Cognition 245 (2024) 105740

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Why are there no girls? Increasing children's recognition of structural causes of the gender gap in STEM

Jamie Amemiya [a,*], Lin Bian [b]

[a] Department of Psychology, Occidental College, USA
[b] Department of Psychology, University of Chicago, USA

ARTICLE INFO

Keywords:
Inequality
Structural explanations
Gender
STEM

ABSTRACT

The gender disparity in STEM fields emerges early in development. This research examined children's explanations for this gap and investigated two approaches to enhance children's structural understanding that this imbalance is caused by societal, systematic barriers. Five- to 8-year-old children ($N = 145$) observed girls' underrepresentation in a STEM competition; the *No Structural Information* condition presented no additional information, the *Structural: Between-Group Comparison (Between)* condition compared boys' greater representation to girls' when boys had more opportunities to practice than girls, and the *Structural: Within-Group Comparison (Within)* condition compared *girls'* greater STEM representation when they had opportunities versus not. Children in the *No Structural* condition largely generated intrinsic explanations; in contrast, children in both structural conditions favored structural explanations for girls' lack of participation (Experiment 1) and achievement (Experiment 2). Importantly, each structural condition also had unique effects: *Between* raised children's fairness concerns, while *Within* increased children's selection of girls as teammates in a competitive STEM activity.

1. Introduction

Women's underrepresentation in major fields of Science, Technology, Engineering, and Mathematics (STEM) is a pervasive phenomenon (Leslie, Cimpian, Meyer, & Freeland, 2015; Schmader, 2023) and takes root in early childhood (Bian, Leslie, & Cimpian, 2017; Master, 2021; Miller, Nolla, Eagly, & Uttal, 2018; Rhodes, Leslie, Yee, & Saunders, 2019). This inequality is strongly influenced by structural factors, which are societal barriers that systematically constrain women's chances to pursue STEM (e.g., unequal distribution of educational resources, gender discrimination, and stereotyping; Bian, 2022; Master, 2021; Tenenbaum & Leaper, 2003). However, little to no research has examined early explanatory frameworks for STEM gender disparities. Children's recognition of structural barriers as causes, as opposed to intrinsic causes (e.g., appealing to girls' natural interests and abilities), not only promotes a more accurate causal framework, but also may encourage them to view current inequalities as unfair and engage in behaviors to mitigate the gender imbalance in STEM. The present research investigated children's explanations for girls' underrepresentation in STEM and developed novel approaches to enhance their attention to underlying structural constraints.

Recent research finds that, when provided with structural information, children's explanations for social group differences can shift from intrinsic to structural (Peretz-Lange, Perry, & Muentener, 2021; Vasilyeva, Gopnik, & Lombrozo, 2018). In these studies, children observed novel, unfamiliar disparities between two groups (e.g., girls are more likely than boys to play a novel game, "Green-Ball") and were presented with structural information indicating that the two groups had differential opportunities. In one study, children learned that the girls' classroom had physical constraints that made it easier to play one game over another, while the setting of the boys' classroom encouraged the opposite choice (Vasilyeva et al., 2018). This type of information, which we refer to as a *between-group comparison*, is able to promote young children's structural reasoning about novel disparities starting early in childhood.

However, there are reasons to believe that promoting children's structural reasoning about real-world, widespread inequalities, such as women's underrepresentation in STEM, may be more challenging. Already by age 6, children believe that girls are less interested in certain STEM fields (Master, Meltzoff, & Cheryan, 2021) and less competent at math than boys (Cvencek, Meltzoff, & Greenwald, 2011). Also by age 6, girls are less likely than boys to believe that members of their own

* Corresponding author at: Department of Psychology, Occidental College, 1600 Campus Rd, Los Angeles, CA 90041, USA.
E-mail address: amemiya@oxy.edu (J. Amemiya).

https://doi.org/10.1016/j.cognition.2024.105740
Received 20 August 2023; Received in revised form 1 February 2024; Accepted 6 February 2024
Available online 15 February 2024

gender are intellectually gifted (Bian et al., 2017). Given children's strong prior intrinsic beliefs about gender and STEM, children may be less receptive to structural information (Amemiya, Mortenson, Ahn, Walker, & Heyman, 2021; Amemiya, Mortenson, Heyman, & Walker, 2023). In line with this possibility, Yang, Naas, and Dunham (2021) found that between-group comparisons increased 7- to 8-year-old children's structural over intrinsic explanations about a *novel* gendered behavior (e.g., girls played with "Green-Ball"), but children maintained intrinsic explanations for a *familiar* gendered behavior (e.g., girls played with dolls). Similarly, Amemiya et al. (2021) found that despite the presence of constraints, 5- to 10-year-old children are more likely to infer a preference when a target made a gender-stereotypical choice (e.g., a girl chose a doll over a truck) than when the target made a gender-neutral choice (e.g., a girl chose a green over a yellow toy). Whether structural information about gender inequalities in STEM can help children adopt a structural explanatory framework, and overturn intrinsic explanations, remains unknown. Addressing this question will begin to answer questions about children's real-world structural reasoning, as much of prior research has focused on novel inequalities (e.g., Peretz-Lange et al., 2021; Peretz-Lange & Muentener, 2019; Vasilyeva et al., 2018).

We examine children's explanations for the STEM gender gap and investigate two theoretically-grounded approaches to promote children's structural reasoning. First, we apply insights from the causal reasoning literature to improve the between-group comparison approach, specifically research indicating that children are sensitive to the *quality* of an explanation (Danovitch, Mills, Sands, & Williams, 2021), including the extent to which the explanation offers new causal information (Mills, Sands, Rowles, & Campbell, 2019). Prior failures of the between-group comparison approach with real-world gender differences (e.g., why girls play with dolls more than boys) may be because the structural information did not offer a compelling alternative explanation to overturn children's prior beliefs about inherent gender differences. For example, one study's structural explanation condition presented a process in which girls threw a pebble into two different-sized buckets that favored playing with dolls (because girls' doll bucket was larger than the truck bucket), while boys' buckets favored trucks (Yang et al., 2021). This type of process is rarely observed in real life and may not serve as a realistic constraint that leads to gendered choices.

Here, we tested whether presenting a more intuitive structural explanation better supports structural over intrinsic reasoning. We informed children that boys (but not girls) were always able to practice their STEM skills with a teacher, and boys (but not girls) were always represented in a subsequent STEM competition. This explanation may be more compelling for two reasons. First, children develop an early understanding of how practice and effort influence performance (Muradoglu & Cimpian, 2020), making the explanation a more plausible mechanism underlying the gender disparity. Second, children are highly attentive to the strength of causal relationships (e.g., if they are deterministic or probabilistic; Schulz & Sommerville, 2006), and may be sensitive to the fact that girls were *always* structurally disadvantaged and subsequently were *always* underrepresented in the STEM competition.

Second, and importantly, we also developed a novel structural approach that drew upon research from the counterfactual reasoning literature (Gopnik & Wellman, 2012), in which children were shown a *within-group comparison* that compares *girls'* STEM outcomes with versus without constraints. Specifically, children were informed of the counterfactual that when girls are given more opportunities to practice, their participation and performance in STEM increases compared to when they have no opportunities. Given that the same group (i.e., girls) is being compared across situations, structural barriers become a clear difference-maker and children may be especially likely to reference constraints in their explanations (see Goddu & Gopnik, 2020; Seiver, Gopnik, & Goodman, 2013). Notably, a within-group comparison removes the gender confound that children may identify in between-group comparisons: boys and girls may also differ on intrinsic factors (e.g., STEM interest) in addition to their constraints (Amemiya et al., 2023; see also Christie & Gentner, 2010 on the importance of minimzing differences for effective comparisons).

2. Differential effects of between-group comparisons and within-group comparisons

We have proposed two approaches—*between-group comparisons* that highlight STEM opportunities that boys but not girls have (here, a STEM teacher), and *within-group comparisons* that show how providing more opportunities for girls makes a difference for girls' STEM outcomes. While we hypothesize that both approaches may increase structural reasoning, we now consider how these approaches may have *differential* effects on other important outcomes: (1) children's concerns about fairness—a major precursor to inequality-rectifying behavior (e.g., allocating resources to disadvantaged groups; Rutland & Killen, 2017)—and (2) their inclusion of girls as teammates in high-stakes STEM activities.

First, we propose that between-group comparisons are more likely to raise fairness concerns about girls lacking a STEM teacher than within-group comparisons. Children are highly sensitive to inequality between individuals and groups, such that they often prefer to distribute resources in a way that all parties have the same amount (Elenbaas, Rizzo, & Killen, 2020; Shaw & Olson, 2012). Between-group comparisons explicitly point out the resource inequality between boys and girls, and may also imply that a STEM teacher is an *expected* resource for each group to have. In contrast, within-group comparisons only focus on girls. Children may be uncertain as to whether or not boys had this resource, and not directly compare girls' and boys' possession of resources, thereby not raising concerns about fairness.

On the other hand, we hypothesized that within-group comparisons would be more effective than between-group comparisons for promoting inclusion of girls in a STEM competition. In high-stakes activities, children may be less concerned about fairness but instead focused on girls' ability to perform. Prior research finds that children show a negative bias toward girls in such contexts: they are less willing to include girls as teammates for activities requiring sheer brilliance (Bian, Leslie, & Cimpian, 2018) and in STEM activities when they believe that boys are more capable than girls (McGuire et al., 2022). We propose that within-group comparisons provide clear evidence that girls have the ability to succeed when given proper training. Although between-group comparisons *suggest* that girls' lack of prior STEM success is due to having less practice than boys, this type of information does not provide direct evidence that girls are capable of succeeding in STEM nor that they perform at boys' level once provided with training opportunities.

3. The present research

The current research examined how children explain girls' underrepresentation in STEM, and whether providing structural information impacts their reasoning about and reactions to this pattern. We tested two theoretically-grounded structural approaches: (1) *between-group comparison*, an approach in which we highlight a more intuitive mechanism to explain gender inequalities in STEM (i.e., girls but not boys lacked opportunities to practice), and (2) *within-group comparison*, a new approach rooted in counterfactual reasoning that reveals the malleability of girls' STEM representation when they are versus are not given training opportunities. We compared these structural conditions to one another and to a *no structural information* condition in order to assess children's spontaneous causal reasoning when the causes are ambiguous.

Experiment 1 focused on children's reasoning about girls' lack of STEM *participation*, while Experiment 2 focused on girls' lack of STEM *achievement*. Girls' lack of STEM participation versus achievement may invoke distinct intrinsic stereotypes (i.e., girls' lack of interest versus ability in STEM, respectively), which allows us to test the robustness of structural approaches to overturn various intrinsic explanations. Experiment 2 also assessed children's perceived fairness of the STEM context and inclusion of girls as teammates in STEM. For both experiments, we hypothesized that the *between-group comparison* and *within-group comparison* conditions would overturn stereotypical reasoning (i.e., privileging structural over intrinsic explanations for girls' underrepresentation) relative to the *no structural information* condition. However, the two approaches may have distinct effects on children's reactions to the STEM gap. We predicted that *between-group comparison* would have the strongest effects on children's perceived (un)fairness of the STEM context, given that it emphasizes an unequal distribution of opportunities between boys and girls. On the other hand, we predicted that *within-group comparison*, which is the only condition to demonstrate girls' abilities, would be most likely to increase children's inclusion of girls in STEM activities.

To our knowledge, this is the first study to examine how children explain gender inequality in STEM and how to improve their recognition of the structural reasons for this disparity. We focused on children ages 5- to 8-years-old for three key reasons. First, the transition from age 5 to 6 appears to be a critical time for the emergence of gendered stereotypes about ability and STEM (Bian et al., 2017; Cvencek et al., 2011). In addition, prior research on structural reasoning has focused on early to middle childhood (i.e., approximately 5 to 8 years old; Peretz-Lange et al., 2021; Vasilyeva et al., 2018; Yang et al., 2021), and thus our findings can speak to this growing body of research. More broadly, our focus on this earlier time period can inform intervention efforts to mitigate gender inequalities in STEM at their roots.

The studies were not pre-registered. All study materials, data, and code are available at the following link: https://osf.io/gx7a5/

4. Experiment 1: Reasoning about girls' STEM participation

Experiment 1 investigated children's explanations for girls' lack of STEM participation. Importantly, we examined how two approaches to providing structural information, *between-group comparison* and *within-group comparison*, affected children's explanations for this pattern.

4.1. Method

4.1.1. Participants

Participants were 73 five- to eight-year-old children (37 girls, 36 boys; 35 White, 14 Multiracial, 10 Asian, 7 Latine/x, 4 Black, 1 Native American, 2 not reported) recruited from a university database and social media advertisements. With respect to socioeconomic status (SES) information, the sample was mostly higher SES as indicated by parents' highest level of education (5% High school diploma, 10% Associate's degree, 18% Bachelor's degree, 34% Master's degree, 30% Professional degree, 3% not reported) and median gross annual income ($95,000). An additional two children were excluded because of attentional difficulties ($n = 1$) or failing to complete the study ($n = 1$). We chose this sample size based on prior developmental studies that indicate a large effect of structural information on how children reason about novel gender inequalities (Vasilyeva et al., 2018; Yang et al., 2021). A sensitivity analysis conducted in G*Power (F-test for one-way ANOVA, $N = 72$, 3 groups, power = 0.80, $\alpha = 0.05$; Faul, Erdfelder, Buchner, & Lang, 2009) indicated that the present sample size would be sufficient to detect an effect of at least $\eta^2 = 0.12$.

4.1.2. Procedure

Children were interviewed by an experimenter over the video conferencing platform, Zoom. The study materials consisted of prerecorded videos embedded within the online survey platform, Qualtrics. Children first heard about an annual robot-building competition, received the experimental manipulation, completed the dependent measures, and finally were debriefed and given a small gift. During the debrief, children were informed that the kids they observed were only pretend kids, that all teachers welcome students to learn robotics, and that it is possible for all kids to learn robotics with practice.

4.1.3. Experimental manipulation

In all experimental conditions, children were introduced to a far, far away town with an all-girls school and an all-boys school. Children were told about an annual robot-building competition in the local park, in which kids are tested on how to build a robot and that there is a prize for building the best one. Children were then presented with information from one of the three conditions depicted in Fig. 1. Note that the actual study conditions used real photographs of children; these materials are available at the OSF study link.

4.1.4. No structural information

In the *no structural information* (*No Structural*) condition, participants were presented with photographs of children who chose to participate in the robot competition for the past four years. Across all years, participants were shown that only boys participated, which aligned with the gender disparities that children observe in everyday life. This condition allowed us to examine how children spontaneously reason about STEM gender disparities when the causes are unclear. Moreover, this condition provided a more stringent comparison to the structural conditions, as prior studies have used control conditions that overtly indicate intrinsic causes (Peretz-Lange et al., 2021; Vasilyeva et al., 2018; Yang et al., 2021).

4.1.5. Structural: Between-group comparison (Between)

Children in this condition were also shown that only boys participated in the robot competition for the past four years. In addition, children were presented information regarding girls' versus *boys'* opportunities. Specifically, children observed that boys had a systematic educational advantage over girls across four years: the boys' school had a robotics teacher and they were able to practice, while the girls' school did not have a teacher and they could not practice. This structural information was presented before each instance of boys' overrepresentation at the robotics competition to reinforce the systematic opportunity difference between boys and girls.

4.1.6. Structural: Within-group comparison (Within)

Children in the within condition received information comparing *girls* across contexts with versus without constraints. Specifically, this condition showed the counterfactual that during the first two years when *girls* had a robotics teacher and could practice, girls were overrepresented in the competition. However, during the final two years when girls did *not* have a teacher and could not practice, boys were overrepresented. To clearly differentiate the *Within* condition from *Between*, there was no information about the boys' school.

4.2. Measures

4.2.1. Open-ended explanation

Children were asked to explain why no girls participated in the most recent competition (i.e., Year 4). All visual information from the structural conditions (i.e., images of the schools and teachers) was removed to examine whether children would spontaneously apply the previous information without explicit instruction to do so. Two independent coders categorized children's responses into one of three explanation types. Explanations that referred to girls' school or educational opportunities as the root cause were coded as *structural* (e.g., "Because I think the science class was cancelled and the girls couldn't participate in the robot competition because they didn't know how to build them"; Cohen's K = 0.86). Explanations were coded as *intrinsic* when children referenced girls' interests or abilities (e.g., "Because they don't like building robots"), or referred to gender categories (e.g., "Because robots

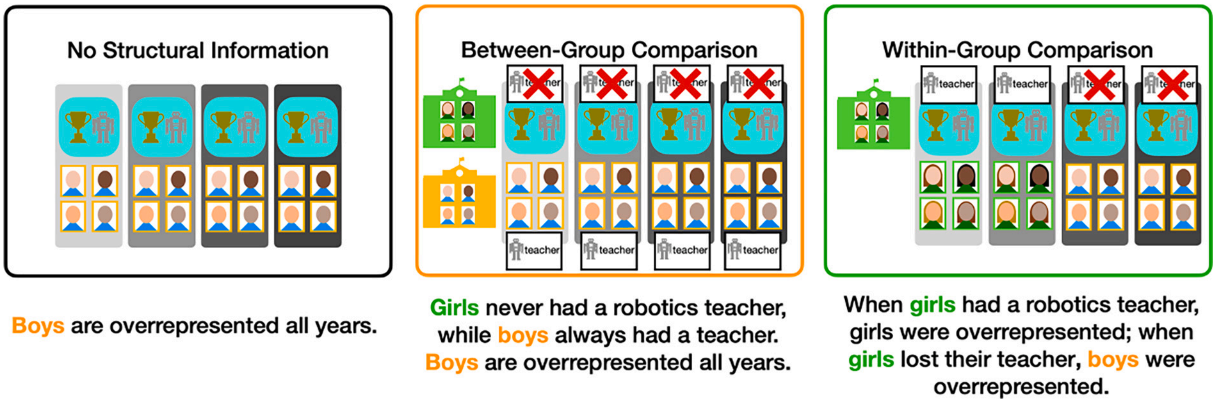

Fig. 1. Overview of Experiment 1 study conditions. Note that the actual study conditions used real photographs of children.

are for boys") (Cohen's K = 0.88). Other miscellaneous answers were coded as *other* (e.g., "There was a bad storm" or "I don't know"; Cohen's K = 0.67).

4.2.2. Explanation ratings task

Children rated explanations offered by three alien characters, presented in random order. Each character was asked, "Why were there no girls who went to the robot building competition?" One character provided a *structural explanation*, "I think it is something about the girls' *school*. Maybe it is because the girls' robot building class was canceled that year." Another character provided an *intrinsic (preference) explanation*, "I think it is something about *girls*. Maybe it is because girls do not really like robot building." Finally, a third character provided a *random explanation* to rule out that children would agree with any explanation, "I think it is something about the *stars* at night. Maybe it is because there are stars in the sky." Following each explanation, children were asked: "Do you think what [character] said is right or not right? Do you think [character] is a little (not) right or really (not) right?" Children's answers were coded such that 1 = *really not right*, 2 = *a little not right*, 3 = *a little right*, and 4 = *really right*.

4.3. Results

4.3.1. Open-ended explanation

Fig. 2 presents children's open-ended explanations (structural, intrinsic, other) by condition. Explanations did not vary by children's gender or age. A chi-square test revealed that children's explanations varied significantly by condition, $\chi^2(4) = 36.22$, $p < .001$, Cramer's $V = 0.50$. Children overwhelmingly generated *intrinsic* explanations in the *No Structural* condition (72% of children). Thus, without structural information, children appear to default to intrinsic causes.

A logistic regression indicated that, compared to the *No Structural* condition, children generated significantly more structural explanations in the *Between* condition, $B = 3.60$, $SE = 0.82$, $p < .001$, $OR = 36.67$, and in the *Within* condition, $B = 3.09$, $SE = 0.78$, $p < .001$, $OR = 22.00$. Moreover, relative to the *No Structural* condition, children generated fewer intrinsic explanations in the *Between* condition, $B = -3.34$, $SE = 0.86$, $p < .001$, $OR = 0.04$, and in the *Within* condition, $B = -3.34$, $SE = 0.86$, $p < .001$, $OR = 0.04$. There were no condition differences in generating "other" explanations. Finally, there were no condition differences when comparing the *Within* condition to *Between* for structural explanations, $B = -0.51$, $SE = 0.72$, $p = .48$, $OR = 0.60$, or for intrinsic explanations, $B = 0.00$, $SE = 0.00$, $p = 1.00$, $OR = 1.00$.

4.3.2. Explanation ratings

Fig. 3 presents children's ratings of the structural, intrinsic, and random explanation by condition. Neither children's gender or age predicted their explanation ratings. One-way analysis of variance indicated that there were condition differences in children's ratings of the structural explanation, $F(2, 70) = 25.47$, $p < .001$, $\eta^2 = 0.42$, the intrinsic explanation, $F(2, 70) = 11.06$, $p < .001$, $\eta^2 = 0.24$, but not the random explanation, $F(2, 70) = 1.34$, $p = .27$, $\eta^2 = 0.04$, in which ratings were low across conditions.

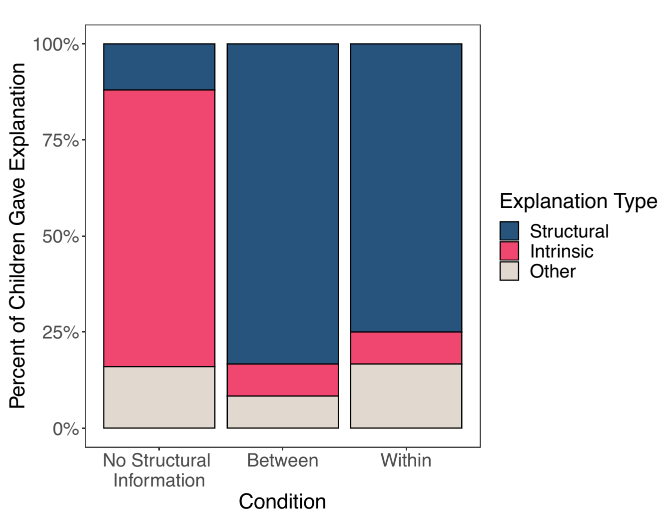

Fig. 2. Children's open-ended explanations for girls' lack of STEM participation.

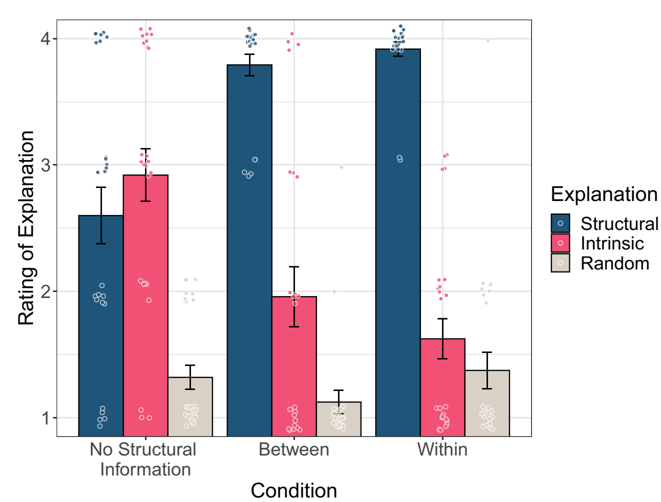

Fig. 3. Children's ratings of explanations for girls' lack of STEM participation. The bars represent means, the error bars represent 95% CIs, and the points indicate participants' raw data.

A linear regression model indicated that compared to the *No Structural* condition, children rated the structural explanation more positively in the *Between* condition, $B = 1.19$, $SE = 0.20$, $p < .001$, $\beta = 0.61$, and in the *Within* condition, $B = 1.32$, $SE = 0.20$, $p < .001$, $\beta = 0.67$. Moreover, compared to the *No Structural* condition, children rated the intrinsic explanation less positively in the *Between* condition, $B = -0.96$, $SE = 0.29$, $p = .001$, $\beta = -0.40$, and in the *Within* condition, $B = -1.30$, $SE = 0.29$, $p < .001$, $\beta = -0.54$. Children in the *Within* versus *Between* conditions did not differ significantly in their ratings of structural explanations, $B = 0.13$, $SE = 0.21$, $p = .55$, $\beta = 0.06$, or in their ratings of intrinsic explanations, $B = -0.33$, $SE = 0.29$, $p = .25$, $\beta = -0.14$.

Importantly, while children in the *No Structural* condition rated both structural and intrinsic explanations similarly (although structural ratings were *numerically* lower than intrinsic ratings), structural-intrinsic difference: $t(24) = -0.93$, $p = .36$, Cohen's $d = -0.19$, children in the structural conditions privileged structural *over* intrinsic explanations, *Between*: structural-intrinsic difference: $t(23) = 6.13$, $p < .001$, Cohen's $d = 1.25$; *Within*: structural-intrinsic difference: $t(23) = 13.08$, $p < .001$, Cohen's $d = 2.67$.

4.4. Experiment 1 Discussion

Experiment 1 indicates that, when the causes are ambiguous in the *No Structural Information* condition, children generate and favor *intrinsic* explanations for girls' STEM underrepresentation. Importantly, *both* structural conditions led children to recognize structural constraints. Thus, in line with our prediction, a between-group comparison *can* overturn intrinsic reasoning about gender inequalities in STEM when it presents a compelling alternative causal mechanism (e.g., unequal opportunities to practice). Moreover, we find support for the novel within-group comparison approach that relies on counterfactual evidence to show that structural barriers make a difference for girls' STEM participation.

In Experiment 2, we sought to replicate these findings when children reason about girls' lack of STEM *achievement*, which may encourage children to draw on gender stereotypes about ability. We focused on the achievement domain because it also allowed us to test for potential key differences between the structural conditions: while *between-group comparisons* may point out the unfairness of who gets more STEM learning opportunities, *within-group comparisons* uniquely reveal girls' ability in STEM and may increase children's selection of girls as teammates in a competitive STEM activity.

5. Experiment 2: Reasoning about girls' STEM achievement and behavioral responses

In Experiment 2, children observed who *won* the robotics competition and were then asked to select a team for an upcoming competition. Children were also asked to reason about girls' lack of achievement in the past competition (i.e., rate causal explanations and judge the fairness of the competition when girls lacked a teacher). We hypothesized that, like Experiment 1, both *Within* and *Between* conditions would increase endorsement of structural over intrinsic explanations for girls' performance. However, we expected that the *Between* condition would lead children to rate the competition as unfair because this condition emphasized girls' disadvantage *relative* to boys' advantage in educational resources. In contrast, we hypothesized that the *Within* condition would increase children's tendency to choose girls as teammates, given that this condition explicitly demonstrated girls' abilities.

5.1. Method

5.1.1. Participants

This study included 72 five- to eight-year-old children (36 girls, 36 boys; 39 White, 13 Latine/x, 6 Asian, 6 Multiracial, 2 Black, 2 Arabic, 4 not reported) recruited from the associated university database, social media advertisements, and the local museum. The sample was mostly higher SES as indicated by parents' highest level of education (14% High school diploma, 4% Associate's degree, 15% Bachelor's degree, 38% Master's degree, 17% Professional degree, 13% not reported) and median gross annual income ($150,000). This sample size was determined following the same criteria as Experiment 1. An additional nine children were excluded from analyses because of failing the comprehension check in the teammate selection task ($n = 6$; see below) or study incompletion ($n = 3$).

5.1.2. Procedure

Children were interviewed by an experimenter either over the video conferencing platform, Zoom, or in-person at the local museum using the same materials.

Across all conditions, children observed the "Program-A-Robot" competition winners for the past four years. In the *No Structural* condition, children observed that a team of four boys won every year and no additional information was provided. In the *Between* condition, children also observed that an all-boy team won the competition for the past four years. In addition, they observed that boys got to talk with a teacher and practice each year, while girls did not get to talk with the teacher and could not practice. In the *Within* condition, children observed that during the first two years, *girls* talked with a teacher and an all-girls team won those years. However, the next two years girls did not get to talk with the teacher and an all-boys team won.

Children then completed the teammate selection task in which they selected four kids who they thought had the best chance to win the upcoming robotics competition. After the selection task, children were shown the manipulation video again (i.e., either *No Structural*, *Between*, or *Within*), and completed in random order (a) causal explanation ratings for girls' lack of achievement in the final year, and (b) judgments of how fair the competition was in the final year. Children were debriefed at the end of the study as in Experiment 1 and given a small gift.

5.2. Measures

5.2.1. Teammate selection task

Children were presented a new group of ten children (5 girls, 5 boys interspersed; children were from diverse racial backgrounds). Critically, prior to making teammate selections, participants were told that *all* of these children had talked with the teacher and had practiced. This allowed us to test children's implicit beliefs about girls' abilities: If children attributed girls' past failures as being caused *solely* by a lack of educational opportunities, they should hold no bias against girls now that they had proper training. To ensure that children encoded this information, the experimenter asked, "Can you remind me, did *all* of these kids in the picture above talk with the teacher to practice, yes or no?" If children failed this check after hearing the information twice, they were excluded from the analyses ($n = 6$ were excluded). Following the question, the experimenter again repeated, "All of these kids got a chance to talk with the teacher to practice." Directly after this statement, participants were asked to choose four teammates from this array of children for the upcoming competition, "You will now pick a team of four kids that you think have the *best* chance at winning Program-A-Robot! Who is the [first/second/third/fourth] kid you want to choose for the team?"

5.2.2. Explanation ratings

Children rated explanations for the final year, in which an all-boys team won. The *intrinsic (ability) explanation* was, "I think it is something about *boys*. Maybe it is because boys are just better at programming robots." We asked about boys' higher ability rather than girls' lower ability given that this framing was less likely to upset participants and to also reduce potential floor effects. The *structural explanation* was, "I think it is something about *girls not having a teacher*. Maybe it is because girls did not get a chance to talk with a teacher and practice programming robots that year." Children were asked, "Do you think it is right or not right? Do you think it is a little (not) right, (not) right, or really (not) right?" Children's responses were coded as 1 = *really not right*, 2 = *not right*, 3 = *a little not right*, 4 = *a little right*, 5 = *right*, 6 = *really right*.

5.2.3. Fairness judgment

Children were asked how fair they believed the competition was the final year in which boys were overrepresented, "Do you think the Program-A-Robot competition that year was fair or not fair? Do you think it was a little (not) fair, (not) fair, or really (not) fair?" Children's fairness evaluations were coded as 1 = *really not fair*, 2 = *not fair*, 3 = *a little not fair*, 4 = *a little fair*, 5 = *fair*, 6 = *really fair*.

5.3. Results

5.3.1. Explanation ratings

There were no gender or age differences in children's endorsement of structural or intrinsic explanations. As depicted in Fig. 4, there were no condition differences in children's ratings of intrinsic (ability) explanations, $F(2, 69) = 1.04$, $p = .36$, $\eta^2 = 0.03$, such that intrinsic ratings were low across conditions. On the other hand, ratings of structural explanations varied significantly by condition, $F(2, 69) = 20.88$, $p < .001$, $\eta^2 = 0.38$. A linear regression model indicated that, compared to the *No Structural* condition, children rated the structural explanations higher in the *Between* condition, $B = 2.67$, $SE = 0.44$, $p < .001$, $\beta = 0.66$, and in the *Within* condition, $B = 2.25$, $SE = 0.44$, $p < .001$, $\beta = 0.56$. Children in the *Within* versus *Between* condition did not differ significantly in their ratings of structural explanations, $B = -0.42$, $SE = 0.44$, $p = .35$, $\beta = -0.10$. Thus, we replicated Experiment 1's results that both structural conditions increase children's endorsement of structural explanations.

We also replicated the findings from Experiment 1 with respect to children's tendency to privilege structural causes *over* intrinsic causes. Specifically, the structural-intrinsic difference was not significant for the *No Structural* condition (although structural ratings were again numerically lower than intrinsic ratings), $t(23) = -1.14$, $p = .27$, Cohen's $d = -0.23$, while the structural-intrinsic difference scores were significant for both structural conditions (*Between*: $t[23] = 4.73$, $p < .001$, Cohen's $d = 0.97$; *Within*: $t[23] = 4.34$, $p < .001$, Cohen's $d = 0.89$).

5.3.2. Fairness judgment

There were no age differences in fairness judgments, but girls rated the competition as less fair than boys, $B = -0.89$, $SE = 0.44$, $p = .046$, $\beta = -0.24$. In line with our hypotheses, children's fairness judgments varied significantly by condition, $F(2, 69) = 12.04$, $p < .001$, $\eta^2 = 0.26$ (see Fig. 5). Compared to the *No Structural* condition, children in the *Between* condition rated the STEM competition as less fair, $B = -1.79$, $SE = 0.48$, $p < .001$, $\beta = -0.45$, while the *Within* condition was not statistically different from the *No Structural* condition, $B = 0.42$, $SE = 0.48$, $p = .39$, $\beta = 0.10$. This is notable given that children in the *Within* condition also rated the final year in which girls did not have a teacher and could not practice. Thus, the *Between* condition appears to have uniquely affected children's moral concerns because it emphasized girls' educational disadvantages *relative* to boys' advantages.

5.3.3. Teammate selections

Because of children's strong ingroup favoritism (e.g., Dunham, Baron, & Carey, 2011), it is possible that children would choose teammates of their own gender in the initial selections, but their beliefs about girls' abilities would appear in later selections. Indeed, prior research using this type of teammate selection task has found that condition effects do not emerge until the later, third trial (Bian et al., 2018). We thus examined initial trials (1 and 2) separately from later trials (3 and 4), given that the experimental manipulation may affect *later* trials more strongly. Furthermore, analyzing across two trials at a time gave us greater power to detect possible condition effects, given the potential noise in any single trial.

Fig. 6 presents children's selections of girls as teammates by condition and split by participant gender. We first ran chi-square tests examining the frequency of selecting girls as teammates by condition for each set of rounds. For the *initial two* selections, children's *own* gender was strongly associated with selections, $\chi^2(1) = 34.05$, $p < .001$, Cramer's V = 0.49, such that girls tended to choose girls (74% of the time) and boys tended to choose boys (76% of the time). Initial selections did not vary by condition, $\chi^2(2) = 1.06$, $p = .59$, Cramer's V = 0.09. For the *last two* rounds, children's selections of girl teammates still varied by their own gender, $\chi^2(1) = 14.80$, $p < .001$, Cramer's V = 0.32, girls tended to choose girls (71% of the time) while boys still favored boys (63% of the time). Critically, teammate selections in the final two

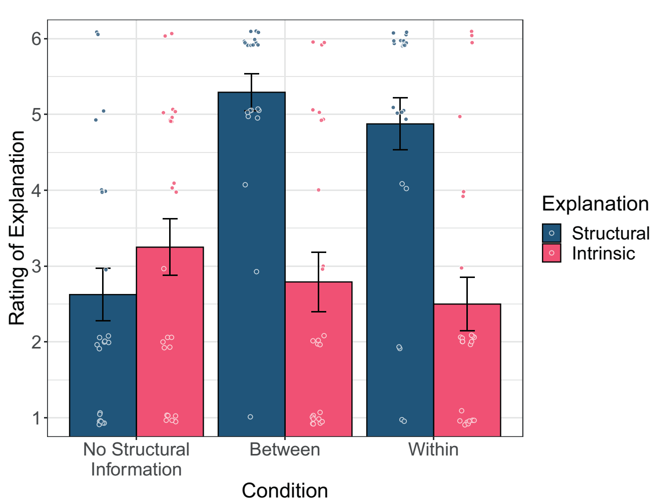

Fig. 4. Children's ratings of explanations for the final year that boys were overrepresented. The bars represent means, the error bars represent 95% CIs, and the points indicate participants' raw data.

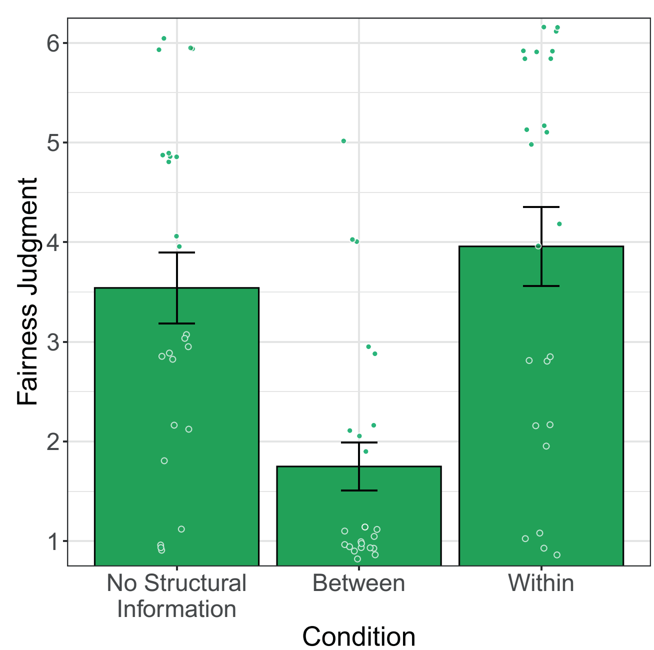

Fig. 5. Children's fairness judgment of the competition during the final year that boys were overrepresented. The bars represent means, the error bars represent 95% CIs, and the points indicate participants' raw data.

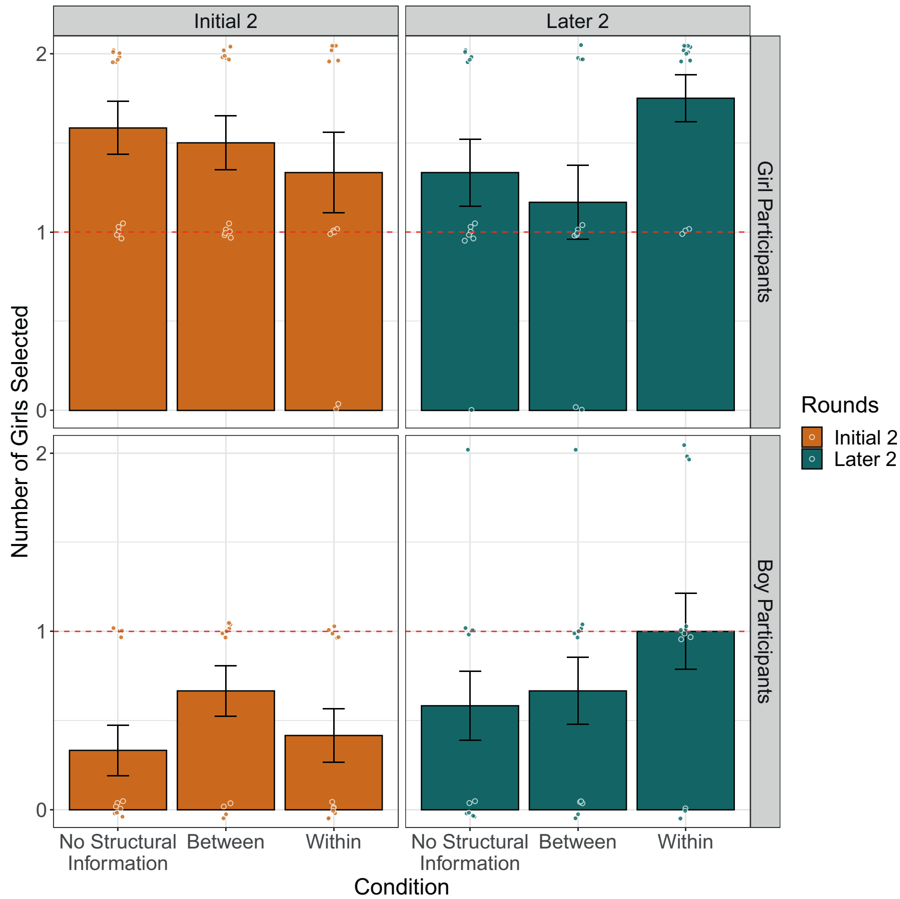

Fig. 6. Number of girls selected as teammates for the initial two rounds (brown bars) and later two rounds (teal bars) split by participant gender (girls in top row, boys in bottom row). The red dotted line represents children's tendency to choose an equal number of boys and girls. The bars represent means, the error bars represent 95% CIs, and the points indicate participants' raw data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

rounds also varied by *condition*, $\chi^2(2) = 6.21$, $p = .04$, Cramer's V = 0.21.

A mixed-effects logistic regression model predicting the final two selections indicated that compared to the *No Structural* condition, children were more likely to choose girls in the *Within* condition, $B = 0.91$, $SE = 0.46$, $p = .048$, $OR = 2.47$. This effect was not moderated by children's own gender; indeed, Fig. 6 shows that the increase in the *Within* condition can be observed in *both* girls' and boys' selections for the final two rounds (teal bars). Notably, the *Between* condition did not differ from the *No Structural* condition, $B = -0.09$, $SE = 0.43$, $p = .84$, $OR = 0.92$. Moreover, compared to the *Between* condition, the *Within* condition led to higher selections of girls in the final two rounds, *Within*: $B = 0.99$, $SE = 0.46$, $p = .03$, $OR = 2.70$. In sum, after children expressed in-group preferences in the initial rounds, within-group comparison predicted higher selections of girls as teammates while between-group comparisons did not.[1]

6. General discussion

The present experiments examined children's explanations for girls' underrepresentation in STEM. We find that children spontaneously generate intrinsic explanations when the causes of this pattern are ambiguous in the *No Structural Information* condition. Importantly, we investigated two approaches to promote structural reasoning: *between-group comparisons* that emphasize girls' educational disadvantages compared to boys, which sought to provide a compelling alternative explanation to innate gender differences, and *within-group comparisons*, which revealed that girls' STEM representation increases once girls have opportunities to learn. Our results indicate the robustness of these structural approaches: both strategies increased children's structural over intrinsic reasoning about girls' lack of STEM participation (Experiment 1) and about girls' lack of STEM achievement (Experiment 2). However, these approaches also had *differential* effects on children's reactions to the gender disparity in STEM: Between-group comparisons led children to evaluate the STEM context as unfair, while within-group comparisons increased children's inclusion of girls as STEM teammates.

While prior research has examined children's early-emerging stereotypes about gender and STEM (Cvencek et al., 2011; Master et al., 2021), our study offers novel insight into how children *explain* the gender disparities they observe. The present results indicate that, when structural information is absent, children generate intrinsic explanations that align with previously documented stereotypes (e.g., girls are less interested in STEM than boys; Master et al., 2021), suggesting that children may readily apply gender stereotypes to make sense of observed inequalities. Moreover, the fact that very few children spontaneously generated structural explanations supports prior research indicating that intrinsic explanations tend to be an intuitive way to explain observed regularities in the world (Cimpian & Salomon, 2014; Rhodes & Mandalaywala, 2017). Given that intrinsic explanations increase children's support for the status quo (Hussak & Cimpian, 2015),

[1] Although children's selections of girls are the focus of the current study, we report children's overall selections at the intersection of gender and race in the Supplemental Materials.

children may be at risk of perpetuating STEM inequalities when left to explain these disparities on their own.

Importantly, our research developed two approaches for providing structural information to reduce children's reliance on inherent reasoning and recognize structural constraints faced by girls. First, we found evidence that it is possible to improve the between-group comparison approach that previously failed to overturn intrinsic reasoning about real-world gender differences (Yang et al., 2021). Specifically, we found that children accepted a between-group comparison as evidence for structural rather than intrinsic causes when they learned that girls never had a chance to practice the relevant STEM skills and boys always did. We posit that the *plausibility* of this structural explanation (i.e., children have prior understanding of how practice impacts performance; Muradoglu & Cimpian, 2020) and the *strength of the relationship* (i.e., girls' relative disadvantage was always followed by their underrepresentation; Schulz & Sommerville, 2006) may have made it a compelling alternative to presumed inherent gender differences as the explanation (Muradoglu & Cimpian, 2020; Schulz & Sommerville, 2006). An open question is the extent to which each of these factors impact structural reasoning or may compensate for the other; for example, if children can learn to accept an unfamiliar structural constraint as causal when they observe that it is routinely followed by a gender disparity.

We also developed a second, novel approach in which children were presented with within-group comparisons of girls' participation in STEM with versus without structural constraints. Given that children observe the *same* group's (i.e., girls') representation change across contexts, this approach provides direct evidence against intrinsic explanations and highlights structural barriers as a clear difference-maker (Amemiya et al., 2023). It is notable that we were able to increase structural reasoning in the within condition without needing to explain anything about boys. Our findings indicate that within-group comparisons about girls are sufficient to increase structural reasoning, however it is not entirely clear if children interpreted this evidence to mean that (a) boys *never* had a teacher and were always intrinsically more interested and capable than girls, or (b) boys *always* had a teacher which facilitated their STEM participation and success. Given that children disagreed with the intrinsic explanations in this condition (agreement with intrinsic explanations—such as "boys are just better at programming robots"—were numerically lowest in the *Within* condition), we posit that the latter interpretation is more likely. Nonetheless, future work could improve our understanding of the *Within* condition by assessing children's inferences about the boys' situation, and also directly assessing whether children changed their beliefs about *girls'* abilities. Future research could also test whether making it explicit that boys also had a teacher when girls were successful with a teacher (i.e., girls did not win just because they had a training advantage over boys) strengthens the effects.

Although this study indicates that both between- and within-group comparisons increase children's structural over intrinsic causal explanations, we anticipate that there may be cases in which within-group comparisons are more effective. Between-group comparisons do not provide counterfactual evidence for what would have happened for girls if they were unconstrained. Without observing girls' STEM representation when given greater opportunities, children could still hold onto intrinsic beliefs and reason that boys would have pursued STEM at higher rates than girls regardless of the societal structure (see Amemiya et al., 2023). The current study likely mitigated this reasoning by presenting a near-deterministic constraint in which girls completely lacked the opportunities to practice, compared to boys who could always practice. However, many societal constraints are more probabilistic and less salient—for example, girls experience psychological constraints such as a lack of belonging (Master, Cheryan, & Meltzoff, 2016) that may be harder for children to recognize as having a strong causal impact on inequality (see Pesowski, Denison, & Friedman, 2016). In cases where children have strong intrinsic beliefs and the constraint is less obvious, within-group comparisons may be more effective than between-group comparisons in showing that these structural factors clearly make a difference for girls' STEM representation.

We also find evidence that the two structural approaches may enhance children's tendency to redress STEM inequalities, though via unique pathways. We note that this is an important theoretical contribution, as prior studies test a *singular* structural approach, leaving open questions about the more fine-grained mechanisms of how the approach operates. In our study, we find that between-group comparisons lead children to evaluate the STEM context as more unfair relative to the *No Structural Information* condition. Interestingly, children in the within-group comparison condition did not rate the final competition as less fair than the *No Structural Information* condition, despite the fact that they were told that girls lost their teacher that year. This suggests that emphasizing girls' *relative* disadvantage to boys may be critical for recognizing unfairness. It is also possible that seeing equal rates of success among girls (in years 1 and 2) and boys (in years 3 and 4) in the within-group comparison buffered concerns about unfairness. Given that fairness concerns motivate children to allocate resources to disadvantaged groups (Rizzo, Elenbaas, & Vanderbilt, 2020), presenting between-group comparisons may be an important means to promote such inequality-rectifying behavior.

On the other hand, only within-group comparisons increased children's selection of girls as teammates in a STEM competition. Replicating prior work using a teammate selection paradigm (Bian et al., 2018), this condition effect emerged in the last rounds of teammate selections after children expressed own-gender preferences in their first two selections. This result indicates an important lever to increase girls' representation in STEM, specifically, presenting children with within-group comparisons may strengthen *both* girls' and boys' desire to include girls by demonstrating girls' ability to perform. Indeed, in competitive contexts, children may be especially concerned about girls' capacity to contribute to a team's performance. Children's focus on this concern may help to explain why the between-group comparison failed to increase children's inclusion of girls: although children learned that girls were given fewer training opportunities, they were still uncertain about their abilities. Instead, observing girls' success in the STEM domain presents direct evidence speaking to their abilities to succeed.

An important next step for future research would be to examine whether *combining* the two structural approaches may be especially powerful for mitigating early STEM inequalities. For example, a full intervention could present evidence that boys historically had greater STEM opportunities than girls, and that once girls were provided with these same level of opportunities, girls' STEM participation and achievement became similar to boys'. We contend that our framework is also relevant for addressing other pervasive societal inequalities beyond gender, such as disparities by race or social class (see Amemiya et al., 2023). Prior research has found that children's causal explanations for inequality relate to racial biases (Mandalaywala, Ranger-Murdock, Amodio, & Rhodes, 2019; Rizzo, Britton, & Rhodes, 2022), and that it is difficult to increase children's structural understanding when explaining real-world inequalities (Mistry, Brown, Chow, & Collins, 2012). We propose that presenting *both* between-group and within-group comparisons may be an especially powerful way to overturn strongly-entrenched stereotypes about these disparities. Another important future direction will be to examine whether our effects, which were observed among a predominantly high-SES sample who likely have greater access to STEM learning opportunities, replicate with a more socioeconomically diverse sample.

Taken together, our results offer new theoretical insights into the different approaches that can promote early structural reasoning and rectification of inequality, and suggest that a multi-faceted approach that includes multiple types of structural information may be the most effective.

Author note

This research and writing of this article were supported by NSF awards granted to Jamie Amemiya (Award: 2203810) and Lin Bian (Award: 2145809), and to both Jamie Amemiya and Lin Bian (Award: 2317714).

CRediT authorship contribution statement

Jamie Amemiya: Writing – original draft, Visualization, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Lin Bian:** Writing – review & editing, Resources, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Data availability

All study materials, data, and code are available at the following link: https://osf.io/gx7a5/

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2024.105740.

References

Amemiya, J., Mortenson, E., Ahn, S., Walker, C. M., & Heyman, G. D. (2021). Children acknowledge physical constraints less when actors behave stereotypically: Gender stereotypes as a case study. *Child Development.*. https://doi.org/10.1111/cdev.13643

Amemiya, J., Mortenson, E., Heyman, G. D., & Walker, C. M. (2023). Thinking structurally: A cognitive framework for understanding how people attribute inequality to structural causes. *Perspectives on Psychological Science, 18*, 259–274. https://doi.org/10.1177/17456916221093593

Bian, L. (2022). Gender stereotypes and education. In D. P. VanderLaan, & W. I. Wong (Eds.), *Gender and sexuality development* (pp. 255–275). Springer International Publishing. https://doi.org/10.1007/978-3-030-84273-4_9.

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

Bian, L., Leslie, S.-J., & Cimpian, A. (2018). Evidence of bias against girls and women in contexts that emphasize intellectual ability. *American Psychologist, 73*(9), 1139–1153. https://doi.org/10.1037/amp0000427

Christie, S., & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development, 11*(3), 356–373. https://doi.org/10.1080/15248371003700015

Cimpian, A., & Salomon, E. (2014). The inherence heuristic: An intuitive means of making sense of the world, and a potential precursor to psychological essentialism. *Behavioral and Brain Sciences, 37*(5), 461–480. https://doi.org/10.1017/S0140525X13002197

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math-gender stereotypes in elementary school children: gender stereotypes. *Child Development, 82*(3), 766–779. https://doi.org/10.1111/j.1467-8624.2010.01529.x

Danovitch, J. H., Mills, C. M., Sands, K. R., & Williams, A. J. (2021). Mind the gap: How incomplete explanations influence children's interest and learning behaviors. *Cognitive Psychology, 130*, Article 101421. https://doi.org/10.1016/j.cogpsych.2021.101421

Dunham, Y., Baron, A. S., & Carey, S. (2011). Consequences of "minimal" group affiliations in children: minimal group affiliations in children. *Child Development, 82*(3), 793–811. https://doi.org/10.1111/j.1467-8624.2011.01577.x

Elenbaas, L., Rizzo, M. T., & Killen, M. (2020). A developmental-science perspective on social inequality. *Current Directions in Psychological Science, 29*(6), 610–616. https://doi.org/10.1177/0963721420964147

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Goddu, M. K., & Gopnik, A. (2020). Learning what to change: Young children use "difference-making" to identify causally relevant variables. *Developmental Psychology, 56*(2), 275–284. https://doi.org/10.1037/dev0000872

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin, 138*(6), 1085–1108. https://doi.org/10.1037/a0028044

Hussak, L. J., & Cimpian, A. (2015). An early-emerging explanatory heuristic promotes support for the status quo. *Journal of Personality and Social Psychology, 109*(5), 739–752. https://doi.org/10.1037/pspa0000033

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

Mandalaywala, T. M., Ranger-Murdock, G., Amodio, D. M., & Rhodes, M. (2019). The nature and consequences of essentialist beliefs about race in early childhood. *Child Development, 90*(4), e437–e453. https://doi.org/10.1111/cdev.13008

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. (2016). Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. *Journal of Educational Psychology, 108*(3), 424–437. https://doi.org/10.1037/edu0000061

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

McGuire, L., Hoffman, A. J., Mulvey, K. L., Hartstone-Rose, A., Winterbottom, M., Joy, A., … 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

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

Mills, C. M., Sands, K. R., Rowles, S. P., & Campbell, I. L. (2019). "I want to know more!": children are sensitive to explanation quality when exploring new information. *Cognitive Science, 43*(1). https://doi.org/10.1111/cogs.12706

Mistry, R. S., Brown, C. S., Chow, K. A., & Collins, G. S. (2012). Increasing the complexity of young adolescents' beliefs about poverty and inequality: Results of an 8th grade social studies curriculum intervention. *Journal of Youth and Adolescence, 41*(6), 704–716. https://doi.org/10.1007/s10964-011-9699-6

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

Peretz-Lange, R., & Muentener, P. (2019). Verbal framing and statistical patterns influence children's attributions to situational, but not personal, causes for behavior. *Cognitive Development, 50*, 205–221. https://doi.org/10.1016/j.cogdev.2019.05.005

Peretz-Lange, R., Perry, J., & Muentener, P. (2021). Developmental shifts toward structural explanations and interventions for social status disparities. *Cognitive Development, 58*, Article 101042. https://doi.org/10.1016/j.cogdev.2021.101042

Pesowski, M. L., Denison, S., & Friedman, O. (2016). Young children infer preferences from a single action, but not if it is constrained. *Cognition, 155*, 168–175. https://doi.org/10.1016/j.cognition.2016.07.004

Rhodes, M., Leslie, S.-J., Yee, K. M., & Saunders, K. (2019). Subtle linguistic cues increase girls' engagement in science. *Psychological Science, 30*(3), 455–466. https://doi.org/10.1177/0956797618823670

Rhodes, M., & Mandalaywala, T. M. (2017). The development and developmental consequences of social essentialism: Social essentialism. *Wiley Interdisciplinary Reviews: Cognitive Science, 8*(4), Article e1437. https://doi.org/10.1002/wcs.1437

Rizzo, M. T., Britton, T. C., & Rhodes, M. (2022). Developmental origins of anti-black bias in white children in the United States: Exposure to and beliefs about racial inequality. *Proceedings of the National Academy of Sciences, 119*(47), Article e2209129119. https://doi.org/10.1073/pnas.2209129119

Rizzo, M. T., Elenbaas, L., & Vanderbilt, K. E. (2020). Do children distinguish between resource inequalities with individual versus structural origins? *Child Development, 91*(2), 439–455. https://doi.org/10.1111/cdev.13181

Rutland, A., & Killen, M. (2017). Fair resource allocation among children and adolescents: The role of group and developmental processes. *Child Development Perspectives, 11*(1), 56–62. https://doi.org/10.1111/cdep.12211

Schmader, T. (2023). Gender inclusion and fit in STEM. *Annual Review of Psychology, 74*(1), 219–243. https://doi.org/10.1146/annurev-psych-032720-043052

Schulz, L. E., & Sommerville, J. (2006). God does not play dice: Causal determinism and Preschoolers' causal inferences. *Child Development, 77*(2), 427–442. https://doi.org/10.1111/j.1467-8624.2006.00880.x

Seiver, E., Gopnik, A., & Goodman, N. D. (2013). Did she jump because she was the big sister or because the trampoline was safe? Causal inference and the development of social attribution. *Child Development, 84*(2), 443–454. https://doi.org/10.1111/j.1467-8624.2012.01865.x

Shaw, A., & Olson, K. R. (2012). Children discard a resource to avoid inequity. *Journal of Experimental Psychology: General, 141*(2), 382–395. https://doi.org/10.1037/a0025907

Tenenbaum, H. R., & Leaper, C. (2003). Parent-child conversations about science: The socialization of gender inequities? *Developmental Psychology, 39*(1), 34–47. https://doi.org/10.1037/0012-1649.39.1.34

Vasilyeva, N., Gopnik, A., & Lombrozo, T. (2018). The development of structural thinking about social categories. *Developmental Psychology, 54*(9), 1735–1744. https://doi.org/10.1037/dev0000555

Yang, X., Naas, R., & Dunham, Y. (2021). Testing the limits of structural thinking about gender. *Developmental Science.*. https://doi.org/10.1111/desc.13169