and boys during childhood in three ways. First, girls' implicit associations of *boys* with *math* and *girls* with *reading* were significantly predicted by national boy advantages in TIMSS-4 math scores and national male advantages in socioeconomic standing (HDI) among adults. Second, these relations were obtained for implicit but not explicit measures

Figure 3
Children's Implicit Math/Reading–Gender Associations as a Function of Data Set Aggregate Mean Age



Note. Model-predicted estimates of implicit math/reading–gender associations as a function of data set-level aggregate age (see the Method section). The *y* axis represents model-predicted implicit association (in *D* scores); positive *D* scores indicate stronger associations of *math* = *boy* and *reading* = *girl*, negative *D* scores indicate stronger associations of *math* = *girl* and *reading* = *boy*, with 0 indicating an equally strong association of *math* with *boys* and *girls*. The *x* axis represents the mean ages of the data sets used in analyses. The regression line represents predicted values derived from model results (shaded regions are 95% CIs), taking into account the intercorrelations among variables, the nested data structure, and the differential sizes of the 16 data sets. Points represent aggregate mean predicted values for each sample. CIs = confidence intervals.

of children's beliefs linking gender and academic subjects. Third, the implicit associations became stronger as a function of children's age. These results were robust across several indices of national gender disparity. Below, we discuss each of these findings and also highlight that children's implicit associations and national gender disparities may bidirectionally reinforce one another over the course of development.

Connecting Societal Disparities and Implicit Associations in Girls

Why did observed national gender disparities relate to implicit associations of girls? Or framed more statistically: Why were the slopes in Figures 1 and 2 significantly positive for girls and relatively flat for boys? One possibility is that the societal stereotypes imputing low math ability or interest to girls and women may be especially salient to young girls due to "negativity biases." Psychological evidence suggests that people rapidly allocate attention to stimuli with negative emotional content, especially to negative information about the self (Baumeister et al., 2001; Soroka et al., 2019; Yiend, 2010). Such "negativity biases" have been found starting very early in childhood (Lagattuta & Kramer, 2017; Repacholi et al., 2016; Vaish et al., 2008). In many countries (including the ones tapped in this study), women and girls are negatively stereotyped for STEM disciplines (Cheryan & Markus, 2020; Cvencek et al., 2014; Leslie et al., 2015; Master et al., 2017; Nosek et al., 2009; Zhao et al., 2022). When these pernicious stereotypes (e.g., "girls are not good at math" or "women are less interested than men in math/engineering") are prevalent in society and carried in the media, girls are confronted with a negative quality about their in-group. Based on the "negativity bias," girls may be particularly likely to perceive and attend to this negative information about their gender in relation to math (even if not veridical). Research and theory suggest that such a pattern of information in the environment could influence the development of implicit associations (Greenwald & Lai, 2020; Meltzoff & Cvencek, 2019; Morehouse & Banaji, 2024; Payne et al., 2019). We acknowledge the speculative nature of this theorizing and encourage further research on this topic.

Implicit Gender-Linked Associations About Academic Subjects Develop Gradually

Using an adult sample, Nosek et al. (2009) showed that national indices of (a) gender diversity in the scientific workforce (e.g., interest, participation, and presence in positions of leadership) more broadly, and (b) gender inequality in STEM achievement more specifically, are related to implicit associations measured by the IAT, but not to explicit stereotypes. Our results align with these effects from adults insofar as national gender disparities were related more strongly to children's IAT scores than to their explicit stereotypes. We endorse the idea that implicit associations are built up through distributed learning experience starting in early childhood and involve deeply entrenched, "automatic" mental links (Baron & Banaji, 2006; Cvencek et al., 2011, 2023). This idea invites further replication and expanded research, but we think it fits in with extant work showing that children's valanced and nonvalanced associations about math and reading are evident at younger ages with implicit than explicit self-report measures (Cvencek, Brečić, et al., 2021).

The present investigation also offered a unique opportunity to examine finer age trends that may not have been detected in earlier individual studies with smaller sample sizes. Our current results pertaining to age effects are highly powered and allowed us to uncover an age-related increase in children's implicit associations between gender and academic subjects (Figure 3). This suggests that children's implicit associations with social groups continue to increase as children approach adolescence (within the ages tested here) and are malleable over a protracted time period (see also Baron, 2015; Halim et al., 2011; Master et al., 2021, for more discussion about developmental patterns in children's stereotypes).

Children Pay Attention to Societal Disparities

Children's acquisition of social knowledge is shaped by the explicit verbal messages from caregivers (Wang et al., 2022). Importantly, however, children also build social knowledge simply through observations of patterns in the social world. Research on social learning demonstrates that young children are intrinsically motivated to attend to and internalize the behavior they see modeled by others (Meltzoff, 2013; Meltzoff & Marshall, 2018; Miller et al., 2018). In particular, young children acquire beliefs, norms, values, and attitudes merely from observing the social interactions and disparities in the surrounding environment (Barragan & Meltzoff, 2021; Eccles & Wigfield, 2020; Martin & Ruble, 2010; Meltzoff & Gilliam, 2024; Skinner et al., 2020).

Although children have this capacity to learn from abstract patterns embodied in the social world, not all environmental cues are equally perceptible or impactful for the observer. For example, children may pay more attention to gender disparities in their immediate environment than they do to other, more distal societal disparities. That is to say, the socioeconomic standing of women in society may be salient to the *adults* in that society but not as salient to young children. For children, gender disparities in their school-related environment may be especially salient, and in the next two subsections, we explore how context, availability, and immediacy matter for children.

Multiple Measures of Academic Achievement

There are multiple ways of measuring children's academic achievement, and this raises interesting issues. Internationally, girls receive higher grades than boys do in all major subjects (Stoet & Geary, 2013; Voyer & Voyer, 2014). In contrast, boys outperform girls on standardized achievement tests and international competitions in math (e.g., Else-Quest et al., 2010; Hyde et al., 1990; but see Lindberg et al., 2010). This boy advantage on standardized tests is evident on both international (e.g., TIMSS, Programme for International Student Assessment) and national (e.g., national public exams) assessments (Cantley & McAllister, 2021). Why might children tend to form their associations about gender and academic subjects based on standardized scores rather than school grades?

We believe that scores on international assessment of math achievement are an especially salient source of associations about gender and academic subjects for children for at least four reasons. First, countries' TIMSS rankings are often covered by national media, and this is true for both high-achieving (e.g., Singapore; Ng, 2020) as well as low-achieving countries (e.g., Chile;

Salgado, 2020). Because of this media coverage, the country's scores may become a topic of everyday conversations, both inside families and inside classrooms. Second, for high-achieving countries, obtaining high TIMSS scores has become a matter of national pride (e.g., Croatia Week, 2020). Students in those countries work hard to prepare for standardized assessments and spend a great deal of time practicing tasks similar to the ones encountered in the actual TIMSS assessment (Holliday & Holliday, 2003). Third, the nation's results on standardized international assessments are often prominently discussed and used by policymakers and educators across nations in assessing their respective nation's comparative standing and developing new curricula for children (Mullis et al., 2016, 2020). Fourth, research has shown that raising the performance stakes can contribute to gender differences in mathematics performance by incentivizing math performance for boys and threatening the math performance of girls (e.g., Lyons et al., 2022). Taken altogether, it is conceivable that, at some level, assessments such as TIMSS may be construed by children as being a "higher stakes" (or a more objective) measure of underlying ability than everyday school grades.

Some Societal Disparities May Be More Evident to Children Than Others

We also found that the national gender gaps in TIMSS scores were linked to girls' implicit associations about gender and academic subjects more strongly than were the national socioeconomic gender inequalities (by comparing Figures 1 and 2, we see that the relation between the national math gender gap and girls' implicit associations was twice as strong as the relation between the national socio-economic gender inequality and girls' implicit associations). At least two speculations can be offered for these patterns.

First, for cultures/environments in which there are larger gender disparities between the representation of men and women in math-intensive fields, it has been theorized that girls do not readily see math achievement as opening up future opportunities (Baker & Jones, 1993; Eccles, 2011). Girls in this situation may perceive math to be less personally useful than boys do (Else-Quest et al., 2010). Thus, for girls, gender-linked differences in math achievement may be a reminder of limitations on the future opportunities within their society, and this is one reason why girls may implicitly believe that math (relative to reading) is more for boys than for them. A second (nonmutually exclusive) possibility is that the socioeconomic differences between men and women are not readily apparent to children (e.g., household finances within the family may be intermixed or combined across the spouses). Both of these possibilities may contribute to why the effects of the national-level disparities in HDI were weaker than the national-level disparities on TIMSS, but these are only two of several possible alternatives, and more research is needed.

Theorizing About Societal Gender Disparities and Children's Implicit Associations: Correlation Versus Causation

Given the correlational nature of the present investigation, we cannot draw conclusions about causal mechanisms. At the same time, the overall pattern of results permits us to offer some ideas which should be empirically tested in the future. We note that national socioeconomic gender inequalities and national math gender gaps are already present in society before any individual child develops his or her own implicit associations about gender and school subjects. Thus, we believe that an individual girl's associations are influenced by the prevailing math gender gaps in her culture, although we also underscore that the relation is likely to be bidirectional in interesting ways. Girls who are reared in cultures with large national math gender gaps favoring boys may acquire implicit associations about gender and math at an early age, and this may, in turn, contribute to gender differences in interests, performance, and participation in math (which further reinforces the associations; Galdi et al., 2014). Other societal-level factors such as differential treatment of the genders by parents and teachers (Eccles, 2011), and more specifically, female teachers' math anxiety (which especially influences young girls; Beilock et al., 2010; Dowker et al., 2016; Gunderson et al., 2012; Levine & Pantoja, 2021) may also come into play. In sum, already existing gender disparities in society may drive the emergence of implicit associations in children, which in turn, as children grow up, feed into maintaining the existing disparities, such that gender disparities and implicit associations about gender and academic subjects reinforce one another over time. Such bidirectional and mutually reinforcing mechanisms could lead some societies to have and maintain larger gender disparities in standardized math achievement than others. This would also be consistent with our finding that implicit associations were stronger in older children, suggesting that longer, more repeated exposures to societal disparities, with attendant "over-learning," may engender stronger implicit associations.

Limitations and Future Research

This study had several strengths, including: (a) a large sample of children, (b) the use of both implicit and explicit measures in the same children, and (c) well-validated measures of national gender disparities. Despite these strengths, we acknowledge four limitations.

First, because we only had data from 16 data sets (from five nations), our power to detect the main effects of our focal national variables and child-level implicit associations was less than optimal. However, post hoc power analyses (see the Results section) showed that the cross-level interaction tests were well-powered for the model of the implicit dependent variable and, to a lesser extent, the explicit dependent variable as well. As such, we were able to detect theorized significant relations between national gender disparities and children's implicit associations. The cross-validation robustness checks suggest, however, that some of the effects were dependent on specific individual samples. Future research should examine the robustness of the findings reported here by replicating the study with more nations.

Second, because the data sets were previously collected as convenience samples, we have no way of knowing whether the implicit association data we analyzed are representative of their nation or if the data happen to come from families who are relatively more or less educated or in some other way different from each nation's sociodemographic composition. (Even if the child samples used in the present study are not completely representative of their nations, it is worth noting that the same theorized relations between national disparities and individual children's associations were found, on average, within each nation.)

Third, without having assessed national gender gaps in reading achievement, it is difficult to know for sure whether the implicit gender-linked associations are uniquely associated with *math* achievement differences. Given the relative nature of the IAT, the implicit associations linking *boys* with *math* could also reflect the associations linking *girls* with *reading* or both associations simultaneously. Future research combining IAT measures with other implicit measures (e.g., the Affective Misattribution Procedure; Vuletich et al., 2020) would enable us to better examine the joint and individual operation of both sides of the association.

Fourth, many of the effect sizes of relations demonstrated here were relatively modest. Nonetheless, understanding these national relations during childhood is of importance because even statistically small effects can have large impacts when they involve meaningful situations that happen repeatedly over time to large numbers of children, such as the framing of certain academic activities as being for one gender and not the other (Martin & Ruble, 2010; Master et al., 2021). Research with implicit cognition has shown that stereotypic or biased events affecting the same person repeatedly, across time and space ("distributed learning"), can have especially strong and meaningful cumulative effects (e.g., Greenwald et al., 2015).

Conclusion

Children link gender to academic subjects as early as elementary school, but the sources of these early associations are understudied. In this article, we found that national patterns of gender disparities in society are related to implicit associations between gender and academic subjects for girls more strongly than for boys. The current work expands our understanding of societal-level contributions to children's implicit associations between gender, math, and reading, and it provides a more detailed analysis of age effects (stronger implicit associations in older children) than has been available in previous studies involving fewer participants.

References

Baker, D. P., & Jones, D. P. (1993). Creating gender equality: Cross-national gender stratification and mathematical performance. *Sociology of Education*, *66*(2), 91–103. https://doi.org/10.2307/2112795

Baron, A. S. (2015). Constraints on the development of implicit intergroup attitudes. *Child Development Perspectives*, *9*(1), 50–54. https://doi.org/10.1111/cdep.12105

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes. Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science*, *17*(1), 53–58. https://doi.org/10.1111/j.1467-9280.2005.01664.x

Barragan, R. C., & Meltzoff, A. N. (2021). Human infants can override possessive tendencies to share valued items with others. *Scientific Reports*, *11*(1), Article 9635. https://doi.org/10.1038/s41598-021-88898-x

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(5), 1860–1863. https://doi.org/10.1073/pnas.0910967107

Bell, A., & Jones, K. (2015). Explaining fixed effects: Random effect modeling of time-series cross-sectional and panel data. *Political Science Research and Methods*, *3*(1), 133–153. https://doi.org/10.1017/psrm.2014.7

Breda, T., Jouini, E., Napp, C., & Thebault, G. (2020). Gender stereotypes can explain the gender-equality paradox. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(49), 31063–31069. https://doi.org/10.1073/pnas.2008704117

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), Article 9. https://doi.org/10.5334/joc.10

Cantley, I., & McAllister, J. (2021). The gender similarities hypothesis: Insights from a multilevel analysis of high-stakes examination results in mathematics. *Sex Roles*, *85*(7–8), 481–496. https://doi.org/10.1007/s11199-021-01234-5

Charlesworth, T. E. S., & Banaji, M. R. (2022). Patterns of implicit and explicit stereotypes III: Long-term change in gender stereotypes. *Social Psychological & Personality Science*, *13*(1), 14–26. https://doi.org/10.1177/1948550620988425

Cheryan, S., & Markus, H. R. (2020). Masculine defaults: Identifying and mitigating hidden cultural biases. *Psychological Review*, *127*(6), 1022–1052. https://doi.org/10.1037/rev0000209

Croatia Week. (2020, December 9). *Croatian fourth-graders outperform in math and science*. https://www.croatiaweek.com/croatian-fourth-graders-outperform-in-math-and-science/

Cvencek, D., Brečić, R., Gaćeša, D., & Meltzoff, A. N. (2021). Development of math attitudes and math self-concepts: Gender differences, implicit–explicit dissociations, and relations to math achievement. *Child Development*, *92*(5), e940–e956. https://doi.org/10.1111/cdev.13523

Cvencek, D., Brečić, R., Sanders, E. A., Gaćeša, D., Skala, D., & Meltzoff, A. N. (2023). Am I a good person? Academic correlates of explicit and implicit self-esteem during early childhood. *Child Development*. Advance online publication. https://doi.org/10.1111/cdev.14052

Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2016). Implicit measures for preschool children confirm self-esteem's role in maintaining a balanced identity. *Journal of Experimental Social Psychology*, *62*, 50–57. https://doi.org/10.1016/j.jesp.2015.09.015

Cvencek, D., Kapur, M., & Meltzoff, A. N. (2015). Math achievement, stereotypes, and math self-concepts among elementary-school students in Singapore. *Learning and Instruction*, *39*, 1–10. https://doi.org/10.1016/j.learninstruc.2015.04.002

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math-gender stereotypes in elementary school children. *Child Development*, *82*(3), 766–779. https://doi.org/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*(1), 73–91. https://doi.org/10.1016/j.jecp.2013.07.018

Cvencek, D., Meltzoff, A. N., Maddox, C. D., Nosek, B. A., Rudman, L. A., Devos, T., Dunham, Y., Baron, A. S., Steffens, M. C., Lane, K., Horcajo, J., Ashburn-Nardo, L., Quinby, A., Srivastava, S. B., Schmidt, K., Aidman, E., Tang, E., Farnham, S., Mellott, D. S., … Greenwald, A. G. (2021). Meta-Analytic use of balanced identity theory to validate the Implicit Association Test. *Personality and Social Psychology Bulletin*, *47*(2), 185–200. https://doi.org/10.1177/0146167220916631

Darlington, R. B., & Hayes, A. F. (2017). *Regression analysis and linear models*. Guilford Press.

Dasgupta, N. (2013). Implicit attitudes and beliefs adapt to situations: A decade of research on the malleability of implicit prejudice, stereotypes, and the self-concept. In P. Devine & A. Plant (Eds.), *Advances in experimental social psychology* (Vol. 47, pp. 233–279). Burlington Academic Press.

De Houwer, J., & Boddez, Y. (2022). Bias in implicit measures as instances of biased behavior under suboptimal conditions in the laboratory. *Psychological Inquiry*, *33*(3), 173–176. https://doi.org/10.1080/1047840X.2022.2106755

Dehdarirad, T., Sotudeh, H., & Freer, J. (2019). Bibliometric mapping of microbiology research topics (2012-16): A comparison by socioeconomic development and infectious disease vulnerability values. *FEMS Microbiology Letters*, *366*(2), Article fnz004. https://doi.org/10.1093/femsle/fnz004

del Río, M. F., Susperreguy, M. I., Strasser, K., Cvencek, D., Iturra, C., Gallardo, I., & Meltzoff, A. N. (2021). Early sources of children's math achievement in Chile: The role of parental beliefs and feelings about math. *Early Education and Development*, *32*(5), 637–652. https://doi.org/10.1080/10409289.2020.1799617

Del Río, M. F., Strasser, K., Cvencek, D., Susperreguy, M. I., & Meltzoff, A. N. (2019). Chilean kindergarten children's beliefs about mathematics: Family matters. *Developmental Psychology*, *55*(4), 687–702. https://doi.org/10.1037/dev0000658

Dovidio, J. F., & Kunst, J. R. (2022). Delight in disorder: Inclusively defining and operationalizing implicit bias. *Psychological Inquiry*, *33*(3), 177–180. https://doi.org/10.1080/1047840X.2022.2106756

Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in Psychology*, *7*, Article 508. https://doi.org/10.3389/fpsyg.2016.00508

Eagly, A. H., & Chaiken, S. (2005). Attitude research in the 21st century: The current state of knowledge. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 743–767). Lawrence Erlbaum.

Early Childhood Science, Technology, Engineering, and Mathematics Working Group. (2017). *Early STEM matters: Providing high-quality experiences for all young learners*. Erikson Institute. https://www.researchconnections.org/childcare/resources/33456

Eccles, J. S. (2011). Understanding women's achievement choices: Looking back and looking forward. *Psychology of Women Quarterly*, *35*(3), 510–516. https://doi.org/10.1177/0361684311414829

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, Article 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*(1), 103–127. https://doi.org/10.1037/a0018053

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*(1), 297–327. https://doi.org/10.1146/annurev.psych.54.101601.145225

Galdi, S., Cadinu, M., & Tomasetto, C. (2014). The roots of stereotype threat: When automatic associations disrupt girls' math performance. *Child Development*, *85*(1), 250–263. https://doi.org/10.1111/cdev.12128

Galdi, S., Mirisola, A., & Tomasetto, C. (2017). On the relations between parents' and children's implicit and explicit academic gender stereotypes. *Psicologia e Sociedade*, *12*(2), 215–238. https://doi.org/10.1482/87248

Gawronski, B., Ledgerwood, A., & Eastwick, P. W. (2022). Implicit bias ≠ bias on implicit measures. *Psychological Inquiry*, *33*(3), 139–155. https://doi.org/10.1080/1047840X.2022.2106750

Gozzi, N., Tizzoni, M., Chinazzi, M., Ferres, L., Vespignani, A., & Perra, N. (2021). Estimating the effect of social inequalities on the mitigation of COVID-19 across communities in Santiago de Chile. *Nature Communications*, *12*(1), Article 2429. https://doi.org/10.1038/s41467-021-22601-6

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Green, P., MacLeod, C. J., & Alday, P. (2023). *Package 'simr'*. https://cran.r-project.org/web/packages/simr/simr.pdf

Greenwald, A. G., & Banaji, M. R. (2017). The implicit revolution: Reconceiving the relation between conscious and unconscious. *American Psychologist*, *72*(9), 861–871. https://doi.org/10.1037/amp0000238

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology*, *108*(4), 553–561. https://doi.org/10.1037/pspa0000016

Greenwald, A. G., & Lai, C. K. (2020). Implicit social cognition. *Annual Review of Psychology*, *71*(1), 419–445. https://doi.org/10.1146/annurev-psych-010419-050837

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/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*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*(1), 17–41. https://doi.org/10.1037/a0015575

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*(3–4), 153–166. https://doi.org/10.1007/s11199-011-9996-2

Halim, M. L., Ruble, D. N., & Amodio, D. M. (2011). From pink frilly dresses to 'one of the boys': A social-cognitive analysis of gender identity development and gender bias. *Social and Personality Psychology Compass*, *5*(11), 933–949. https://doi.org/10.1111/j.1751-9004.2011.00399.x

Hamaker, E. L., & Muthén, B. (2020). The fixed versus random effects debate and how it relates to centering in multilevel modeling. *Psychological Methods*, *25*(3), 365–379. https://doi.org/10.1037/met0000239

Harter, S., & Pike, R. (1984). The pictorial scale of perceived competence and social acceptance for young children. *Child Development*, *55*(6), 1969–1982. https://doi.org/10.2307/1129772

Holliday, W. G., & Holliday, B. W. (2003). Why using international comparative math and science achievement data from TIMSS is not helpful. *The Educational Forum*, *67*(3), 250–257. https://doi.org/10.1080/00131720309335038

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*(2), 139–155. https://doi.org/10.1037/0033-2909.107.2.139

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist*, *73*(1), 1–2. https://doi.org/10.1037/amp0000263

Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes, gender identification, and math-related outcomes: A prospective study of female college students. *Psychological Science*, *18*(1), 13–18. https://doi.org/10.1111/j.1467-9280.2007.01841.x

Klasen, S. (2017). *UNDP's gender-related measures: Current problems and proposals for fixing them* (Discussion papers, no. 220). https://hdl.handle.net/10419/157265

Klasen, S., & Schüler, D. (2011). Reforming the gender-related development index and the gender empowerment measure: Implementing some specific proposals. *Feminist Economics*, *17*(1), 1–30. https://doi.org/10.1080/13545701.2010.541860

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lagattuta, K. H., & Kramer, H. J. (2017). Try to look on the bright side: Children and adults can (sometimes) override their tendency to prioritize negative faces. *Journal of Experimental Psychology: General*, *146*(1), 89–101. https://doi.org/10.1037/xge0000247

Leslie, S.-J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *347*(6219), 262–265. https://doi.org/10.1126/science.1261375

Levine, S. C., & Pantoja, N. (2021). Development of children's math attitudes: Gender differences, key socializers, and intervention approaches. *Developmental Review*, *62*, Article 100997. https://doi.org/10.1016/j.dr.2021.100997

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin*, *136*(6), 1123–1135. https://doi.org/10.1037/a0021276

Lyons, E., Mesghina, A., & Richland, L. E. (2022). Complicated gender gaps in mathematics achievement: Elevated stakes during performance as one explanation. *Mind, Brain and Education*, *16*(1), 36–47. https://doi.org/10.1111/mbe.12312

Marsh, H. W., Parker, P. D., Guo, J., Basarkod, G., Niepel, C., & Van Zanden, B. (2021). Illusory gender-equality paradox, math self-concept, and frame-of-reference effects: New integrative explanations for multiple paradoxes. *Journal of Personality and Social Psychology*, *121*(1), 168–183. https://doi.org/10.1037/pspp0000306

Martin, C. L., & Ruble, D. N. (2010). Patterns of gender development. *Annual Review of Psychology*, *61*(1), 353–381. https://doi.org/10.1146/annurev.psych.093008.100511

Master, A., Cheryan, S., Moscatelli, A., & Meltzoff, A. N. (2017). Programming experience promotes higher STEM motivation among first-grade girls. *Journal of Experimental Child Psychology*, *160*, 92–106. https://doi.org/10.1016/j.jecp.2017.03.013

Master, A., Meltzoff, A. N., & Cheryan, S. (2021). Gender stereotypes about interests start early and cause gender disparities in computer science and engineering. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(48), Article e2100030118. https://doi.org/10.1073/pnas.2100030118

Mathieu, J. E., Aguinis, H., Culpepper, S. A., & Chen, G. (2012). Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. *Journal of Applied Psychology*, *97*(5), 951–966. https://doi.org/10.1037/a0028380

Meltzoff, A. N. (2013). Origins of social cognition: Bidirectional self-other mapping and the "Like-Me" hypothesis. In M. R. Banaji & S. A. Gelman (Eds.), *Navigating the social world: What infants, children, and other species can teach us* (pp. 139–144). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199890712.003.0025

Meltzoff, A. N., & Cvencek, D. (2019). How stereotypes shape children's STEM identity and learning. In P. K. Kuhl, S.-S. Lim, S. Guerriero, & D. Van Damme (Eds.), *Developing minds in the digital age: Towards a science of learning for 21st century education* (pp. 37–47). OECD Publishing.

Meltzoff, A. N., & Gilliam, W. S. (2024). Young children & implicit racial biases. *Daedalus*, *153*(1), 65–83. https://doi.org/10.1162/daed_a_02049

Meltzoff, A. N., & Marshall, P. J. (2018). Human infant imitation as a social survival circuit. *Current Opinion in Behavioral Sciences*, *24*, 130–136. https://doi.org/10.1016/j.cobeha.2018.09.006

Miller, D. I., Nolla, K. M., Eagly, A. H., & Uttal, D. H. (2018). The development of children's gender-science stereotypes: A meta-analysis of 5 decades of U.S. Draw-a-Scientist studies. *Child Development*, *89*(6), 1943–1955. https://doi.org/10.1111/cdev.13039

Morehouse, K. N., & Banaji, M. R. (2024). The science of implicit race bias: Evidence from the Implicit Association Test. *Daedalus*, *153*(1), 21–50. https://doi.org/10.1162/daed_a_02047

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/isc/publications.html

Mullis, I. V. S., Martin, M. O., & Loveless, T. (2016). *20 years of TIMSS: International trends in mathematics and science achievement, curriculum, and instruction*. TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/isc/publications.html

Muradoglu, M., Cimpian, J. R., & Cimpian, A. (2023). Mixed-effects models for cognitive development researchers. *Journal of Cognition and Development*, *24*(3), 307–340. https://doi.org/10.1080/15248372.2023.2176856

Ng, K. G. (2020, December 8). Singapore students top maths, science rankings for second consecutive edition of international study. *The Straits Times*. https://www.straitstimes.com/singapore/singapore-students-top-maths-science-rankings-for-second-consecutive-edition-of

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. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10593–10597. https://doi.org/10.1073/pnas.0809921106

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

Payne, B. K., Vuletich, H. A., & Brown-Iannuzzi, J. L. (2019). Historical roots of implicit bias in slavery. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(24), 11693–11698. https://doi.org/10.1073/pnas.1818816116

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*(4), 233–248. https://doi.org/10.1080/1047840X.2017.1335568

Ratliff, K. A., & Smith, C. T. (2022). Implicit bias as automatic behavior. *Psychological Inquiry*, *33*(3), 213–218. https://doi.org/10.1080/1047840X.2022.2106764

Régner, I., Thinus-Blanc, C., Netter, A., Schmader, T., & Huguet, P. (2019). Committees with implicit biases promote fewer women when they do not believe gender bias exists. *Nature Human Behaviour*, *3*(11), 1171–1179. https://doi.org/10.1038/s41562-019-0686-3

Repacholi, B. M., Meltzoff, A. N., Toub, T. S., & Ruba, A. L. (2016). Infants' generalizations about other people's emotions: Foundations for trait-like attributions. *Developmental Psychology*, *52*(3), 364–378. https://doi.org/10.1037/dev0000097

Salgado, F. A. (2020, December 16). *Resultados TIMSS 2019: ¿Qué podemos hacer para mejorar?* [TIMSS 2019 results: What can we do to improve?]. Noticias Los Ríos. https://www.noticiaslosrios.cl/2020/12/16/resultados-timss-2019-que-podemos-hacer-para-mejorar/

Schmader, T., Bareket-Shavit, C., & Baron, A. S. (2022). Beyond awareness: The many forms of implicit bias and its implications. *Psychological Inquiry*, *33*(3), 156–161. https://doi.org/10.1080/1047840X.2022.2106752

Skinner, A. L., Olson, K. R., & Meltzoff, A. N. (2020). Acquiring group bias: Observing other people's nonverbal signals can create social group biases. *Journal of Personality and Social Psychology*, *119*(4), 824–838. https://doi.org/10.1037/pspi0000218

Soroka, S., Fournier, P., & Nir, L. (2019). Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(38), 18888–18892. https://doi.org/10.1073/pnas.1908369116

Stoet, G., & Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within- and across-nation assessment of 10 years of Program for International Student Assessment data. *PLOS ONE*, *8*(3), Article e57988. https://doi.org/10.1371/journal.pone.0057988

Troumbis, A. Y. (2020). Testing the socioeconomic determinants of COVID-19 pandemic hypothesis with aggregated Human Development Index. *Journal of Epidemiology and Community Health*, *75*(4), 414–415. https://doi.org/10.1136/jech-2020-215986

United Nations Development Programme. (2019). *Human development report 2019: Beyond income, beyond averages, beyond today: Inequalities in human development in the 21st century*.

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social–emotional development. *Psychological Bulletin*, *134*(3), 383–403. https://doi.org/10.1037/0033-2909.134.3.383

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, *140*(4), 1174–1204. https://doi.org/10.1037/a0036620

Vuletich, H. A., Kurtz-Costes, B., Cooley, E., & Payne, B. K. (2020). Math and language gender stereotypes: Age and gender differences in implicit biases and explicit beliefs. *PLOS ONE*, *15*(9), Article e0238230. https://doi.org/10.1371/journal.pone.0238230

Wang, M. M., Cardarelli, A., Leslie, S.-J., & Rhodes, M. (2022). How children's media and teachers communicate exclusive and essentialist views of science and scientists. *Developmental Psychology*, *58*(8), 1455–1471. https://doi.org/10.1037/dev0001364

Yiend, J. (2010). The effects of emotion on attention: A review of attentional processing of emotional information. *Cognition and Emotion*, *24*(1), 3–47. https://doi.org/10.1080/02699930903205698

Zhao, S., Setoh, P., Storage, D., & Cimpian, A. (2022). The acquisition of the gender-brilliance stereotype: Age trajectory, relation to parents' stereotypes, and intersections with race/ethnicity. *Child Development*, *93*(5), e581–e597. https://doi.org/10.1111/cdev.13809

Received November 24, 2022
Revision received May 3, 2024
Accepted May 3, 2024 ■

Cvencek, D., Sanders, E. A., del Río, M. F., Susperreguy, M. I., Strasser, K., Brečić, R., Gaćeša, D., Skala, D., Tomasetto, C., Galdi, S., Cadinu, M., Kapur, M., Passolunghi, M. C., Rueda Ferreira, T. I., Mirisola, A., Mariani, B., & Meltzoff, A. N. (2024). National disparities favoring males are reflected in girls' implicit associations about gender and academic subjects. *Developmental Psychology*. https://doi.org/10.1037/dev0001797

Supporting Information for

"National Disparities Favoring Males Are Reflected in Girls' Implicit Associations About Gender and Academic Subjects"

Developmental Psychology

2024

1. Datasets

1.1. Datasets for Implicit Association Measures. Descriptive statistics of each of the 16 datasets used in analyses are described in Table S1.

Table S1

Descriptive Statistics of Analytic Sample Dataset Characteristics

Dataset	Year of implicit data collection	*N*	Age		Child IAT *D*-score		National standardized math test gender gap (TIMSS–4)		National socioeconomic gender inequality	
			M	(*SE*)	*M*	(*SE*)	*M*	(*SE*)	HDI	GGI
Chile										
del Río et al. (2019)	2016	179	5.61	(0.03)	0.04	(0.03)	1	(3.2)	1.04	.30
del Río et al. (2021)	2018	363	7.30	(0.06)	0.07	(0.02)	1	(3.2)	1.04	.28
Croatia										
Cvencek et al. (2018)	2018	96	7.12	(0.12)	0.04	(0.04)	12	(2.7)	1.01	.29
Cvencek et al. (2020)	2020	335	8.39	(0.07)	0.10	(0.02)	12	(3.1)	1.01	.28
Italy										
Galdi et al. (2014)	2011	80	6.00	(0.00)	0.07	(0.05)	9	(3.0)	1.03	.32
Galdi et al. (2017)	2012	43	6.65	(0.06)	0.15	(0.09)	9	(3.0)	1.03	.33
Mariani (2012)	2010	14	6.66	(0.07)	0.31	(0.22)	15	(2.5)	1.03	.32
Passolunghi et al. (2014)	2012	241	10.22	(0.14)	0.19	(0.03)	9	(3.0)	1.03	.33
Tomasetto et al. (2012)	2010	153	6.71	(0.02)	0.12	(0.05)	15	(2.5)	1.03	.32
Singapore										
Cvencek et al. (2014)	2011	147	9.57	(0.13)	0.10	(0.03)	-4	(3.0)	1.03	.31
Cvencek et al. (2015)	2013	120	9.38	(0.16)	0.05	(0.03)	-4	(3.0)	1.03	.30
United States										
Cvencek & Meltzoff (2018)	2018	32	7.33	(0.30)	0.08	(0.06)	7	(1.9)	1.01	.28
Cvencek & Meltzoff (2019a)	2019	293	7.82	(0.07)	0.09	(0.02)	11	(2.9)	1.01	.28
Cvencek & Meltzoff (2019b)	2019	125	7.19	(0.08)	0.07	(0.04)	11	(2.9)	1.01	.28
Cvencek & Meltzoff (2019c)	2019	292	9.52	(0.15)	0.13	(0.03)	11	(2.9)	1.01	.28
Cvencek et al. (2011)	2006	243	8.65	(0.09)	0.16	(0.02)	8	(1.6)	1.01	.30

Note. Child IAT *D*-score = Implicit math/reading–gender association strength. TIMSS–4 = Trends in International Math and Science Study, Grade 4. HDI = Human Development Index male-to-female ratio. GGI = Global Gender Gap Index. Positive values for national math gender gap (TIMSS Grade 4 boy–girl differences) indicate that boys scored higher than girls. Larger values for national socioeconomic gender inequality (HDI, GGI) indicate larger male advantage in socioeconomic development over females.

1.2. Sources for National Math Gender Gap Data. National data for math achievement gender gaps were obtained from the Trends in International Math and Science Study (TIMSS) report published for each cycle of the test (Mullis et al., 2004, 2008, 2012, 2016, 2020). All reports were accessed on April 23, 2020. As described in the main text, we identified TIMSS data from the same year as each dataset's math/reading–gender association data collection. If TIMSS data were unavailable for the exact year of the child-level implicit/explicit data collection, the closest year prior was substituted (all national data used was collected within three years of each dataset's child-level data collection year). For example, this meant that for IAT data collected in 2020, TIMSS data published in 2020 were used, and for IAT data collected in 2018, TIMSS data published in 2016 were used, and so forth.

1.3. Sources for National Socioeconomic Gender Inequality Data Reported in the Main Text. U.N. Human Development Index (HDI) data used for the national socioeconomic gender inequality were obtained from the UNDP Download Center on May 4, 2021 (UNDP, 2020).

1.4. Using Global Gender Gap Index Data as a Robustness Check (See Section 3.2 for results). The Global Gender Gap Index (GGI) was also used (in place of the HDI M/F ratio) in alternative analyses to test the robustness of our findings on national socioeconomic gender inequality reported in the main text (see Section 3.2 below). GGI scores were integrated into the analyzed datasets matched by the year of math/reading–gender association data collection. The World Economic Forum computes the national GGI scores annually as an average of female-to-male ratios on four sub-indexes: (a) economic participation and opportunity (labor force participation rate, wage equality, estimated earned income, legislators and senior officials/managers, and professional and technical workers), (b) educational attainment (literacy rate and enrollment in primary, secondary, and tertiary education), (c) health and survival (sex ratio at birth and healthy life expectancy), and (d) political empowerment (women in parliament, women in ministerial positions, and years with female head of state). GGI data were obtained from the World Economic Forum's Global Gender Gap Reports on March 31, 2023 (World Economic Forum, 2006, 2010, 2011, 2012, 2013, 2016, 2018, 2019). Data were linked to IAT data using the same principle described for TIMSS data above.

As computed by the World Economic Forum, the GGI can be interpreted as a percentage of the gender gaps that have closed within a given country. In other words, higher GGI scores indicate greater gender *parity* (i.e., smaller gender gap, or women = men). For our analyses to be directionally consistent with the HDI M/F ratio, we recoded the national GGI scores. For example, the original GGI score for Italy in 2011 was 0.68, indicating that the gender gap was 68% *closed*; we recoded that GGI score to be 0.32 (1.0 – 0.68). All GGI scores in our analyses were recoded in this fashion, so that higher scores indicated greater gender *non-parity* (i.e., larger gender gap of men > women). Therefore, the GGI used in all of our analyses ranged from 0 (indicating gender parity) to 1 (indicating non-parity favoring men over women). That way, the HDI and GGI scores were coded in the same direction.

2. Multilevel Modeling

2.1. Predictor Coding. For ease of results interpretation (particularly so that the intercept could be interpreted as the mean association strength across the dataset), we effect-coded both binary predictors (1, -1), and standardized remaining predictors as *z*-scores prior to analysis. Similarly, the 2-way interactions were created using the product of the effect-coded gender variable (female = 1) and the standardized versions of the gender gap predictors.

2.2. Software. For all models, we used maximum likelihood estimates from the *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017) packages in *R*.

2.3. Effect Sizes. The interpretation of the approximate *d*-values for coefficient estimates in our results is as follows. (a) For the intercept, *d* is the model-predicted number of *standard deviations* the grand mean of the dependent variable (DV) is from zero. This is meaningful in the case of modeling implicit association measures because it provides a way to quantify the magnitude of the association. (b) For the slope of a binary predictor that is effect-coded (as was the case in the present study for time of data collection and gender), *d* is the predicted difference between the group coded as 1 and the grand mean of the DV *in standard deviations*, holding all else constant (the coefficient is doubled to obtain the difference between two groups). (c) For the slope of a predictor that is continuous, *d* is the predicted change in the DV per standard deviation increase in the predictor, all other things held constant. (d) For the 2-way gender interactions tested in the present study, *d* is the difference between the predictor–DV relation for the group coded 1 (girls) and the predictor–DV relation (in standard deviations), again holding all else constant; yet, with this noted, we also note that because the meaning of interactions can be difficult to conceptualize, we have also visualized them as part of our reporting of results (see Figures 1 and 2 in the main text, as well as Figure S1 in Section 5.2 below).

2.4. Missingness. A very small proportion of the original sample of $N = 2{,}925$ was missing predictor information on age (5.2%) or gender (0.6%) variables. We excluded these children from analyses on the assumption that data were missing at random, because there was no information in any of the studies we drew from indicating that missingness could be related to children's actual ages or genders. Further, our own preliminary tests showed that missingness levels were not significantly correlated with either dependent variable (*r*s = -.01 and .00 for missingness on gender with implicit and explicit measures, respectively; *r*s = -.01 and -.03 for missingness on age with implicit and explicit measures, respectively). Thus, excluding these children from analyses should have no biasing impact on model results' point estimates. Further, given the small proportion of missingness and the large sample size at Level 1, excluding these children from analyses should also have no biasing impact on model results precision.

3. Multiple Robustness Checks

In addition to the robustness checks reported in the main text, we conducted further robustness checks on our model results testing the effects of the gender disparity predictors. These are listed in the Sections 3.1 to 3.3.

3.1. Controlling for TIMSS and HDI M/F Ratio Means. We evaluated the stability of the gender disparity effects (on implicit associations and explicit stereotypes) after controlling for mean levels of national achievement and socioeconomic indices. To that end, we re-specified the same models shown in the main text, but with *mean* dataset TIMSS–4 math scores and *mean* dataset HDI levels included in the models (standardized into *z*-scores). Findings from these models showed that the main effect of TIMSS–4 math gender gaps on implicit association strength was no longer significant (but same positive pattern with the dependent variable remained); however, the focal interaction effects between gender and the TIMSS–4 math gender gap and the interaction between gender and the HDI M/F ratio remained significant and in the same direction. Similar model results were found for the explicit stereotype: HDI M/F ratio still uniquely negatively predicted explicit stereotype (and as in the original model results, there were no gender interactions with TIMSS–4 math gender gap or HDI M/F ratio). For clarity and focus, and because mean TIMSS–4 math scores and mean HDI scores were strongly related to the mean gender disparity variables (i.e., inducing multicollinearity), we present the simpler model results in the main text.

3.2. Interchanging Global Gender Gap Index (GGI) for HDI M/F Ratio as Socioeconomic Indicator. To ensure that the socioeconomic gender disparity effect (and interaction with child gender) was not solely due to the unique properties of the HDI M/F measure, we re-specified our models with GGI as the socioeconomic disparity predictor (standardized into *z*-scores) instead of the HDI M/F predictor. Those results are provided in Table S2 and show that there is no change to the pattern of findings. We retain the HDI M/F ratio as our predictor in the main text for consistency with prior research.

Table S2

Robustness Check: Multilevel Model Fixed Effect Results With GGI as Socioeconomic Gender Disparity Predictor

Fixed effect	Implicit math/reading–gender association				Explicit gender stereotype			
	Coeff	(*SE*)	*p*	*d*	*Coeff*	(*SE*)	*p*	*d*
Intercept (mean association/stereotype)	**0.100**	**(0.009)**	**<.001**	**0.25**	**0.091**	**(0.026)**	**<.001**	**0.08**
Time period (1 = after 2014, dataset level)	-0.010	(0.017)	.553	-0.03	**-0.248**	**(0.052)**	**<.001**	**-0.21**
Aggregate child age (dataset level) (*z*)	**0.025**	**(0.009)**	**.007**	**0.06**	-0.046	(0.031)	.136	-0.04
Mean-centered child age (child level) (*z*)	**0.034**	**(0.008)**	**<.001**	**0.08**	-0.017	(0.027)	.524	-0.01
TIMSS-4 math boy–girl gap (dataset level) (*z*)	**0.019**	**(0.009)**	**.026**	**0.05**	0.038	(0.026)	.139	0.03
GGI ratio (dataset level) (*z*)	0.010	(0.017)	.569	0.02	**-0.189**	**(0.052)**	**<.001**	**-0.16**
Gender (1 = girl, child level)	**0.032**	**(0.008)**	**<.001**	**0.08**	-0.044	(0.025)	.079	-0.04
Child gender × TIMSS–4 boy–girl gap	**0.031**	**(0.008)**	**<.001**	**0.08**	-0.025	(0.025)	.325	-0.02
Child gender × GGI ratio	**0.031**	**(0.009)**	**<.001**	**0.08**	-0.021	(0.027)	.434	-0.02

Note. *N* = 2,756 children from 16 datasets across five nations. TIMSS–4 = Trends in International Math and Science Study, Grade 4. GGI = Global Gender Gap Index. All predictors standardized into *z*-scores (*z*) except for the two binary predictors, Time period and Gender, which were effect-coded (see Method in the main text). Parameter estimates derived from 3-level random intercept models estimated with full information maximum likelihood using lme4 in *R*. Effect size *d* = approximate Cohen's *d*, computed as the coefficient divided by the pooled *SD*, with the pooled *SD* = square root of the sum of the variance components. Observed *p*-values reported; significant coefficients at the .05 level are boldfaced.

3.3. Interchanging Continuous With Dichotomized Dataset-Level National Disparity Predictors. Although there was no significant skew in TIMSS–4 math gender gap and HDI M/F ratio at the dataset level (Level 2), at the child level (Level 1) significant skew was present in each (*p*s < .001). Because skew remained present in both predictors after typical data transformations (i.e., logged values, inverse of values, square root of values), as another check on the robustness of our results, we re-specified models using dichotomized versions of both the TIMSS–4 math gender gap and HDI M/F ratio predictors. Results for the implicit associations were substantively the same, with the exception that the main effect of TIMSS–4 math gender gap on implicit association levels was no longer significant (but the same positive relation was found); however, the interaction effects between gender and the TIMSS–4 math gender gap and the interaction between gender and the HDI M/F ratio remained significant and in the same direction. Similar model results were found for the explicit stereotype: With dichotomized versions of the TIMSS–4 gender gap and HDI M/F ratio predictors, HDI M/F ratio was still uniquely negatively predictive of explicit stereotype (and, like the original model results, there were no gender interactions with TIMSS–4 math gender gap or HDI M/F ratio).

4. Statistical Significance of National Gender Disparities

To assess whether there were significant national gender disparities in the current study's pool of 16 datasets, we used OLS multiple linear regression to estimate the significance of the intercept for each of the two gender disparity variables, controlling for nation membership using effect coding ($n = 5$ nations; the U.S. was used as the reference category as it had the largest representation in the dataset). Specifically, the intercept, adjusted for nation membership, indicated that the national boy–girl gap in math (i.e., TIMSS–4) averaged 6.00 points ($SE = 0.64$), which was significantly different from zero (i.e., different from a null of no boy–girl gap), $t(11) = 9.45$, $p < .001$, $d = 2.60$. Similarly, the difference in HDI M/F ratio had an adjusted average of 1.02 ($SE = 0.0003$), which was also significantly different from perfect equality of 1.00, $t(11) = 71.95$, $p < .001$, $d = 19.83$. These analyses establish that, for the 16 datasets included here, there were statistically significant national gender disparities favoring boys in math achievement and men in socioeconomic standing.

5. Explicit Stereotype

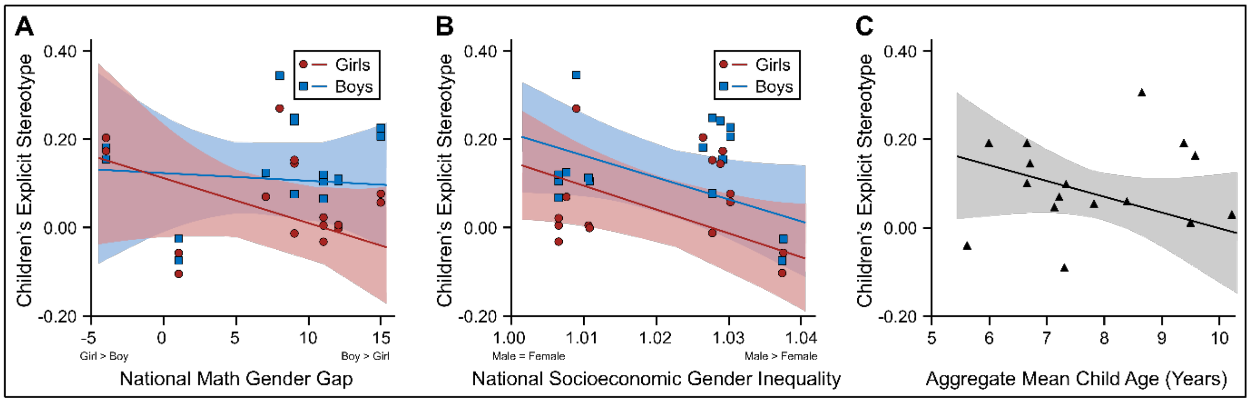
5.1. Explicit Stereotype Measures Used Specifically With Italian Datasets. As noted in the main text, the Italian datasets used slight variations of the explicit stereotype measure than were used with the other datasets. Across the five Italian datasets, the response options varied somewhat, as did the scoring procedure (see below). In some of their work, the authors performed a designed intervention. For any studies involving an experimental manipulation to influence stereotypes (e.g., stereotype activation), the authors specifically provided us with the data only for participants in the control condition (i.e., participants who did *not* undergo the manipulation).

In four Italian datasets (Galdi et al., 2014, 2017; Mariani, 2012; Tomasetto et al., 2012), children were shown a picture of a girl and a boy described as "good at school" and asked which child they thought was better at math (and separately, language), or if they were equally good. The scores on the math item and the language item were combined to arrive at the explicit stereotype score, which ranged from -2, indicating that girls are better at math than boys, to +2, indicating that boys are better at math than girls. A score of 0 indicated a belief that girls and boys are equally good at math and reading. In the fifth Italian dataset (Passolunghi et al., 2014), three items with a 5-point Likert response scale were used to assess the endorsement of gender stereotypes about math (e.g., "In your opinion, who is better at math between girls and boys?"). Explicit stereotype scores were computed so they ranged from -2, indicating that girls are better at math than boys, to +2, indicating that boys are better at math than girls. A score of 0 indicated a belief that girls and boys are equally good at math.

5.2. Explicit Stereotype Results. As reported in the main text, there was a statistically significant explicit gender stereotype favoring boys when it came to math relative to reading, $p < .003$, $d = 0.08$. In contrast with the results for implicit associations, only two variables uniquely predicted explicit stereotypes. First, as shown in Table 2 of the main text, data collection time period was negatively predictive of the explicit stereotype level, $p = .003$: datasets collected more recently (2014–2020) were predicted to be 0.121 lower in stereotype levels compared to average (in other words, 0.242 points per year lower than datasets collected between 2006–2013) suggesting a notable *decrease* in gender stereotypes over time when assessed using explicit self-report measures (see below for further discussion). Second, although math gender gaps (TIMSS–4) were not predictive of explicit stereotypes (see Figure S1, Panel A), higher national socioeconomic gender inequality (HDI M/F ratio) was negatively related to explicit stereotype level, $p = .030$ (Figure S1, Panel B). For every standard deviation increase in HDI M/F ratio, there was a predicted decrease of 0.098 points in explicit stereotype levels. (These results are interpreted in Section 5.3 in the context of a well-known "gender-equality paradox.") As noted above (see Section 3.1), when we controlled for mean levels of TIMSS and HDI M/F ratio, the effect of socioeconomic gender inequality favoring men on stereotype persisted. (For completeness, Figure S1, Panel C provides plots the relation between explicit stereotype and age, which was not statistically significant.)

Figure S1

Children's Explicit Stereotypes as a Function of National and Dataset Means



Note. Model-predicted estimates of explicit stereotypes as a function of: (A) national math gender gaps and child gender, (B) national socioeconomic gender inequalities and child gender, and (C) dataset-level aggregate age. In all panels, the *y*-axis represents model predicted explicit stereotype (see above and Method in main text); positive scores indicate stronger stereotypes favoring boys, and negative scores indicate stronger stereotypes favoring girls, with 0 indicating the absence of stereotype favoring one gender over another. In Panel A, the *x*-axis represents national math gender gap (TIMSS–4 boy–girl difference; 0 indicates no difference, see Method in main text); in Panel B, the *x*-axis represents national socioeconomic gender inequality (HDI M/F ratio; 1 indicates no difference, see Method in main text); in Panel C, the *x*-axis represents the mean child ages of the datasets. Regression lines represent predicted values derived from model results, taking into account the intercorrelations among variables, the nested data structure, and the differential sizes of the 16 datasets. Shaded regions indicate 95% CIs for each line's slope, and points represent aggregate predicted values for each dataset.

5.3. Explicit Stereotype Supplemental Discussion. Two results with explicit measures warrant further discussion, which is provided in the following two sub-sections.

5.3.1. Differences in Explicit Stereotypes. In the current study, explicit gender stereotypes favored boys, although with a small effect size. This result should be considered alongside results in a child dataset collected more recently and reported in Tang et al. (2024), who found that explicit stereotypes about math ability and math interest significantly favored girls. This raises the idea that children's explicit gender stereotypes about math are malleable and may be shifting over historical time. Datasets gathered between 2006 and 2014—as used in the current study—may reflect explicit stereotypes favoring boys because this was a pervasive societal stereotype at that time in history (and in the contexts studied), whereas the newer data collected by Tang et al. (2024) may be capturing a genuine (and more recent) change in children's stereotypes in the United States. For completeness, we also note that there are two technical differences between the current study and Tang et al.'s that may also be useful to bear in mind: (a) the cross-cultural nature of our study in contrast to the exclusively United States sample used by Tang et al. (2024), and (b) procedural differences (e.g., we used the Harter and Pike picture set to present the stereotype problems and Tang et al. asked a purely verbal questions).

5.3.2. Gender-Equality Paradox. A second result warranting discussion is that higher national socioeconomic gender inequality predicted a *lower* magnitude of *explicit* stereotypes for both boys and girls. This fits with the "gender-equality paradox" (Breda et al., 2020) whereby gender gaps in math achievement and occupational interests in math-intensive fields are stronger in more developed and more egalitarian societies when explicit measures are used. The most comprehensive work on this gender-equality paradox was reported by Breda et al. (2020), using individual-level data on the explicit math attitudes of 15-year-old students. They found that the magnitude of the stereotype linking math with men was positively associated with national indices of socioeconomic gender equality (including the Gross Domestic Product and the overall Human Development Index).

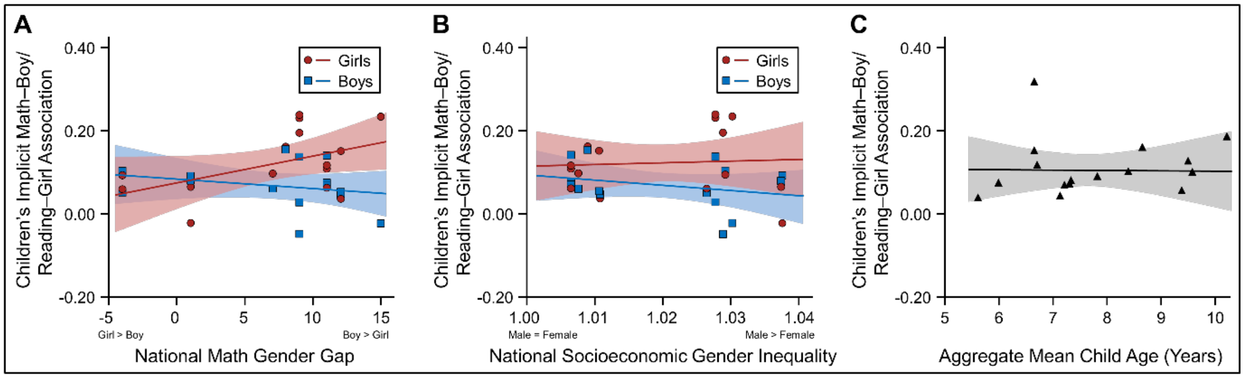
In the published literature, three explanations have been offered for this paradox (which arises with explicit measures). First, students in more developed nations may not need to study as hard to have satisfactory career prospects in math-related fields (which will undergird their material security), thus affording less instrumental value to math in more developed societies (Breda et al., 2020; Goldman & Penner, 2016). Second, high-income parents may be more involved with their children's educational choices (Williams & Best, 1990), which has been offered as an explanation for why parents in more developed nations may transmit gender norms regarding educational aptitudes and choices—including gender stereotypes about math—earlier than parents from economically less developed nations (Reardon et al., 2019). Finally, with regards to STEM in particular, more developed societies also tend to have more difficult math and science curricula and higher performance standards in those fields, both of which may heighten gender stereotypes about STEM (Breda et al., 2020; Mann & DiPrete, 2016). There is not a settled opinion on the reasons for the gender equality paradox, and more research is needed.

6. Graphic Illustrations Using Raw Data

The figures in the main text all present model-predicted estimates of implicit math/reading–gender associations, which we believe is the most appropriate statistical representation of the data because it takes all dependencies into account. For the sake of transparency, we are also presenting the raw mean relations in Figs S2 and S3.

Figure S2

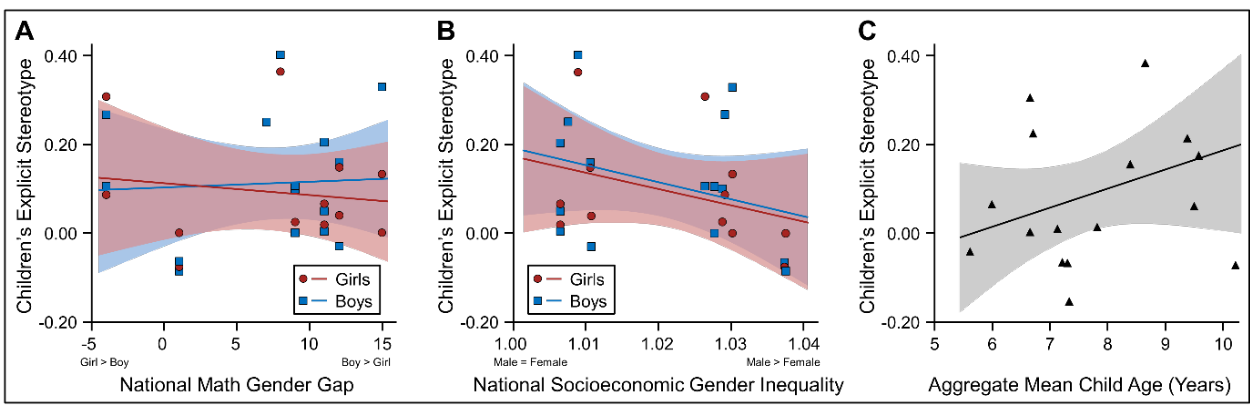
Children's Raw Implicit Math/Reading–Gender Associations as a Function of National and Dataset Means



Note. Raw means of implicit stereotypes (*D*-scores; see Methods) as a function of: (A) national math gender gaps and child gender, (B) national socioeconomic gender inequalities and child gender, and (C) dataset-level aggregate age. In all panels, the *y*-axis represents raw (not the modeled-predicted estimates of) implicit stereotypes as indicated by *D*-scores; positive *D*-scores indicate stronger associations of *math* = *boy* and *reading* = *girl*, negative *D*-scores indicate stronger associations of *math* = *girl* and *reading* = *boy*, with 0 indicating an equally strong association of *math* with *boys* and *girls*. In Panel A, the *x*-axis represents national math gender gap (TIMSS–4 boy–girl difference; 0 indicates no difference); in Panel B, the *x*-axis represents national socioeconomic gender inequality (HDI M/F ratio; 1 indicates no difference); in Panel C, the *x*-axis represents the mean child ages of the datasets. Regression lines represent simple linear associations. Shaded regions indicate 95% CIs and points represent mean values for each dataset.

Figure S3

Children's Raw Explicit Stereotypes as a Function of National and Dataset Means



Note. Raw means of explicit stereotypes as a function of: (A) national math gender gaps and child gender, (B) national socioeconomic gender inequalities and child gender, and (C) dataset-level aggregate age. In all panels, the *y*-axis represents raw explicit stereotype scores; positive scores indicate stronger stereotypes favoring boys, and negative scores indicate stronger stereotypes favoring girls, with 0 indicating the absence of stereotype favoring one gender over another. In Panel A, the *x*-axis represents national math gender gap (TIMSS–4 boy–girl difference; 0 indicates no difference); in Panel B, the *x*-axis represents national socioeconomic gender inequality (HDI M/F ratio; 1 indicates no difference); in Panel C, the *x*-axis represents the mean child ages of the datasets. Regression lines represent simple linear associations. Shaded regions indicate 95% CIs and points represent mean values for each dataset.

7. References

An asterisk (*) indicates studies included in the analyses reported in this Supplement.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Breda, T., Jouini, E., Napp, C., & Thebault, G. (2020). Gender stereotypes can explain the gender-equality paradox. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(49), 31063–31069. https://doi.org/10.1073/pnas.2008704117

*Cvencek, D., Brečić, R., Gaćeša, D., & Meltzoff, A. N. (2018). *Math–gender stereotypes of Croatian children and their parents* [Unpublished raw data]. University of Washington.

*Cvencek, D., Brečić, R., Gaćeša, D., & Meltzoff, A. N. (2020). *Gender stereotypes about math in Croatian elementary-school children* [Unpublished raw data]. University of Washington.

*Cvencek, D., Kapur, M., & Meltzoff, A. N. (2015). Math achievement, stereotypes, and math self-concepts among elementary-school students in Singapore. *Learning and Instruction*, *39*, 1–10. https://doi.org/10.1016/j.learninstruc.2015.04.002

*Cvencek, D., & Meltzoff, A. N. (2018). *Inter-generational transfer of beliefs about math* [Unpublished raw data]. University of Washington.

*Cvencek, D., & Meltzoff, A. N. (2019a). *Developmental emergence of math–gender stereotypes and math self-concepts* [Unpublished raw data]. University of Washington.

*Cvencek, D., & Meltzoff, A. N. (2019b). *Development of math–gender stereotypes in elementary-school children* [Unpublished raw data]. University of Washington.

*Cvencek, D., & Meltzoff, A. N. (2019c). *Science in action: Math–gender stereotypes of child visitors to a science museum* [Unpublished raw data]. University of Washington.

*Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child Development*, *82*(3), 766–779. https://doi.org/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. https://doi.org/10.1016/j.jecp.2013.07.018

*del Río, M. F., Strasser, K., Cvencek, D., Susperreguy, M. I., & Meltzoff, A. N. (2019). Chilean kindergarten children's beliefs about mathematics: Family matters. *Developmental Psychology*, *55*(4), 687–702. https://doi.org/10.1037/dev0000658

*del Río, M. F., Susperreguy, M. I., Strasser, K., Cvencek, D., Iturra, C., Gallardo, I., & Meltzoff, A. N. (2021). Early sources of children's math achievement in Chile: The role of parental beliefs and feelings about math. *Early Education and Development*, *32*(5), 637–652. https://doi.org/10.1080/10409289.2020.1799617

*Galdi, S., Cadinu, M., & Tomasetto, C. (2014). The roots of stereotype threat: When automatic associations disrupt girls' math performance. *Child Development*, *85*(1), 250–263. https://doi.org/10.1111/cdev.12128

*Galdi, S., Mirisola, A., & Tomasetto, C. (2017). On the relations between parents' and children's implicit and explicit academic gender stereotypes. *Psicologia Sociale*, *12*(2), 215–238. https://doi.org/10.1482/87248

Goldman, A. D., & Penner, A. M. (2016). Exploring international gender differences in mathematics self-concept. *International Journal of Adolescence and Youth*, *21*(4), 403–418. https://doi.org/10.1080/02673843.2013.847850

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Mann, A., & DiPrete, T. A. (2016). The consequences of the national math and science performance environment for gender differences in STEM aspiration. *Sociological Science*, *3*(25), 568–603. https://doi.org/10.15195/v3.a25

*Mariani, B. (2012). *Maschi & matematica, femmine & Italiano: Associazioni implicite ed esplicite tra materie scolastiche e genere a sei anni* [Unpublished master dissertation]. University of Bologna.

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/isc/publications.html

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 international results in mathematics*. TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/isc/publications.html

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/isc/publications.html

Mullis, I. V. S., Martin, M. O., Foy, P., Olson, J. F., Preuschoff, C., Erberber, E., Arora, A., & Galia, J. (2008). *TIMSS 2007 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/isc/publications.html

Mullis, I. V. S., Martin, M. O., Gonzales, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 International mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/isc/publications.html

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

Reardon, S. F., Fahle, E. M., Kalogrides, D., Podolsky, A., & Zárate, R. C. (2019). Gender achievement gaps in US school districts. *American Educational Research Journal*, *56*(6), 2474–2508. https://doi.org/10.3102/0002831219843824

Tang, D., Meltzoff, A. N., Cheryan, S., Fan, W., & Master, A. (2024). Longitudinal stability and change across a year in children's gender stereotypes about four different STEM fields [Advance online publication]. *Developmental Psychology*. https://doi.org/10.1037/dev0001733

*Tomasetto, C., Galdi, S., & Cadinu, M. (2012). Quando l'implicito precede l'esplicito: Gli stereotipi di genere sulla matematica in bambine e bambini di 6 anni [When the implicit precedes the explicit: Gender stereotypes about math in 6-year-old girls and boys]. *Psicologia Sociale*, *7*(2), 169–186. https://doi.org/10.1482/37693

UNDP. (2020). *Table 1: Human Development Index and its components*. https://hdr.undp.org/en/content/download-data

Williams, J. E., & Best, D. L. (1990). *Measuring sex stereotypes: A multination study*. Sage.

World Economic Forum. (2006). *The Global Gender Gap Report 2006*. https://www3.weforum.org/docs/WEF_GenderGap_Report_2006.pdf

World Economic Forum. (2010). *The Global Gender Gap Report 2010*. https://www3.weforum.org/docs/WEF_GenderGap_Report_2010.pdf
World Economic Forum. (2011). *The Global Gender Gap Report 2011*. https://www3.weforum.org/docs/WEF_GenderGap_Report_2011.pdf
World Economic Forum. (2012). *The Global Gender Gap Report 2012*. https://www3.weforum.org/docs/WEF_GenderGap_Report_2012.pdf
World Economic Forum. (2013). *The Global Gender Gap Report 2013*. https://www3.weforum.org/docs/WEF_GenderGap_Report_2013.pdf
World Economic Forum. (2016). *The Global Gender Gap Report 2016*. https://www3.weforum.org/docs/GGGR16/WEF_Global_Gender_Gap_Report_2016.pdf
World Economic Forum. (2018). *The Global Gender Gap Report 2018*. https://www3.weforum.org/docs/WEF_GGGR_2018.pdf
World Economic Forum. (2019). *The Global Gender Gap Report 2020*. https://www3.weforum.org/docs/WEF_GGGR_2020.pdf