AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Bulletin

ISSN: 0033-2909

2024, Vol. 150, No. 12, 1363–1396
https://doi.org/10.1037/bul0000456

The Development of Children's Gender Stereotypes About STEM and Verbal Abilities: A Preregistered Meta-Analytic Review of 98 Studies

David I. Miller[1], Jillian E. Lauer[2], Courtney Tanenbaum[1], and Lauren Burr[1]
[1] Human Services Division, American Institutes for Research
[2] Faculty of Education, Cambridge University

This meta-analysis studied the development of ability stereotypes that could limit girls' and women's participation in science, technology, engineering, and mathematics (STEM) fields, as well as contribute to boys' underachievement in reading and writing. We integrated findings from 98 studies measuring children's gender stereotypes about STEM and verbal abilities. The data comprised 145,204 children (ages 4–17) from 33 nations across more than 40 years (1977–2020). Preregistered analyses showed why prior researchers have reached diverging conclusions about the onset, change, and extent of these stereotypes in childhood and adolescence. Contrary to some prior conclusions, math stereotypes favoring male ability were minimal on average (0.11 *SD*s from gender neutrality). Stereotypes were instead far stronger for computer science, engineering, and physics (0.51 *SD*s), which favored male ability by age 6. Girls increasingly endorsed pro-male STEM stereotypes with age. Pro-female verbal ability stereotypes were also substantial (0.46 *SD*s), emerging by age 8 and becoming more female-biased with age. Additionally, STEM stereotypes were weaker for Black than White U.S. participants, as predicted. Unexpectedly, however, boys' STEM stereotypes declined before age 13 but increased thereafter, revealing an *asymmetric* development across STEM versus verbal domains. We integrated developmental intergroup theory and social role theory to explain this asymmetry, considering both cognitive and sociocultural processes. The early emergence of verbal stereotypes and certain STEM stereotypes (e.g., engineering) means that they have ample time to affect later downstream outcomes such as domain-specific confidence and interests.

Public Significance Statement
This quantitative review of nearly 100 studies shows that, by age 6, children already think that boys are better than girls at computer science and engineering. With age, girls increasingly believe in male superiority in these technical fields—a stereotype that could potentially limit girls' future aspirations. In contrast, children hold far more gender-neutral beliefs about math ability. Children also think that girls are much better in verbal domains like reading and writing, which could contribute to boys' underachievement in those domains.

Keywords: gender stereotypes, child development, meta-analysis, STEM abilities, verbal abilities

Supplemental materials: https://doi.org/10.1037/bul0000456.supp

Rong Su served as action editor.
David I. Miller https://orcid.org/0000-0003-3457-7994
Jillian E. Lauer https://orcid.org/0000-0002-4859-3576
This study's inclusion criteria, confirmatory hypotheses, analysis plan, and other protocol details were preregistered on the Open Science Framework and can be accessed at https://osf.io/7mnsw. All data, analysis scripts, and study materials are also available on the Open Science Framework and can be accessed at https://osf.io/29egh/. Portions of this article's findings were presented at the 2022 Network Gender & STEM Conference in Munich, Germany.

The authors have no conflicts of interest to disclose. The National Science Foundation (Grant DUE-1920401) funded this project awarded to David I. Miller and Courtney Tanenbaum. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

The authors thank Abigail Jefferys, Alberto Guzman-Alvarez, Eben Witherspoon, and Robert Schwarzhaupt for their assistance in screening and coding studies; Megha Joshi and Charlie Ebersole for the support in creating the interactive data tool; Martha Ramirez for administrative support; Martyna Citkowicz and Ryan Williams for methodological advice; Andrei Cimpian, Beth Kurtz-Costes, Catherine Riegle-Crumb, Jo Boaler, and Larry Hedges for the guidance throughout as advisory board members; the 2018 Meta-Analysis Training Institute instructors and workshop attendees for their contributions to the early methodological and conceptual design; the many study authors who responded to the requests for information; and methodological input from scholars from the Methods of Synthesis and Integration Center at the American Institutes for Research.

continued

Children's gender stereotypes about abilities in science, technology, engineering, and mathematics (STEM) fields have been a focus of research for several decades, and with good reason. Despite earning most of the U.S. bachelor's degrees in life science and social sciences, women are underrepresented in other STEM fields, both in college and the workforce (Cheryan et al., 2017). In 2021, women were 40% of mathematicians, 33% of physical scientists, 25% of computer scientists, and 16% of engineers in the United States (National Center for Science and Engineering Statistics [NCSES], 2023, Table 1-2). Although many factors likely explain these postsecondary and career gaps, ability stereotypes are a foundational construct that drive other psychological phenomena (e.g., gender bias, motivational processes) that could shape educational trajectories and career plans starting at early ages (Wang & Degol, 2013). In a large, nationally representative U.S. sample of 15-year-olds, only 3% of girls expected to work in a computing or engineering career at age 30, compared to 16% of boys (Organisation for Economic Co-operation and Development, 2016). These aspirational differences may partly arise from gender stereotypes that form early in children's development and strengthen with age (Arthur et al., 2008).

This article presents findings from a comprehensive meta-analysis of more than 4 decades of studies on children's gender stereotypes about STEM abilities. We aimed to bring clarity to an extensive body of research that has found mixed empirical results. These mixed results arise in part from variation across STEM fields (McGuire et al., 2022; Tang et al., 2024). For instance, in one study, high schoolers believed in gender equality in biology ability but male superiority in physics ability (Lerdpornkulrat et al., 2012). Results are mixed even within the widely studied domain of math stereotypes. Although some studies have found the expected stereotype of superior male math ability (e.g., Hargreaves et al., 2008; Keller, 2001; Muzzatti & Agnoli, 2007), others have found that "perceived gender differences were minimal for math" (Heyman & Legare, 2004, p. 227; see also Martinot & Désert, 2007; Plante et al., 2009). Some have even found that, "girls were viewed as more competent in maths/science than boys" (Rowley et al., 2007, p. 157; see also Evans et al., 2011; Kurtz-Costes et al., 2014).

These mixed results likely reflect differences in how children in different communities and of different ages construct beliefs based on the cultural messages they receive. For instance, children's stereotypes may strengthen with age as their cultural exposure grows, contributing to differences across studies with varying average participant ages (e.g., D. I. Miller, Nolla, et al., 2018). Heterogeneity could also arise from differences in cultural environments and social identities (e.g., communities of color may hold different stereotypes; Skinner et al., 2021). Even among the same participants, children may provide different answers based on the measurement features (e.g., item wording). Instead of emphasizing aggregate means, this meta-analysis focused on investigating how cultural, demographic, and measurement factors may explain heterogeneity.

This meta-analysis also aimed to understand gender stereotype development in a contrasting domain: verbal abilities. Studying verbal ability stereotypes, which tend to favor girls and women (e.g., Rowley et al., 2007), was important for three main reasons:

- *boys' underperformance in reading and writing:* Despite small to no gender gaps in average math test scores, girls robustly outperform boys in reading and writing by about 0.3 to 0.5 *SD*s across the world (D. I. Miller & Halpern, 2014; Reilly et al., 2019). This gender gap underscores the importance of understanding possible contributors such as children's gender stereotypes about verbal abilities.

- *contrast to STEM stereotypes:* Contrasting STEM stereotypes with verbal stereotypes informs understanding of variation across male-typed and female-typed domains. As a preview of later results, STEM and verbal stereotypes showed an unexpected *asymmetry* in age-related differences (which helped to extend theories of stereotype development).

- *potential contribution to STEM gender gaps:* Students often compare their strengths across domains when evaluating their abilities. Among students with equal math performance, those with higher verbal performance tend to have lower math confidence (Möller & Marsh, 2013; Wigfield et al., 2020). Relatedly, pro-female verbal ability stereotypes might draw girls away from quantitative fields and attract boys to them, contributing to STEM gender gaps (Kurtz-Costes et al., 2014).

We define domain-specific *ability stereotypes* as beliefs about the competence of different social groups to accomplish tasks or duties in an academic, cognitive, or occupational domain. As such, the term captures a wide, multifaceted set of beliefs, including stereotypes about "raw" or innate intelligence (e.g., "Who is more naturally gifted in physics?") and stereotypes about learned competencies that depend on environment and experience (e.g., achievement stereotypes such as "Who does better on physics tests?"). We also define ability stereotypes as distinct from *cultural fit* stereotypes about interests (e.g., "Who likes physics?"; Master et al., 2021a) and associations (e.g., "Is math for boys or girls?"; Cvencek et al., 2011), which this meta-analysis did not include (but for reviews on cultural fit stereotypes, see Cheryan et al., 2015; Master & Meltzoff, 2020).

We draw on theories from developmental science, social psychology, and sociology to frame understanding of how and when gender stereotypes about STEM and verbal abilities emerge and evolve through childhood and adolescence. These frameworks aim to explain broad patterns of gender development across many domains (e.g., sports, clothing, toys). Our empirical investigation informs these broader theories via two routes. First, STEM and verbal stereotypes can inform broader understanding of why women

David I. Miller played a lead role in conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing–original draft, and writing–review and editing. Jillian E. Lauer played a supporting role in conceptualization, data curation, funding acquisition, investigation, methodology, and writing–review and editing. Courtney Tanenbaum played a supporting role in conceptualization, funding acquisition, project administration, supervision, writing–original draft, and writing–review and editing. Lauren Burr played a supporting role in data curation, investigation, methodology, project administration, resources, and writing–review and editing.

Correspondence concerning this article should be addressed to David I. Miller, Human Services Division, American Institutes for Research, 10 South Riverside Plaza, Suite 600, Chicago, IL 60606, United States. Email: dimiller@air.org

and men pursue different fields of education and employment (Eccles & Wigfield, 2020; Su & Rounds, 2015). Second, focused domain-specific investigations can identify nuances and contextual influences not presently accounted for in broader theories of stereotype development, ultimately extending those theories.

Theoretical Perspectives on Gender Stereotype Development

Three prominent frameworks informed the meta-analytic predictions: (a) cognitive theories of gender development, (b) social role theory, and (c) intersectionality. Though overlapping in some of their philosophical underpinnings, these frameworks also offered unique perspectives that informed distinct predictions (e.g., age vs. cultural variation).

Cognitive Theories of Gender Development

Cognitive theories of gender development emphasize that children are active participants in their gender development (Martin & Ruble, 2004). As "gender detectives," children seek to identify typical female and male behaviors, motivated by a desire to interpret the social world around them. Specific theories vary some in their underlying assumptions and relative focus on cognitive versus environmental processes:

- *cognitive-developmental theory:* Per Kohlberg's (1966) cognitive-developmental theory, children's understanding of gender progresses through distinct stages, which then drives other aspects of children's gendered behavior and thinking. These stages develop during ages 2 to 6 as children recognize that (a) they are a boy or girl (*gender identity*), (b) this identity does not change over time (*gender stability*), and (c) changes in gender-typed appearances and behaviors do not affect this identity (*gender consistency*).

- *gender schema theory:* Compared to cognitive-developmental theory, Bem's (1983) gender schema theory deemphasizes a stage-like model of gender identity, while more heavily drawing from information processing theories of basic cognitive processes (e.g., memory distortions; see also Martin & Halverson, 1981). Children develop organizing cognitive structures (schemas) used to search for, encode, and remember cultural information about gender, often in stereotype-reinforcing ways.

- *developmental intergroup theory:* Developmental intergroup theory aims to integrate and extend prior theories in several ways, such as explaining why children attend to certain social categories like gender and ignore others (Bigler & Liben, 2006). In this theory, both *internal* cognitive factors (e.g., multiple classification skills) and *external* environmental factors (e.g., teachers' use of gender labels) interact to shape children's stereotyping and prejudice (for applications to gender, see Arthur et al., 2008; Liben, 2014). As discussed later, this theory also shares many similar assumptions with social role theory, which we later exploit when building bridges across traditionally unconnected perspectives.

Despite their differences, these theories all assume that children's gender identity motivates their active construction of gender stereotypes. Young children are eager to find, remember, and construct positive information about their gender, driven by in-group gender bias as early as ages 3–5 (Dunham et al., 2016; Halim et al., 2017; Kurtz-Costes, Defreitas, et al., 2011). Young children can even spontaneously form counter-cultural beliefs: As Martin (1993) hypothesized, "before children know much about trait stereotypes, they claim the positive traits for their own sex" (p. 80).

Change in Childhood. With age, young children begin to integrate this in-group/out-group mentality with stereotype knowledge that children actively construct based on increased environmental exposure (e.g., with mass media, toys, parents, peers). Cognitive theories of gender development broadly agree that children's gender identity and stereotype knowledge start to emerge around ages 2–3 and strengthen until ages 5–6 (Martin et al., 2002). Though in-group bias remains strong at ages 5–6, children's stereotypes also start to more closely resemble the normative views of adults (e.g., boys begin to learn pro-female trait stereotypes; Serbin et al., 1993).

After ages 5–6, children undergo cognitive maturation that allows for more sophisticated reasoning about gender (e.g., successfully remembering counterstereotypic examples like female engineers; Bigler & Liben, 2006, pp. 75–78). Developmental scholars have described ages 5–6 as a period of "peak gender rigidity" for measures assessing *prescriptive stereotypes* about how the world should be (e.g., "Who *should* study math?"; Ruble et al., 2007). That is, children aged 5–6 are especially likely to say that certain roles or traits are appropriate for only one gender. But this rigidity wanes during elementary school (Halim, 2016; Trautner et al., 2005).

Different Patterns for Descriptive Stereotypes. Compared to prescriptive stereotypes, however, developmental patterns are different for *descriptive stereotypes* about how the world currently is[1] (Leaper, 2015, p. 813). Descriptive stereotypes are far more relevant to this article's focus because STEM and verbal ability stereotypes are descriptive (e.g., "Who is good at math?"), not prescriptive (e.g., "Who should do math?"). Unlike prescriptive stereotypes, gender-stereotyped responses tend to strengthen from ages 5–6 to ages 10–11 for descriptive stereotypes, as found in a prior meta-analysis (Signorella et al., 1993; see results on p. 165 for descriptive stereotypes about "Who is usually … ?" or "Who is better … ?"). Signorella et al. attributed this shift to an increased knowledge of cultural stereotypes. Hence, children appear to continually learn about stereotypes with age, but children at ages 10–11 may be more flexible than children at ages 5–6 in applying those stereotypes to saying who should do certain activities.

Change in Adolescence. Descriptive and prescriptive stereotypes continue to show distinct developmental patterns in adolescence. Of most relevance to our meta-analysis, studies on descriptive gender stereotypes have typically found increasing traditionality in adolescence from ages 11 to 16[2] (Alfieri et al., 1996;

[1] Other scholars have made the same theoretical distinction using different words such as *personal beliefs* versus *cultural knowledge* (Halim & Ruble, 2010, p. 501) or *attitudes toward stereotypes* versus *knowledge of stereotypes* (Signorella et al., 1993). We prefer the terms *prescriptive* and *descriptive* as more informative labels that are less prone to misinterpretation, consistent with some other scholars (Koenig, 2018; Leaper, 2015; Tobin et al., 2010).

[2] Studies on *prescriptive* stereotypes have found more mixed results in adolescence, including curvilinear change (Crouter et al., 2007) or decreasing traditionality (Katz & Ksansnak, 1994; Klaczynski et al., 2020; Liben & Bigler, 2002; Marcell et al., 2011).

Neff & Terry-Schmitt, 2002; Skinner et al., 2021; Starr et al., 2023; L. A. Wood et al., 2022). The exact mechanisms for this increasing traditionality are unclear. Some scholars have proposed that *gender intensification*—an increased pressure to conform traditional gender roles in adolescence (Hill & Lynch, 1983)—should strengthen traditional stereotypes, despite the counteracting force of cognitive maturation (see Galambos et al., 2009, for a review). Yet, other scholars have argued that gender intensification might not directly apply to stereotypes (Klaczynski et al., 2020, p. 428), and others have noted the mixed evidence for domain-general gender intensification (Perry & Pauletti, 2011; Priess et al., 2009). Regardless of the exact mechanism, these findings suggest that ability stereotypes might also increase in traditionality in adolescence (a later section considers other factors beyond gender intensification).

Development of In-Group Bias. We expected that increasing traditionality in ability stereotypes would co-occur with declining in-group bias. This bias tends to wane after ages 5–6 (e.g., Dunham et al., 2016), including as expressed in measures of descriptive stereotypes (e.g., Powlishta et al., 1994). In our context, we defined in-group bias as generally assigning more ability to one's own gender more than others (e.g., a boy who says boys are good at math *and* reading).

Developmental scholars often attribute declining in-group bias to three factors: (a) a growing ability to classify others along multiple dimensions (Bigler & Liben, 2006); (b) an increasing focus on treating others as individuals rather than groups (Aboud, 2008; Misch et al., 2022); and (c) advances in moral reasoning (e.g., increased attention to fairness and equity; Leaper, 2015, p. 814). These changes should theoretically weaken in-group biases from ages 5 to 11 across multiple domains (e.g., also racial or ethnic prejudice; Raabe & Beelmann, 2011).

A later section unpacks how the expected age-related weakening of in-group bias, coupled with increasing traditionality, leads to specific Age × Gender predictions for this present meta-analysis. The theories described next (e.g., social role theory) help place these predictions in a broader cultural context, while attending to other social identities such as race.

Social Role Theory

Eagly and Wood's (2012) social role theory integrates social psychological perspectives on stereotyping with biological, economic, developmental, sociological, and evolutionary perspectives to explain the origins of sex differences and similarities in behavior. Beyond its common application to research on adults, this theory also attends to socialization mechanisms (W. Wood & Eagly, 2012, pp. 65–70), placing children's stereotype development in a changing cultural context (D. I. Miller, Nolla, et al., 2018; Wilbourn & Kee, 2010).

A core tenet of social role theory is that gender stereotypes form based on observing the behaviors of men and women enacting distinct social roles (e.g., occupational roles). People often assume that differences in external behaviors (e.g., working as an engineer) reflect differences in internal traits (e.g., math ability) via a cognitive process called *correspondent inference* (Gilbert & Malone, 1995). For instance, when observing women caring for children (an *external* behavior), perceivers often infer that women possess *internal* communal traits such as warmth and nurturance (Koenig & Eagly, 2014).

Social role theory aligns with and extends the previously reviewed cognitive theories. For instance, correspondence inference closely aligns with what developmental intergroup theory labels *implicit attribution* (Bigler & Liben, 2006, Figure 1). Social role theory extends these developmental perspectives by attending to the distal causes of gender stereotypes (e.g., division of labor, cultural and evolutionary forces).

Considering the distal causes of gender stereotypes offers a unique perspective on the accuracy of ability stereotypes. Per social role theory, ability stereotypes partly reflect the reality of career role differences (e.g., more male than female engineers). However, perceivers may not always account for other constraints on behavior, such as other reasons for the paucity of women in STEM (e.g., interests, discrimination, child-rearing duties; Ceci et al., 2014; Cheryan et al., 2017). Observing real career differences could produce stereotypes about male math ability, even if those inferences conflict with small to no gender gaps in average math test scores (e.g., Else-Quest et al., 2010). Career differences could also indirectly shape children's stereotypes via transmission by adults. That is, gender stereotypes tend to be shared among members of society (Eagly et al., 2020) who could transmit these beliefs to children (e.g., via mass media, parenting, teaching), even if children lack direct exposure to STEM workers.

Per social role theory, changes in occupational roles should impact societal gender stereotypes. Supporting this view, a meta-analysis of 16 nationally representative public opinion polls found change in U.S. gender stereotypes since the 1940s (Eagly et al., 2020). Of most relevance to our meta-analysis, general intelligence stereotypes favored men in the 1940s but shifted toward equality over time. In the most recent poll in 2018, 86% of adults said women and men were equally intelligent, with more saying that women are more intelligent (9%) than those saying men are (5%). Eagly et al. argued this shift was likely rooted in dramatic changes since the 1940s in U.S. women's educational attainment and labor force participation. Such changes provided increased opportunities to observe women in cognitively demanding roles (ultimately shifting societal stereotypes in this explanation). These societal considerations informed several predictions about variation across time, cultures, and disciplinary fields.

Intersectional Frameworks

The intersection of gender with race and ethnicity is a critical consideration when examining the development of gender identity and stereotypes. Gender is not a homogenous construct: Intersectional frameworks posit that people experience the world uniquely based on how their multiple social identities such as gender, race, and class simultaneously intersect to shape bias and inequality (Cole, 2009; Crenshaw, 1991; Else-Quest & Hyde, 2016; Ghavami & Peplau, 2013; Lei & Rhodes, 2021; Purdie-Vaughns & Eibach, 2008).

For this meta-analysis, intersectionality informed how different communities of color construct and endorse gender stereotypes. Fisk and Ridgeway (2018) argued that the traditional stereotypes seen in the media and other cultural institutions tend to resemble the perspectives of those in the dominant culture (e.g., White, middle-class Americans). Communities of color may instead construct alternate understandings. For instance, compared to White U.S. adults, Black U.S. adults tend to have stereotypes more favorable to women with respects to STEM (vs. liberal arts) associations (O'Brien et al., 2015) and general intelligence stereotypes (Eagly et al., 2020, Figure 4). These prior findings motivated a specific meta-analytic prediction about participants' race/ethnicity, as described later (see also Skinner et al., 2021).

Gender stereotypes can also vary based who the stereotype targets (Ghavami & Peplau, 2013). Negative stereotypes exist specific to Black and Hispanic/Latinx boys' academic aptitude and engagement, for instance (Hudley & Graham, 2001). In one study, children aged 5–6 saw Black men as less intellectually brilliant than Black women, contrary to the findings for stereotypes of White targets (Jaxon et al., 2019). As noted later, we were unable to test hypotheses about intersectional targets, but this point nevertheless informs the theoretical framing.

Moderator Hypotheses

The theoretical frameworks in the prior section formed a foundation for preregistered moderator predictions. The following section explains the rationale for each prediction, while situating each in STEM- and verbal-specific literatures. As Tobin et al. (2010) argued (p. 610), focusing on specific attributes like STEM and verbal ability can identify contextual influences that are often lost in a domain-general approach that aggregates across many stereotype domains.

Participant Age and Gender

Per cognitive theories of gender development, we predicted that children's STEM and verbal ability stereotypes would reflect in-group bias at ages 5–6 (boys favoring boys and girls favoring girls). This in-group bias would then wane with age, as children also learn stereotypic messages in their environments, leading to traditional stereotypes in older ages (e.g., favoring female verbal ability). Figure 1 depicts these implied Age × Gender interactions (copied from the preregistration; see https://osf.io/8ktnj). These predictions also imply stronger age effects for girls' (vs. boys') STEM stereotypes and boys' (vs. girls') verbal stereotypes.

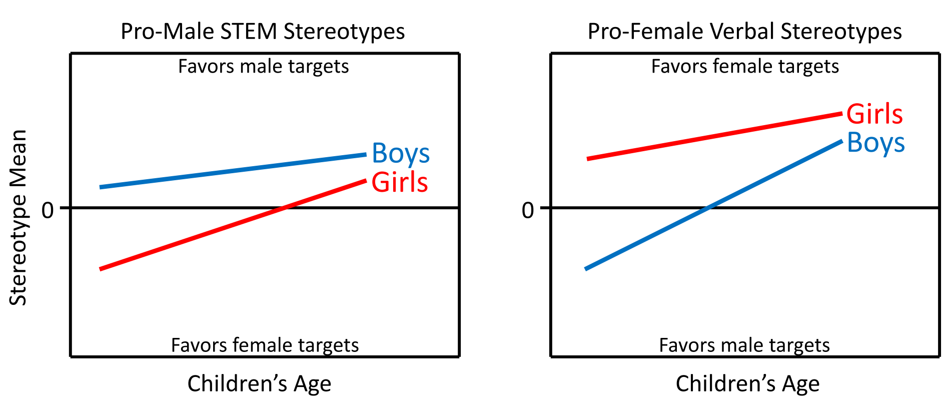
Figure 1 shows a general increase of traditionality with age (aligning with an earlier point that descriptive stereotypes often reflect increased knowledge with age). A prior meta-analysis of 78 Draw-A-Scientist studies supported this prediction for male–science associations; children increasingly depicted more male than female scientists with age (D. I. Miller, Nolla, et al., 2018). Children could construct these stereotypes based on stereotypic messages (e.g., male–science, female–reading) in child-directed conversations, books, and television (Charlesworth et al., 2021).

We predicted the general increase in traditionality would continue in adolescence. Beyond gender intensification (reviewed earlier), relevant external observations and internal observations could strengthen traditional STEM and verbal ability stereotypes in adolescence:

- *external observations (correspondent inference):* Academic course-taking becomes more gender-differentiated in high school with the introduction of advanced elective courses (e.g., Advanced Placement Computer Science; Bahar et al., 2022). Per social role theory, observing more boys than girls in these advanced STEM course electives could strengthen STEM ability stereotypes (via correspondent inference). Gender differences in college majors and occupations may also become more salient as high school students explore their career interests and plan for the future, further increasing traditionality.
- *internal observations (self-to-group pathway):* Gender differences in STEM and verbal interests and confidence (e.g., "I am good at science") tend to strengthen before or during adolescence (Levine & Pantoja, 2021; Liou et al., 2023; Parker et al., 2020). Adolescents may then align these beliefs about the self with their gender ("I am a boy") to actively construct stereotypes about their gender ("Therefore, boys must be good at science"; Master, 2021), strengthening traditional ability stereotypes in adolescence (for illustrative studies of this self-to-group pathway, see Patterson & Bigler, 2018; Starr & Simpkins, 2021, Figure 1).

Consistent with these hypotheses, three large U.S. longitudinal studies (all *N*s > 1,000) have found increasing traditionality in math and science ability stereotypes from grades 8 to 11 (Starr et al., 2023),

Figure 1
Preregistered Predictions for Age, Gender, and Stereotype Domain



Note. STEM = science, technology, engineering, and mathematics. See the online article for the color version of this figure.

grades 9 to 11 (Starr & Simpkins, 2021), and grades 10 to 12 (Skinner et al., 2021).

Confirmatory hypotheses therefore predicted general increases during the ages examined (99% of the meta-analyzed studies had mean ages between 6.0 and 18.0). However, exploratory analyses also examined nonlinear age effects given related past findings (e.g., Crouter et al., 2007) and the above considerations (which might suggest increased age effects in adolescence).

Magnitude of STEM Versus Verbal Stereotypes

We predicted that verbal stereotypes would overall favor female ability to a greater extent than STEM stereotypes favoring male ability. This asymmetric magnitude might arise from children generally associating girls with superior academic performance (Hartley & Sutton, 2013; Heyder & Kessels, 2013). These general academic stereotypes reflect the reality of girls' superior grades in all school subjects (Voyer & Voyer, 2014), potentially reinforcing pro-female verbal ability stereotypes and mitigating pro-male STEM ability stereotypes. Gender gaps in academic performance are also larger in verbal than STEM subjects, which might contribute to asymmetric magnitudes in verbal versus STEM ability stereotypes if children are aware (D. I. Miller & Halpern, 2014; Reilly, 2012; Reilly et al., 2019). This prediction considered STEM fields as a whole (e.g., combining math with engineering). But stereotypes could also vary widely across STEM fields, which would qualify interpretation of this magnitude difference.

STEM Domain Differences

The wide variability in gender diversity across STEM fields may partly derive from and lead to field-specific beliefs (Cheryan et al., 2017). We similarly predicted that stereotypes about the most male-dominated STEM fields (computer science, engineering, physics) would favor male ability more strongly than math or science stereotypes (for related prior evidence, see McGuire et al., 2022; Tang et al., 2024).

Computer science, engineering, and physics show starkly different patterns of women's representation than other STEM fields. For instance, among U.S. bachelor's degree earners in 2020, women's representation was 21% for computer science, 24% for engineering, 24% for physics, compared to 42% for mathematics, 53% for chemistry, and 64% for life sciences (NCSES, 2023, Table 2-2). Per social role theory, these educational and occupational differences could lead to stereotype differences. In observing more male than female engineers, for instance, perceivers might most directly infer that men have more engineering ability than women (in contrast, an inference to math ability would be more tenuous). Once broader society (e.g., adults, cultural artifacts) establishes these cross-field differences in stereotypes, children could learn them through various means (e.g., mass media, educational materials, parents, teachers, peers).

Beyond the confirmatory contrasts, we conducted exploratory analyses to examine stereotypes for less commonly studied STEM domains (e.g., biology, chemistry, spatial ability) and disaggregate results for male-dominated STEM fields (e.g., physics vs. engineering).

Cultural Context

Per social role theory, we predicted that children's pro-male STEM ability stereotypes would be weaker for (a) nations with a greater representation of women among STEM college graduates and (b) more recent decades of data collection within the United States.

Related prior research has supported both hypotheses for cultural fit stereotypes. In one study of 66 nations, adults had weaker male–science associations in nations with more women in STEM (D. I. Miller et al., 2015). In a meta-analysis with data since the 1960s, U.S. children's male–science associations were weaker in more recent decades (D. I. Miller, Nolla, et al., 2018). Like D. I. Miller, Nolla, et al. (2018), we restricted our cross-temporal hypothesis to U.S. samples, given that U.S. women's representation has greatly increased since the 1960s across multiple STEM fields, but progress has been uneven in other nations (Ceci et al., 2014; D. I. Miller & Wai, 2015).

We preregistered both cultural variation hypotheses as confirmatory, but as explained later, we ultimately presented them as exploratory analyses, due to proven limitations in the available data to reliably test these hypotheses (based on power analyses).

Participant Race and Ethnicity

Drawing on intersectional frameworks, we predicted that pro-male STEM ability stereotypes would be weaker for Black than White U.S. children. Related past studies have found that STEM-related and intelligence-related stereotypes were more favorable to female targets among Black than White perceivers (Eagly et al., 2020; O'Brien et al., 2015), including in studies of children's STEM ability stereotypes (Skinner et al., 2021).

This hypothesis focused on the participant's race (who endorses the stereotype), not the stereotype target's race (who the stereotype concerns), due to limitations in the available evidence. Studies on children's ability stereotypes tend to use generic labels like "girls" and "boys" (e.g., "Are girls or boys better in mathematics?"), not intersectional targets. These considerations may be intertwined, however, if children default to thinking of their own race/ethnicity when asked about generic "boys" and "girls." Hence, negative messages about Black boys' academic engagement (Hudley & Graham, 2001) could still contribute to participant race differences.

Two other factors further motivated this hypothesis's specificity to Black versus White U.S. children. First, we restricted this hypothesis to the United States, given this nation's unique history and cultural context surrounding race and ethnicity (Omi & Winant, 2014). Second, this hypothesis was neutral with respect to children of other racial/ethnic identities (e.g., Asian children, multiracial children); prior evidence has found more consistent differences in gender ability stereotypes for Black versus White perceivers than for other racial/ethnic comparisons (e.g., White vs. Hispanic/Latinx perceivers; Eagly et al., 2020; Riegle-Crumb & Peng, 2021).

Measurement Characteristics

Ability stereotype measures in this literature tend to explicitly mention gender, use child targets, and concern average or typical abilities (e.g., "Are girls or boys typically better readers?"). Compared to these typical features, we predicted that pro-male STEM ability stereotypes would be stronger for measures that do not explicitly mention gender, use adult targets, and concern exceptional abilities, as detailed next. These features denote the measure's *directness*, *target age*, and *target exceptionality*, respectively.

Measure Directness. We predicted that indirect measures would show stronger pro-male STEM ability stereotypes than direct measures due to social desirability bias. Direct stereotype measures mention gender explicitly (e.g., "Are boys or girls better at math?"), which might introduce social desirability bias if children perceive that saying boys are better in STEM is socially unacceptable. Indirect measures, sometimes called implicit measures,[3] use more covert methods that do not explicitly mention gender, theoretically mitigating the potential for social desirability to affect participants' responses. One common example is asking children to select the student good at math (or another subject) among pictures of girls and boys, without calling attention to the genders depicted (e.g., Heyman & Legare, 2004; Martinot et al., 2012). Another common example is to ask children to draw a picture of someone good at math and then record the gender of the person drawn (e.g., Beilock et al., 2010; Steele, 2003). A few other indirect measurement approaches exist but are less common (e.g., the affect misattribution procedure in Vuletich et al., 2020). A related method is reaction-time measures, such as the Implicit Association Test (IAT), but those typically assess associative (cultural fit) stereotypes, not ability stereotypes, making them out of scope for this meta-analysis[4] (e.g., Cvencek et al., 2011).

Target Age. Children may develop stereotypes about adult targets (e.g., "Men are better at math") because adult women are underrepresented in STEM careers. Yet, children might not always apply those beliefs to child targets (e.g., "Boys are better at math"), as some studies have suggested (Martinot et al., 2012; Steele, 2003; though null findings were found in Hildebrand et al., 2021). Children may observe the reality that girls earn better grades (even in math and science) and are less disruptive in classrooms, leading to beliefs that girls are academically superior across subjects (Hartley & Sutton, 2013; Voyer & Voyer, 2014). We therefore predicted that STEM stereotypes favoring male ability will be stronger for adult than child targets.

Target Exceptionality. We predicted that pro-male STEM ability stereotypes would be stronger for exceptional targets or traits (e.g., "Who can be really, really good at math?") than average targets or traits (e.g., "On average, are boys or girls better at math?"). Several recent studies have found "brilliance" stereotypes about exceptional intelligence favoring male targets among both child and adult participants in multiple cultures (Bian et al., 2017; Jaxon et al., 2019; Okanda et al., 2022; Shu et al., 2022; Storage et al., 2020; S. Zhao et al., 2022). In contrast, stereotypes about average intelligence now mostly show gender neutrality among U.S. adult participants (with some female advantage; Eagly et al., 2020). This exceptional-average distinction could arise from observing women in cognitively demanding jobs in general yet scarce in roles seen to require brilliance (e.g., Nobel Prize winners; Storage et al., 2020, p. 11).

Exploratory Moderators

Exploratory analyses investigated additional moderators for which we had less strong hypotheses or expected less available data. One such moderator was item wording related to children's conceptions of effort and innate ability (e.g., Muradoglu & Cimpian, 2020). Some measures used wording such as "talent," "gifted," or "aptitude" that suggest innate abilities (e.g., "Do girls or boys have stronger natural aptitudes for math?"), whereas others concerned school performance (e.g., "Do girls or boys get better math grades?"), which likely have greater effort-related connotations. This distinction is relevant given that effortful school achievement is often stereotyped as feminine, whereas innate intellectual ability is stereotyped as relatively more masculine (e.g., Bian et al., 2017; Heyder & Kessels, 2017).

Summary of Predictions

In sum, we made these confirmatory predictions for STEM and verbal domains:

- *overall means:* Both domains would show traditionality overall (e.g., verbal stereotypes favoring female ability), but with a larger magnitude for verbal than STEM stereotypes.
- *participant age:* Both domains would show increased traditionality with increased participant age.
- *participant gender:* Both domains would show participant gender differences in line with in-group bias (e.g., stronger pro-male STEM stereotypes among boys than girls).
- *interactions between age and gender:* Participant gender differences would decline with age, leading to the predicted Age × Gender interactions shown in Figure 1.

Pro-male STEM stereotypes should also be stronger for (a) computer science, engineering, and physics versus math and general science, (b) White versus Black U.S. children, (c) indirect versus direct measures, (d) adult versus child targets, and (e) exceptional versus average targets.

Method

We tested how the preceding hypotheses can help explain mixed empirical findings in a comprehensive meta-analysis spanning more than 4 decades of data collection. Though meta-analyses often focus on mean differences, they can also analyze means as effect sizes when applying appropriate scaling (e.g., Buecker et al., 2021), as detailed later in this section.

Transparency, Openness, and Reproducibility

All data, analysis code, and review procedures are publicly available on the Open Science Framework (OSF) site at https://osf.io/29egh/ (D. I. Miller & Lauer, 2024), including the R Markdown code to reproduce the results text, tables, and figures (enabling transparent linking of any reported statistic to the R code that generated it). We preregistered the review's protocol[5] within that OSF project site at

[3] We prefer to call these measures "indirect" because the term *implicit* is more theoretically laden with assumptions (e.g., assuming that "implicit" measures capture unconscious associations; Corneille & Hütter, 2020).

[4] We found one reaction-time measure that explicitly concerned ability stereotypes (Nowicki & Lopata, 2017), but we decided to not include the measure because (a) we could not compute our focal stereotype score metric for it (there was no theoretical minimum and maximum) and (b) the measure was unlike all other included indirect measures (lacking comparability). However, we included the study's explicit stereotype measure, which met other eligibility criteria.

[5] The preregistration describes two meta-analyses planned as separate articles on (a) stereotype development and (b) correlations with STEM outcomes. The current article summarizes the first meta-analysis.

https://osf.io/7mnsw, including preregistering the inclusion criteria, codebook, moderator hypotheses, analysis plan, and other review procedures. We first preregistered the protocol in February 2020 (before completing eligibility screening) and then updated it in January 2022 (before analyzing any effect size data but after initial data extraction). The updated protocol transparently notes the changes (see pages 5–7 at https://osf.io/8ktnj). There were no deviations from that updated protocol for confirmatory analyses, except for one minor change to imputing missing moderator data for participant race/ethnicity.[6] This article followed the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA 2020) reporting guidelines (Page et al., 2021); see Supplemental Material A for the completed PRISMA checklist. Lastly, a web-based interactive data tool (D. I. Miller, Joshi, et al., 2024) allows users to explore the evidence base, subset by key moderators, and understand variation in effect sizes: https://d-miller.shinyapps.io/STEM_verbal_stereotypes/.

Inclusion Criteria

This section briefly summarizes the inclusion criteria (for granular detail, see Appendix B of the preregistration). We included studies that met all the following criteria, without any restrictions on the publication year, publication language,[7] publication status, or study location.

Criterion 1: Mean Age

We included samples with a mean age less than 18.0 years.

Criterion 2: Measured Gender Stereotypes About STEM or Verbal Abilities

Aligning with our earlier definition of ability stereotypes, we included STEM and verbal measures about raw or "innate" intelligence (e.g., "Are girls or boys more naturally gifted in math?"), academic or cognitive performance (e.g., "Who gets better grades?" "Who is better at math?"), or job performance or abilities to pursue a STEM career. We defined a gender stereotype measure as one that captures beliefs about the attributes of novel or general targets (e.g., "boys" vs. "girls"), excluding measures about specific targets who the respondents already knew (e.g., "Is your mom or dad better at math?"). We included measures of self-endorsement (e.g., "Do you think that … ?") and perception of others' stereotypes (e.g., "Do teachers think that … ?").

We excluded measures about representation in a field (e.g., "Are engineers typically men or women?"), gender role attitudes (e.g., "Is science more appropriate for boys or girls?"), or domain-general ability stereotypes (e.g., brilliance stereotypes; Bian et al., 2017). We generally excluded associative measures (e.g., Draw-A-Scientist Test, IAT), except if they concerned abilities or performance (e.g., "Draw a student who is good at math").

We included measures of computing stereotypes about computer programming or coding as well as general computer use[8] (e.g., "good at computers"). We included spatial ability (e.g., mental rotation) as a "STEM field" given the close relationship of spatial skills to success in STEM fields (Wai et al., 2009). We did not include social sciences in the STEM fields examined.

Criterion 3: Permitted a Directional Assessment of Stereotypes

We included measures whose response structure permitted a directional assessment of stereotypes such as:

- *direct comparisons:* measures comparing female and male targets (e.g., "Are girls or boys better at math?")
- *separate ratings:* measures with separate ratings of female and male targets (e.g., ask children to rate how good boys and girls are at math on a visual analog scale)
- *agreement ratings with analogously worded reverse-coded items:* agreement ratings with male-biased items (e.g., "Boys are better at math") and analogously worded female-biased items (e.g., "Girls are better at math")

We excluded response structures that provided more ambiguous results, such as agreement ratings to statements about equality (e.g., "females are as good as males in geometry"; Fennema & Sherman, 1976); disagreements with such statements could indicate pro-male or pro-female beliefs (Forgasz et al., 2004). The Limitations section further discusses this point, including detailing the rationale for excluding agreement ratings that lacked items in the reverse direction.

Criterion 4: Timing of Stereotype Measurement

To study naturalistic variation in children's stereotypes, we included (a) cross-sectional studies (multiple age groups or a single age group), (b) data from the first time when stereotypes were measured in a longitudinal study, or (c) measures collected before a researcher-developed manipulation or intervention.

Other Inclusion Criteria

We also required that studies (a) reported means for ability-related items separately from other items, (b) included 10 or more participants, and (c) reported sufficient statistics to extract an effect size (we emailed study authors if the study report was insufficient).

Literature Search

We gathered records for published and unpublished studies using (a) keyword searches of literature databases, (b) gray literature searches, and (c) citation tracking.

[6] In the imputation models for missing moderator data, we specified the Level-1 moderator covariance matrix as common across studies (setting meth = "common" in the *jomo* R package). We did not follow the preregistered specification of allowing the covariances to vary (setting meth = "random") because of the nonsensical imputations it yielded (e.g., produced extreme values such as values less than −200 or greater than +200 for proportion variables that were otherwise restricted to a 0–1 range). This change only applies to participant race because no other confirmatory moderator had missing data.

[7] Though the keyword searches were English-based, many literature databases included translated English abstracts for non-English reports. We also found non-English reports through citation tracking in Google Scholar.

[8] We found nearly identical means (e.g., change by 0.01) when applying a more restrictive definition of computing stereotypes that excluded stereotypes about general computer use or technology, indicating robustness.

Database Searches

We conducted the most recent keyword searches in May 2022 using six databases: Education Resources Information Center, Education Source, ProQuest Dissertations and Theses Global, APA PsycInfo, Scopus, and Web of Science Core Collection.[9] We searched for reports whose titles, abstracts, or author-provided keywords included at least one keyword in each of the four following categories: (a) domain-related keyword (e.g., *math**, *verbal abilit**), (b) gender-related keyword (e.g., *gender**, *girl**), (c) stereotype-related keyword (e.g., *stereotyp**, *gender* belief**), and (d) age-related keyword (e.g., *child**, *middle school**). We also customized the search for APA PsycInfo to leverage its controlled vocabulary (i.e., its Thesaurus of Psychological Index Terms). Table 3 and Appendix A of the preregistration include all search terms and the exact database-specific strings to replicate the search.

We identified the search terms through a five-step process. First, we used the *litsearchr* R package (Grames et al., 2019) to identify common terms from the titles and abstracts from 37 example eligible studies. Second, an experienced research librarian recommended other terms and advised on translating the search syntax across databases. Third, we conducted test searches to examine the number and relevance of additional search hits; we translated general phrases that had far too many irrelevant hits (e.g., *beliefs*) into more specific variants (e.g., *belie* about gender**). Fourth, we applied the *litsearchr* package again to all retrieved titles and abstracts from a revised search. Fifth, we added keywords in updated literature searches to capture eligible studies found via citation tracking but not initial keyword searching.

Gray Literature Searches

Beyond searching databases that index gray literature (e.g., ProQuest Dissertations and Theses Global), we sought unpublished studies by searching (a) conference programs, (b) federal grant abstracts, and (c) additional websites (e.g., OSF).

First, we searched 80 conferences programs from eight societies (American Educational Research Association, Cognitive Development Society, Gender Development Research Conference, Society for Personality and Social Psychology, Society for Research on Adolescence, Society for Research on Child Development, Society for the Psychological Study of Social Issues, Society of Experimental Social Psychology). For promising presentation titles and abstracts, we emailed the first author for the full presentation or related reports.

Second, we found relevant federal research grants through keyword searches of project-level abstracts from three funding agencies: Institute of Education Sciences, National Institutes for Health, and National Science Foundation. We searched for associated publications from relevant grants by searching for the award number in Google Scholar, which had the potential to yield additional unpublished studies not indexed in other literature databases.

Third, we conducted keyword searches on four additional websites: (a) EdWorkingPapers (https://edworkingpapers.com/), (b) OSF website (https://osf.io/search/), (c) OSF preprints (https://osf.io/search/, which include PsyArXiv and EdArXiv, among others), and (d) Think Tank Search (https://guides.library.harvard.edu/c.php?g=310680&p=2072552).

Forward and Backward Citation Tracking

We used Google Scholar and Scopus to examine citations to and from eligible studies.[10] Google Scholar especially helped to identify unpublished studies through faculty websites and other websites (Haddaway et al., 2015).

Number of Records

These search methods yielded 18,482 unique records, after removing duplicates. These records included 11,324 from keyword searches of literature databases, 7,118 from citation tracking, and at least 40 records[11] from gray literature searches (see Figure 2 for a PRISMA chart depicting the flow of records; Page et al., 2021). We follow PRISMA's terminology in which a *report* is a document (e.g., a journal article); a *record* is metadata about a report (e.g., title, abstract); and a *study* is an investigation with a group of participants (which could yield multiple reports, such as a dissertation and a journal article).

Screening and Coding

Screening Procedures

A team of four screeners reviewed records for eligibility in three stages: abstract screening (Stage 1), full-text screening (Stage 2), and data availability screening (Stage 3). A written protocol provided detailed guidance for answering stage-specific screener questions, including eligible and ineligible examples (see Appendix B of the preregistration). The first author, an experienced systematic reviewer, was one of the four screeners and met weekly with the screening team to discuss and resolve all discrepancies for dual-screened records. We used Google Translate to translate non-English abstracts or reports.

In Stage 1 (abstract screening), screeners examined titles and abstracts using the semiautomated screening tool Abstrackr (Rathbone et al., 2015). The tool used machine learning to continuously learn from screeners' decisions, sorting abstracts with the highest probability of inclusion to appear earlier in the queue. At least one human screener reviewed all 18,482 abstracts. From these, two humans dual screened 3,617 of the more relevant abstracts. Hence, the machine learning in Abstrackr did not replace human judgment but instead helped us prioritize which records to dual screen. Of the 18,482 abstracts, 1,240 proceeded to full-text screening.

In Stage 2 (full-text screening), screeners examined full-text reports, judging them against all inclusion criteria except data availability and recording specific exclusion reasons for ineligible articles (see Figure 2 for counts of these exclusion reasons[12]). We

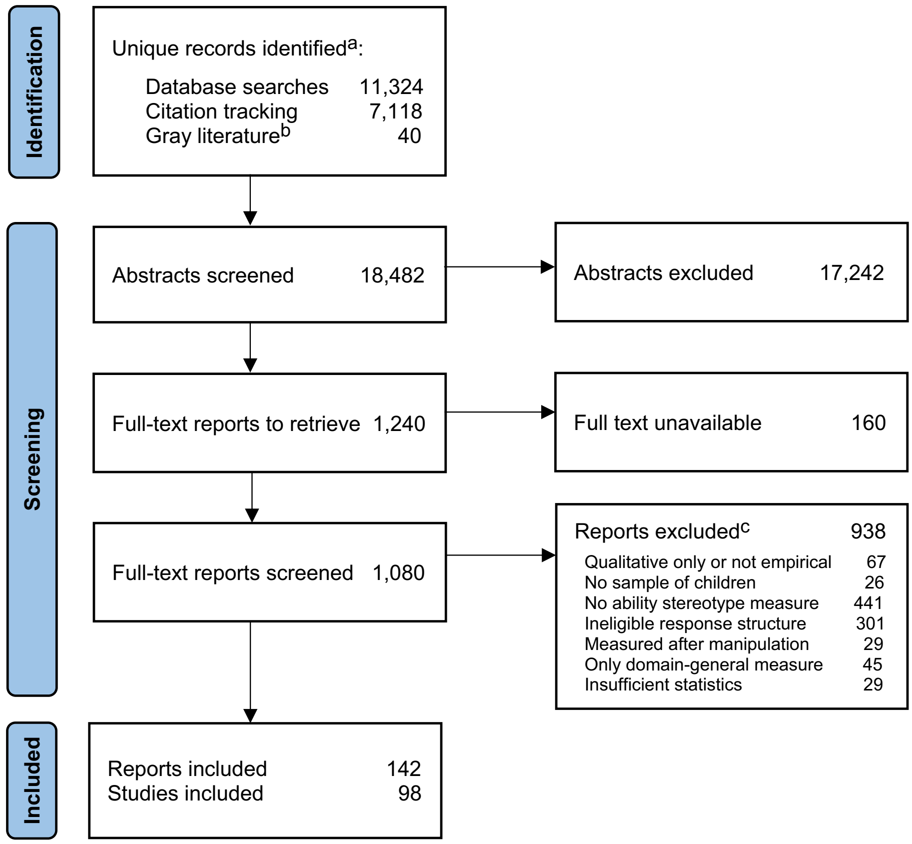
[9] We conducted the first search in November 2019 using 12 databases, but we dropped the following six databases in later searches because they did not uniquely find eligible studies in earlier searches: Academic Search Complete, Education Full Text, Education Research Complete, GenderWatch, Social Science Full Text, and SocINDEX.

[10] For feasibility, we used eligible studies identified through the first November 2019 database searches (i.e., we did not update citation searching based on studies from the May 2022 database search update).

[11] In practice, the number reviewed for gray literature searches was higher due to sources like conference programs where tracking the exact number of records reviewed was intractable.

[12] The OSF site also includes a data file with the screening decisions for all 18,482 citations (see https://osf.io/bhvg5), including specific exclusion reasons for ineligible records excluded at Stage 2 or Stage 3.

Figure 2
Flow of Records Through the Literature Search and Screening



Note. See the online article for the color version of this figure.
[a] These counts already remove exact duplicates of reports (i.e., each row is the number of unique new records identified compared to the previous rows). [b] In practice, the number reviewed for gray literature searches was higher due to sources like conference programs where tracking the exact number of records reviewed was intractable. [c] Reports ineligible for multiple reasons were grouped into a single reason based on the order listed here.

were unable to retrieve the full text for 160 of the 1,240 reports. Of the remaining 1,080 full-text reports reviewed (of which 604 were dual screened), 171 proceeded to data availability screening.

In Stage 3, the lead author reviewed whether the report provided sufficient information to extract at least one effect size and emailed study authors for missing information if needed. Of the 171 eligible full-text reports reviewed, 29 did not provide sufficient statistics in the report or via author query. Hence, 142 full-text reports proceeded to study coding. These 142 reports yielded 98 unique studies after accounting for the same study appearing in multiple reports.

Coding Procedures

A team of three coders, including the lead author, coded study characteristics and extracted quantitative information from eligible studies. The preregistration details variable-specific coding rules, definitions, and assumptions (e.g., if participants' age was not reported, we inferred it from the participants' grade level). Of the 98 studies coded, 18 were dual coded by two independent coders; the lead author resolved discrepancies after discussion with the other two coders. For any study not dual coded, the lead author reviewed the data entries of the other coder, checking for accuracy against the full-text report and providing continuous weekly feedback to the rest of the coding team. This process helped prevent coding drift over time and ensure consistent interpretation of coding procedures and definitions.

Whenever possible, we coded the overall study sample and demographic subsamples (e.g., girls vs. boys; third vs. fifth graders) separately for each stereotype measure (e.g., physics vs. biology stereotypes). We coded subsamples that differed by the confirmatory moderators of participant age, gender, race/ethnicity, and nationality when possible. We did not try to code subsamples differing in other respects (e.g., high vs. low academic achievement).

We used a relational database (Microsoft Access) to efficiently code and link nested levels of study, sample, measure, and effect size information. For studies reported in multiple reports (e.g., a dissertation and journal article), we examined all reports for relevant information and combined as needed. If any statistics differed across a study's multiple reports, we first attempted to contact the study authors for clarification. If unsuccessful, we then generally relied on the peer-reviewed report (e.g., journal article) or most recent report.

Author Queries

We emailed the first authors of 125 reports for additional information (or other authors if we could not find the first author's email address). We sent these author queries for (a) eligible studies that did not provide sufficient information to compute any effect size (47 reports), (b) eligible studies that provided sufficient data for some but not all effect sizes or had conflicting information (55 reports), and (c) studies that lacked sufficient information to determine eligibility (23 reports). We incentivized authors with a $100 https://Amazon.com e-gift card for replying with the requested information in one of two ways: (a) send us the raw data or (b) fill out custom tables our team prepared. We sent up to two reminders after the initial email. These emails also included a two-page data-sharing agreement, which has been shown to increase successful author query rates (Polanin & Terzian, 2019; Polanin & Williams, 2016).

We successfully received the needed additional information for 50 reports, yielding a 40% success rate. We received participant-level data sets for 19 of these reports, for which we developed reproducible R code to extract summary statistics, allowing for more disaggregated statistics than often available in the original study reports.

Effect Size Metrics

Stereotypes Scores (Confirmatory)

Our confirmatory effect size metric was *stereotype scores* quantified on a common scale from −1.0 to +1.0. Positive scores indicate traditional stereotypes (favoring male STEM ability or female verbal ability). A score of +1 is the maximum possible stereotype mean in the traditional direction (e.g., all children selected the most extreme pro-male STEM endpoint). For instance, a score of +0.35 indicates 35% of the maximum possible pro-male STEM stereotype. Negative scores indicate stereotypes in the reverse direction (e.g., favoring female STEM ability).

This metric of stereotype scores is conceptually similar to Cohen et al.'s (1999) percent of maximum possible scoring method, given that both use the theoretical minimum and maximum of the response scale to compare across measures. These types of metrics have two main advantages: (a) They avoid conflating changes in means with changes in standard deviations, and (b) they enhance interpretability by allowing for translation back to the original response scales (e.g., 1–5 scales). For these reasons, several prior meta-analyses have used percent of maximum possible scores as the main effect size metric, such as to study relationship satisfaction across age (Bühler et al., 2021) or loneliness over time (Buecker et al., 2021). The preregistration provides further justification and effect size extraction details (e.g., standard error equations).

Standardized Means (Exploratory)

We also computed *standardized means* for a robustness check (i.e., standard deviations away from gender neutrality). Standardized means have the same directional meaning as for stereotype scores (e.g., positive values for pro-male STEM stereotypes). But the magnitude scale is based on the observed variability, not the response scale per se.

Qualitative Responses (Exploratory)

As exploratory metrics, we also examined response frequencies for three qualitative responses (traditional, nontraditional, or gender neutral). Consider if the measure asked children, "Are girls or boys better at engineering?" A response of "boys" would be traditional, "girls" would be nontraditional, and "the same" would be gender neutral. We analyzed these qualitative responses by analyzing the log odds for providing (a) traditional versus nontraditional responses and (b) gender-neutral versus directional responses.

Analytic Approach

Mixed-effects meta-regression models investigated how STEM and verbal ability stereotypes varied across child demographics, cultural contexts, and measurement characteristics. These models, estimated using restricted maximum likelihood, assumed that variation in effect sizes was due to fixed effects of moderators (e.g., age), random effects of residual between-study heterogeneity, and within-study sampling variance (Borenstein et al., 2009).

Robust Variance Estimation

Most studies had multiple samples (Mdn = 3, M = 4.91) and multiple stereotype measures (Mdn = 2, M = 3.38), often leading to multiple effect sizes per study (Mdn = 8, M = 13.91). We used robust variance estimation to account for these effect size dependencies, applying the small-sample correction based on the Satterthwaite approximation (Pustejovsky & Tipton, 2018; Tipton, 2015; Tipton & Pustejovsky, 2015). A "correlated and hierarchical effects" model accounted for both the nesting of effect sizes within samples and nesting of samples within studies (Pustejovsky & Tipton, 2022). This approach estimated coefficients using the *rma.mv()* function in the *metafor* R package (Viechtbauer, 2010) and then adjusted the other values (e.g., standard errors, degrees of freedom) using the *coef_test()* function in the *clubSandwich* R package (Pustejovsky, 2018). We assumed a correlation of $r = .5$ for effects from the same sample (but examined $r = .2$ and $r = .8$ in sensitivity analyses).

Characterizing Heterogeneity

We quantified heterogeneity in ability stereotypes by presenting 90% prediction intervals, which is the estimated middle 90% of true underlying effects, after removing dispersion due to within-study sampling error. Prediction intervals provide an absolute measure of heterogeneity using the original effect size scale (Borenstein et al., 2017), unlike other metrics like I^2 statistics (percentage of total variation in effect size due to heterogeneity rather than chance). Another heterogeneity metric was the estimated standard deviation τ of true underlying effects in random-effects, intercept-only models.

Confirmatory Moderator Analyses

We examined each confirmatory moderator in separate mixed-effects models and in one multivariable model that simultaneously adjusted for all confirmatory moderators. For parsimony, the results section focuses on models that included multiple confirmatory moderators simultaneously (results were nearly identical when testing one confirmatory moderator at a time). To control for nuisance methods variance, all moderator models included four dummy-coded covariates for the scale type: (a) two or three response options, (b) forced-choice scale with no gender-neutral response option for individual items, (c) continuous scale (vs. discrete ratings), and (d) separate ratings (vs. comparative scales). We analyzed STEM and verbal stereotypes in separate statistical models, except when directly comparing their overall magnitude.

Conditional Means

Beyond testing moderator differences, we also computed covariate-adjusted conditional means at specific moderator values (e.g., age 6) to aid interpretation (for details, see the supplemental materials of Williams et al., 2022, for the approach we adopted). For instance, a conditional mean for age 6 represents the model-predicted mean if the entire sample of effect sizes came from 6-year-olds, while holding constant other moderators (e.g., STEM domain). These predicted values enable comparison of means while adjusting for potential confounds.

Moderator Power Analyses

Before conducting moderator analyses, we first conducted simulation-based power analyses to inform the final selection of confirmatory moderators (see Appendix E of the preregistration for details). These power analyses aimed to understand which moderator tests had such little statistical power that their results should be interpreted especially cautiously due to lack of sensitivity (i.e., label as exploratory). We retained confirmatory moderators that demonstrated at least 50% power based on a priori optimistic assumptions, such as detecting an effect size difference of 0.36 *SD*s (the median effect size found in a review of meta-analyses in social psychology; Lovakov & Agadullina, 2021).

Based on these a priori rules, three moderators for which we had confirmatory hypotheses became exploratory moderators instead (i.e., still analyzed but only in exploratory models): (a) women's national representation among female STEM majors, (b) women's national representation among employed researchers, and (c) data collection year for U.S. samples.

Missing Moderator Data

The only confirmatory moderator with missing data was participant race/ethnicity, which was missing for 24 of 559 effect sizes for U.S. samples. We used multiple imputation to account for these missing moderator data, which often allows for less biased estimates compared to listwise deletion in meta-analysis (Pigott, 2001, 2012). We used the *jomo* R package to account for the multilevel structure of effect sizes nested within studies, including the effect size and other analyzed moderators in the imputation model (Quartagno et al., 2019). We aggregated results across 80 imputations, accounting for both the within- and between-imputation variance (Barnard & Rubin, 1999; Pustejovsky, 2017).

Selective Reporting Bias

Although we found several unpublished studies (32% of the included studies), selective reporting bias could nevertheless affect the conclusions via publication bias or outcome reporting bias (Carter et al., 2019; Flore & Wicherts, 2015). We diagnosed and adjusted for such biases through (a) selection modeling, (b) comparison of unpublished versus published studies, and (c) meta-regression to assess small-study effects. Appendix F in the preregistration details these analyses and justifies using these methods over others (e.g., trim-and-fill, *p*-curve analysis).

Addressing Cohort Confounds

Supplemental Material C of this article details how we controlled for cohort effects to address potential concerns about cohort confounds, given that the literature synthesized has overwhelmingly used cross-sectional designs to estimate age effects.

Other Robustness Checks

Supplemental Material D details these other robustness checks: (a) using standardized means rather than stereotype scores, (b) including versus excluding outliers, (c) restricting analyses to data from the year 2000 or later, (d) using different values for the assumed correlation between dependent effects, (e) group-mean centering age effects, and (f) repeating analyses separately by participant gender.

Exploratory Analyses

Exploratory analyses examined (a) STEM domain differences that were not preregistered, (b) age differences that were not preregistered (nonlinear effects), and (c) other exploratory moderators or metrics (e.g., item wording, qualitative responses).

Results

Descriptive Frequencies

We extracted 1,363 effect sizes from 98 studies, representing 145,204 children, including 95 studies of STEM stereotypes and 46 studies of verbal stereotypes. Confirmatory moderators for STEM stereotypes had the following frequencies (see Supplemental Material B for sample sizes, other characteristics, and descriptive frequencies for verbal stereotypes):

- *participant age:* Mean sample ages ranged from 4.2 to 17.7 years ($M = 12.4$; $SD = 3.0$). Samples between upper elementary school to high school (9 to <18 years) were common (86% of effect sizes), but less common were samples of preschoolers (<6 years, 1% of effect sizes) and lower elementary school students (6 to <9 years, 13% of effect sizes).

- *participant gender:* Most effect sizes came from disaggregated results for boy participants (46%) and girl participants (46%), rather than aggregated by gender (8%).

- *STEM domain:* The most commonly studied STEM subdomain was mathematics, comprising 55% of STEM stereotype effect sizes. General science was the next most common (17%), followed by computer science/technology (14%), engineering (7%), biology/life science (2%), physics (2%), spatial ability (2%), and chemistry (1%).

- *cultural context:* Studies' data collection spanned more than 4 decades (1977–2020) across 31 nations (two more nations contributed data for verbal, but not STEM, stereotypes). However, most effect sizes came from data in the 2000s (40%) and 2010s (49%), and U.S. samples provided about half of the effect sizes (53%).

- *participant race/ethnicity:* For U.S. samples, participants were 51% White, 26% Black, 11% Hispanic/Latinx, 6% multiracial, 4% Asian, 1% American Indian/Alaskan Native, and 1% Native Hawaiian/Pacific Islander on average. The representation of White children roughly matched population percentages of U.S. children in 2020 (53% in population). Hispanic/Latinx children were underrepresented (26% in population), and Black children were overrepresented (14% in population); other racial/ethnic groups were within two percentage points of population values (U.S. Census Bureau, 2021).
- *measurement characteristics:* Nearly all STEM stereotype effect sizes represented direct measures (95%), child targets (95%), and average targets (98%), rather than indirect measures (5%), adult targets (5%), or exceptional targets (2%).

Aggregate Means and Heterogeneity

STEM stereotypes overall favored male ability, and verbal stereotypes overall favored female ability (see Table 1 for specific statistics). As predicted, aggregate verbal stereotypes were also larger in magnitude than aggregate STEM stereotypes (by a factor of 2.16 for stereotype scores and 2.72 for standardized means).

To illustrate the aggregate magnitudes in Table 1, consider a STEM stereotype measure that asked children, "Are girls or boys better at mathematics?" on a 5-point scale (1 = *girls much better*, 2 = *girls better*, 3 = *equally good*, 4 = *boys better*, 5 = boys much better). The overall mean of 0.09 for STEM stereotype scores would map to a raw mean of $M = 3.17$, which is only slightly different from "equally good," illustrating the small magnitude of aggregate STEM stereotypes on average. The overall standardized mean of 0.17 *SD*s would also imply a raw standard deviation of 1.01 for this example.

Though informative, these aggregate means should be interpreted cautiously as they are only a starting point for understanding a wide distribution of effect sizes (see Figure 3). The estimated middle 90% of underlying effects (i.e., prediction interval) was −0.25 to 0.43 for STEM stereotypes and −0.07 to 0.45 for verbal stereotypes. This wide variability emphasizes the need to understand how moderators (e.g., participant age) contribute to heterogeneity.

Moderator Analyses of Participant Age and Gender

Confirmatory analyses of participant age and gender examined the linear effects of participant age and their interaction with gender, whereas exploratory analyses examined nonlinear age effects and a direct measure of in-group bias.

Confirmatory Analyses

Qualitatively, Figure 4 illustrates how STEM and verbal stereotypes showed in-group bias at age 6 (boys favoring boys, girls favoring girls) but traditionality at age 16 (favoring male STEM ability, favoring female verbal ability). Quantitatively, Table 2 presents the regression coefficients for confirmatory hypothesis tests.

As predicted, the results showed robust participant gender differences (all ps $< .001$), such that children strongly endorsed positive stereotypes about their own gender (e.g., stronger pro-female verbal stereotypes for girls than boys). Also as predicted, these participant gender differences robustly declined with age, becoming near 0 by age 16 (see Figure 4), as indicated by significant Age × Gender interactions (all ps $< .001$). Lastly, verbal stereotypes overall more strongly favored female ability with age (as predicted), but the overall age effect was not significant for STEM stereotypes (counter to hypotheses); see Table 2 for specific statistics.

The Age × Gender interactions also imply that the age slopes differed by participant gender. Girls' STEM stereotypes became more male-biased with age, switching from pro-female to pro-male beliefs at ages 10–12; the predicted 10-year age difference[13] was $b = 0.19$, $SE = 0.04$, $t(14.44) = 4.73$, $p < .001$. Counter to hypotheses, boys' STEM stereotypes showed a slight, but not significant, decline with age, $b = -0.10$, $SE = 0.05$, $t(14.26) = -2.06$, $p = .058$, which the following exploratory analyses further examined.

Relatedly, boys' verbal stereotypes became more female-biased with age, $b = 0.34$, $SE = 0.06$, $t(10.89) = 5.37$, $p < .001$, switching from pro-male to pro-female beliefs at ages 8–10. For girls, this increase was smaller and not significant, $b = 0.08$, $SE = 0.06$, $t(10.97) = 1.30$, $p = .219$, though it was significant for standardized means in sensitivity analyses (see Supplemental Material D). When averaged across genders, verbal stereotypes significantly favored female ability at age 8.

Exploratory Analyses

Boys' STEM stereotypes showed nonlinear change ($p < .001$), weakening before age 13 ($p < .001$) but strengthening after age 13 ($p = .015$). In contrast, girls' STEM stereotypes increasingly favored male ability both before and after age 13 (Figure 5). No evidence was found for nonlinear change in boys' or girls' verbal stereotypes (see Supplemental Material E for more detail on these nonlinear analyses and results).

To better understand the unexpected age findings, we computed a direct measure of in-group bias of how much boys or girls generally assigned more ability (STEM or verbal) to their in-group (see Supplemental Material E). As an example, the in-group bias score would be greater than 0 if a boy expressed pro-male stereotypes for both STEM *and verbal* domains, whereas the score would be 0 if STEM and verbal stereotypes were equally strong in opposite directions. Results showed an asymmetric development of in-group bias: Boys' in-group bias greatly declined during ages 6–13, whereas girls' in-group bias was stable (see Supplemental Figure S6).

Moderator Analyses of STEM Domain Differences

Confirmatory analyses of STEM domain differences examined means for the three most male-dominated STEM fields (computer science, engineering, physics) as a single aggregate category. Exploratory analyses disaggregated this category, examined less commonly studied STEM fields, and investigated how age differences varied by STEM domain.

[13] For scaling presentation purposes, we present age coefficients as the predicted difference per 10 years of age.

Table 1
Random-Effects Models of Overall Stereotype Means and Heterogeneity

Model	b	m	k	SE	df	p	τ	90% prediction interval
				Stereotype scores				
STEM stereotypes	0.09	95	1,053	0.02	91.40	<.001	0.21	[−0.25, 0.43]
Boys	0.16	78	480	0.02	73.70	<.001	0.20	[−0.16, 0.48]
Girls	0.02	82	484	0.02	77.06	.292	0.22	[−0.34, 0.38]
Verbal stereotypes	0.19	46	310	0.02	43.56	<.001	0.16	[−0.07, 0.45]
Boys	0.09	35	133	0.03	32.44	.002	0.16	[−0.17, 0.35]
Girls	0.26	38	135	0.02	35.74	<.001	0.14	[0.03, 0.49]
				Standardized means				
STEM stereotypes	0.17	90	993	0.04	86.42	<.001	0.46	[−0.59, 0.93]
Boys	0.36	75	461	0.05	71.67	<.001	0.43	[−0.34, 1.06]
Girls	0.02	77	458	0.05	72.58	.726	0.48	[−0.77, 0.81]
Verbal stereotypes	0.46	42	295	0.05	39.39	<.001	0.38	[−0.17, 1.09]
Boys	0.23	34	130	0.06	32.29	.001	0.38	[−0.39, 0.85]
Girls	0.69	36	131	0.05	34.05	<.001	0.33	[0.15, 1.23]

Note. These statistics come from random-effects meta-analyses estimated separately by type of effect size metric and stereotype domain. These intercept-only models are intended for descriptive purposes only, not as tests of confirmatory hypotheses. The standard errors were adjusted for effect size dependencies using robust variance estimation (RVE). b = average effect size; m = number of studies; k = number of effect sizes; SE = RVE-adjusted standard error of the average effect sizes; df = RVE-adjusted degrees of freedom; p = RVE-adjusted significance level for the mean being different from 0; τ = estimated standard deviation of true underlying effect sizes; 90% prediction interval = estimated middle 90% of true underlying effect sizes; STEM = science, technology, engineering, and mathematics.

Confirmatory Analyses

As predicted, STEM stereotypes much more strongly favored male ability for the most male-dominated STEM fields than for mathematics and general science (see Table 2 for confirmatory hypothesis tests). Computer science, engineering, and physics stereotypes ($b = 0.25$, $m = 27$ studies) were comparable in magnitude to verbal stereotypes ($b = 0.19$, $m = 46$ studies). In contrast, the means for general mathematics ($b = 0.06$, $m = 69$ studies) and general science ($b = 0.09$, $m = 26$ studies) only slightly differed from 0.

Exploratory Analyses

Even when disaggregated, the male-dominated fields of computer science (17 studies), engineering (eight studies), and physics (10 studies) each had much larger means than mathematics (69 studies); see Figure 6. General science stereotypes (26 studies) masked considerable variation across science fields: Unlike physics stereotypes, biology stereotypes (eight studies) favored female ability. Chemistry stereotypes (six studies) and spatial ability stereotypes (five studies) did not significantly differ from gender neutrality.

We also examined age differences for computer science, engineering, and physics stereotypes separately (27 studies), given their exceptionally large means. Using mathematics as a comparison, Figure 7 shows two key findings:

- *early-emerging STEM domain differences:* Even young children had diverging views across STEM domains. For instance, at age 6, girls' math stereotypes significantly favored female ability, in line with in-group bias. Yet, their stereotypes about computer science, engineering, and physics were in the opposite direction, nonsignificantly favoring male ability at age 6. When averaged across both genders, male-dominated STEM stereotypes significantly favored male ability at age 6 (math stereotypes did not do so until age 12).
- *similar age slopes:* After age 6, stereotypes shifted with age in nearly identical ways across STEM domains. That is, the age slopes and interactions with gender were similar (Supplemental Table S14), even though the starting intercepts differed at age 6.

Other Moderator Analyses

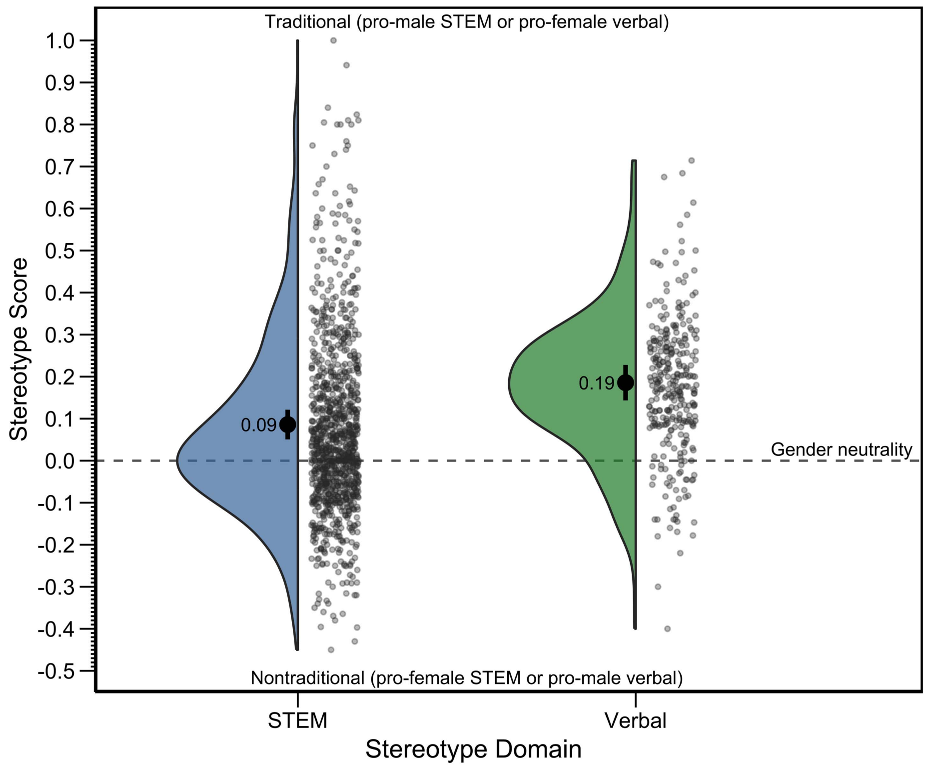
Confirmatory analyses of other moderators examined participant race and three confirmatory measurement characteristics (measure directness, target age, and exceptionality), whereas exploratory analyses examined many other moderators (e.g., item wording, cultural characteristics, nation-specific comparisons, verbal domain differences).

Confirmatory Analyses

As predicted, Black U.S. children had weaker STEM stereotypes than White U.S. children, $b = -0.05$, $SE = 0.02$, $t(4.34) = -3.37$, $p = .025$. Black children's STEM stereotypes were gender neutral on average ($b = -0.01$), whereas White children's STEM stereotypes slightly favored male ability ($b = 0.04$). This finding for participant race was sensitive to participant gender: Black girls' STEM stereotypes were weaker than White girls', but Black boys' and White boys' stereotypes did not significantly differ (see Supplemental Material D).

Counter to hypotheses, the differences were not significant for indirect versus direct measures, adult versus child stereotype targets, and exceptional versus average stereotype targets (Table 2). These null findings should be interpreted somewhat cautiously given the large standard errors (i.e., large uncertainty) for indirect measures (5% of effect sizes), adult targets (5% of effect sizes), and exceptional targets (2% of effect sizes). Among these contrasts, the largest difference was between adult targets ($b = 0.21$) versus child targets ($b = 0.09$), which was in the predicted direction but not significant

Figure 3
Distribution of STEM and Verbal Stereotypes



Note. The solid shapes on the left-hand sides show the effect size distributions, and the dots on the right-hand sides plot the effect sizes. The black dots inside the solid shapes indicate the weighted mean effect size (the error bars represent 95% confidence intervals). STEM = science, technology, engineering, and mathematics. See the online article for the color version of this figure.

(p = .079); in sensitivity analyses, this difference was significant for girl, but not boy, participants (see Supplemental Material D).

Exploratory Analyses

Two exploratory moderators showed significant differences: item wording and certain nation-specific comparisons.

Regarding item wording, measures with wording about innate abilities (e.g., "natural aptitude" or "more talented"; 16% of effect sizes) had more male-biased STEM stereotypes than measures explicitly mentioning school performance (e.g., "math grades"; 18% of effect sizes). Measures with more generic language like "better" or "good at math" (54% of effect sizes) had means in between these two categories (see Supplemental Material F).

Regarding nation-specific comparisons, samples from Germany ($b = 0.28$) and the United Kingdom ($b = 0.18$) had stronger pro-male STEM stereotypes than U.S. samples (b = 0.02); other nation-specific comparisons were not significant or had insufficient data to test.

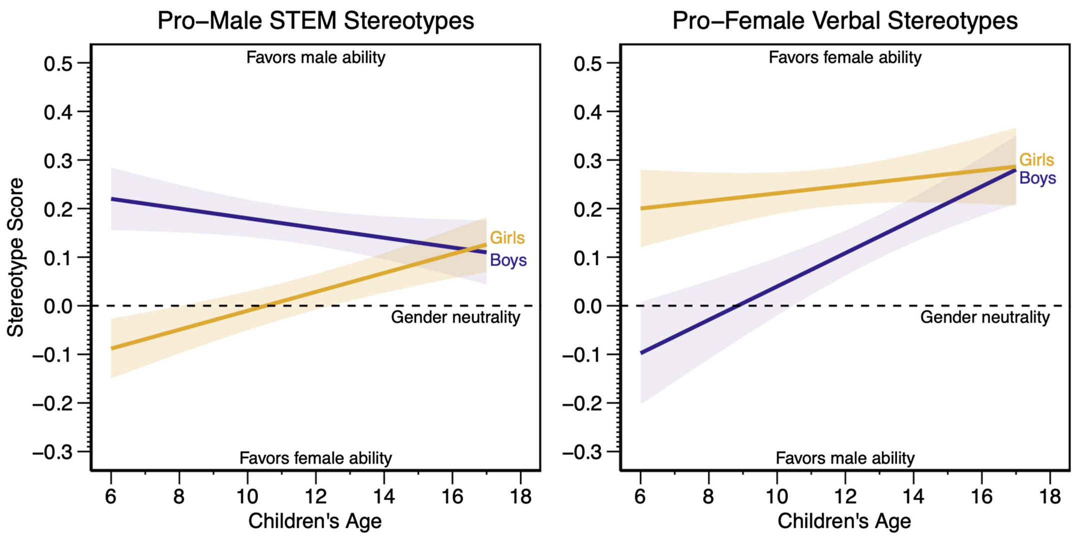
No significant differences were found for other exploratory moderators, including (a) cultural characteristics (e.g., women's national representation in STEM, data collection year), (b) exploratory participant race/ethnicity comparisons (other than the confirmatory Black–White contrast for STEM stereotypes), (c) perception of others' stereotypes (vs. self-endorsement), and (d) verbal domain (e.g., reading vs. writing); see Supplemental Material F.

Qualitative Response Frequencies

Figure 8 further characterizes aggregate means by grouping responses into three qualitative categories (e.g., pro-male responses). For STEM stereotypes, children reported pro-male beliefs more often than pro-female beliefs (30% vs. 19%, respectively), though gender-neutral responses were most common (52%). For verbal stereotypes, pro-female responses were most common (52%), followed by gender-neutral responses (36%) and pro-male responses (12%).

These qualitative responses shifted with participant age, both conceptually replicating earlier findings (e.g., boys' verbal stereotypes becoming more female-biased with age) and adding nuance about gender-neutral responding (see Supplemental Figure S7). Gender-neutral responding significantly decreased with age for verbal stereotypes but not for STEM stereotypes (Supplemental Material E further discusses in relation to *stereotype flexibility*).

Figure 4
Confirmatory Analysis of Participant Age and Gender



Note. These trends came from mixed-effects models that controlled for methods covariates and other confirmatory moderators. The shaded areas represent 95% confidence intervals. STEM = science, technology, engineering, and mathematics. See the online article for the color version of this figure.

Robustness Checks of Confirmatory Analyses

Selective Reporting Bias

Analyses for selective reporting bias reinforced the following conclusion: STEM stereotypes favored male ability on average but with a small magnitude. For instance, adjusted mean estimates ranged from 0.05 to 0.15 in selection models of STEM stereotypes depending on the model specification, providing both upward and downward adjustments to the unadjusted mean of 0.09. However, most adjustments (30 out of 34) increased the mean estimate (see Supplemental Table S6), indicating minimal impact on our conclusions.

Controlling for Cohort Confounds

The estimated age effects for STEM and verbal stereotypes were much larger than the effects for birth cohort or data collection year, which were small in magnitude and not significant. Hence, cohort confounds cannot explain the large age effects found, supporting an interpretation of developmental change (see Supplemental Material C).

Additional Robustness Checks

As detailed in Supplemental Material D, other moderator results were robust to several sensitivity analyses (e.g., using standardized means, excluding outliers, excluding older studies). However, as noted previously, some slight sensitivities were found for participant race, target age, and age effects disaggregated by participant gender.

Discussion

This meta-analysis integrated more than 4 decades of research, spanning 33 nations, on children's gender stereotypes about STEM and verbal abilities. Despite the substantial empirical base, drawing firm conclusions about when and how these stereotypes develop has been formidable in prior narrative reviews, given the wide variability in results. This synthesis explains prior enigmas by uncovering how features such as stereotype domain, age, and gender contribute to the mixed findings across nearly 100 studies with more than 140,000 children.

As expected, verbal stereotypes robustly favored female ability, with an aggregate mean that was 19% of the maximum possible stereotype (0.46 *SD*s away from gender neutrality). Findings for STEM domain, age, gender, and race/ethnicity largely aligned with hypotheses and were robust to many sensitivity analyses (e.g., effect size metrics). However, as discussed next, two especially surprising findings were as follows: (a) STEM domain differences emerged remarkably *early* in development, and (b) age differences were *asymmetric* across STEM versus verbal domains. Additionally, no significant differences were found for confirmatory measurement characteristics (e.g., measure directness; see Table 3 for a summary).

Beyond these specific findings, one broader takeaway is that these stereotypes emerged early in development, including by age 8 for verbal ability and by age 6 for computer science and engineering ability (on average). As such, these beliefs have ample time to affect later downstream outcomes (e.g., confidence, interests) via multiple mechanisms (e.g., incorporating beliefs about one's group into beliefs about one's self; Master & Meltzoff, 2020). This overall point further emphasizes the need to consider how children learn these stereotypes with age and why these stereotypes can also vary widely along other dimensions (e.g., across STEM fields).

Discussion of STEM Domain Differences

Our discussion starts with STEM domain differences, given how the robust findings offer especially clear implications for changing the status quo of current research practice.

Table 2
Confirmatory Moderator Results

Moderator	Simple				Multivariable			
	b	*SE*	*df*	*p*	*b*	*SE*	*df*	*p*
STEM stereotypes								
Age (in decades[a])	0.05	0.04	13.99	.213	0.05	0.04	11.21	.277
Proportion male	0.12	0.01	28.40	<.001	0.12	0.01	28.28	<.001
Age × Male[b]	−0.30	0.03	23.52	<.001	−0.29	0.03	23.41	<.001
Proportion Black (United States only)	−0.06	0.01	3.75	.022	−0.05	0.02	4.34	.025
Male-dominated versus math[c]	0.19	0.03	7.74	<.001	0.19	0.03	7.05	<.001
Male-dominated versus science[c]	0.17	0.03	7.07	.001	0.16	0.03	6.52	.001
Indirect measure	−0.04	0.06	6.15	.471	−0.03	0.05	5.60	.523
Adult target	0.12	0.05	4.00	.071	0.12	0.05	3.21	.079
Exceptional target	0.00	0.13	5.24	.995	−0.02	0.09	4.87	.856
Verbal stereotypes								
Age (in decades[a])	0.22	0.05	10.71	.002	0.21	0.06	8.41	.007
Proportion female	0.13	0.02	11.41	<.001	0.14	0.01	11.96	<.001
Age × Female[b]	−0.27	0.03	17.49	<.001	−0.27	0.03	17.49	<.001
Cross-domain comparison[d]								
Verbal versus STEM	0.16	0.02	10.91	<.001	0.16	0.02	10.89	<.001

Note. The simple models (left-hand side) tested each confirmatory moderator one-by-one in separate models, but controlling for four methods covariates: (a) 2 or 3 discrete response options, (b) forced-choice scale with no gender-neutral option, (c) continuous scale, and (d) separate ratings (vs. comparative scale). The multivariable models (right-hand side) controlled for all confirmatory moderators and methods covariates simultaneously. Supplemental Table S8 provides the analogous table for the exploratory standardized mean metric. STEM = science, technology, engineering, and mathematics; *SE* = standard error.
[a] Age was entered as decades (e.g., 1.7 for 17 years of age) for scaling presentation purposes. Hence, the regression coefficient is the predicted difference per 10 years of age. [b] Age and gender were grand mean centered so that the age and gender effect in the multivariable models can be interpreted as overall average effects. [c] "Male-dominated" refers to computer science, engineering, and physics, which have the most extreme gender imbalance compared to other STEM fields. [d] We analyzed STEM and verbal ability stereotypes in separate statistical models, except when testing the bottom hypothesis about overall magnitude difference; in that one case, we included the stereotype domain as a dummy code (1 = verbal; 0 = STEM).

STEM Domain Differences Were Large

As expected, computer science, engineering, and physics stereotypes robustly favored male ability (25% of the maximum possible; 0.51 *SD*s away from neutrality; based on 27 studies), showing far more male-biased stereotypes compared to other STEM fields. These STEM domain differences were robust, appearing in *every* model that tested for them (all *p*s < .005), even when disaggregating these three male-dominated fields and separately comparing each to mathematics.

In contrast, stereotypes favoring male math ability were weak on average (6% of the maximum possible; 0.11 *SD*s away from neutrality), even in late adolescence (Figure 7). This finding, supported by 69 studies with *N* = 49,502 children, challenges the claim that children hold the "pervasive cultural stereotype" of "males are better at math" as early as "early elementary school" (Levine & Pantoja, 2021, p. 23).

General "science" stereotypes also masked considerable variation across science fields. While physics stereotypes strongly favored male ability, biology stereotypes strongly favored female ability, mirroring variation in adult women's representation across fields (Cheryan et al., 2017). Chemistry stereotypes were gender neutral on average but imprecisely estimated.

STEM Domain Differences Emerged Early

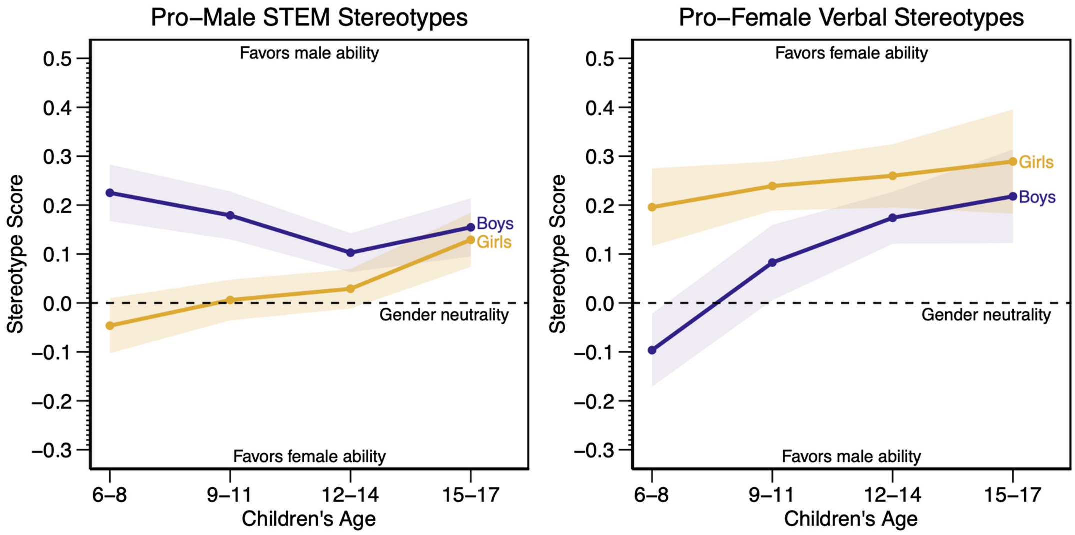
The stark differences by STEM domain emerged remarkably early by age 6. When averaged across boys and girls, computer science and engineering stereotypes[14] already favored male ability at age 6, whereas overall math stereotypes did not do so until age 12. Example measures used between ages 5 and 8 included questions about being good *computer programmers* or *engineers* (David & Tompa, 2018; C. F. Miller, Wheeler, & Woods, 2024) as well as being good at *computer coding*, *technology*, or *engineering* (Master et al., 2021a; McGuire et al., 2020, 2022). At these early ages, boys robustly show computing and engineering stereotypes favoring male ability, whereas girls often show neutrality or even sometimes male bias (contrasting with girls' typical in-group bias at age 6). We find these results surprising given children's presumably limited exposure to computer science and engineering in early childhood.

The early onset of these domain-specific stereotypes may partly reflect how children interpret terms like "engineer." Young children often have misperceptions. For instance, some young English-speaking children assume that engineers must fix car engines because "engineer" contains the word "engine" (Lampley et al., 2022). In a study of 504 children in grades 1–5, most said that engineers repair cars (78%), install wiring (75%), drive machines (71%), and construct buildings (70%), but only 32% said that engineers design things (Cunningham et al., 2005).

These misperceptions of engineers are relevant given that even young children likely have direct exposure to blue-collar jobs such as auto mechanics and construction workers. Young children may therefore extend stereotypes about other male-dominated fields

[14] Physics stereotypes were not measured until ages 13–14 (see Supplemental Material B), so we focus on computing and engineering stereotypes instead when discussing early emerging STEM domain differences.

Figure 5
Exploratory Analysis of Nonlinear Age Effects



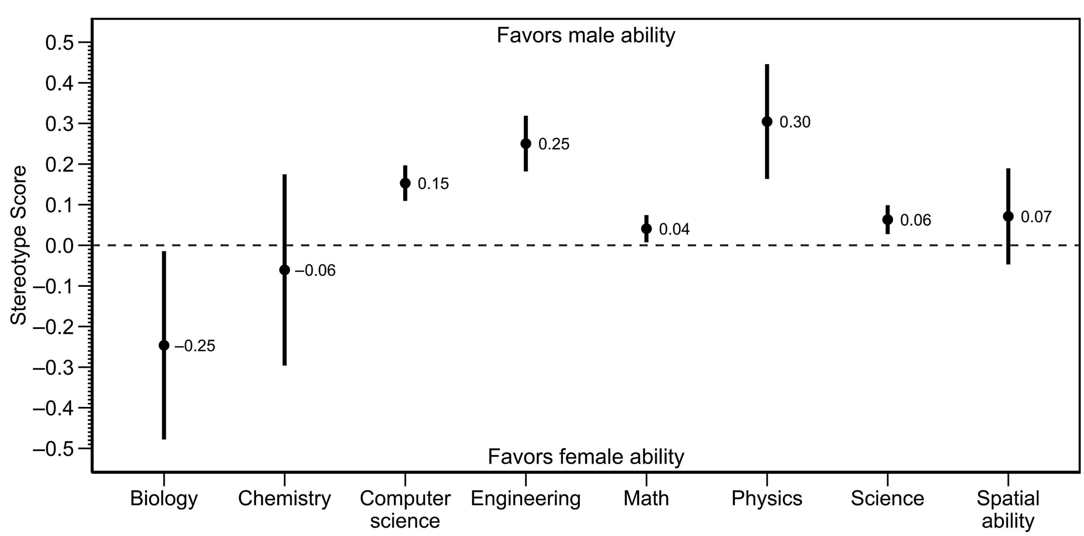
Note. These trends came from mixed-effects models that included methods covariates, confirmatory moderators, and dummy codes for age-gender groups (for quadratic models, see Supplemental Figure S5). The shaded areas represent 95% confidence intervals. STEM = science, technology, engineering, and mathematics. See the online article for the color version of this figure.

(e.g., auto repair) when constructing answers to measures about engineering. Relatedly, young children might have limited direct exposure to computer programming but could learn about male computer nerd stereotypes through mass media (e.g., television, movies; Cheryan et al., 2015; Hansen et al., 2017; Pantic et al., 2018). These broader stereotypes (e.g., about auto mechanics or computer nerd stereotypes) could shape children's early understanding of and engagement with computing and engineering, even when children later learn more field-specific information (e.g., receiving concrete definitions of "computing coding" or "engineering"; Master et al., 2021a).

The Results Call for Rethinking Research on STEM Ability Stereotypes

The substantial divergences across STEM fields call into question the field's predominant focus on stereotypes about math and general science (72% of the STEM effect sizes). In our view, computer

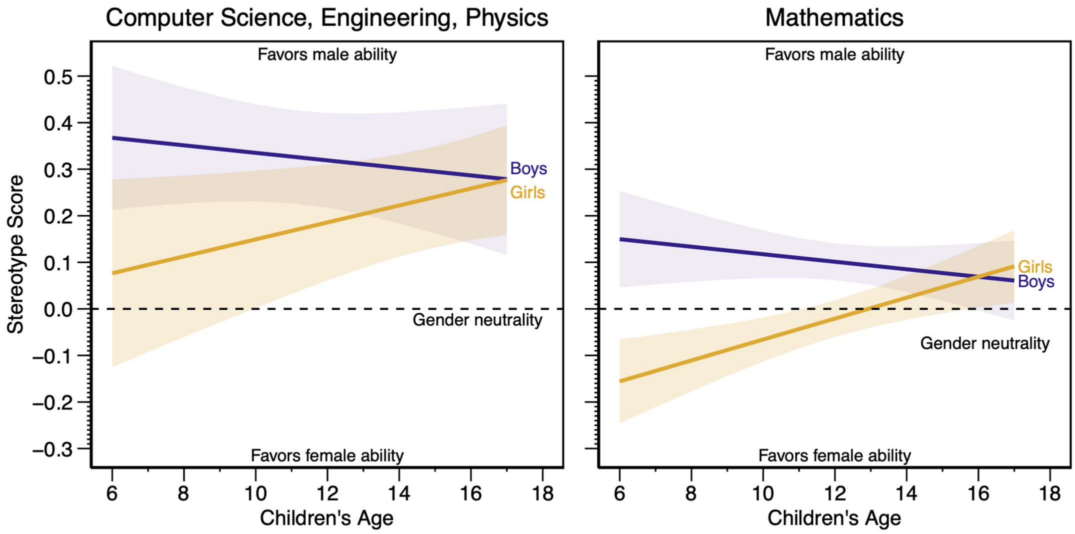
Figure 6
Moderation by Detailed STEM Domain



Note. Data points represent conditional means from models controlling for methods covariates and confirmatory moderators. Error bars denote 95% confidence intervals. STEM = science, technology, engineering, and mathematics.

Figure 7

Moderation by Age and Gender, Separately by STEM Domain



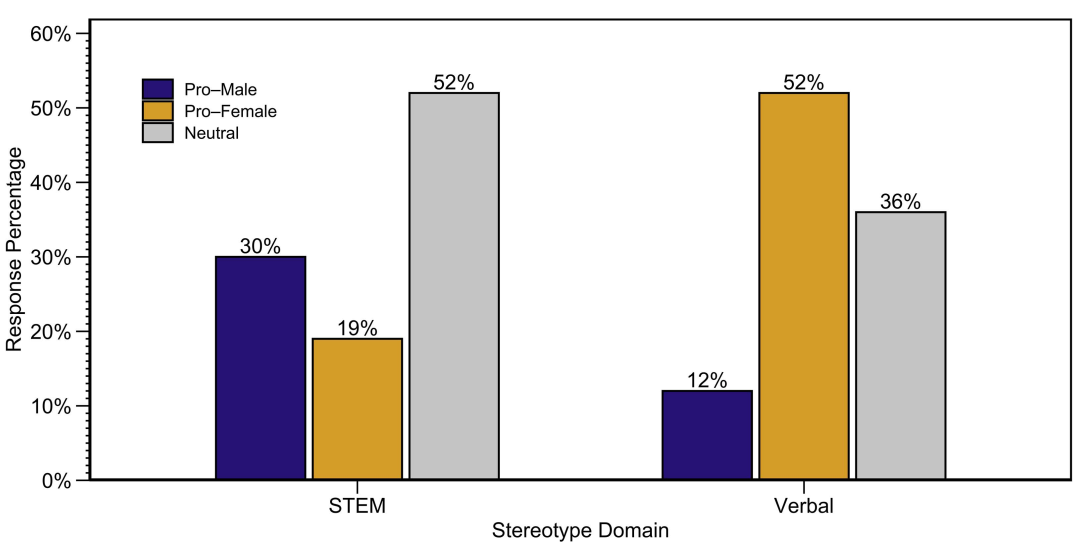
Note. These trends were estimated separately for computer science, engineering, and physics (left side) versus mathematics (right side), while controlling for methods covariates and other confirmatory moderators. The shaded areas represent 95% confidence intervals. STEM = science, technology, engineering, and mathematics. See the online article for the color version of this figure.

science, engineering, and physics stereotypes should instead take center stage in future research on children's gender stereotypes about STEM abilities.

Focusing on these male-dominated STEM fields offers a firmer empirical, theoretical, and policy-relevant foundation for understanding and addressing how stereotypes may contribute to STEM participation gaps (Master & Meltzoff, 2020). These stereotypes may both reflect and maintain the variation among adults: Gender gaps in college and the workforce are far larger for computer science, engineering, and physics relative to other STEM fields (Cheryan et al., 2017). These gender gaps are concerning from an equity perspective given that computer science and engineering jobs alone account for 78% of the college-educated U.S. STEM workforce (NCSES, 2023, Table 1-2,

Figure 8

Aggregate Qualitative Response Frequencies



Note. The figure represents the subset of studies in which those qualitative response frequencies were reported (see Supplemental Table S3 for more details). STEM = science, technology, engineering, and mathematics. See the online article for the color version of this figure.

Table 3
Summary of Major Findings for Confirmatory Moderators

Prediction	Empirical finding	Future research direction
	Age and gender differences	
Verbal ability stereotypes will generally strengthen with age	Confirmed (increase with age for girls and boys)	How do children learn verbal stereotypes? How do these beliefs affect other outcomes?
STEM ability stereotypes will generally strengthen with age	Increase for girls; nonlinear change for boys	What explains the unexpected nonlinearity for boys?
Participant gender differences will exist but decline with age	Confirmed (but see the discussion about asymmetries)	What explains why in-group bias developed asymmetrically across boys versus girls?
	Other predictions for STEM stereotypes	
Stronger for computer science, engineering, physics than other STEM fields	Confirmed	Devote far more attention to computer science, engineering, and physics stereotypes
Stronger for White than Black U.S. participants	Confirmed	Study intersectional targets (with diverse samples)
Indirect > direct measures Adult > child targets Exceptional > average targets	Nonsignificant findings (but imprecisely estimated)	Need more research on indirect measures, adult targets, and exceptional targets

Note. STEM = science, technology, engineering, and mathematics.

excluding health and social science), emphasizing their societal relevance.

If this research field continues to focus its primary empirical attention on children's math stereotypes, one risk is concluding that children do not hold STEM ability stereotypes on average. Statistical power calculations illustrate this point. Detecting the average math stereotype of 0.11 *SD*s at 80% power requires a sample size of $n = 651$ children, whereas detecting the average of 0.51 *SD*s for male-dominated STEM fields requires only $n = 32$ children. These demanding sample size requirements, along with substantial heterogeneity, illustrate why math ability stereotypes have shown such muddled findings in past research.

Results for Age Differences Mostly Aligned With Hypotheses

The results for age differences (Figure 5) largely aligned with preregistered hypotheses (Figure 1). Ability stereotypes shifted from in-group bias at age 6 (boys favoring boys, girls favoring girls) to traditional stereotypes by age 16 (favoring male STEM ability, female verbal ability). Beyond this qualitative change, verbal stereotypes generally increased with age and so did girls' STEM stereotypes. Participant gender differences robustly declined with age (typically $p < .001$ across many sensitivity analyses). All of these findings aligned with hypotheses.

Three points are important to note about the scope of these conclusions. First, these findings largely describe differences between ages 6 to 17, but not earlier, given that samples younger than 6 years were rare (1% of effect sizes). Second, these age differences were similar across STEM domains (even though the starting intercepts greatly differed), so we consider age differences for STEM fields as a whole to streamline the following discussion. Third, these age differences were far larger than birth cohort or data collection year effects, which were small in magnitude and not significant; we therefore interpret the age differences as *developmental change*, given that cohort confounds cannot explain the large age effects.

Unexpectedly, the development of boys' STEM stereotypes reversed with age, declining before age 13 but increasing after age 13. What explains the early decline for boys and later reversal in adolescence? Follow-up exploratory analyses on children's in-group bias provided empirical clues for answering these questions, as discussed next. In short, the early decline in boys' STEM stereotypes aligns with broader declines in boys' in-group bias during middle childhood.

Understanding an Asymmetric Development of In-Group Bias

Our initial hypotheses assumed that (a) in-group bias would characterize children's ability stereotypes at age 6 and (b) this bias would decline by similar amounts for boys and girls in later ages. The results supported assumption (a) but not (b). Instead, the results showed an *asymmetric* development of in-group bias, as this following section further discusses.

Definition of In-Group Bias

As mentioned earlier, we defined in-group bias as a *general* positivity toward one's own gender (e.g., "Boys rule, girls drool!"), aligning with other scholars (e.g., Powlishta et al., 1994). A boy expresses in-group bias when saying that boys are better at math *and* reading (i.e., boys are generally better). A boy who says "boys are good at math" but "girls are good at reading" expresses traditional stereotypes, not in-group bias per se (i.e., boys and girls have their respective strengths, rather than boys are generally better).

Empirical Differences

With this definition in mind, empirical analyses showed the following patterns about in-group bias: (a) similarity in boys' and girls' in-group bias at age 6, (b) decline for boys from ages 6 to 13, and (c) stability for girls. In addition to this asymmetric development, these patterns also imply a main effect of participant gender, such that girls generally showed more in-group bias than boys after ages 6–8 (see Supplemental Figure S6).

The main effect of participant gender (i.e., greater in-group bias for girls) closely aligns with many findings in other developmental literatures beyond ability stereotypes. For instance, girls like girls more than boys like boys (e.g., Dunham et al., 2016; Verkuyten & Thijs, 2001), a pattern also found among adult participants (e.g., Rudman & Goodwin, 2004). Compared to boys, girls also tend to show stronger in-group bias in assigning positive traits (e.g., *confident*, *affectionate*) to their own gender (Carver et al., 2003; Egan & Perry, 2001; C. F. Miller, Wheeler, & Woods, 2024; Powlishta, 1995; Powlishta et al., 1994; Serbin et al., 1993; Silvern, 1977; Susskind & Hodges, 2007; Zalk & Katz, 1978), as reviewed elsewhere (Powlishta, 2004, pp. 117–120).

The asymmetric development (i.e., greater decline with age for boys) also aligns with broader literatures. For instance, a study with 441 U.S. children also found asymmetric development across ages 4–17 for general in-group preferences, as assessed by both explicit measures (e.g., "Do you like girls or boys more?") and implicit measures (e.g., automatic gendered associations with good vs. bad; Dunham et al., 2016). Like our findings, this study found that boys' in-group bias sharply declined with age, whereas girls' in-group bias did not.

Speculative Mechanisms for Asymmetric Development

Why would in-group bias develop asymmetrically? The introduction section briefly cited three cognitive factors that might spur a general decline of in-group bias (e.g., growth in multiple classification skills, advances in moral reasoning), but these factors do not account for an *asymmetric* decline in bias.

One possible explanation concerns general academic stereotypes favoring girls. That is, when children enter formal schooling at ages 5–6, they may increasingly observe girls' superior grades in school (even in math and science; Voyer & Voyer, 2014). Children may then associate school as feminine (e.g., Hartley & Sutton, 2013; Heyder & Kessels, 2013), contributing to asymmetry across STEM versus verbal domains. Supporting this idea, exploratory analyses found that measures with school-based wording (e.g., "Who gets better math grades?") showed weaker pro-male STEM stereotypes than measures with wording about innate ability (e.g., "Who has more natural talent in math?"). Though this finding does not directly speak to age-related shifts, it illustrates how general associations of school as feminine could seep into domain-specific stereotypes, which could buffer against declines in girls' in-group bias that might occur otherwise.

Of course, other factors are possible too. For instance, other scholars have proposed that an awareness of greater male *social status* may explain asymmetry in ability stereotypes (Rowley et al., 2007, p. 152). As these scholars have argued, girls may resent their lower status in society and may seek to "self-enhance" when responding to ability stereotype measures, protecting their self-esteem and buffering against an age-related decline in their in-group bias (Kurtz-Costes et al., 2008, p. 390; Vuletich et al., 2020, p. 2). More generally, several other factors may contribute to greater in-group bias in girls than boys, such as positive stereotypes of female communion (e.g., warmth, Eagly & Mladinic, 1994, p. 28), positive maternal attachment (Rudman & Goodwin, 2004, p. 496), and negative experiences with male aggression and dominance (Dunham et al., 2016, p. 786), as reviewed elsewhere (Powlishta, 2004, pp. 117–120).

Explaining Unexpected Declines in Boys' STEM Stereotypes

This broader perspective about in-group bias helps to demystify the unexpected decline in boys' STEM stereotypes during middle childhood (ages 6–13). During that period, boys increasingly express more female positivity with age in other domains (e.g., general liking of girls vs. boys, verbal ability stereotypes), making their increasing female positivity for STEM ability stereotypes less surprising, despite pro-male STEM messages in their cultural environments. In contrast, for girls, stable in-group bias along with cultural learning of pro-male STEM messages would contribute to increasing male bias in girls' STEM stereotypes during this developmental period.

Changes in Adolescence

Despite earlier declines, boys' in-group bias showed stability during adolescence (ages 13–17; see Supplemental Figure S6), suggesting that some of the relevant constructs may stabilize by adolescence (e.g., multiple classification skills). Consistent with this idea, a prior meta-analysis on ethnic, racial, and national prejudice found stable bias for high-status groups in adolescence, despite earlier declines in childhood (Raabe & Beelmann, 2011).

With stable in-group bias, several other factors may explain why both boys' and girls' STEM stereotypes became more male-biased in adolescence, as found in our exploratory analyses (for longitudinal evidence, see also Skinner et al., 2021; Starr et al., 2023). The introduction section noted how relevant external observations (e.g., more boys than girls in Advanced Placement Computer Science) and internal observations (e.g., boys' self-confidence in math) could become salient in adolescence, strengthening traditional stereotypes. As another factor, the meaning of "math" or "science" could shift in adolescence too: As high school students explore future career paths, they may increasingly view math and science as fields dominated by adult men, rather than feminine school subjects.[15]

Summary

Though the exact underlying mechanisms are presently unclear, the asymmetric development of in-group bias (as expressed in children's STEM and verbal ability stereotypes) aligns with other

[15] Consistent with this idea, some tentative evidence suggested that pro-male STEM stereotypes may be stronger for adult than child targets ($p = .079$ in the confirmatory model). The effect was significant for girls, but not boys, in sensitivity analyses, consistent with some individual studies (Conlon, 2020; Martinot et al., 2012; Steele, 2003; though see Hildebrand et al., 2021). Though these findings do not directly speak to age-related shifts, they support that STEM stereotypes might strengthen as adolescents increasingly think of the adult-related connotations of "math" and "science."

developmental literatures and helps demystify other age-related findings.

Theoretical Implications for Cognitive Theories of Gender Development

The findings on age differences extend cognitive theories of gender development by stressing the need to explain an asymmetric development of in-group bias. Though developmental scholars have long found stronger in-group bias in girls than boys (e.g., Silvern, 1977), this empirical conclusion typically has not featured prominently in theories of stereotype development. For theories like Bem's (1983) gender schema theory, the reader is left to assume that the cognitive processes apply equally across girls and boys (the theory does not state otherwise). Our initial hypotheses (e.g., Figure 1) reflect the tacitly implied symmetry.

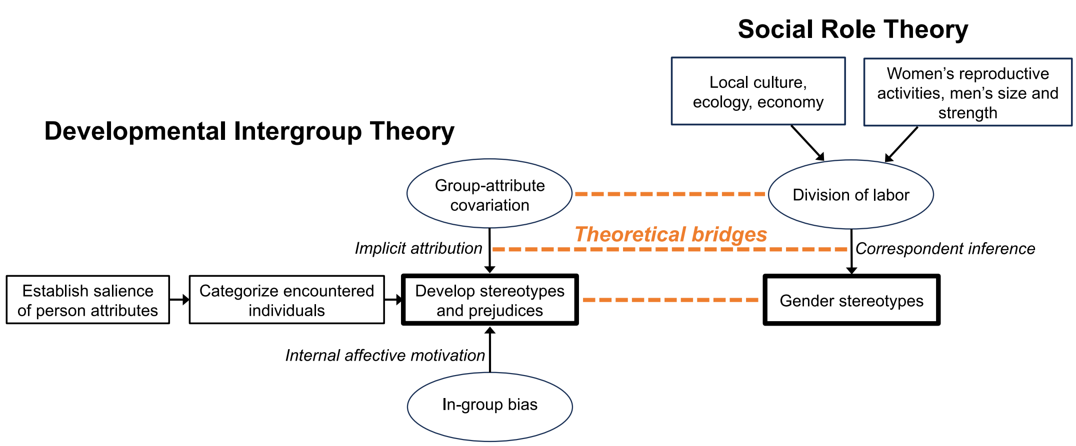
Though surprising at first, these asymmetries make sense in a broader societal context. In most societies, men hold greater power than women across several domains such as business, politics, control of resources, and property ownership. With this broader perspective in mind, asymmetries should be expected for at least some areas of gender development. Bridging social role theory (W. Wood & Eagly, 2012) with developmental theories provides one promising path to better connect these societal and cognitive perspectives. Though a full integration is clearly beyond our present scope, we briefly outline this direction here for others to pursue, inspired by our empirical findings and Leaper's (2011) recommendations for *theory bridging*.

Social role theory complements and aligns especially well with developmental intergroup theory (Bigler & Liben, 2006), which provides an integrative account of how children learn stereotypes with age, while accounting for cognitive developmental changes (see also Arthur et al., 2008; Liben, 2014, 2017). The central bridge between these two theories is the assumed cognitive process of inferring internal traits from external behaviors (Figure 9). That is, *implied attribution* in developmental intergroup theory maps to *correspondent inference* in social role theory (likewise, *group-attribute covariation* maps to *division of labor*). In both theories, perceivers actively construct stereotypes. For instance, high school students may infer male math ability (an internal trait) from observing a man teaching math (an external behavior; Sansone, 2019).

Developmental intergroup theory helps frames how in-group bias depends on age-related changes in cognitive maturation (e.g., multiple classification skills; Bigler & Liben, 2006, p. 77), as well as environmental features that establish the psychological salience of gender (the far-left box in Figure 9). A social role perspective extends this understanding by looking to a society's division of labor to explain asymmetries. For instance, women were 89% of U.S. elementary school teachers in 2021 (Irwin et al., 2022), and men have been 100% of U.S. presidents so far. Observing these extreme gender imbalances could lead to stereotypes of greater female academic ability and stereotypes of greater male power and agency. Hence, the development of children's stereotypes is not random or arbitrary. Rather, it reflects the asymmetric distribution of roles in society (e.g., women in roles that convey competence but not power; Eagly et al., 2020). Both of these cognitive and societal perspectives are needed to more fully account for the complex processes of stereotype development in future theoretical integrations and extensions.

Figure 9

Bridges Between Developmental Intergroup Theory and Social Role Theory



Note. This figure presents simplified versions of the two theories to emphasize the theoretical bridges between them. Developmental intergroup theory also details two other contributors to stereotype development (essentialism and explicit attributions). Social role theory also details the downstream consequences of gender stereotypes on regulatory processes (hormonal, social, and self). Essentialism and explicit attributions adapted from "A Developmental Intergroup Theory of Social Stereotypes and Prejudice," by R. S. Bigler and L. S. Liben, in R. V. Kail (Ed.), *Advances in Child Development and Behavior* (p. 54), 2006, Elsevier. Copyright 2006 by Elsevier. Hormonal, social, and self adapted from "Biosocial Construction of Sex Differences and Similarities in Behavior," by W. Wood and A. H. Eagly, in M. P. Zanna and J. M. Olson (Eds.), *Advances in Experimental Social Psychology* (p. 58), 2012, Elsevier. Copyright 2012 by Elsevier. See the online article for the color version of this figure.

Further Extensions With Intersectional Perspectives

Characterizing the complex processes of stereotype development also requires understanding variation within a society (e.g., across race or ethnicity). As predicted, U.S. Black children had weaker pro-male STEM stereotypes than U.S. White children, aligning with past related findings for adults (Eagly et al., 2020, p. 310; O'Brien et al., 2015). This participant race difference was also sensitive to participant gender (i.e., strongest for Black vs. White girl participants, not significant for Black vs. White boy participants). The direction of this sensitivity aligns with a recent nationally representative analysis (Starr et al., 2023), which found that Black girls had the weakest pro-male math ability stereotypes compared to any other participant group (e.g., Black boys, White boys, White girls).

Why do Black girls show this nontraditionality, even in late adolescence? Starr et al. (2023) hypothesized that, as a disenfranchised group, Black girls have "less to gain by upholding the status quo and may be more aware of social justice issues" (p. 19). This hypothesis is plausible, but other factors may be at play too. Theoretical perspectives on *intersectional invisibility* argue that individuals with multiple subordinate-group identities (e.g., Black girls) can experience both distinct advantages and disadvantages (Purdie-Vaughns & Eibach, 2008). Black girls may be overlooked and marginalized because they are not seen as prototypical of "girls" or "Black children," rendering them metaphorically invisible. But this invisibility can also sometimes protect racial minority girls and women by making them less prototypical targets of some common biases and stereotypes (e.g., Biernat & Sesko, 2013; Ghavami & Peplau, 2013).

We hesitate to make strong theoretical claims from the meta-analyzed data given a major evidence gap: Almost no evidence exists for intersectional targets in this literature. That is, the underlying measures used generic targets like "girls" and "boys" without specifying the targets' race or ethnicity. The meta-analyzed evidence therefore speaks to intersectionality among participants (who endorses the stereotype), not stereotype targets (who the stereotype concerns). Also, less is known about other racial/ethnic groups beyond Black and White children (e.g., estimates were less precise for Asian or Hispanic/Latinx children; though see Starr et al., 2023). Overall, more research is needed on intersectional targets, while using racially and ethnically diverse samples.

Nonsignificant Findings for Confirmatory Measurement Characteristics

Like intersectional targets, another major evidence gap was the field's limited use of adult targets (5% of STEM effect sizes), indirect measures (5%), and exceptional targets (2%) in ability stereotype measures. The null findings for the confirmatory measurement characteristics (e.g., target age) should be interpreted somewhat cautiously given those limitations. Nevertheless, the findings for indirect measures still warrant further discussion about how they inform hypotheses about social desirability bias (e.g., if saying "boys are good at math" is a social taboo).

The Findings Mitigate Concern About Social Desirability Bias

In total, 10 studies used at least one indirect measure that did not explicitly call attention to gender (e.g., "Draw a student who is good at math"), indicating a small but nontrivial evidence base (Supplemental Table S1). Theoretically, if children "hide" their bias on direct measures due to social desirability, we should expect to see much larger means for indirect measures that "uncover" that bias. But the results did not align with that hypothesis: If anything, the (nonsignificant) trend was in the opposite direction (smaller mean for indirect measures; Supplemental Figure S8), mitigating concern about social desirability bias.

Can Social Desirability Bias Explain the Age-Related Findings?

Despite the somewhat reassuring findings for indirect measures, is it still possible that social desirability bias might account for some of the age-related findings? For instance, boys might increasingly learn that saying "boys are better at STEM" is a social taboo, explaining their declining STEM stereotypes during ages 6–13. In our view, this explanation raises far more questions than it answers. For instance, if expressing stereotyped beliefs is such a taboo, then why would girls and boys be so comfortable with stating strong pro-male beliefs about computer science and engineering? Does the reversal of boys' STEM stereotypes in adolescence mean that social desirability bias somehow lessens after age 13? Do the same processes not apply to girls' STEM stereotypes? And why would verbal stereotypes generally become more traditional with age if children are learning to hide stereotyped beliefs? Though answers to these questions may exist,[16] we overall do not view social desirability bias as a plausible explanation for the observed age differences.

Implications for "Implicit" Associations

Our results for indirect measures contrast with those for "implicit" associations (cultural fit stereotypes that our meta-analysis did not include). For instance, IATs have found gendered associations with STEM versus verbal subjects among elementary school children (e.g., Cvencek et al., 2011, 2014; del Río et al., 2021). These IAT measures have a serious problem though: They confound STEM with verbal stereotypes. That is, IATs assess gendered associations with math *versus language* (but typically not gendered associations with math *only*). This confound is crucial given our findings about strong verbal stereotypes, yet weak math stereotypes. Many past findings on children's implicit "math" stereotypes might instead reflect gender–verbal associations, not gender–math associations, as others have cautioned (Levine & Pantoja, 2021, p. 25; Vuletich et al., 2020, p. 5). For instance, girls have not shown math–male bias in reaction-time studies that separately assessed math and verbal stereotypes (Nowicki & Lopata, 2017; Steffens & Jelenec, 2011).

[16] For instance, maybe stating beliefs about gender differences is more socially sensitive for STEM than verbal ability. And maybe girls' STEM stereotypes are less suspectable to social desirability if stating negative stereotypes is more socially acceptable for in-group members (i.e., girls) than out-group members (i.e., boys). These hypotheses are highly speculative. Empirical evidence suggests that young children are sensitive to social norms about expressing racial prejudice (e.g., Rutland et al., 2005), but gender differences are presumably less socially sensitive.

Summary

The available evidence for indirect measures indicates two key conclusions. First, the concern about social desirability bias for direct measures is likely not extreme.[17] Second, previous studies on children's implicit "math" stereotypes might instead reflect math–verbal confounds. We urge the field to use more precise terminology like *math-reading IAT* or *associations with science versus liberal arts* when contrast domain effects cannot be ruled out.

Limitations

Limitations in the Evidence Base

As previously noted, limitations in the evidence base constrained the conclusions, especially regarding adult targets, exceptional targets, indirect measures, intersectional targets, and preschool children. Relatedly, power analyses indicated serious data limitations for detecting systematic cross-national or cross-temporal effects.[18] The wide variation in measures and constructs used across studies, time points, and nations likely contributed to the low power. Related past studies that used a single measure to examine variation across 66 nations (D. I. Miller et al., 2015) or changes since the 1960s (D. I. Miller, Nolla, et al., 2018) have found weakened STEM stereotypes in cultures with more women in STEM nationally.

Limitations in the Review Methods Used

Due to resource constraints, we dual screened only 20% of abstracts and 56% of full-text articles, which introduced some risk of missing relevant studies. We also dual coded only 18% of eligible studies, which introduced some risk of error. However, several other methodological strengths mitigated those risks (e.g., using Abstrackr to prioritize dual screening the more relevant abstracts, checking for consistent interpretation of eligibility criteria and coding guidelines in weekly meetings, the lead author checking the entered data for any study that was not dual coded). We are confident that such limitations would not change the review's overall conclusions, especially given the risk mitigation strategies.

The literature search likely also missed some relevant non-English reports, given that our keyword searches were English-based. Using Google Translate as the primary translation method could also have led to some misinterpretations. Nevertheless, the searches found many non-English reports via (a) literature databases that had translated English abstracts and (b) citation tracking in Google Scholar (e.g., non-English reports citing eligible studies). Methodological research has also shown Google Translate to generally yield accurate and reliable translations for the purposes of systematic reviews (Jackson et al., 2019), in part due to major improvements in its algorithm in recent years (Turner, 2016).

Excluding Measures Without a Meaningful Zero Point

While not a limitation per se, we included only measures that allowed for computing effect sizes with a meaningful zero point (see Criterion 3 in the Method section). This criterion meant needing to exclude Likert agreement ratings that lacked analogously worded items in the reverse direction (e.g., rate agreement to "boys are better," without an analogous item for "girls are better"). This exclusion is not a major constraint to the conclusions for two reasons. First, even if we had included such measures, we would have analyzed them in separate meta-analytic models given the large structural differences (keeping results for the measures we included intact). Second, the excluded Likert measures raise concern about the question structure biasing children's responses (called *acquiescence bias*). A child might agree that "boys are better in math" based on a general tendency to agree with a researcher's statements, even if the child does not strongly hold such beliefs (Saris et al., 2010).

Constraints Due to Construct Focus

Lastly, the construct focus on STEM and verbal ability stereotypes constrains our focus to *descriptive stereotypes about positive traits*, which can differ from other gender stereotypes in several critical ways as follows (see also C. F. Miller, Wheeler, & Woods, 2024):

- *descriptive stereotypes:* As reviewed previously, prescriptive stereotypes about how the world *should* be (e.g., "Who should study engineering?") often show radically different patterns of development than descriptive stereotypes (e.g., Signorella et al., 1993).
- *trait stereotypes:* Findings for descriptive stereotypes about traits (e.g., ability, warmth) may not directly generalize to descriptive stereotypes about highly gendered activities, occupations, or toys (e.g., "Who plays with dolls?"), which develop far more rapidly and often reach ceiling levels of knowledge by age 6 (e.g., see Serbin et al., 1993, Figure 1).
- *positive traits:* In-group bias often emerges far more starkly for stereotypes about positive traits (e.g., confident) than negative traits (e.g., aggressive; Susskind & Hodges, 2007, Figure 1). Hence, the evaluative processes that shape in-group bias may operate differently for stereotypes of positive versus negative traits.

This focus constrains the conclusions but also suggests how they might generalize. For instance, the present results provide hypotheses to test in future research on descriptive gender stereotypes about positive *agentic* traits (e.g., ambitious) and *communal* traits (e.g., kind). This focus also implies that the results might not necessarily generalize to STEM and verbal stereotypes in nonability domains (e.g., associations or interest stereotypes; D. I. Miller, Nolla, et al., 2018; Tang et al., 2024).

Recommendations for Future Research

Understanding the complexity of how these ability stereotypes develop and shape children's lives requires collaborative efforts across multiple fields such as developmental science, social psychology,

[17] Nevertheless, indirect and direct measures likely assess different constructs (e.g., associative vs. propositional processes; Gawronski & Bodenhausen, 2006, 2011). Hence, the field's limited use of indirect measures constrains our main conclusions to accounts of *explicit* social cognition, which is an important phenomenon in its own right.

[18] To illustrate, the analysis did not even have 50% power to detect a theoretical change in U.S. STEM stereotypes from −0.18 standard deviations in 1977 to +0.18 standard deviations in 2019.

sociology, and education. Our synthesis guides research in this direction by integrating diverse empirical literatures with multiple perspectives from psychology and beyond. An interactive data tool allows scholars to find specific prior studies to help guide their future research in these directions: https://d-miller.shinyapps.io/STEM_verbal_stereotypes/.

For STEM stereotypes, the findings indicate a clear, pressing need for this field to shift far more attention to computer science, engineering, and physics stereotypes. Responding to this need can occur in tandem with addressing other critical evidence gaps (i.e., studying with indirect measures, adult targets, exceptional targets, intersectional targets, or preschool children).

For verbal stereotypes, the findings call for investigating (a) the cues that children use in learning verbal ability stereotypes with age, (b) other moderators that may further explain variability, and (c) the downstream effects on reading and writing outcomes. These directions are especially important given the large magnitudes for verbal ability stereotypes, their clear increase with age, and their relevance for understanding boys' underachievement in reading and writing.

For intervention research, the findings raise the question: Should applied interventions aim to directly *challenge* these stereotypes? Some interventions have found short-term change in stereotypes by countering negative messages about girls' STEM abilities (e.g., Cyr et al., 2024; F. Zhao et al., 2018). However, other scholars have argued that gender stereotypes are likely to reemerge later, due to the difficulty of long-lasting stereotype change (Eagly & Koenig, 2021). Given this tension, some scholars have instead suggested to *broaden* stereotypes (e.g., highlight diversity) or *mitigate the effects* of stereotypes (e.g., with growth mindset messages; Master & Meltzoff, 2020). Regardless of the exact approach, our results provide future intervention researchers with key developmental context about when children tend to endorse certain beliefs.

Conclusions and Implications

The findings overall emphasize the variability in children's ability stereotypes, while also calling for far more research on computing, engineering, physics, and verbal stereotypes. Age-related findings showed a simple qualitative conclusion and a more complex quantitative conclusion. Qualitatively, STEM and verbal stereotypes shifted from in-group bias at age 6 to traditional stereotypes by age 16. Quantitatively, the development of STEM versus verbal stereotypes was *asymmetric*, with a developmental reversal for boys' STEM stereotypes.

These empirical findings provided opportunities to build novel theoretical bridges. Integrating cognitive theories with social role theory implicates broader developmental and societal factors that may extend beyond ability stereotypes. Specifically, in-group bias declined faster for boys than girls, as empirically found in other domains too. This theoretical analysis offers new testable hypotheses for other descriptive gender stereotypes about positive traits (e.g., agency, communion), while better accounting for both cognitive and societal considerations.

More proximally, the early onset of computing, engineering, and verbal ability stereotypes echoes that young children develop a wealth of understanding about gender even before formal schooling. With this early onset, multiple mechanisms exist for how these beliefs could contribute to boys' underachievement in reading or variation in women's participation across STEM fields. For instance, at age 6, children already believe that boys and men are better at computing and engineering; girls increasingly endorse this belief in older ages. These stereotypes therefore have ample time to affect children's motivations and behaviors as children engage with these subjects in formal schooling and select diverging pathways in later grades.

References

References marked with an asterisk indicate reports included in the meta-analysis.

Aboud, F. E. (2008). A social-cognitive developmental theory of prejudice. In S. M. Quintana & C. McKown (Eds.), *Handbook of race, racism, and the developing child* (pp. 55–71). Wiley.

*Abramovich, S. (1997). *The relation between gender-role stereotypes and math performance as mediated by math anxiety* (Publication No. 1384573) [Master's thesis, Eastern Michigan University]. ProQuest Dissertations and Theses.

Alfieri, T., Ruble, D. N., & Higgins, E. T. (1996). Gender stereotypes during adolescence: Developmental changes and the transition to junior high school. *Developmental Psychology*, *32*(6), 1129–1137. https://doi.org/10.1037/0012-1649.32.6.1129

*Alkhadrawi, A. (2015). *Gender differences in math and science choices and preferences* [Unpublished dissertation]. University of Toledo.

Arthur, A. E., Bigler, R. S., Liben, L. S., Gelman, S. A., & Ruble, D. N. (2008). Gender stereotyping in young children: A developmental intergroup perspective. In S. Levy & M. Killen (Eds.), *Intergroup attitudes and relations in childhood through adulthood* (pp. 66–86). Oxford University Press. https://doi.org/10.1093/oso/9780195189742.003.0005

Bahar, A. K., Kaya, E., & Zhang, X. (2022). Gender disparities in AP Computer Science exams: Analysis of trends in participation and top achievement. *Journal of Advanced Academics*, *33*(4), 574–603. https://doi.org/10.1177/1932202X221119499

*Barkatsas, A., Forgasz, H., & Leder, G. (2001). The gender stereotyping of Mathematics: Cultural dimensions. In J. Bobis, B. Perry, & M. Mitchelmore (Eds.), *Proceedings of the twenty-fourth annual conference of the Mathematics Education Research Group of Australasia Inc*, (pp. 74–81). Mathematics Education Research Group of Australasia.

Barnard, J., & Rubin, D. B. (1999). Miscellanea. Small-sample degrees of freedom with multiple imputation. *Biometrika*, *86*(4), 948–955. https://doi.org/10.1093/biomet/86.4.948

*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

Bem, S. L. (1983). Gender schema theory and its implications for child development: Raising gender-aschematic children in a gender-schematic society. *Signs: Journal of Women in Culture and Society*, *8*(4), 598–616. https://doi.org/10.1086/493998

Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, *355*(6323), 389–391. https://doi.org/10.1126/science.aah6524

Biernat, M., & Sesko, A. K. (2013). Evaluating the contributions of members of mixed-sex work teams: Race and gender matter. *Journal of Experimental Social Psychology*, *49*(3), 471–476. https://doi.org/10.1016/j.jesp.2013.01.008

Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudice. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 34, pp. 39–89). Elsevier. https://doi.org/10.1016/S0065-2407(06)80004-2

*Bogdanova, O. Y., Miklashevsky, A., Bogdanova, E. L., & Soldatenkova, O. B. (2019). Academic achievement in math and foreign language: Individual characteristics and gender stereotypes. *Sibirskiy Psikhologi-*

cheskiy Zhurnal, *73*(73), 176–196. https://doi.org/10.17223/17267080/73/11

*Bömer, A. T., & Imhof, M. (2011). Computer sind doch nichts für Mädchen!? Computer-bezogene Geschlechterstereotypen bei Grundschülerinnen und Grundschülern [Computers aren't for girls, are they? Computer-related gender stereotypes in elementary school students]. *Psychologie in Erziehung und Unterricht*, *3*(3), 189–197. https://doi.org/10.2378/peu2011.art12d

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley. https://doi.org/10.1002/9780470743386

Borenstein, M., Higgins, J. P., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, *8*(1), 5–18. https://doi.org/10.1002/jrsm.1230

*Brandell, G., Larsson, S., Nystrom, P., & Palbom, A. (2005). *Kön och matematik*. https://www.diva-portal.org/smash/get/diva2:998356/FULLTEXT01.pdf

*Brandell, G., Leder, G., & Nystrom, P. (2007). Gender and Mathematics: Recent development from a Swedish perspective. *ZDM—International Journal on Mathematics Education*, *39*(3), 235–250. https://doi.org/10.1007/s11858-007-0025-4

*Brandell, G., Nyström, P., & Staberg, E.-M. (2003). *Matematik i gymnasieskolan*. https://www.maths.lth.se/GeMa/reports/Skolrappgymn.pdf

*Brandell, G., Nystrom, P., & Sundqvist, C. (2004). *Mathematics-a male domain?* https://www.diva-portal.org/smash/get/diva2:1013433/FULLTEXT01.pdf

*Brandell, G., & Staberg, E. (2008). Mathematics: A female, male or gender-neutral domain? A study of attitudes among students at secondary level. *Gender and Education*, *20*(5), 495–509. https://doi.org/10.1080/09540250701805771

*Brown, B. A. (2007). *Student perceptions regarding gender-related learning differences and their correlations to achievement outcomes* (Publication No. 3283868) [Doctoral dissertation, Boston College]. ProQuest Dissertations and Theses Global.

Buecker, S., Mund, M., Chwastek, S., Sostmann, M., & Luhmann, M. (2021). Is loneliness in emerging adults increasing over time? A preregistered cross-temporal meta-analysis and systematic review. *Psychological Bulletin*, *147*(8), 787–805. https://doi.org/10.1037/bul0000332

Bühler, J. L., Krauss, S., & Orth, U. (2021). Development of relationship satisfaction across the life span: A systematic review and meta-analysis. *Psychological Bulletin*, *147*(10), 1012–1053. https://doi.org/10.1037/bul0000342

*Cansız, M., Cansız, N., Tas, Y., & Yerdelen, S. (2017). Turkish version of students' ideas about nature of science questionnaire: A validation study. *International Journal of Progressive Education*, *13*(1), 42–51. https://eric.ed.gov/?id=EJ1132069

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, *2*(2), 115–144. https://doi.org/10.1177/2515245919847196

Carver, P. R., Yunger, J. L., & Perry, D. G. (2003). Gender identity and adjustment in middle childhood. *Sex Roles*, *49*(3/4), 95–109. https://doi.org/10.1023/A:1024423012063

Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interest*, *15*(3), 75–141. https://doi.org/10.1177/1529100614541236

Charlesworth, T. E. S., Yang, V., Mann, T. C., Kurdi, B., & Banaji, M. R. (2021). Gender stereotypes in natural language: Word embeddings show robust consistency across child and adult language corpora of more than 65 million words. *Psychological Science*, *32*(2), 218–240. https://doi.org/10.1177/0956797620963619

*Chatard, A., Guimond, S., & Selimbegovic, L. (2007). "How good are you in math?" The effect of gender stereotypes on students' recollection of their school marks. *Journal of Experimental Social Psychology*, *43*(6), 1017–1024. https://doi.org/10.1016/j.jesp.2006.10.024

*Chen, S., Chang, W.-H., Lieu, S.-C., Kao, H.-L., Huang, M.-T., & Lin, S.-F. (2013). Development of an empirically based questionnaire to investigate young students' ideas about nature of science. *Journal of Research in Science Teaching*, *50*(4), 408–430. https://doi.org/10.1002/tea.21079

Cheryan, S., Master, A., & Meltzoff, A. N. (2015). Cultural stereotypes as gatekeepers: Increasing girls' interest in computer science and engineering by diversifying stereotypes. *Frontiers in Psychology*, *6*, Article 49. https://doi.org/10.3389/fpsyg.2015.00049

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin*, *143*(1), 1–35. https://doi.org/10.1037/bul0000052

*Chikuvadze, P., & Matswetu, V. S. (2013). Gender stereotyping and female pupils' perception of studying Advanced Level Sciences: A survey of one province in Zimbabwe. *Gender & Behaviour*, *11*(1), 5285–5296. https://hdl.handle.net/10520/EJC136361

Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research*, *34*(3), 315–346. https://doi.org/10.1207/S15327906MBR3403_2

Cole, E. R. (2009). Intersectionality and research in psychology. *American Psychologist*, *64*(3), 170–180. https://doi.org/10.1037/a0014564

*Conlon, R. (2020). *Examining the relationship between teacher math-gender stereotypes and students' math outcomes* (Publication No. 28155989) [Doctoral dissertation, The Florida State University]. ProQuest Dissertations and Theses Global.

*Cooper, I. (2013). *The waiting time: Student perceptions of gender bias in middle school mathematics* [Master's thesis]. ProQuest Dissertations and Theses Global.

*Copping, K. E. (2008). *Sex and race differences in the development of gender stereotypes* [Master's thesis]. ProQuest Dissertations and Theses Global.

*Copping, K. E. (2011). *High school students' career aspirations: Influences of gender stereotypes, parents, and the school environment* [Doctoral dissertation]. ProQuest Dissertations and Theses Global.

*Copping, K. E., Kurtz-Costes, B., Wood, D., & Skinner, O. D. (2013). *Race differences in the development of academic gender stereotypes of Black and White adolescents* [Paper presentation]. The 2013 meeting for the Society for Research on Child Development, Seattle, WA, United States.

Corneille, O., & Hütter, M. (2020). Implicit? What do you mean? A comprehensive review of the delusive implicitness construct in attitude research. *Personality and Social Psychology Review*, *24*(3), 212–232. https://doi.org/10.1177/1088868320911325

*Cousins, A., & Mills, M. (2015). Gender and high school chemistry: Student perceptions on achievement in a selective setting. *Cambridge Journal of Education*, *45*(2), 187–204. https://doi.org/10.1080/0305764X.2014.934202

Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and the violence against women of color. *Stanford Law Review*, *43*(6), 1241–1299. https://doi.org/10.2307/1229039

Crouter, A. C., Whiteman, S. D., McHale, S. M., & Osgood, D. W. (2007). Development of gender attitude traditionality across middle childhood and adolescence. *Child Development*, *78*(3), 911–926. https://doi.org/10.1111/j.1467-8624.2007.01040.x

Cunningham, C., Lachapelle, C., & Lindgren-Streicher, A. (2005, June). *Assessing elementary school students' conceptions of engineering and technology* [Conference session]. Proceedings of the 2005 annual conference proceedings for the American Society for Engineering Education, Portland, OR, United States.

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 Experi-*

mental Child Psychology, *117*, 73–91. https://doi.org/10.1016/j.jecp.2013.07.018

Cyr, E. N., Kroeper, K. M., Bergsieker, H. B., Dennehy, T. C., Logel, C., Steele, J. R., Knasel, R. A., Hartwig, W. T., Shum, P., Reeves, S. L., Dys-Steenbergen, O., Litt, A., Lok, C. B., Ballinger, T., Nam, H., Tse, C., Forest, A. L., Zanna, M., Staub-French, S., … Spencer, S. J. (2024). Girls are good at STEM: Opening minds and providing evidence reduces boys' stereotyping of girls' STEM ability. *Child Development*, *95*(2), 636–647. https://doi.org/10.1111/cdev.14007

*David, L. T., & Tompa, A. (2018). Are gender stereotypes still recognized in young children of new generation? *Journal Plus Education*, *21*, 109–116. https://www.ceeol.com/search/article-detail?id=726652

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

*Diez, J. L., Ramos, A., & Candela, C. (2022). Static and dynamic assessment of STEM gender stereotypes in secondary education using a novel cluster-based analysis. *International Journal of Technology and Design Education*, *33*, 749–774. https://doi.org/10.1007/s10798-022-09746-1

*Dryden, K. L. (2013). *Implicit and explicit gender stereotypes and their relationship to self-concept* [Unpublished undergraduate thesis]. University of North Carolina at Chapel Hill.

Dunham, Y., Baron, A. S., & Banaji, M. R. (2016). The development of implicit gender attitudes. *Developmental Science*, *19*(5), 781–789. https://doi.org/10.1111/desc.12321

Eagly, A. H., & Koenig, A. M. (2021). The vicious cycle linking stereotypes and social roles. *Current Directions in Psychological Science*, *30*(4), 343–350. https://doi.org/10.1177/09637214211013775

Eagly, A. H., & Mladinic, A. (1994). Are people prejudiced against women? Some answers from research on attitudes, gender stereotypes, and judgments of competence. *European Review of Social Psychology*, *5*(1), 1–35. https://doi.org/10.1080/14792779543000002

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. *American Psychologist*, *75*(3), 301–315. https://doi.org/10.1037/amp0000494

Eagly, A. H., & Wood, W. (2012). Social role theory. In P. van Lange, A. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories in social psychology* (pp. 458–476). Sage Publications. https://doi.org/10.4135/9781446249222.n49

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

Egan, S. K., & Perry, D. G. (2001). Gender identity: A multidimensional analysis with implications for psychosocial adjustment. *Developmental Psychology*, *37*(4), 451–463. https://doi.org/10.1037/0012-1649.37.4.451

*Eisen, S., Matthews, S. E., & Jirout, J. (2021). Parents' and children's gendered beliefs about toys and screen media. *Journal of Applied Developmental Psychology*, *74*, Article 101276. https://doi.org/10.1016/j.appdev.2021.101276

*Elliott, A. (1988, June). *Sex typing of young children's behaviours in a computer active preschool classroom. Project report: A working paper* [Paper presentation]. The National Educational Computing Conference, Dallas, TX, United States. https://eric.ed.gov/?id=ED302334

Else-Quest, N. M., & Hyde, J. S. (2016). Intersectionality in quantitative psychological research: I. Theoretical and epistemological issues. *Psychology of Women Quarterly*, *40*(2), 155–170. https://doi.org/10.1177/0361684316629797

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

*Emran, A., Spektor-levy, O., Paz Tal, O., & Ben Zvi Assaraf, O. (2020). Understanding students' perceptions of the nature of science in the context of their gender and their parents' occupation. *Science & Education*, *29*(2), 237–261. https://doi.org/10.1007/s11191-020-00103-z

*Espinoza, A. M. (2021). *Gender identity and stereotypes as key factors in reading motivation: Three studies in secondary school in Chile* [Doctoral dissertation]. ProQuest Dissertation and Theses.

*Espinoza, A. M., & Strasser, K. (2020). Is reading a feminine domain? The role of gender identity and stereotypes in reading motivation in Chile. *Social Psychology of Education*, *23*(4), 861–890. https://doi.org/10.1007/s11218-020-09571-1

*Evans, A. B., Copping, K., Rowley, S. J., & Kurtz-Costes, B. (2011). Academic self-concept in black adolescents: Do race and gender stereotypes matter? *Self and Identity*, *10*(2), 263–277. https://doi.org/10.1080/15298868.2010.485358

Fennema, E., & Sherman, J. A. (1976). Fennema-Sherman mathematics attitudes scales: Instruments designed to measure attitudes toward the learning of mathematics by females and males. *Journal for Research in Mathematics Education*, *7*(5), 324–326. https://doi.org/10.2307/748467

Fisk, S. R., & Ridgeway, C. L. (2018). Framing gender. In B. Risman, C. Froyum, & W. Scarborough (Eds.), *Handbook of the sociology of gender* (pp. 157–171). Springer. https://doi.org/10.1007/978-3-319-76333-0_12

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology*, *53*(1), 25–44. https://doi.org/10.1016/j.jsp.2014.10.002

*Forgasz, H. J. (2001, April). *Mathematics: Still a male domain? Australian findings* [Paper presentation]. The 2001 Annual Meeting of the American Educational Research Association, Seattle, WA, United States.

*Forgasz, H. J., Leder, G., & Kaur, B. (1999). *Who can('t) do maths—Boys/girls? An international comparison* [Paper presentation]. The combined Annual Meeting of the Australian Association for Research in Education and the New Zealand Association for Research in Education, Melbourne, Australia.

*Forgasz, H. J., Leder, G. C., & Barkatsas, T. (1999). Of course I can['t] do mathematics: Ethnicity and the stereotyping of mathematics. In J. M. Truran & K. M. Truran (Eds.), *Making the difference: Proceedings of the twenty-second annual conference of the Mathematics Educational Research Group of Australasia* (pp. 200–206). Mathematics Education Research Group of Australasia.

*Forgasz, H. J., Leder, G. C., & Gardner, P. L. (1996). The Fennema-Sherman 'mathematics as a male domain' scale: A re-examination. In L. Puig & A. Gutierrez (Eds.), *Proceedings of the conference of the international group for the psychology of mathematics education, Valencia, Spain* (Vol. 2, pp. 361–368). International Group for the Psychology of Mathematics Education.

*Forgasz, H. J., Leder, G. C., & Gardner, P. L. (1999). The Fennema-Sherman mathematics as a male domain reexamined. *Journal for Research in Mathematics Education*, *30*(3), 342–348. https://doi.org/10.2307/749839

*Forgasz, H. J., Leder, G. C., & Kaur, B. (2001). Who can('t) do maths—Boys/girls? Beliefs of Australian and Singaporean secondary school students. *Asia Pacific Journal of Education*, *21*(2), 106–116. https://doi.org/10.1080/02188791.2001.10600197

*Forgasz, H. J., Leder, G. C., & Kloosterman, P. (2004). New perspectives on the gender stereotyping of mathematics. *Mathematical Thinking and Learning*, *6*(4), 389–420. https://doi.org/10.1207/s15327833mtl0604_2

*Forgasz, H. J., Leder, G. C., & Taylor, L. (1996). *Mathematics and language studies: A cross-national comparison* [Paper presentation]. Annual Meeting of the American Educational Research Association, New York, NY, United States.

*Forgasz, H. J., & Markovits, Z. (2018). Elementary students' views on the gendering of mathematics. *European Journal of Educational Research*, *7*(4), 867–876. https://doi.org/10.12973/eu-jer.7.4.867

*Forgasz, H. J., & Mittelberg, D. (2008). Israeli Jewish and Arab students' gendering of mathematics. *ZDM Mathematics Education*, *40*(4), 545–558. https://doi.org/10.1007/s11858-008-0139-3

*Frid, S., Nortvedt, G., & Sumpter, L. (2020). Who is best in mathematics? Grade nine students' attitudes about boys, girls and mathematics. In M. Inprasitha, N. Changsri, N. Boonsena, & K. Kaen (Eds.), *Proceedings of the 44th conference of the international group for the psychology of mathematics education* (pp. 152–161). Psychology of Mathematics Education.

*Frome, P. M. (1999). *The influence of girls' gender-linked beliefs on their educational and occupational aspirations* [Doctoral dissertation]. ProQuest Dissertations and Theses.

Galambos, N. L., Berenbaum, S. A., & McHale, S. M. (2009). Gender development in adolescence. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology: Vol. 1. Individual bases of adolescent development* (3rd ed., pp. 305–357). Wiley.

*Galligan, K. M. (2012). *Examining adolescents' gender stereotypes and ingroup biases about academics, classroom regulatory behavior, and occupations* [Doctoral dissertation]. ProQuest Dissertations and Theses.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731. https://doi.org/10.1037/0033-2909.132.5.692

Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: Theory, evidence, and open questions. In M. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology* (Vol. 44, pp. 59–127). Elsevier. https://doi.org/10.1016/B978-0-12-385522-0.00002-0

Ghavami, N., & Peplau, L. A. (2013). An intersectional analysis of gender and ethnic stereotypes: Testing three hypotheses. *Psychology of Women Quarterly*, *37*(1), 113–127. https://doi.org/10.1177/0361684312464203

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*(1), 21–38. https://doi.org/10.1037/0033-2909.117.1.21

*Google. (2015). *Images of computer science: Perceptions among students, parents and educators in the U.S*. Gallup. https://services.google.com/fh/files/misc/images-of-computer-science-report.pdf

Grames, E. M., Stillman, A. N., Tingley, M. W., & Elphick, C. S. (2019). An automated approach to identifying search terms for systematic reviews using keyword co-occurrence networks. *Methods in Ecology and Evolution*, *10*(10), 1645–1654. https://doi.org/10.1111/2041-210X.13268

*Guimond, S., & Roussel, L. (2001). Bragging about one's school grades: Gender stereotyping and students' perception of their abilities in science, mathematics, and language. *Social Psychology of Education*, *4*(3/4), 275–293. https://doi.org/10.1023/A:1011332704215

Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PLOS ONE*, *10*(9), Article e0138237. https://doi.org/10.1371/journal.pone.0138237

Halim, M. L. D. (2016). Princesses and superheroes: Social-cognitive influences on early gender rigidity. *Child Development Perspectives*, *10*(3), 155–160. https://doi.org/10.1111/cdep.12176

Halim, M. L. D., & Ruble, D. (2010). Gender identity and stereotyping in early and middle childhood. In J. Chrisler & D. McCreary (Eds.), *Handbook of gender research in psychology* (pp. 495–525). Springer. https://doi.org/10.1007/978-1-4419-1465-1_24

Halim, M. L. D., Ruble, D. N., Tamis-LeMonda, C. S., Shrout, P. E., & Amodio, D. M. (2017). Gender attitudes in early childhood: Behavioral consequences and cognitive antecedents. *Child Development*, *88*(3), 882–899. https://doi.org/10.1111/cdev.12642

*Hall, J. (2011, October). Elementary students' view of mathematics and mathematicians: Boys' and girls' perspectives. In L. R. Wiest & T. Lamberg (Eds.), *Proceedings of the thirty-third annual conference of the North American chapter of the international group for the psychology of mathematics education* (p. 1876). North American Chapter of the International Group for the Psychology of Mathematics Education.

*Hall, J. (2015). *Societal views of mathematics and mathematicians and their influence on elementary students* [Doctoral dissertation]. ProQuest Dissertations and Theses.

*Hand, S., Rice, L., & Greenlee, E. (2017). Exploring teachers' and students' gender role bias and students' confidence in STEM fields. *Social Psychology of Education*, *20*(4), 929–945. https://doi.org/10.1007/s11218-017-9408-8

Hansen, A. K., Dwyer, H. A., Iveland, A., Talesfore, M., Wright, L., Harlow, D. B., & Franklin, D. (2017, March). Assessing children's understanding of the work of computer scientists: The draw-a-computer-scientist test. *Proceedings of the 2017 ACM SIGCSE technical symposium on computer science education* (pp. 279–284). https://doi.org/10.1145/3017680.3017769

*Hargreaves, M., Homer, M., & Swinnerton, B. (2008). A comparison of performance and attitudes in mathematics amongst the "gifted." Are boys better at mathematics or do they just think they are? *Assessment in Education: Principles, Policy & Practice*, *15*(1), 19–38. https://doi.org/10.1080/09695940701876037

Hartley, B. L., & Sutton, R. M. (2013). A stereotype threat account of boys' academic underachievement. *Child Development*, *84*(5), 1716–1733. https://doi.org/10.1111/cdev.12079

*Heard, B. (2019). *The role of race and gender in determining college major choice* [Master's thesis, University of West Georgia]. ProQuest Dissertations and Theses.

Heyder, A., & Kessels, U. (2013). Is school feminine? Implicit gender stereotyping of school as a predictor of academic achievement. *Sex Roles*, *69*(11–12), 605–617. https://doi.org/10.1007/s11199-013-0309-9

Heyder, A., & Kessels, U. (2017). Boys don't work? On the psychological benefits of showing low effort in high school. *Sex Roles*, *77*(1–2), 72–85. https://doi.org/10.1007/s11199-016-0683-1

*Heyman, G. D., & Legare, C. H. (2004). Children's beliefs about gender differences in the academic and social domains. *Sex Roles*, *50*(3/4), 227–239. https://doi.org/10.1023/B:SERS.0000015554.12336.30

*Hildebrand, L., Jirout, J., Newcombe, N. S., & Gunderson, E. (2019). *The development of gender stereotypes about spatial, math, and reading domains* [Paper presentation]. The 2019 Meeting for the Society for Research on Child Development, Baltimore, MD, United States.

*Hildebrand, L., Jirout, J., Newcombe, N. S., & Gunderson, E. (2021). *Children's implicit, indirect, and explicit gender stereotypes of spatial, mathematical, and verbal achievement across the first years of formal education* [Unpublished manuscript]. Department of Psychology, Boston College.

Hill, J., & Lynch, M. (1983). The intensification of gender-related role expectations during early adolescence. In J. Brooks-Gunn & A. Petersen (Eds.), *Girls at puberty* (pp. 201–228). Plenum Press. https://doi.org/10.1007/978-1-4899-0354-9_10

*Hill, P. W., McQuillan, J., Talbert, E., Spiegel, A., Gauthier, G. R., & Diamond, J. (2017). Science possible selves and the desire to be a scientist: Mindsets, gender bias, and confidence during early adolescence. *Social Sciences*, *6*(2), Article 55. https://doi.org/10.3390/socsci6020055

*Hong, Z.-R. (2002). *An investigation of self-esteem and school achievement of Taiwanese secondary students* (UMI No. 3039640) [Doctoral dissertation]. ProQuest Theses and Dissertations Global.

*Hong, Z.-R. (2003). An investigation of the gender stereotyped thinking of Taiwanese secondary school boys and girls. *Sex Roles*, *48*(11/12), 495–504. https://doi.org/10.1023/A:1023571111617

*Hong, Z.-R., Veach, P. M., & Lawrenz, F. (2004). Psychosocial predictors of Taiwanese secondary students' self-esteem. *The Alberta Journal of Educational Research*, *50*(4), 430–446. https://doi.org/10.55016/ojs/ajer.v50i4.55081

*Hong, Z.-R., Veach, P. M., & Lawrenz, F. (2005). Psychosocial predictors of psychological distress in Taiwanese secondary school boys and girls. *Sex Roles*, *53*(5–6), 419–431. https://doi.org/10.1007/s11199-005-6764-1

*Hreno, N. (2019). *Feminist-infused participatory action research to promote female students' self-efficacy in mathematics* [Doctoral dissertation]. University of South Carolina.

Hudley, C., & Graham, S. (2001). Stereotypes of achievement striving among early adolescents. *Social Psychology of Education*, *5*(2), 201–224. https://doi.org/10.1023/A:1014438702266

*Illi, F. (2018). *Genus i mellanstadiet-Hur uppfattningar om genus kan paverka elevernas agerande* [Gender in middle school—How perceptions about gender can influence students' actions]. Malmo Universitet.

*Ingels, S. J., Pratt, D. J., Herget, D. R., Burns, L. J., Dever, J. A., Ottem, R., Rogers, J. E., Jin, Y., & Leinwand, S. (2011). *High School Longitudinal Study of 2009 (HSLS:09). Base-year data file documentation (NCES 2011-328)*. U.S. Department of Education; National Center for Education Statistics. https://nces.ed.gov/pubsearch

Irwin, V., De La Rosa, J., Wang, K., Hein, S., Zhang, J., Burr, R., Roberts, A., Barmer, A., Bullock Mann, F., Dilig, R., & Parker, S. (2022). *Report on the Condition of Education 2022* (NCES 2022-144). U.S. Department of Education; National Center for Education Statistics. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2022144

Jackson, J. L., Kuriyama, A., Anton, A., Choi, A., Fournier, J.-P., Geier, A.-K., Jacquerioz, F., Kogan, D., Scholcoff, C., & Sun, R. (2019). The accuracy of Google Translate for abstracting data from non–English-language trials for systematic reviews. *Annals of Internal Medicine*, *171*(9), 677–679. https://doi.org/10.7326/M19-0891

*Jacobs, J. E. (1987). *Parents' gender stereotypes and perceptions of their child's abilities: Influences on the child* [Doctoral dissertation]. ProQuest Theses and Dissertations.

*Jang, S. T. (2018). *Student experiences and educational outcomes of Southeast Asian female secondary school students in the United States—A critical quantitative intersectionality analysis* [Doctoral dissertation, University of Minnesota]. ProQuest Dissertations and Theses.

Jaxon, J., Lei, R. F., Shachnai, R., Chestnut, E. K., & Cimpian, A. (2019). The acquisition of gender stereotypes about intellectual ability: Intersections with race. *Journal of Social Issues*, *75*(4), 1192–1215. https://doi.org/10.1111/josi.12352

*Jelenec, P. (2008). *Girls and the leaky math pipeline implicit math-gender stereotypes and math withdrawal in female adolescents and women* [Doctoral dissertation]. Friedrich Schiller University Jena. https://core.ac.uk/download/pdf/224760265.pdf

*Jones, E. (2019). *Gender stereotype endorsement predicts engineering competency* [Paper presentation]. The annual Fulton Schools of Engineering Undergraduate Research Symposium, Tempe, AZ, United States.

*Jones, E. F., & Nelson-Le Gall, S. (2009). Black children's judgments of in-group and out-group students' academic achievement, motivation, and behavior. *Negro Educational Review*, *60*(1–4), 53–69. https://www.proquest.com/openview/98d356d417246ab3b404531662f6421f/

*Jugović, I. (2010). Uloga motivacije i rodnih stereotipa u objanjenju namjere odabira studija u stereotipno mukom podruju [The role of motivation and gender stereotypes in explanation of the intention to select studies in stereotypically male area]. *Sociology & Space*, *48*(1), 77–98. https://hrcak.srce.hr/55439

*Jugović, I. (2017). Students' gender-related choices and achievement in physics. *Center for Educational Policy Studies Journal*, *7*(2), 71–95. https://doi.org/10.26529/cepsj.170

*Kaiser, G., Hoffstall, M., & Orschulik, A. B. (2012). Gender role stereotypes in the perception of mathematics: An empirical study with secondary students in Germany. In H. Forgasz & F. Rivera (Eds.), *Towards equity in mathematics education. Gender, culture, and diversity* (pp. 115–140). Springer. https://doi.org/10.1007/978-3-642-27702-3_11

*Kassahun, T., & Kedir, B. (2006). Girls' performance in mathematics in upper primary schools of Addis Ababa. *Indian Journal of Gender Studies*, *13*(3), 401–424. https://doi.org/10.1177/097152150601300304

Katz, P. A., & Ksansnak, K. R. (1994). Developmental aspects of gender role flexibility and traditionality in middle childhood and adolescence. *Developmental Psychology*, *30*(2), 272–282. https://doi.org/10.1037/0012-1649.30.2.272

Keller, C. (2001). Effect of teachers' stereotyping on students' stereotyping of mathematics as a male domain. *The Journal of Social Psychology*, *141*(2), 165–173. https://doi.org/10.1080/00224540109600544

Klaczynski, P. A., Felmban, W. S., & Kole, J. (2020). Gender intensification and gender generalization biases in pre-adolescents, adolescents, and emerging adults. *British Journal of Developmental Psychology*, *38*(3), 415–433. https://doi.org/10.1111/bjdp.12326

*Kloosterman, P., Tassell, J., Ponniah, A. G., & Essex, N. K. (2008). Perceptions of mathematics and gender. *School Science and Mathematics*, *108*(4), 149–162. https://doi.org/10.1111/j.1949-8594.2008.tb17821.x

*Kloosterman, P., Tassell, J. H., Ponniah, A. G., & Essex, N. K. (2001). *Mathematics as a gendered domain in the United States* [Paper presentation]. The 2001 Annual Meeting of the American Educational Research Association, Seattle, WA, United States.

Koenig, A. M. (2018). Comparing prescriptive and descriptive gender stereotypes about children, adults, and the elderly. *Frontiers in Psychology*, *9*, Article 1086. https://doi.org/10.3389/fpsyg.2018.01086

Koenig, A. M., & Eagly, A. H. (2014). Evidence for the social role theory of stereotype content: Observations of groups' roles shape stereotypes. *Journal of Personality and Social Psychology*, *107*(3), 371–392. https://doi.org/10.1037/a0037215

Kohlberg, L. A. (1966). A cognitive-developmental analysis of children's sex role concepts and attitudes. In E. Maccoby (Ed.), *The development of sex differences* (pp. 82–173). Stanford University Press.

*Kombe, D., Che, S. M., & Bridges, W. (2019). Students' gendered perceptions of mathematics in middle grades single-sex and coeducational classrooms. *School Science and Mathematics*, *119*(7), 417–427. https://doi.org/10.1111/ssm.12363

*Koul, R., Lerdpornkulrat, T., & Poondej, C. (2016). Gender compatibility, math-gender stereotypes, and self-concepts in math and physics. *Physical Review Physics Education Research*, *12*(2), Article 020115. https://doi.org/10.1103/PhysRevPhysEducRes.12.020115

*Kurtz-Costes, B., Copping, K. E., Rowley, S. J., & Kinlaw, C. R. (2014). Gender and age differences in awareness and endorsement of gender stereotypes about academic abilities. *European Journal of Psychology of Education*, *29*(4), 603–618. https://doi.org/10.1007/s10212-014-0216-7

Kurtz-Costes, B., Defreitas, S. C., Halle, T. G., & Kinlaw, C. R. (2011). Gender and racial favouritism in black and white preschool girls. *British Journal of Developmental Psychology*, *29*(2), 270–287. https://doi.org/10.1111/j.2044-835X.2010.02018.x

*Kurtz-Costes, B., Patel, N., & Wood, D. (2011). Academic gender stereotypes and academic self-concept of Indian adolescents. *Journal of All India Association for Educational Research*, *23*, 1–10.

*Kurtz-Costes, B., Rowley, S. J., Harris-Britt, A., & Woods, T. A. (2008). Gender stereotypes about mathematics and science and self-perceptions of ability in late childhood and early adolescence. *Merrill-Palmer Quarterly*, *54*(3), 386–409. https://doi.org/10.1353/mpq.0.0001

Lampley, S. A., Dyess, S. R., Benfield, M. P., Davis, A. M., Gholston, S. E., Dillihunt, M. L., & Turner, M. W. (2022). Understanding the conceptions of engineering in early elementary students. *Education Sciences*, *12*(1), Article 43. https://doi.org/10.3390/educsci12010043

Leaper, C. (2011). More similarities than differences in contemporary theories of social development? A plea for theory bridging. In J. B. Benson (Ed.), *Advances in child development and behavior* (Vol. 40, pp. 337–378). Elsevier. https://doi.org/10.1016/B978-0-12-386491-8.00009-8

Leaper, C. (2015). Gender and social-cognitive development. In R. M. Lerner, L. S. Liben, & U. Muller (Eds.), *Handbook of child psychology and developmental science* (Vol. 2, 7th ed., pp. 806–853). Wiley. https://doi.org/10.1002/9781118963418.childpsy219

*Leaper, C., & Starr, C. (2017). *Gender stereotyping, self-objectification, and same-gender peer academic support in relation to girls' and boys' motivation in mathematics and language arts* [Unpublished manuscript].

*Leder, G. C., & Brandell, G. (2004). *Mathematics still a male domain? Preliminary data from a Swedish study*. https://www.maths.lth.se/GeMa/reports/KoM5.pdf

Lei, R. F., & Rhodes, M. (2021). Why developmental research on social categorization needs intersectionality. *Child Development Perspectives*, *15*(3), 143–147. https://doi.org/10.1111/cdep.12421

*Lerdpornkulrat, T., Koul, R., & Sujivorakul, C. (2012). The influence of ability beliefs and motivational orientation on the self-efficacy of high school science students in Thailand. *Australian Journal of Education*, *56*(2), 163–181. https://doi.org/10.1177/000494411205600205

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

*Li, J., & McLellan, R. (2021). Is language learning a feminine domain? Examining the content and stereotype threat effect of female-language stereotypes among EFL learners in China. *Contemporary Educational Psychology*, *66*, Article 101991. https://doi.org/10.1016/j.cedpsych.2021.101991

Liben, L. S. (2014). The individual ↔ context nexus in developmental intergroup theory: Within and beyond the ivory tower. *Research in Human Development*, *11*(4), 273–290. https://doi.org/10.1080/15427609.2014.967048

Liben, L. S. (2017). Gender development: A constructivist ecological perspective. In N. Budwig, E. Turiel, & P. D. Zelazo (Eds.), *New perspectives on human development* (pp. 145–164). Cambridge University Press. https://doi.org/10.1017/CBO9781316282755.010

Liben, L. S., & Bigler, R. S. (2002). The developmental course of gender differentiation: Conceptualizing, measuring, and evaluating constructs and pathways. *Monographs of the Society for Research in Child Development*, *67*(2), i–viii. https://doi.org/10.1111/1540-5834.t01-1-00187

Liou, P. Y., Lin, Y. M., Huang, S. C., & Chen, S. (2023). Gender differences in science motivational beliefs and their relations with achievement over grades 4 and 8: A multinational perspective. *International Journal of Science and Mathematics Education*, *21*(1), 233–249. https://doi.org/10.1007/s10763-021-10243-5

Lovakov, A., & Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology*, *51*(3), 485–504. https://doi.org/10.1002/ejsp.2752

*Lummis, M., & Stevenson, H. W. (1990). Gender differences in beliefs and achievement: A cross-cultural study. *Developmental Psychology*, *26*(2), 254–263. https://doi.org/10.1037/0012-1649.26.2.254

*Lundmark, V. A. (1996). *Sex-typing and computing among American upper secondary students: Preserving a male domain softly?* [Doctoral dissertation]. University of Minnesota.

*Magnenat, E., & Courtine, N. (2015). *Stereotype de genre a l'ecole obligatoire: Adhésion, fréquence et liens avec le sentiment d'auto-efficacité chez des élèves de 10 à 16 ans* [Gender stereotype in compulsory school: Membership, frequency and links to feelings of self-efficacy in students from 10 to 16 years old; Unpublished dissertation]. Haute Ecole Pédagogique.

*Makrakis, V., & Sawada, T. (1996). Gender, computers and other school subjects among Japanese and Swedish students. *Computers & Education*, *26*(4), 225–231. https://doi.org/10.1016/0360-1315(95)00085-2

Marcell, A. V., Eftim, S. E., Sonenstein, F. L., & Pleck, J. H. (2011). Associations of family and peer experiences with masculinity attitude trajectories at the individual and group level in adolescent and young adult males. *Men and Masculinities*, *14*(5), 565–587. https://doi.org/10.1177/1097184X11409363

Martin, C. L. (1993). The development of sex typing in middle childhood. *Monographs of the Society for Research in Child Development*, *58*(2), 75–85. https://doi.org/10.2307/1166118

Martin, C. L., & Halverson, C. F. (1981). A schematic processing model of sex typing and stereotyping in children. *Child Development*, *52*(4), 1119–1134. https://doi.org/10.2307/1129498

Martin, C. L., & Ruble, D. (2004). Children's search for gender cues: Cognitive perspectives on gender development. *Current Directions in Psychological Science*, *13*(2), 67–70. https://doi.org/10.1111/j.0963-7214.2004.00276.x

Martin, C. L., Ruble, D. N., & Szkrybalo, J. (2002). Cognitive theories of early gender development. *Psychological Bulletin*, *128*(6), 903–933. https://doi.org/10.1037/0033-2909.128.6.903

*Martinot, D., Bagès, C., & Désert, M. (2012). French children's awareness of gender stereotypes about mathematics and reading: When girls improve their reputation in math. *Sex Roles*, *66*(3–4), 210–219. https://doi.org/10.1007/s11199-011-0032-3

*Martinot, D., & Désert, M. (2007). Awareness of a gender stereotype, personal beliefs and self-perceptions regarding math ability: When boys do not surpass girls. *Social Psychology of Education*, *10*(4), 455–471. https://doi.org/10.1007/s11218-007-9028-9

*Mason, S. L. (2019). *Teaching K-6 computer science: Teacher and student attitudes and self-efficacy* [Doctoral dissertation]. ProQuest Dissertations and Theses Global.

*Mason, S. L., & Rich, P. J. (2020). Development and analysis of the elementary student coding attitudes survey. *Computers & Education*, *153*, Article 103898. https://doi.org/10.1016/j.compedu.2020.103898

Master, A. (2021). Gender stereotypes influence children's STEM motivation. *Child Development Perspectives*, *15*(3), 203–210. https://doi.org/10.1111/cdep.12424

*Master, A., Cheryan, S., & Meltzoff, A. N. (2019). *Differential power of distinct kinds of STEM stereotypes: Interest versus ability stereotypes* [Paper presentation]. The 2019 meeting for the Society for Research on Child Development, Baltimore, MD, United States.

Master, A., & Meltzoff, A. N. (2020). Cultural stereotypes and sense of belonging contribute to gender gaps in STEM. *International Journal of Gender, Science and Technology*, *12*(1), 152–198.

*Master, A., Meltzoff, A. N., & Cheryan, S. (2020). *Stereotypes contribute to gender imbalances in STEM fields* [Paper presentation]. Presentation at the 2020 meeting for the American Educational Research Association, San Francisco, CA, United States.

*Master, A., Meltzoff, A. N., & Cheryan, S. (2021a). 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

*Master, A., Meltzoff, A. N., & Cheryan, S. (2021b). *The emergence of ability stereotypes and gender differences across STEM fields* [Paper presentation]. The 2021 meeting for the American Educational Research Association.

*Mateos-Jimenez, A., Torres-Martinez, B., & Garcia-Fernandez, B. (2018). Impacto de un relato sobre las percepciones de racismo y sexismo en escolares de primaria [Impact of a story on the perceptions of racism and sexism in primary school students]. *Ocnos*, *17*(2), 55–66. https://doi.org/10.18239/ocnos_2018.17.2.1570

McGuire, L., Hoffman, A. J., Mulvey, K. L., Hartstone-Rose, A., Winterbottom, M., Joy, A., Law, F., Balkwill, F., Burns, K. P., Butler, L., Drews, M., Fields, G., Smith, H., & Rutland, A. (2022). Gender stereotypes and peer selection in STEM domains among children and adolescents. *Sex Roles*, *87*(9–10), 455–470. https://doi.org/10.1007/s11199-022-01327-9

*McGuire, L., Monzavi, T., Hoffman, A. J., Law, F., Irvin, M. J., Winterbottom, M., Hartstone-Rose, A., Rutland, A., Burns, K. P., Butler, L., Drews, M., Fields, G. E., & Mulvey, K. L. (2021). Science and math interest and gender stereotypes: The role of educator gender in informal science learning sites. *Frontiers in Psychology*, *12*, Article 503237. https://doi.org/10.3389/fpsyg.2021.503237

*McGuire, L., Mulvey, K. L., Goff, E., Irvin, M. J., Winterbottom, M., Fields, G. E., Hartstone-Rose, A., & Rutland, A. (2020). STEM gender stereotypes from early childhood through adolescence at informal science centers. *Journal of Applied Developmental Psychology*, *67*, Article 101109. https://doi.org/10.1016/j.appdev.2020.101109

*Mencia-Ripley, A., Sanchez-Vincitore, L. V., & Garrido, L. D. (2015). *Baseline report for the gender, inclusion, and safe school components of Project Read* (Report No. AID-517-A-15-00005). USAID.

*Miller, C. F., Wheeler, L. A., & Reisslein, M. (2018). *Rigidity in gender-stereotyped attitudes: Ability versus category stereotypes* [Paper presentation]. The 2018 Gender Development Research Conference, San Francisco, CA, United States.

Miller, C. F., Wheeler, L. A., & Woods, B. (2024). A multidimensional examination of children's endorsement of gender stereotypes. *Social Development*, *33*(2), Article e12725. https://doi.org/10.1111/sode.12725

Miller, D. I., Eagly, A. H., & Linn, M. C. (2015). Women's representation in science predicts national gender-science stereotypes: Evidence from 66 nations. *Journal of Educational Psychology*, *107*(3), 631–644. https://doi.org/10.1037/edu0000005

Miller, D. I., & Halpern, D. F. (2014). The new science of cognitive sex differences. *Trends in Cognitive Sciences*, *18*(1), 37–45. https://doi.org/10.1016/j.tics.2013.10.011

Miller, D. I., Joshi, M., & Ebersole, C. R. (2024). *Explore research on children's gender stereotypes about STEM and verbal abilities* [Interactive data tool application]. https://d-miller.shinyapps.io/STEM_verbal_stereotypes/

Miller, D. I., & Lauer, J. E. (2024, August 31). *Gender stereotypes about STEM and verbal abilities meta-analysis* [Online data and code repository]. https://osf.io/29egh

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 five decades of U.S. Draw-A-Scientist studies. *Child Development*, *89*(6), 1943–1955. https://doi.org/10.1111/cdev.13039

Miller, D. I., & Wai, J. (2015). The bachelor's to Ph.D. STEM pipeline no longer leaks more women than men: A 30-year analysis. *Frontiers in Psychology*, *6*, Article 37. https://doi.org/10.3389/fpsyg.2015.00037

*Mills, D. C. (2018). *Examining the associations between parental socialization and science motivation in rural African American adolescents: A study of expectancy-value theory* [Doctoral Dissertation, University of North Carolina at Chapel Hill]. ProQuest Dissertations and Theses.

Misch, A., Dunham, Y., & Paulus, M. (2022). The developmental trajectories of racial and gender intergroup bias in 5- to 10-year-old children: The impact of general psychological tendencies, contextual factors, and individual propensities. *Acta Psychologica*, *229*, Article 103709. https://doi.org/10.1016/j.actpsy.2022.103709

*Moè, A. (2018). Mental rotation and mathematics: Gender-stereotyped beliefs and relationships in primary school children. *Learning and Individual Differences*, *61*, 172–180. https://doi.org/10.1016/j.lindif.2017.12.002

Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, *120*(3), 544–560. https://doi.org/10.1037/a0032459

*Morgade, G., & Bonder, G. (1995). Women, mathematics and natural sciences: Objectives and results of learning among girls and boys. In J. Sirinelli & M. Grunberg-Manago (Eds.), *The scientific education of girls: Education beyond reproach?* (pp. 124–138). United Nations Educational, Scientific, and Cultural Organization. https://files.eric.ed.gov/fulltext/ED399147.pdf

*Muntoni, F., Wagner, J., & Retelsdorf, J. (2021). Beware of stereotypes: Are classmates' stereotypes associated with students' reading outcomes? *Child Development*, *92*(1), 189–204. https://doi.org/10.1111/cdev.13359

Muradoglu, M., & Cimpian, A. (2020). Children's intuitive theories of academic performance. *Child Development*, *91*(4), e902–e918. https://doi.org/10.1111/cdev.13325

*Muzzatti, B. (2005). Alunni e discipline scolastiche: Differenze di genere negli atteggiamenti espressi [Pupils and subject matters: Gender differences in attitudes]. *Età Evolutiva*, *82*, 5–17.

*Muzzatti, B. (2007). Differenze di genere negli atteggiamenti verso le discipline scolastiche in preadolescenza [Gender differences in attitudes towards school subjects in pre-adolescence]. *Età Evolutiva*, *29*(88), 53–62.

*Muzzatti, B., & Agnoli, F. (2007). Gender and mathematics: Attitudes and stereotype threat susceptibility in Italian children. *Developmental Psychology*, *43*(3), 747–759. https://doi.org/10.1037/0012-1649.43.3.747

National Center for Science and Engineering Statistics. (2023). *Diversity and STEM: Women, minorities, and persons with disabilities 2023. Special Report NSF 23-315*. National Science Foundation. https://ncses.nsf.gov/wmpd

Neff, K. D., & Terry-Schmitt, L. N. (2002). Youths' attributions for power-related gender differences: Nature, nurture, or God? *Cognitive Development*, *17*(2), 1185–1202. https://doi.org/10.1016/S0885-2014(02)00094-1

*Neuburger, S., Ruthsatz, V., Jansen, P., Heil, M., & Quaiser-Pohl, C. (2013). Acceptance and effects of role models in the spatial domain: A storytelling experiment with fourth graders. *Frontiers in Psychological and Behavioral Science*, *2*(3), 73–88. https://www.researchgate.net/publication/283671733_Acceptance_and_Effects_of_Role_Models_in_the_Spatial_Domain_A_Storytelling_Experiment_with_Fourth_Graders

*Nowicki, E. A., & Lopata, J. (2017). Children's implicit and explicit gender stereotypes about mathematics and reading ability. *Social Psychology of Education*, *20*(2), 329–345. https://doi.org/10.1007/s11218-015-9313-y

O'Brien, L. T., Blodorn, A., Adams, G., Garcia, D. M., & Hammer, E. (2015). Ethnic variation in gender-STEM stereotypes and STEM participation: An intersectional approach. *Cultural Diversity & Ethnic Minority Psychology*, *21*(2), 169–180. https://doi.org/10.1037/a0037944

Okanda, M., Meng, X., Kanakogi, Y., Uragami, M., Yamamoto, H., & Moriguchi, Y. (2022). Gender stereotypes about intellectual ability in Japanese children. *Scientific Reports*, *12*(1), Article 16748. https://doi.org/10.1038/s41598-022-20815-2

Omi, M., & Winant, H. (2014). *Racial formation in the United States*. Routledge. https://doi.org/10.4324/9780203076804

Organisation for Economic Co-operation and Development. (2016). *PISA 2015 results: Vol. I. Excellence and equity in education*. OECD Publishing. https://doi.org/10.1787/9789264266490-en

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, *372*, Article n160. https://doi.org/10.1136/bmj.n160

Pantic, K., Clarke-Midura, J., Poole, F., Roller, J., & Allan, V. (2018). Drawing a computer scientist: Stereotypical representations or lack of awareness? *Computer Science Education*, *28*(3), 232–254. https://doi.org/10.1080/08993408.2018.1533780

Parker, P. D., Van Zanden, B., Marsh, H. W., Owen, K., Duineveld, J. J., & Noetel, M. (2020). The intersection of gender, social class, and cultural context: A meta-analysis. *Educational Psychology Review*, *32*(1), 197–228. https://doi.org/10.1007/s10648-019-09493-1

*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

Patterson, M. M., & Bigler, R. S. (2018). Effects of consistency between self and in-group on children's views of self, groups, and abilities. *Social Development*, *27*(1), 154–171. https://doi.org/10.1111/sode.12255

*Pelgrum, W. J., Janssen Reinen, I. A. M., & Plomp, T. J. (1993). *Schools, teachers, students and computers: A cross-national perspective*. University of Twente. https://files.eric.ed.gov/fulltext/ED372734.pdf

*Pennington, C. R., Kaye, L. K., Qureshi, A. W., & Heim, D. (2020). Do gender differences in academic attainment correspond with scholastic attitudes? An exploratory study in a UK secondary school. *Journal of Applied Social Psychology*, *51*(1), 3–16. https://doi.org/10.1111/jasp.12711

Perry, D. G., & Pauletti, R. E. (2011). Gender and adolescent development. *Journal of Research on Adolescence*, *21*(1), 61–74. https://doi.org/10.1111/j.1532-7795.2010.00715.x

Pigott, T. D. (2001). Missing predictors in models of effect size. *Evaluation & the Health Professions*, *24*(3), 277–307. https://doi.org/10.1177/01632780122034920

Pigott, T. D. (2012). Missing data in meta-analysis: Strategies and approaches. In T. D. Pigott (Ed.), *Advances in meta-analysis* (pp. 79–107). Springer. https://doi.org/10.1007/978-1-4614-2278-5_7

Plante, I., Théorêt, M., & Favreau, O. E. (2009). Student gender stereotypes: Contrasting the perceived maleness and femaleness of mathematics and language. *Educational Psychology*, *29*(4), 385–405. https://doi.org/10.1080/01443410902971500

Polanin, J. R., & Terzian, M. (2019). A data-sharing agreement helps to increase researchers' willingness to share primary data: Results from a randomized controlled trial. *Journal of Clinical Epidemiology*, *106*, 60–69. https://doi.org/10.1016/j.jclinepi.2018.10.006

Polanin, J. R., & Williams, R. T. (2016). Overcoming obstacles in obtaining individual participant data for meta-analysis. *Research Synthesis Methods*, *7*(3), 333–341. https://doi.org/10.1002/jrsm.1208

Powlishta, K. K. (1995). Gender bias in children's perceptions of personality traits. *Sex Roles*, *32*(1–2), 17–28. https://doi.org/10.1007/BF01544755

Powlishta, K. K. (2004). Gender as a social category: Intergroup processes and gender-role development. In M. Bennett & F. Sani (Eds.), *The development of social self* (pp. 103–133). Psychology Press. https://doi.org/10.4324/9780203391099_chapter_4

Powlishta, K. K., Serbin, L. A., Doyle, A. B., & White, D. R. (1994). Gender, ethnic, and body type biases: The generality of prejudice in childhood. *Developmental Psychology*, *30*(4), 526–536. https://doi.org/10.1037/0012-1649.30.4.526

Priess, H. A., Lindberg, S. M., & Hyde, J. S. (2009). Adolescent gender-role identity and mental health: Gender intensification revisited. *Child Development*, *80*(5), 1531–1544. https://doi.org/10.1111/j.1467-8624.2009.01349.x

*Puckett, S. E. (2009). *The effect of dating activity and school environment on adolescent girls' explicit and implicit math attitude* [Undergraduate thesis]. University of Virginia.

Purdie-Vaughns, V., & Eibach, R. P. (2008). Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex Roles*, *59*(5–6), 377–391. https://doi.org/10.1007/s11199-008-9424-4

Pustejovsky, J. E. (2017). *Pooling clubSandwich results across multiple imputations* [Blog post]. https://jepusto.com/posts/MI-with-clubSandwich/

Pustejovsky, J. E. (2018). *clubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections* (R package Version 0.3.2). https://CRAN.R-project.org/package=clubSandwich

Pustejovsky, J. E., & Tipton, E. (2018). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, *36*(4), 672–683. https://doi.org/10.1080/07350015.2016.1247004

Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science*, *23*(3), 425–438. https://doi.org/10.1007/s11121-021-01246-3

Quartagno, M., Grund, S., & Carpenter, J. (2019). jomo: A flexible package for two-level joint modelling multiple imputation. *The R Journal*, *11*(2), 205–228. https://doi.org/10.32614/RJ-2019-028

Raabe, T., & Beelmann, A. (2011). Development of ethnic, racial, and national prejudice in childhood and adolescence: A multinational meta-analysis of age differences. *Child Development*, *82*(6), 1715–1737. https://doi.org/10.1111/j.1467-8624.2011.01668.x

Rathbone, J., Hoffmann, T., & Glasziou, P. (2015). Faster title and abstract screening? Evaluating Abstrackr, a semi-automated online screening program for systematic reviewers. *Systematic Reviews*, *4*(1), Article 80. https://doi.org/10.1186/s13643-015-0067-6

Reilly, D. (2012). Gender, culture, and sex-typed cognitive abilities. *PLOS ONE*, *7*(7), Article e39904. https://doi.org/10.1371/journal.pone.0039904

Reilly, D., Neumann, D. L., & Andrews, G. (2019). Gender differences in reading and writing achievement: Evidence from the National Assessment of Educational Progress (NAEP). *American Psychologist*, *74*(4), 445–458. https://doi.org/10.1037/amp0000356

*Riegle-Crumb, C., & Peng, M. (2021). Examining high school students' gendered beliefs about math: Predictors and implications for choice of STEM college majors. *Sociology of Education*, *94*(3), 227–248. https://doi.org/10.1177/00380407211014777

*Rowley, S. J., Kurtz-Costes, B., Mistry, R., & Feagans, L. (2007). Social status as a predictor of race and gender stereotypes in late childhood and early adolescence. *Social Development*, *16*(1), 150–168. https://doi.org/10.1111/j.1467-9507.2007.00376.x

*Royal Witte, E. A. (2017). *The relationship between gender stereotypes and academic performance for rural eighth grade students: A mixed methods ethnographic case study* [Doctoral dissertation]. University of South Carolina. https://scholarcommons.sc.edu/etd/4438

Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2007). Gender development. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 858–932). Wiley.

Rudman, L. A., & Goodwin, S. A. (2004). Gender differences in automatic in-group bias: Why do women like women more than men like men? *Journal of Personality and Social Psychology*, *87*(4), 494–509. https://doi.org/10.1037/0022-3514.87.4.494

Rutland, A., Cameron, L., Milne, A., & McGeorge, P. (2005). Social norms and self-presentation: Children's implicit and explicit intergroup attitudes. *Child Development*, *76*(2), 451–466. https://doi.org/10.1111/j.1467-8624.2005.00856.x

*Sansone, D. (2019). Teacher characteristics, student beliefs, and the gender gap in STEM fields. *Educational Evaluation and Policy Analysis*, *41*(2), 127–144. https://doi.org/10.3102/0162373718819830

Saris, W. E., Revilla, M., Krosnick, J. A., & Shaeffer, E. M. (2010). Comparing questions with agree/disagree response options to questions with construct-specific response options. *Survey Research Methods*, *4*(1), 61–79. https://doi.org/10.18148/srm/2010.v4i1.2682

*Selimbegovic, L., Chatard, A., & Mugny, G. (2007). Can we encourage girls' mobility towards science-related careers? Disconfirming stereotype belief through expert influence. *European Journal of Psychology of Education*, *22*(3), 275–290. https://doi.org/10.1007/BF03173426

Serbin, L. A., Powlishta, K. K., Gulko, J., Martin, C. L., & Lockheed, M. E. (1993). The development of sex typing in middle childhood. *Monographs of the Society for Research in Child Development*, *58*(2), 1–99. https://doi.org/10.2307/1166118

Shu, Y., Hu, Q., Xu, F., & Bian, L. (2022). Gender stereotypes are racialized: A cross-cultural investigation of gender stereotypes about intellectual talents. *Developmental Psychology*, *58*(7), 1345–1359. https://doi.org/10.1037/dev0001356

*Shutts, K. (2017). *Young children's academic gender stereotypes* [Paper presentation]. The 2017 Meeting for the Society for Research on Child Development, Austin, TX, United States.

Signorella, M. L., Bigler, R. S., & Liben, L. S. (1993). Developmental differences in children's gender schemata about others: A meta-analytic review. *Developmental Review*, *13*(2), 147–183. https://doi.org/10.1006/drev.1993.1007

Silvern, L. E. (1977). Children's sex-role preferences: Stronger among girls than boys. *Sex Roles*, *3*(2), 159–171. https://doi.org/10.1007/BF00288666

*Singh, K. K. (2020). Traditional gender ideology on boys' participation in biology post-GCSE. *Journal of Biological Education*, *56*(2), 174–189. https://doi.org/10.1080/00219266.2020.1757488

*Skinner, O. D. (2010). *Academic race and gender stereotypes and adolescents' self-perceptions* [Master's thesis]. ProQuest Dissertations and Theses Global.

*Skinner, O. D., Kurtz-Costes, B., Vuletich, H., Copping, K., & Rowley, S. J. (2021). Race differences in Black and white adolescents' academic gender stereotypes across middle and late adolescence. *Cultural Diversity & Ethnic Minority Psychology*, *27*(3), 537–545. https://doi.org/10.1037/cdp0000384

*Smead, V. S. (1976). *Pygmalion vs. Galatea: Expectations of eighth grade girls and boys and their significant others as they relate to achievement in mathematics class* [Doctoral dissertation]. ProQuest Dissertations and Theses.

*Smead, V. S., & Chase, C. I. (1981). Student expectations as they relate to achievement in eighth grade mathematics. *The Journal of Educational Research*, *75*(2), 115–120. https://doi.org/10.1080/00220671.1981.10885365

*Starr, C., & Leaper, C. (2017). *Stereotyping, self-objectification, & perceived support in relation to girls' and boys' math motivation* [Paper presentation]. The 2017 Meeting for the Society for Research on Child Development, Austin, TX, United States.

*Starr, C., & Simpkins, S. (2021). High school students' math and science gender stereotypes: Relations with their STEM outcomes and socializers' stereotypes. *Social Psychology of Education*, *24*(1), 273–298. https://doi.org/10.1007/s11218-021-09611-4

Starr, C. R., Gao, Y., Rubach, C., Lee, G., Safavian, N., Dicke, A. L., Eccles, J. S., & Simpkins, S. D. (2023). "Who's better at bath, boys or girls?": Changes in adolescents' math gender stereotypes and their motivational beliefs from early to late adolescence. *Education Sciences*, *13*(9), Article 866. https://doi.org/10.3390/educsci13090866

*Steele, J. (2003). Children's gender stereotypes about math: The role of stereotype stratification. *Journal of Applied Social Psychology*, *33*(12), 2587–2606. https://doi.org/10.1111/j.1559-1816.2003.tb02782.x

*Steffens, M. C., & Jelenec, P. (2011). Separating implicit gender stereotypes regarding math and language: Implicit ability stereotypes are self-serving for boys and men, but not for girls and women. *Sex Roles*, *64*(5–6), 324–335. https://doi.org/10.1007/s11199-010-9924-x

*Steffens, M. C., Jelenec, P., & Noack, P. (2010). On the leaky math pipeline: Comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *Journal of Educational Psychology*, *102*(4), 947–963. https://doi.org/10.1037/a0019920

Storage, D., Charlesworth, T. E., Banaji, M. R., & Cimpian, A. (2020). Adults and children implicitly associate brilliance with men more than women. *Journal of Experimental Social Psychology*, *90*, Article 104020. https://doi.org/10.1016/j.jesp.2020.104020

Su, R., & Rounds, J. (2015). All STEM fields are not created equal: People and things interests explain gender disparities across STEM fields. *Frontiers in Psychology*, *6*, Article 189. https://doi.org/10.3389/fpsyg.2015.00189

Susskind, J. E., & Hodges, C. (2007). Decoupling children's gender-based in-group positivity from out-group negativity. *Sex Roles*, *56*(11–12), 707–716. https://doi.org/10.1007/s11199-007-9235-z

*Swinton, A. D. (2011). *African American boys' and girls' causal attributions about math, English, and science are shaped by gender stereotypes, influence classroom engagement, and change across the high school years* [Doctoral dissertation]. University of North Carolina at Chapel Hill.

*Takahira, S. (1996). *A cross-cultural study on variables influencing gender differences in mathematics performance* [Doctoral dissertation]. ProQuest Dissertations and Theses.

*Takwe, M.-A. A. (2019). Gender disparity, implications to students' academic performance in science subjects in secondary schools in Buea sub-division, Cameroon. *International Journal of Trend in Scientific Research and Development*, *3*(5), 1541–1567. https://www.ijtsrd.com/papers/ijtsrd26763.pdf

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. *Developmental Psychology*, *60*(6), 1109–1130. https://doi.org/10.1037/dev0001733

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*(3), 375–393. https://doi.org/10.1037/met0000011

Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, *40*(6), 604–634. https://doi.org/10.3102/1076998615606099

Tobin, D. D., Menon, M., Menon, M., Spatta, B. C., Hodges, E. V. E., & Perry, D. G. (2010). The intrapsychics of gender: A model of self-socialization. *Psychological Review*, *117*(2), 601–622. https://doi.org/10.1037/a0018936

*Tomasetto, C., Galdi, S., & Cadinu, M. (2012). Quando I'implicito precede I'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 e Sociedade*, *7*(2), 169–185.

Trautner, H. M., Ruble, D. N., Cyphers, L., Kirsten, B., Behrendt, R., & Hartmann, P. (2005). Rigidity and flexibility of gender stereotypes in childhood: Developmental or differential? *Infant and Child Development*, *14*(4), 365–381. https://doi.org/10.1002/icd.399

*Turner, A. (2003). *The attitudes of high school pupils to technology* [Unpublished doctoral dissertation]. University of Huddersfield.

Turner, K. (2016, July 9). Google Translate is getting really, really accurate. *Washington Post*. https://www.washingtonpost.com/news/innovations/wp/2016/10/03/google-translate-is-gettingreally-really-accurate/

U.S. Census Bureau. (2021). *Improved race and ethnicity measures reveal U.S. population is much more multiracial*. https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html

*Ulmer, E. (2018). *Girls' decision to study STEM as influenced by maternal views on math ability based on gender* [Master's thesis, Georgetown University]. ProQuest Dissertations and Theses.

*Vander Heyden, K. M., van Atteveldt, N. M., Huizinga, M., & Jolles, J. (2016). Implicit and explicit gender beliefs in spatial ability: Stronger stereotyping in boys than girls. *Frontiers in Psychology*, *7*, Article 1114. https://doi.org/10.3389/fpsyg.2016.01114

*Verger, H. (2014). *Comment les émotions sont-elles impliquées chez les élèves stigmatisés en situation de menace du stéréotype?* [How are emotions involved in stigmatized students in a situation of stereotypical threat? Master's thesis]. Universite d'Orleans.

Verkuyten, M., & Thijs, J. (2001). Ethnic and gender bias among Dutch and Turkish children in late childhood: The role of social context. *Infant and Child Development*, *10*(4), 203–217. https://doi.org/10.1002/icd.279

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

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

*Vuletich, H. A., Kurtz-Costes, B., Perkins, K. A., Rowley, S., & Hoffman, A. J. (2017). *Longitudinal analysis of the effect of gender stereotype endorsement on success ability attributions in African American youth* [Paper presentation]. The 2017 meeting for the Society for Research on Child Development, Austin, TX, United States.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*(4), 817–835. https://doi.org/10.1037/a0016127

*Wang, J., Hong, H., Ravitz, J., & Hejazi Moghadam, S. (2016). Landscape of K-12 computer science education in the U.S.: Perceptions, barriers, and access. In C. Alphonce, J. Tims, M. Caspersen, & S. Edwards (Eds.), *U.S. proceedings of the 47th ACM technical symposium on computing science education—SIGCSE '16* (pp. 645–650). Association for Computing Machinery. https://doi.org/10.1145/2839509.2844628

Wang, M. T., & Degol, J. (2013). Motivational pathways to STEM career choices: Using expectancy–value perspective to understand individual and gender differences in STEM fields. *Developmental Review*, *33*(4), 304–340. https://doi.org/10.1016/j.dr.2013.08.001

*White, A. (2015). *Gender differences in elective science course enrollment: Influences of stereotypes and causal attributions* [Bachelor's thesis]. University of North Carolina.

*Whitehead, J. M. (1996). Sex stereotypes, gender identity and subject choice at A-level. *Educational Research*, *38*(2), 147–160. https://doi.org/10.1080/0013188960380203

Wigfield, A., Eccles, J. S., & Möller, J. (2020). How dimensional comparisons help to understand linkages between expectancies, values, performance, and choice. *Educational Psychology Review*, *32*(3), 657–680. https://doi.org/10.1007/s10648-020-09524-2

Wilbourn, M. P., & Kee, D. W. (2010). Henry the nurse is a doctor too: Implicitly examining children's gender stereotypes for male and female occupational roles. *Sex Roles*, *62*(9–10), 670–683. https://doi.org/10.1007/s11199-010-9773-7

*Wille, E., Gaspard, H., Trautwein, U., Oschatz, K., Scheiter, K., & Nagengast, B. (2018). Gender stereotypes in a children's television program: Effects on girls' and boys' stereotype endorsement, math performance, motivational dispositions, and attitudes. *Frontiers in Psychology*, *9*, Article 2435. https://doi.org/10.3389/fpsyg.2018.02435

Williams, R., Citkowicz, M., Miller, D. I., Lindsay, J., & Walters, K. (2022). Heterogeneity in mathematics intervention effects: Evidence from a meta-analysis of 191 randomized experiments. *Journal of Research on Educational Effectiveness*, *15*(3), 584–634. https://doi.org/10.1080/19345747.2021.2009072

Wood, L. A., Hutchison, J., Aitken, M., & Cunningham, S. J. (2022). Gender stereotypes in UK children and adolescents: Changing patterns of knowledge and endorsement. *British Journal of Social Psychology*, *61*(3), 768–789. https://doi.org/10.1111/bjso.12510

Wood, W., & Eagly, A. H. (2012). Biosocial construction of sex differences and similarities in behavior. In M. P. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology* (Vol. 46, pp. 55–123). Elsevier. https://doi.org/10.1016/B978-0-12-394281-4.00002-7

Zalk, S. R., & Katz, P. A. (1978). Gender attitudes in children. *Sex Roles*, *4*(3), 349–357. https://doi.org/10.1007/BF00287287

Zhao, F., Zhang, Y., Alterman, V., Zhang, B., & Yu, G. (2018). Can math-gender stereotypes be reduced? A theory-based intervention program with adolescent girls. *Current Psychology*, *37*(3), 612–624. https://doi.org/10.1007/s12144-016-9543-y

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 March 23, 2023
Revision received September 5, 2024
Accepted September 9, 2024 ■