



RESEARCH ARTICLE

Testing the limits of structural thinking about gender

Xin Yang¹ | Ragnhild Naas² | Yarrow Dunham¹¹ Department of Psychology, Yale University, New Haven, Connecticut, USA² Department of Psychology, University of Bath, Bath, UK**Correspondence**

Xin Yang, Department of Psychology, Yale University, 2 Hillhouse Ave., New Haven, CT 06511, USA.

Email: xin.yang@yale.edu

Accepted at Developmental Science

Abstract

When seeking to explain social regularities (such as gender differences in the labor market) people often rely on internal features of the targets, frequently neglecting structural and systemic factors external to the targets. For example, people might think women leave the job market after childbirth because they are less competent or are better suited for child-rearing than men, thereby eliding socio-cultural and economic factors that disadvantage women. Across two studies (total $N = 192$) we probe 4- and 5-year-olds and 7- and 8-year-olds' internal versus structural reasoning about gender. We explore the evaluative and behavioral implications of this reasoning process with both novel gendered behaviors that were experimentally created and familiar gendered behaviors that exist outside of a lab context. We show that children generate more structural explanations, evaluate the structural explanation more positively, expect behaviors to be more mutable, and evaluate gender non-conforming behaviors more positively when structural cues are provided. However, we also show that such information may be of limited effectiveness at reducing pre-existing group-based discriminatory behaviors: children continue to report less willingness to affiliate with peers who display non-conforming behaviors even in the presence of structural cues. Taken together, these results provide evidence concerning children's structural reasoning about gender categories and shed new light on how such reasoning might affect social evaluations and behavioral intentions.

KEYWORDS

behavioral intentions, essentialism, evaluations, explanations, gender, structural thinking

1 | INTRODUCTION

Why do more women than men leave the job market after the arrival of children? Why are many societies characterized by large wealth disparities? To explain these regularities, people frequently look to internal factors, that is, factors intrinsic to the targets in question (e.g., Cimpian & Salomon, 2014). For example, women might be thought to leave the job market because they are less competent or better suited for child-rearing than men. Perhaps people differ in wealth because some people are smarter or work harder than others. A less common explanatory strategy focuses instead on factors outside the targets in question (e.g., Haslanger, 2016). For example, women might leave the job market because the socioeconomic system puts them at a

disadvantage, and people might acquire different levels of wealth for historical or systemic reasons that disproportionately benefit some and disadvantage others. In this paper, we focus on one subtype of external explanation, namely *structural explanation*. Structural explanations consider the larger structure that targets are situated in (e.g., social systems) and structural constraints that shape behaviors (e.g., policies that advantage or disadvantage some people over others; Haslanger, 2016).

Using gender as a case study, we explore how structural thinking about social categories develops in childhood. In so doing we follow recent work finding that when young children are presented with plausible cues to a structural explanation they do endorse structural reasoning, at least with respect to novel behaviors that are not previously familiar or internalized as gender-stereotypical (i.e., girls play



with Yellow-Ball more than Green-Ball because Yellow-Ball is more available in their classroom; Vasilyeva et al., 2018). However, being able to endorse structural reasoning about novel behaviors that were experimentally created does not necessarily guarantee an equivalent ability to think about familiar behaviors that exist outside of a lab context in a structural way. We argue that this is an important consideration given how much past work has employed novel behaviors (see also Peretz-Lange et al., 2021; Van Wye et al., 2020). Notably, many familiar gendered behaviors are already entrenched as stereotypes (e.g., boys do not cry, girls play with dolls), and may already pull for salient internal or essentialist explanations (e.g., girls play with dolls because girls have a caring nature). Such established views might hamper the endorsement of a competing structural explanation (e.g., dolls are more available in girls' environment). Therefore, structural reasoning about familiar, everyday behaviors could be more difficult to induce in an experimental context, warranting further investigation.

More importantly, one aim of work in this area is to cultivate structural reasoning in the children of the context more routinely encounter in their daily lives, namely familiar behaviors that are the locus of discrimination and the usual target of intervention efforts. We thus probed a potential boundary condition of structural thinking by shifting the focus from novel behaviors (Study 1) to familiar, everyday behaviors (Study 2). Novel to the present inquiry, we explored the downstream consequences of internal versus structural reasoning in terms of its effect on evaluations of and behavioral intentions directed towards non-conforming members, thereby hoping to learn more about the promise of structural interventions in mitigating the stigmatization of gender non-conforming peers.

1.1 | Non-structural reasoning about social categories

The vast literature on social category reasoning generally documents people's tendency to focus on inherent, non-structural features rather than external, structural features. One influential theoretical approach is based on the notion of psychological essentialism (Gelman, 2004), that is, representing categories in terms of unobservable underlying "essences" that causally determine their features. Relatedly, the inheritance heuristic perspective (Cimpian & Salomon, 2014) posits that people explain observed patterns through an implicit cognitive process that directs attention to predominately inherent features (e.g., girls wear pink because pink is a delicate, girlish color). These and related approaches (Prasada & Dillingham, 2009; Ross, 1977) emphasize that people reason about social categories in terms of their internal, putatively stable features.

Experimental evidence suggests that the reliance on internal or essentialist reasoning is early emerging. For example, children from the preschool age view category properties as innate, inductively rich, and reflecting underlying structures (for a review, see Gelman, 2004). Such tendencies often persist even in the presence of information that suggests otherwise, such as the fact that social categories are randomly

Research Highlights

- Two studies probed children's structural reasoning about gender norms and explored evaluative and behavioral implications of this reasoning process with both novel and familiar gendered behaviors.
- Children as young as 4 attend to structural cues to explain gendered behaviors and adjust their judgments about mutability and acceptability of those behaviors accordingly.
- However, they still have internalist tendencies in explaining familiar gendered behaviors and show group-based discrimination.
- These findings advance the scientific understanding of structural reasoning in children and could potentially shed light on interventions to combat prejudice against non-typical members.

assigned (Yang & Dunham, 2019). Relevant to our current inquiry, gender is a salient social category that has been associated with essentialist thought (e.g., Haslam et al., 2000; Prentice & Miller, 2007; for recent developmental work, see Gülgöz et al., 2019, 2021). Such essentialist thought is also associated with negative outcomes including prejudice and stereotypes (e.g., Chen & Ratliff, 2018). Further, combined with people's tendency to conflate "what it is" with "what ought to be" (Roberts et al., 2017), these forms of reasoning may cause stronger negativity towards non-typical category members because the norms they violate are treated as innate and critical features of the category membership, perhaps especially towards ingroup members (Goldring & Heiphetz, 2020).

1.2 | Structural reasoning about social categories

When and how people also reason about social categories in structural terms is much less documented in psychological research. In a philosophical approach, Haslanger (2016) develops an account of social structure and connects it to structural explanations, while sociologists have long emphasized structural interpretations of racism and sexism (e.g., Bonilla-Silva, 1997; Merolla & Jackson, 2019; Ridgeway & Smith-Lovin, 1999). A related literature on critical consciousness has explored similar ground, investigating adolescents' structural understanding of social injustice (e.g., Diemer et al., 2016; Godfrey et al., 2019; Watts et al., 2011). Generally speaking, "social structures" refer to the social spaces individuals are situated within and constrained by. Social structures can be broad and abstract, like the wage-labor system of capitalism; they can also be local and concrete, like a particular institution (e.g., a school). Yet, to date, very little work on this topic has been conducted



with children. Following Kraus and Park (2017) we would argue that more work of this sort is needed throughout the psychological sciences.

There are several reasons why structural explanations might be more difficult to generate than internal ones. To begin with, people may preferentially attend to internal features or emphasize essences from a young age (see the internalist approaches reviewed above). People also tend to assume that everyone is responsible for their own deeds and outcomes, leading them to over-emphasize agency, motivation, and responsibility (Ross, 1977, 2018). Relatedly, people show the “fundamental attribution error,” the tendency to favor dispositional explanations over situational ones (Ross, 1977, 2018). Second, cues to the broader structures are often abstract, lost to history, or hidden behind other more readily available explanations. Supporting this, recent work finds that when discussing incarceration and law-breaking both children and adults favor internal and behavioral explanations, and they rarely reference societal factors or endorse structural explanations (Dunlea & Heiphetz, 2020). And in a study involving competition between novel social groups, even when both internal (differences in physical strength) and structural (forms of competition) factors were equally present, children viewed the losing group as inferior rather than structurally disadvantaged except when structural factors were verbally emphasized repeatedly (Peretz-Lange et al., 2021). Further, for some phenomena such as social inequalities, people might be motivated to emphasize internal factors (e.g., merit) when they can be used to justify social structures that are to their benefit (see literature on system justification theory and just-world beliefs; for example, Jost et al., 2004; Rubin & Peplau, 1975). Somewhat more speculatively, structural reasoning might share some cognitive prerequisites with counterfactual and/or relational reasoning, as coming up with structural explanations might require reasoning about how things would happen differently if structures were different (and/or reasoning about relations among people in those social structures). Importantly for young children, these abilities are not well developed until after the preschool years (Rafetseder et al., 2013; for a review, see Rafetseder & Perner, 2014; but see Nyhout & Ganea, 2019). Taken together, these all speak to the possibility that young children may endorse non-structural explanations more frequently than structural ones.

We are aware of two recent studies that experimentally manipulate structural versus non-structural cues for observed differences and probe children's own explanations (e.g., Peretz-Lange et al., 2021; Vasilyeva et al., 2018). These studies suggest that with salient information about the underlying structure of novel, previously *unfamiliar* and non-stereotyped behaviors, children endorse structural reasoning more compared to when such information is absent. They also reveal important developmental changes, with stronger forms of structural reasoning observed in adults than older and younger children (they included 3–4 year and 5–6 year, or 5–6 year and 9–10 year). What remains unknown, and important to the current inquiry concerns whether structural reasoning about familiar, presumably more stereotyped behaviors in children's everyday experiences can also be induced in a similar experimental setting. If so, it opens the door to potential interventions targeting the downstream consequences of structural reasoning in the real world, which we discuss in more detail below.

1.3 | Can structural reasoning influence evaluation and treatment of others?

More important, little is known about whether structural reasoning induced from simple experimental manipulations goes on to affect children's evaluation and treatment of others. As shown in a separate but related line of work, children from a young age negatively evaluate those who deviate from group-typical behaviors (hereafter non-conforming peers; see Gülgöz et al., 2018), and actively enforce norms through spontaneous protests (for a review, see Schmidt & Tomasello, 2012). These patterns might be especially pronounced when the groups or categories are thought of in a more essentialized manner, which could further lead to negative consequences like prejudice and discrimination (Chen & Ratliff, 2018), as also alluded to above. But will evaluation and treatment of others grow more positive in the presence of structural cues?

Here we consider two possibilities. First, reflecting the potential value of structural explanation, it is possible that when children endorse structural cues showing that social norms are not determined by features of the person or category (but rather the surrounding structural forces) they also come to understand that non-conforming individuals should not be judged negatively. For instance, once people endorse the view that women have been historically disadvantaged by the labor market and other systemic forces, they would more likely appreciate women who manage to climb up the social ladder rather than discriminate against them. If so, promoting structural reasoning could lead to promising social changes. Indeed, there is some evidence for this in some other domain: children rate norm violations and incarcerations more positively when they are caused by situations and societal factors rather than traits or behaviors (Dunlea & Heiphetz, *in press*; Van Wye et al., 2020), and children think inequality is more unfair when it is caused by gender discrimination or other external constraints than when it is caused by difference in merit (Hussak & Cimpian, 2015; Rizzo et al., 2018; Rizzo & Killen, 2020). These studies together suggest that structural reasoning (if successfully induced) influences children's reasoning and evaluations of others in a positive way—reducing negativity toward norm violators and promoting fairness.

Second and less optimistically, children's endorsement of structural reasoning and their evaluations could diverge: they understand the structural causes of norms, but still negatively evaluate and sanction non-conforming individuals, for example merely because they are non-conforming. This would imply that negativity towards non-conformity is difficult to mitigate and that structural information is not on its own sufficient to do so. This is perhaps because children do not integrate explanatory frameworks and normative judgments, or that they assume the “structure” must have been there for a good reason (and therefore they should conform to the norms more; see Eidelman & Crandall, 2014). The present work examines these possibilities.

1.4 | The present study

Two studies experimentally manipulate different structural cues underlying gender norms (in this case gendered play norms) and



measure children's structural reasoning and their evaluation and treatment of gender non-conforming peers. Study 1 is a pre-registered close replication and extension of Vasilyeva et al. (2018) with the addition of a new measure to assess how children evaluate conforming and non-conforming peers. Study 2 further explore whether structural reasoning can also be induced with presumably more gender-stereotyped behaviors that more clearly reflect children's everyday experiences (i.e., girls mostly play with pink dolls and boys with blue trucks). Because of familiarity and existing beliefs (that are essentialist or at least non-structural), it provides us with a potential boundary condition of probing the limits of structural reasoning. Besides evaluations, Study 2 also investigates children's *behavioral intentions* (whether they choose to affiliate with non-conforming peers); these two critical attitudinal and behavioral choice measures probe the potential consequences of structural versus non-structural reasoning. Unlike most past work on this topic that either did not probe the consequences of structural reasoning (Vasilyeva et al., 2018) or only probed evaluative judgments (Dunlea & Heiphetz, *in press*; Peretz-Lange et al., 2021; Van Wye et al., 2020), our work examined both evaluative judgments and behavioral intentions, providing a test of the effects of structural cues on a wider range of outcomes. As stated above, our goal is to test whether structural reasoning is impactful on attitudes and behavioral intentions even when pre-existing beliefs are strong, as in many cases outside of laboratory settings where researchers and practitioners seek to reduce the negative consequences of stereotyping. We conducted these studies with 4- to 5-year-olds and 7- to 8-year-olds, ages that in prior work were linked to important developmental shifts in structural thinking (Peretz-Lange et al., 2021; Vasilyeva et al., 2018). Building on these findings as well as the developmental patterns of its potential prerequisites reviewed above (e.g., counterfactual and relational reasoning; Rafetseder & Perner, 2014; Rafetseder et al., 2013), we hypothesized that 7- to 8-year-olds would show stronger forms of structural thinking, and stronger evaluative consequences of it, than 5- to 6-year-olds.

2 | STUDY 1

We pre-registered this study at <https://aspredicted.org/3wq8i.pdf>. All materials, data, and analysis code to replicate all findings and create all figures for this study and Study 2 can be found in online supplemental materials at: https://osf.io/6kx4e/?view_only=33413fde2df346328f32e0bc5e72d11e.

2.1 | Method

2.1.1 | Participants

There were 48 four- to five-year-olds ($M = 5.05$, $SD = .64$, range 4.03–5.99, 24 females, 24 males) and 48 seven- to eight-year-olds ($M = 7.91$, $SD = .58$, range 7.01–9.00, 24 females, 24 males). Among the 71% ($n = 68$) participants whose parents reported race information,

there were 45 White participants, 8 Hispanic/Latinx participants, 8 Multiracial participants, 6 Asian participants, and 1 Black participant. This pre-registered sample size was determined because a power analysis indicated that at least $n = 89$ total was required to detect a small to medium effect with >80% power (linear multiple regression in a random model, two tails, up to 10 predictors, $\alpha = 0.05$, $\rho^2 = 0.2$; using G*Power). In each age group, half of the participants were assigned to the structural condition in a counterbalanced manner and the other half were in the non-structural condition. An additional 10 children were tested but excluded from data analyses due to parent or sibling interference ($n = 2$), failure to complete the study ($n = 2$), failure to pass comprehension check questions ($n = 3$), or improper consent procedure ($n = 3$). Participants were tested in the lab, at local museums, at festivals, or at a summer camp by the first author or trained research assistants. Most participants came from middle-class backgrounds. Both studies reported in this paper were approved by Yale University Institutional Review Boards, project title "Development of Social Category Knowledge," protocol #1305012100. Written parental consent was obtained in advance of all testing; children also provided verbal assent prior to beginning the procedures.

2.1.2 | Materials, design, and procedure

We adapted the same illustrated storybook (presented on a laptop using Microsoft PowerPoint) from Vasilyeva et al. (2018). As described above, participants heard a story where girls and boys go to different classrooms and play two different novel games, Yellow-Ball or Green-Ball, by throwing a pebble into the yellow or the green bucket (they later played the game that matched the color of the bucket in which they threw the pebble). The detailed story and visual stimuli are provided in online supplemental materials. In the structural condition, the classrooms were set up such that it was easier for students to end up engaging in one of the games (one bucket was larger than the other bucket), thus placing a structural constraint on students' behaviors, while in the non-structural condition there was no such structural constraint (see Figure 1). Note that we did not explicitly emphasize internal cues (e.g., gender preferences) in the non-structural condition. The only difference between the two conditions was whether the two buckets were of different or similar sizes. The logic is that in the absence of structural constraints (i.e., different bucket sizes) observed behavioral patterns are more likely to result from internal factors such as personal choice, while in the presence of structural factors an optimal reasoner should be less likely to make such attributions but instead attend to structural constraints. We asked participants memory check questions about bucket sizes and about the games most girls or boys played in the story. Children were never provided with any explanations, and they needed to reason about explanations themselves. Three participants failed the memory check and were excluded from data analyses. Next, we administered a series of measures targeting internal versus structural construal of the property-category association and subsequent norm enforcement¹.

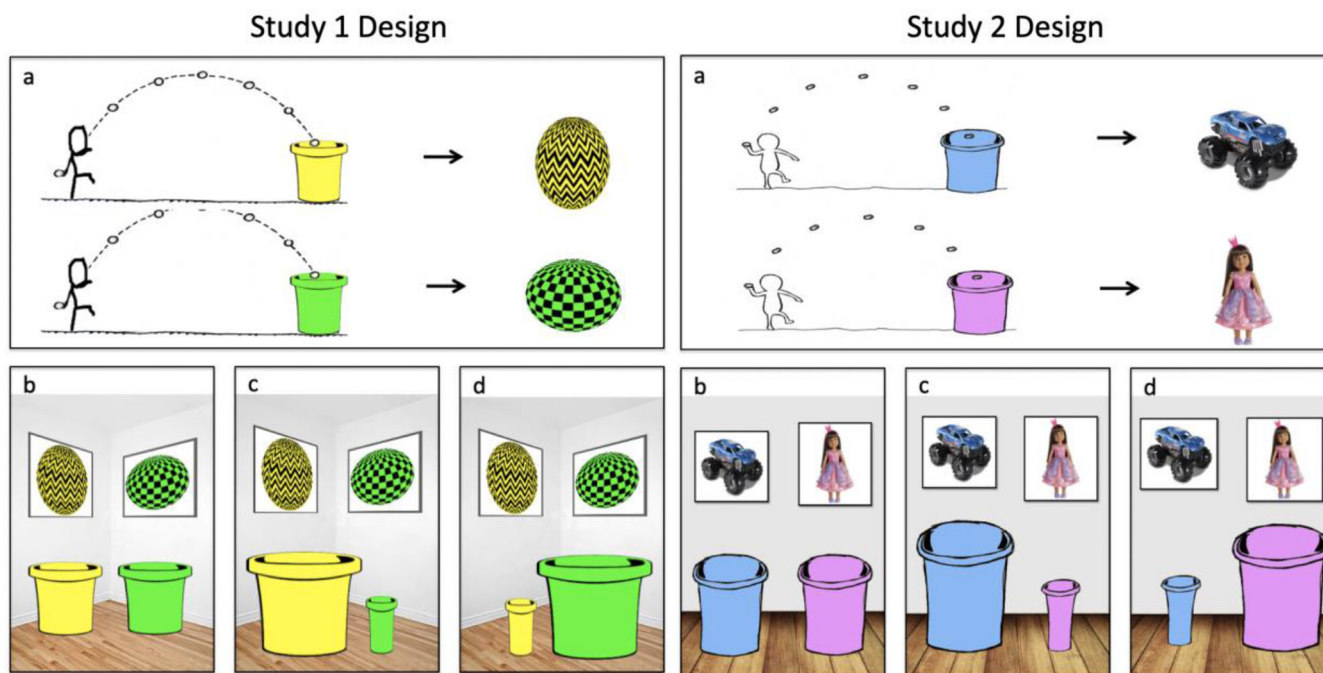


FIGURE 1 Study 1 (left) and 2 (right) Design: (a) Illustrations of the pebble-throwing procedure determining which game each student played in the story; (b) Two buckets in the non-structural condition; (c and d) Two buckets in the structural condition. See the online article for the color version of this figure. Study 1 materials were borrowed with permission from Vasilyeva et al. (2018)

2.1.3 | Measures

Participants completed three sets of measures in a fixed order: *explanations*, *mutability judgments*, and *evaluations*. There were two versions of scripts to counterbalance which game was played by most girls or boys. In what follows, we describe test measures in one of the versions. In this version, most girls played Yellow-Ball while most boys played Green-Ball. Correspondingly, in the structural condition, girls' classroom featured a bigger yellow bucket while the boys' classroom featured a bigger green bucket (in the non-structural condition, the two buckets were of the same size in both girls' and boys' classrooms).

Explanations. We first asked an open-ended question on why girls in the girls' classroom play Yellow-Ball a lot at their school. Then, regardless of their answers, we told them about three different explanations offered by three puppets. The explanations were "because girls like playing Yellow-Ball" (internal explanation), "because in the girls' classroom, it's easier to throw a pebble in the yellow bucket" (structural explanation), and "because they got sprinkled with water" (incidental explanation stating an irrelevant fact from the cover story; included addressing the possibility those young children might rate all explanations very positively, e.g., Amsterlaw, 2006; following Vasilyeva et al., 2018). Participants rated how good these explanations were one by one following a two-step, four-point thumb scale (from very bad to very good). The internal and structural explanations were presented in counterbalanced order with the incidental explanation presented in the end. After they evaluated all three explanations, those who gave the same evaluation ratings for the internal and structural explanations ($n = 41$) were asked to indi-

cate which explanation was better in a forced-choice manner². Finally, we asked participants a new explorative open-ended question on why the classrooms were set up this way ("Why do classrooms have same-sized/different-sized buckets?"), probing their explanations further.

Mutability judgments. We told participants that the school rule just changed and girls could go to the boys' classroom and vice versa, and introduced them to a girl named "Suzy" whose parents transferred her to the boys' classroom where most boys played Green-Ball. Thus, in the structural condition, the buckets switched sizes for Suzy, implying a new structural constraint that now favored Green-Ball. In the non-structural condition, however, the two buckets were of the same sizes in both classrooms, posing no structural constraint. We first made sure participants successfully recalled which game Suzy would play if the pebble went into the yellow or the green bucket (all participants passed these questions), and then as the test question, we asked them to predict which game they thought Suzy would play and how sure they were of their prediction (a little sure or very sure).

Evaluations. We added this new measure to examine possible consequences of internal versus structural thinking in children's evaluations of gender-conforming and non-conforming behaviors. We first made sure participants successfully recalled the games girls and boys played most often (they all passed these questions). Then, as the test question, we introduced participants to two new girls, one at a time, who also transferred to the boys' classroom (in a counterbalanced order). One girl played the game that most boys played, thus violating the gender norm (non-conforming behavior), while the other girl stuck with the gender norm (conforming behavior). We asked participants whether

TABLE 1 Coding scheme and example answers for the first open-ended explanation question in Study 1 (“Why do girls play Yellow-Ball/Green-Ball a lot?”), presented by condition and explanation types

Condition	Coded explanation types		
	Internal explanation (mentioning category members' properties)	Structural explanation (mentioning the accessibility of games)	Other (restatements, “I don’t know,” irrelevant/classifiable answers)
Non-structural	“Because that’s a girls’ color” “Because girls like yellow/green balls”	“Because It’s much easier to aim for”	“It’s equal for all people” “So it would be easier”
Structural	“Because yellow/green is their favorite color” “Because they like it”	“The bucket is easier to get in” “Because the yellow/green bucket is bigger”	“Because they have long hair” “Because they want to play outside all day”

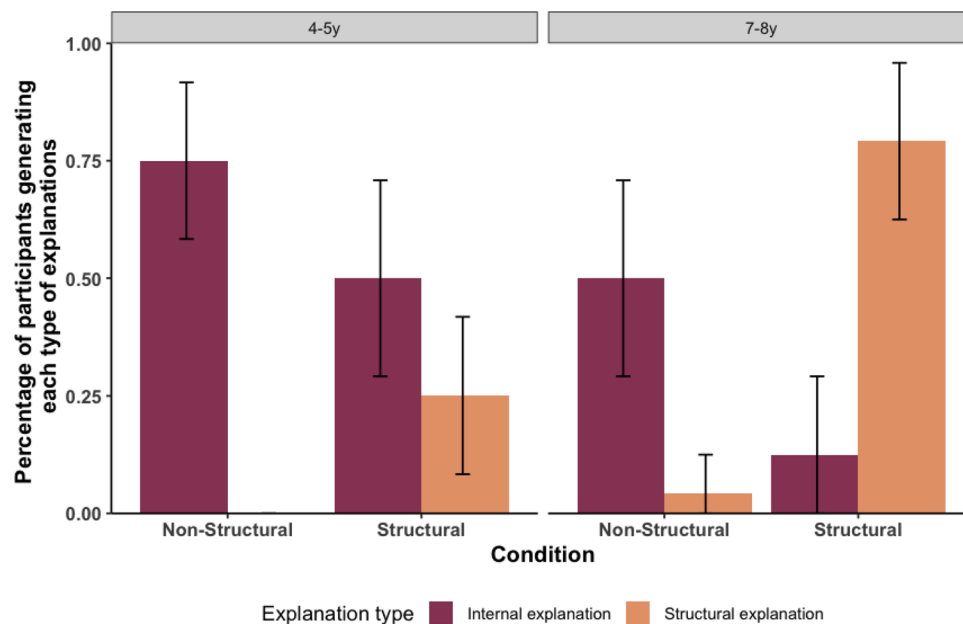


FIGURE 2 Study 1: Distribution of internal and structural explanations for the first open-ended explanation question (“Why do girls play Yellow-Ball/Green-Ball a lot?”), as a function of condition and age group. Error bars represent 95% bootstrapped confidence intervals. Percentages in each condition did not add up to one because some of the responses were not coded as internal or structural. Note that in the non-structural condition, four- to five-year-olds did not generate any structural explanations

they thought each behavior was right or wrong, followed by a scale (a little right/wrong or very right/wrong).

2.2 | Results

Unless otherwise noted, preliminary analyses revealed no effect of order (mentioning which explanation, behavior, or character first) or gender, so these factors were not discussed further. Unless otherwise noted, all analyses were carried out in the same way as specified in the pre-registered analysis plan.

2.2.1 | Explanations

For the first open-ended explanation question, “why do girls play (e.g.) Yellow-Ball a lot at their school,” participants’ answers were coded into

three distinct categories, internal, structural, and other (see Table 1 for coding scheme). The majority of the answers (87.5%) were coded directly from video by two trained independent coders (agreement rate 98%). The rest were coded from experimenter notes by the first author because we did not get permission for videotaping or video quality was poor.

As expected, explanations differed between conditions, with more structural explanations produced in the structural condition, as well as between two age groups, with older children showing stronger structural reasoning. Chi-square tests comparing explanation type distributions as a function of condition were significant, $\chi^2(2, N = 96) = 30.39$, $p < 0.001$ (for all participants); with stronger effects in older children, $\chi^2(2, N = 48) = 27.83$, $p < 0.001$, than in younger children, $\chi^2(2, N = 48) = 7.20$, $p = 0.03$. As more clearly shown in Figure 2, the distribution of the two critical explanation types (internal vs. structural explanations) was also affected by condition. Similarly, Fisher’s exact tests (non-pre-registered) comparing this distribution as a function of



condition were significant, $OR = 47.27$, $p < 0.001$ (for all participants); again with stronger effects in older children: $OR = 61.11$, $p < 0.001$ than in younger children: $p = 0.02$ (Odds Ratio was not estimable because children in the non-structural condition made 100% internal explanations). Results for the open-ended explanation question on the classroom setup are provided in supplemental materials.

For explanation evaluations, we fit a linear mixed-effects model predicting evaluation ratings (from 1 = very bad to 4 = very good) as a function of explanation type, condition, age group, and their interactions, controlling for participant gender, color (most girls play Yellow-Ball or Green-Ball), and order, with trials clustered within participants.

A similar effect of condition was also found on explanation evaluations, that is, evaluating the structural explanation more positively in the structural condition; however, this effect was only significant in older children. There was a marginally significant three-way interaction ($F(2, 184) = 3.04$, $p = 0.0501$), suggesting different patterns of results for the two age groups. To more clearly reveal patterns for each age group, we decomposed the model for each age group, as specified in our pre-registration. For 7-8-year-olds, we found a significant explanation type by condition interaction (see Figure 3), $F(2, 92) = 8.46$, $p < 0.001$: they were sensitive to structural cues, rating the structural explanation higher, $B = 0.75$, 95% CI [0.20, 1.30], $p = 0.008$, $R^2 = 0.05$, and the internal explanation lower, $B = -0.67$, 95% CI [-1.21, -0.12], $p = 0.02$, $R^2 = 0.04$, in the structural condition than in the non-structural condition. The condition difference for the incidental explanation was not significant, $p = 0.46$ ³. For 4-5-year-olds, there was only a significant effect of explanation type, $F(2, 94) = 5.44$, $p = 0.006$. Regardless of condition, they rated both internal and structural explanations as better than the incidental explanation (see Figure 3), $B = 0.54$, 95% CI [0.21, 0.87], $p = 0.002$, $R^2 = 0.05$, and $B = 0.40$, 95% CI [0.06, 0.73], $p = 0.02$, $R^2 = 0.03$, respectively, but they did not differentially evaluate the former two explanations, $p = 0.39$.

2.2.2 | Mutability judgments

We fit a linear model predicting mutability (from 1 = low mutability, i.e., for sure Suzy would play Yellow-Ball, just as most girls, to 4 = high mutability, i.e., for sure Suzy would play Green-Ball; mean-centered prior to regressions to interpret intercepts), as a function of condition, age group, and their interaction, controlling for participant gender and color.

As predicted, we found higher mutability judgments in the structural condition compared to the non-structural condition, that is, believing that properties or behaviors that are construed as structural are more mutable as structural cues changed. As shown in Figure 4, we found a significant effect of condition ($B = 0.66$, 95% CI [0.21, 1.11], $p = 0.005$, $R^2 = 0.08$) while the effect of age group and the interaction were not significant ($ps > 0.16$). Children gave higher mutability judgments in the structural condition compared to the non-structural condition. More specifically, in the structural condition, children predicted that gender non-conforming behaviors were more likely to happen, $M = 3.17$, $SD = 1.12$, $B = 0.66$, 95% CI [0.34, 0.98], $p < 0.001$, while in the non-

structural condition, they did not predict either gender-conforming or non-conforming behaviors, $M = 2.50$, $SD = 1.13$, $B = 0.00$, 95% CI [-0.32, 0.32], $p = 1.00$.

2.2.3 | Evaluations

Turning to the measure novel to our study, we computed evaluation scores by deducting the evaluations of the non-conforming character from the evaluations of the gender-conforming character (raw ratings ranged from 1 = very wrong to 4 = very right, so the scores ranged from -3 to 3). Higher scores indicate relatively more positivity towards gender-conforming behaviors compared to non-conforming behaviors. We fit a linear model⁴ predicting evaluation score (scale mid-point = 0) as a function of condition, age group, and their interaction, controlling for participant gender, color, and order (asking about the conforming or non-conforming character first).

We predicted that children would sanction gender non-conforming behaviors (i.e., evaluation scores > 0), but perhaps less so in the structural condition (i.e., lower evaluation scores). These predictions were partially confirmed. As shown in Figure 5, we found a marginally significant effect of condition, while the effect of age group and the interaction were not significant ($ps > 0.10$). Overall children scored slightly lower in the structural condition compared to the non-structural condition, $B = -0.67$, 95% CI [-1.36, 0.03], $p = 0.06$, $R^2 = 0.04$. However, children in both conditions actually judged gender non-conforming behaviors more positively than conforming behaviors (evaluation scores < 0 ; $M = -0.74$, $SD = 1.76$, $B = -0.75$, 95% CI [-1.09, -0.40], $p < 0.001$). Partially confirming our prediction, children were somewhat more lenient towards non-conforming behaviors when they were given structural cues, but unexpectedly, they did not seem to sanction gender non-conforming behaviors in this novel context.

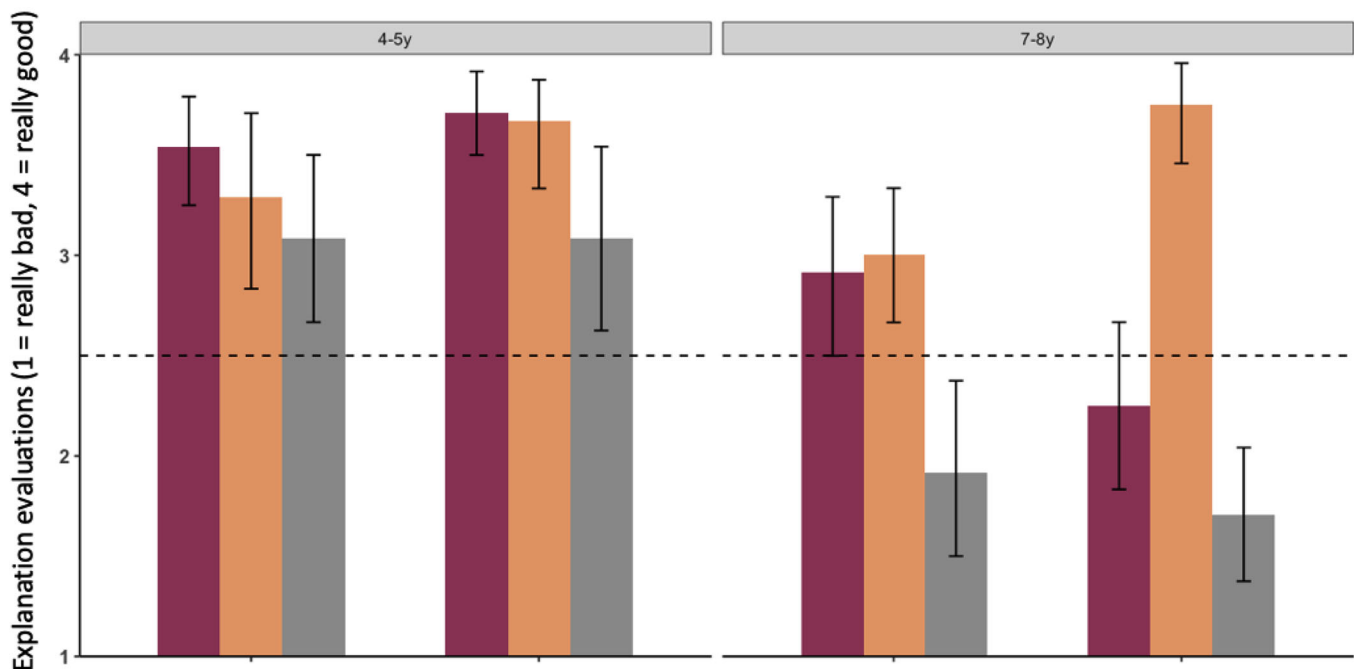
2.3 | Discussion

Replicating prior work (Vasilyeva et al., 2018), here we showed that children by ages 4 and 5 could reason about novel gendered behaviors in a structural way when provided with salient structural cues, with stronger forms of structural reasoning emerging in 7-8-year-olds. Specifically, when structural cues were available compared to when they were not, children generated more structural explanations themselves (by age 4-5), evaluated the structural explanation more positively than the internal one (by age 7-8), and thought people were more likely to change to gender non-conforming behaviors (as the structure changed in the same direction; by age 4-5).

Novel to the current inquiry, we further explored the effect of inducing structural thinking on one important potential downstream consequence of such reasoning—the extent to which children negatively evaluated gender non-conforming peers. We found that children evaluated non-conformity (somewhat) more positively when it was justified by a structural reason, suggesting that the structural framing might reduce negativity towards non-conformity. However, in this study



Study 1



Study 2

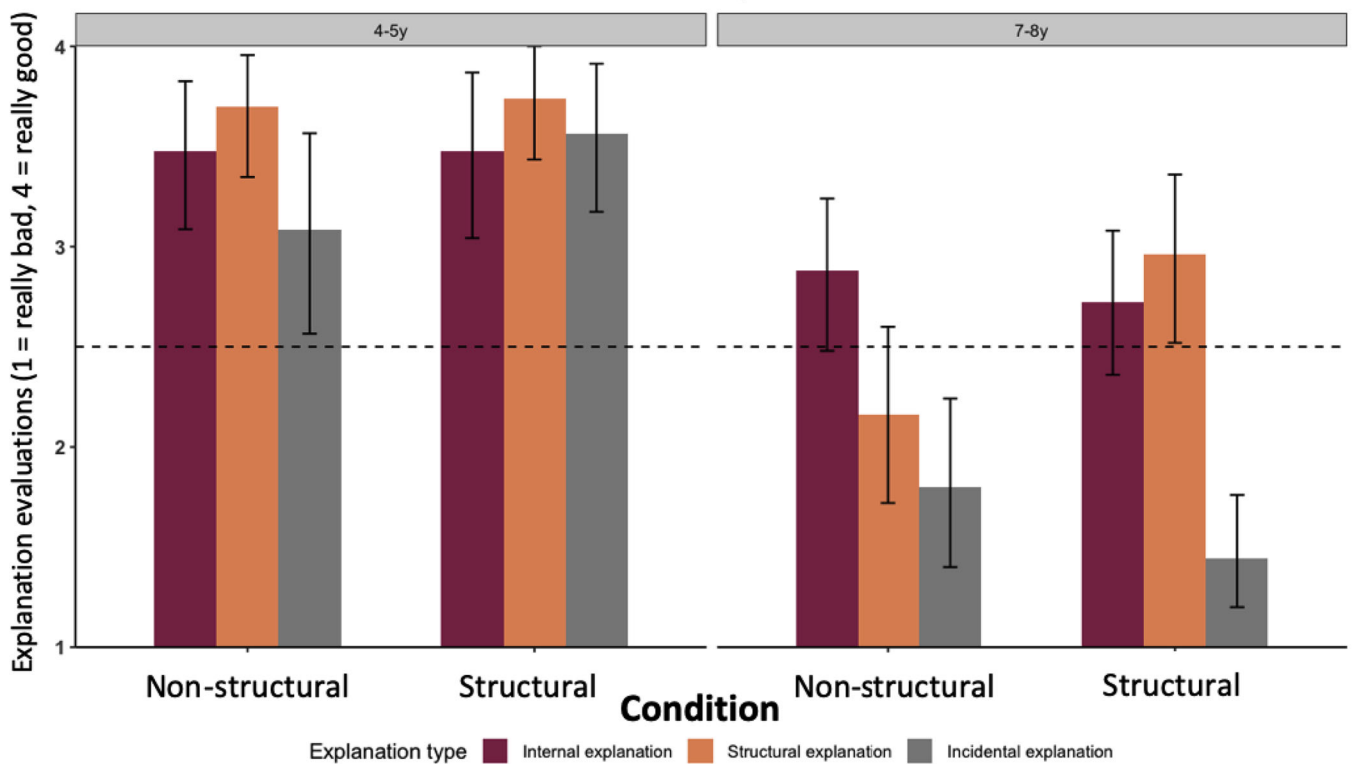


FIGURE 3 Explanation evaluations as a function of condition and age group in Study 1 (top) and Study 2 (bottom). Error bars represent 95% bootstrapped confidence intervals

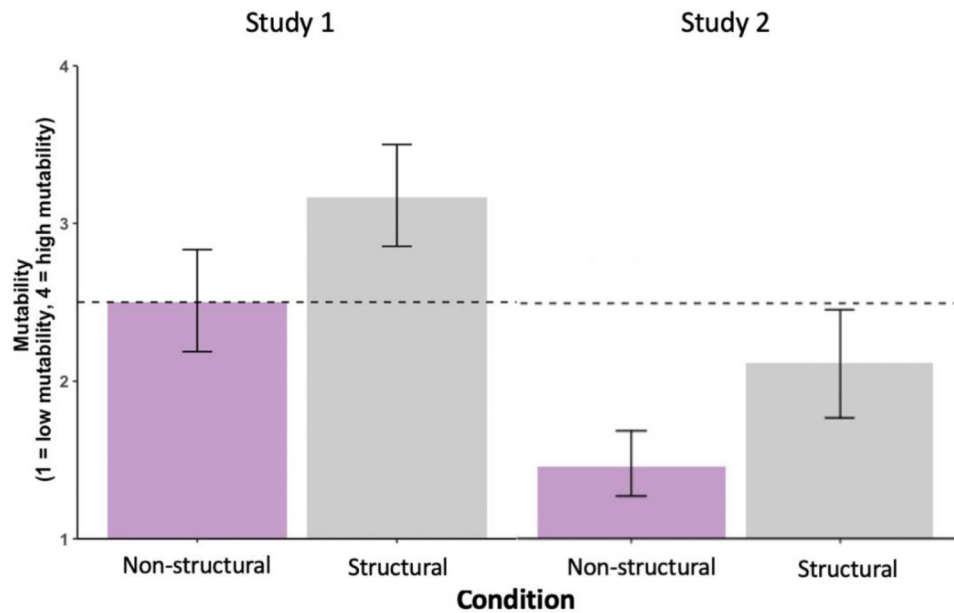


FIGURE 4 Mutability judgments (1 = low mutability, i.e., for sure this character would play the gender-conforming game, 4 = high mutability, i.e., for sure this character would play the non-conforming game), as a function of condition in Study 1 (left) and Study 2 (right). Error bars represent 95% bootstrapped confidence intervals

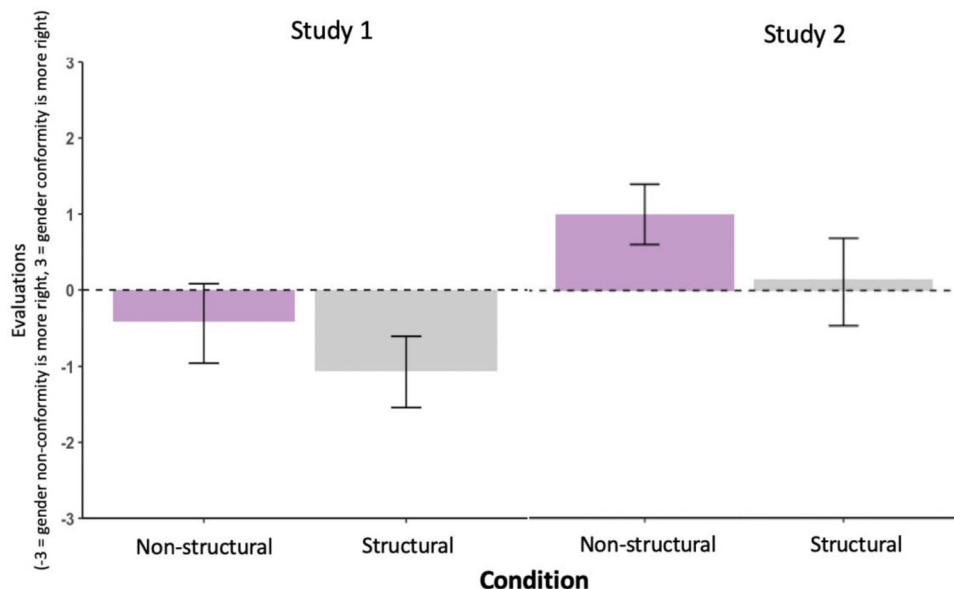


FIGURE 5 Evaluations (-3 = gender non-conformity is more right, 3 = gender conformity is more right) as a function of condition in Study 1 (left) and Study 2 (right). Error bars represent 95% bootstrapped confidence intervals

children did not seem to sanction gender non-conforming behaviors even when structural cues were absent. We suspect this emerged because, upon transferring to the new classroom, the student found themselves in the midst of the new alternative norm, and thus the same behavior that was norm-deviant with respect to the original classroom was norm-compliant in the new classroom. In other words, our design created a potential norm conflict, perhaps muddying judgments about the “non-conforming” character. We address this issue with a modified design in Study 2.

We note that children in the non-structural condition did not *evaluate* the internal explanation more positively than the structural explanation (although their open-ended explanations did shift in the direction of internal explanations). This is somewhat puzzling, and might imply that children’s endorsement of the internal explanation might not always be as strong a default as we might have assumed based on past work, at least in the case of unfamiliar gendered behaviors (cf. Vasilyeva et al., 2018). That said, the main focus here is the effect of condition, and across measures children shifted to structural



reasoning more under the structural framing than the default framing. In Study 2, we further explore these questions and also extended this work to test children's reasoning about familiar, everyday behaviors. Most familiar gendered behaviors (e.g., girls play with dolls while boys play with trucks) are stereotyped, and are more likely to invite internal reasoning and sanctioning of non-conformity (Rhodes & Mandalaywala, 2017). Investigating structural reasoning and its downstream consequences with familiar behaviors is important in and of itself, as eventually any intervention designed to mitigate biases against the non-typical has to apply to real-world cases.

3 | STUDY 2

Study 2 is pre-registered at <https://aspredicted.org/qr2ky.pdf>.

3.1 | Method

3.1.1 | Participants

There were 46 four- to five-year-olds ($M = 4.95$, $SD = 0.59$, range 4.01–5.94, 24 females, 22 males) and 50 seven- to eight-year-olds ($M = 7.98$, $SD = 0.58$, range 7.01–8.99, 25 females, 25 males)⁵. Among the 68% ($n = 65$) participants whose parents reported race information, there were 54 White participants, 4 Multiracial/Other participants, 3 Asian participants, 2 Hispanic/Latinx participants, and 2 Black participants. In each age group half of the participants were assigned to the structural condition in a counter-balanced manner and the other half were in the non-structural condition. Participants were tested in the lab, at local schools, or at local museums, by the first or second author or trained research assistants. An additional 20 children (11 females, 9 males) were pretested on study materials ("Which toy do you think Lucy/Tom will play with," contrasting a pink doll and a blue truck, on a 6-point scale of preference; 1 = prefer the non-typical toy a lot, 6 = prefer the gender-typical toy a lot; higher scores indicate greater gender-typicality). Both the girl game (pink doll) and the boy game (blue truck) were considered highly gender-typical ($M_{girl} = 5.7$, $SD_{girl} = 0.57$; $M_{boy} = 5.8$, $SD_{boy} = 0.41$).

3.1.2 | Materials, design, procedure

We used Qualtrics (on iPads or laptops) instead of PowerPoint to eliminate the need for data entry, adopted highly gender-typical stimuli (pink dolls and blue trucks; pretested to ensure strong gender-typicality), and made a few changes from Study 1, described here (for illustrations, see Figure 1).

First and foremost, we used familiar, everyday gender-typical behaviors, that is, focusing on girls playing with pink dolls and boys playing with blue trucks. We coupled gender-typical colors and gender-typical toys to create a context in which observed regularities are more readily explicable via internal or stereotyped reasoning, thus creating a

more stringent test of whether structural cues reduce reliance on internal explanations. There were two versions of scripts, gender-matched to participants such that we could probe how boys and girls reason about their own gender. Second, to avoid the potential confound of peer pressure and the new norms in the new classroom (e.g., Costanzo & Shaw, 1966; Haun & Tomasello, 2011; Walker & Andrade, 1996), here we manipulated the structural change in the structural condition not through students transferring to different classrooms but rather with buckets switching sizes in the same classroom (in the non-structural condition, buckets changed sides rather than sizes).

3.1.3 | Measures

Measures were similar to those in Study 1 except two changes. First, to more closely examine how structural thinking affects children's behaviors, we added a behavioral intention (*playmate choice*) measure after the evaluation measure. We asked participants to choose between the conforming child and the non-conforming child as their playmate ("Can you point to the kid you would rather play with?" followed by a memory check question "Do you remember what he/she played with?"). Second, we removed the open-ended explanation questions to simplify data collection and administered the forced-choice explanation question to all participants (not just those who gave same ratings for internal and structural explanations) for a more balanced design.

3.2 | Results

We conducted similar analyses as described in Study 1. Preliminary analyses revealed that there was no effect of order (mentioning which explanation, behavior, or character first) on the explanations and mutability measures, but there were order differences on the other two measures. We report these unexpected results in supplemental materials. Additionally, on the mutability judgment where we had >80% power to detect gender effects (post hoc test using the mixed model in G*Power with our sample size), we found an effect of gender ($B = -0.53$, 95% CI $[-0.92, -0.14]$, $p = 0.01$, $R^2 = 0.07$): boys predicted fewer gender non-conforming behaviors in boys than girls did for gender non-conforming behaviors in girls. Similar patterns (i.e., boys showing weaker structural reasoning than girls did about their own gender) also appeared on other measures. However, we were not powered to conclusively detect gender differences, and even more importantly our design, in which participants observed same-gender peers, does not allow us to disentangle participant and target gender effects; thus, we do not focus on this finding here. Interested readers are referred to the supplemental materials. Unless otherwise noted, all analyses described below were carried out as specified in the pre-registered analysis plan.

3.2.1 | Explanations

We predicted higher endorsement of the structural explanation in the structural condition than in the non-structural condition. We found a



significant three-way interaction involving age group, condition, and explanation type ($F(2, 184) = 4.74, p = 0.01$). For 7-8-year-olds, we found a significant type by condition interaction (see Figure 3), $F(2, 96) = 5.48, p = 0.006, R^2 = 0.04$, which was primarily driven by the increased positive ratings of the structural explanation in the structural condition compared to the non-structural condition, $B = 0.80, 95\% \text{ CI } [0.25, 1.35], p = 0.005, R^2 = 0.04$, suggesting that participants were sensitive to structural cues. However, ratings for the internal explanations did not differ between conditions, nor did ratings for the incidental explanation ($ps > 0.20$). These findings suggest that unlike in Study 1 where the internal explanation was rated less positively in the structural condition, internalist/essentialist reasoning is persistent and hard to downplay here when the behavior in question is already associated with gender. For 4-5-year-olds, similar to Study 1, there was a marginally significant effect of explanation type, $F(2, 90) = 2.43, p = 0.09$. Regardless of condition, they rated the structural explanation as better than the incidental explanation (see Figure 3), $B = 0.39, 95\% \text{ CI } [0.04, 0.74], p = 0.03, R^2 = 0.03$, but the other two pairwise comparisons were both not significant, $ps > 0.18$. These results suggest that as in Study 1, younger children did not fully differentiate these three different explanations; instead, they rated all three explanations in a positive way. The forced-choice measure (non-pre-registered) yielded largely similar results; since we did not preregister an analysis plan for this measure, we report these results in supplemental materials.

3.2.2 | Mutability judgment

We predicted higher mutability judgments in the structural condition compared to the non-structural condition, and overall lower mutability judgments than that in Study 1; results confirmed these predictions. We found a significant effect of condition ($B = 0.68, 95\% \text{ CI } [0.29, 1.07], p < 0.001, R^2 = 0.11$). As shown in Figure 4, children indeed gave higher mutability judgments in the structural condition ($M = 2.13, SD = 1.21$) compared to the non-structural condition ($M = 1.46, SD = 0.74$). Overall children predicted low mutability (predicting more gender-conforming behaviors than non-conforming behaviors), $B = -0.72, 95\% \text{ CI } [-0.91, -0.53], p < 0.001$. We again did not find a significant effect of age group ($p = 0.16$).

3.2.3 | Evaluations

With familiar gendered behaviors, we expected negativity towards non-conforming peers, but sought to investigate whether this would be reduced in the presence of structural cues. As shown in Figure 5, overall children showed negativity towards gender non-conforming peers compared to conforming peers (evaluation scores > 0), $B = 0.58, 95\% \text{ CI } [0.23, 0.94], p = 0.002$. However, this negativity was mitigated in the structural condition: there was a significant effect of condition, $B = -0.85, 95\% \text{ CI } [-1.56, -0.14], p = 0.02, R^2 = 0.06$. Importantly, under the structural framing, children were no longer biased against

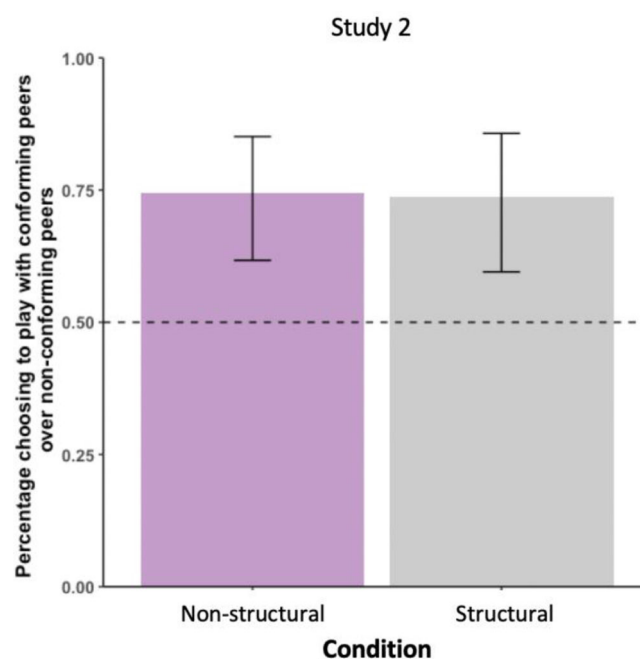


FIGURE 6 Behavioral intention (playmate forced-choice) question in Study 2 (coded 1 = choose conforming kid, 0 = choose non-conforming kid), as a function of condition. Error bars represent 95% bootstrapped confidence intervals

non-conformity (evaluation scores did not differ from 0), $p = 0.55$. The effect of age group was not significant, $p = 0.22$.

3.2.4 | Behavioral intentions

For this measure ($n = 89$; excluding participants who failed to remember what the targets played with and for whom we did not have data, following the pre-registration), we fit a binomial linear model predicting behavioral intention (1 = choose to play with the conforming kid, 0 = choose to play with the non-conforming kid) based on condition, age group, gender, and their interactions. We did not find a significant effect of condition ($p = 0.90$) or age group ($p = 0.21$): as shown in Figure 6, children regardless of conditions and age groups chose to play with conforming peers at significantly above-chance rates, $B = 1.29, 95\% \text{ CI } [0.72, 1.87], p < 0.001$. The structural framing did not mitigate children's behavioral sanctions of gender non-conformity (in terms of behavioral intentions).

3.3 | Discussion

Targeting familiar, everyday gendered behaviors, Study 2 shows that when provided with structural cues children aged 4 to 8 endorsed more structural explanations, predicted that behaviors were more mutable (and subject to structural changes), and reduced negativity towards non-conforming peers in terms of evaluations. However, even when salient structural cues were present, internal reasoning remained



stable, and *behavioral* sanctioning of gender-atypical behaviors (in terms of behavioral intentions to affiliate with peers) was not mitigated. We discuss these findings in more detail in the General Discussion.

4 | GENERAL DISCUSSION

Across two studies we probed 4-5-year-olds' and 7-8-year-olds' structural reasoning about gender and explored the implications of this reasoning process with both novel (Study 1) and familiar (Study 2) gendered behaviors. When being provided cues to structural explanations (but not structural explanations themselves), children in both studies showed the effect of structural framing across multiple measures: they endorsed the structural explanation more, expected behaviors to be more mutable, and evaluated gender non-conforming behaviors more leniently in the structural condition compared to the non-structural one. Importantly, and novel to the current work, we explored the boundary conditions of structural reasoning. We found that the effect of structural cues remained across explanations, mutability judgments, and evaluations even when internalized gender stereotypes were pre-existing and strong. We also provided evidence as to the potential mitigating effect of structural framing in children's evaluations of gender non-conforming behaviors, implying that structural thinking could be used as an intervention to reduce stereotyping and prejudice.

Dovetailing past work (Peretz-Lange et al., 2021; Vasilyeva et al., 2018), we found developmental changes in children's structural thinking. Older children differentiated structural and internal explanations more than younger children. This age pattern is consistent with the development of children's counterfactual and relational reasoning (Rafetseder & Perner, 2014; Rafetseder et al., 2013), which we speculate might be an underlying cognitive prerequisite for structural thinking (see also Vasilyeva et al., 2018). These developmental changes in children's ratings of various explanations in our studies also parallel the finding that children develop the ability to distinguish others' "good" reasoning from "bad" reasoning later in childhood (Amsterlaw, 2006).

One thought-provoking finding in Study 2 is that although children showed the effect of structural framing in their *reasoning* and *evaluations* (similar to e.g., Hussak & Cimpian, 2015; Van Wye et al., 2020, in which children were explicitly provided with different types of explanations), they still did not choose non-conforming peers over conforming ones as playmates. This finding suggests that merely inducing structural thinking might not be enough to mitigate children's *behavioral* sanctioning (or *discrimination*) of non-conforming peers, possibly because behavioral consequences are particularly hard to mitigate (see also Yang & Dunham, 2019). Future work may need to explore interventions beyond the provision of structural cues. For example, one potential intervention might be to introduce children to social norms that facilitate diversity in gendered behaviors (e.g., Killen, 2007). Additionally, the fact that we successfully altered children's evaluations but not behavioral intentions demonstrate a discrepancy between attitudes and behaviors (as widely documented across different domains, see Juvan & Dolnicar, 2014; Kollmuss & Agyeman, 2010). It is consistent

with past work showing that children do not want to associate with non-conforming peers even when they do not judge non-conformity as wrong (Carter & McCloskey, 1984).

There are interesting differences in children's structural thinking about novel versus familiar gendered behaviors. To begin with, when reasoning about novel behaviors, 7-8-year-olds increased reliance on structural explanations and reduced reliance on internal explanations under a structural framing; however, in terms of familiar behaviors, they only increased reliance on structural explanations without reducing reliance on internal ones. Our finding also implies that though researchers can induce structural thinking, this may not always be sufficient to reduce internal reasoning, especially in cases when internal reasoning was presumably strong at baseline. While one might have thought that these two types of reasoning necessarily trade-off against one another, our results suggest that needs not be the case. This result is also reminiscent of previous work finding that children sometimes simultaneously endorse contradictory explanations (Legare et al., 2012), though it is not clear that structural and internalist reasoning are necessarily contradictory.

Children also expected familiar gendered behaviors to be more constrained by gender norms than novel behaviors: children in Study 2 judged behaviors as less mutable and evaluated gender non-conforming behaviors more negatively than did children in Study 1. We expect that these differences are due to the heavier reliance on internal reasoning when thinking about familiar gendered behaviors as opposed to novel, unfamiliar behaviors. These findings also suggest that intervening on how children reason about familiar behaviors with strong prior beliefs is more difficult than it might seem when using novel behaviors as a proxy, though we acknowledge that pioneering work with novel behaviors is of vital importance. As the ultimate goal lies in the interventions on behaviors with real-world relevance, more future work is needed in this area.

Our study is among a small group of studies looking at children's structural reasoning capabilities (e.g., Dunlea & Heiphetz, 2020; Peretz-Lange et al., 2021; Vasilyeva et al., 2018). Most past research focuses on revealing children's strong essentialist or internalist tendencies (Cimpian & Salomon, 2014; Gelman, 2004; Gülgöz et al., 2019, 2021)—but here we provide clear evidence that despite this, when different cues are available, children can be flexible in their reasoning and can show structuralist thinking. However, it should be emphasized that we provided children with highly salient structural cues. Could we induce structural reasoning with less salient cues? This also leads to a fundamental question about whether children have a natural tendency to generate internal explanations (Cimpian & Salomon, 2014), or if they live in an environment where internal explanations are more frequently provided or easily observable, whereas structural explanations are less salient. If the latter, could we shift children's default explanatory strategies by providing a greater range of contexts with structural explanations? Future research will be necessary to address these questions.

This work also has implications for prejudice reduction. As shown in the present work, with the provision of structural cues, children viewed behaviors as more mutable and they were more accepting of non-conforming or category-atypical behaviors in terms of evaluations.



These reasoning and evaluative changes might lead to reduced stereotyping and prejudice against category-atypical individuals. Our findings also align with new work on a related topic (“winners” and “losers” of a novel game) which finds that children reduce negativity towards losers when they endorse that losers are structurally disadvantaged as compared to the view that losers are inherently inferior (Peretz-Lange et al., 2021). Such a possibility stands in contrast to the documented effect of internalist or essentialist thinking, which is generally considered prejudice/negativity-enhancing (e.g., Chen & Ratliff, 2018). That said, we do note that structural interventions might not always be desirable compared to, for example, internalist reasoning (as domains or situations matter, see Peretz-Lange, 2021). It is possible that there are cases in which structural reasoning will not be wholly beneficial, for example if it leads individuals to feel helpless in the face of powerful social structures that constrain their life possibilities but which they have little hope of changing.

To conclude, across two pre-registered studies we show that children as young as four attend to structural cues to explain gendered behaviors and adjust their judgments about mutability and acceptability of those behaviors accordingly. Some of these abilities continue to improve by 8 years of age. However, they still have internalist tendencies in explaining familiar gendered behaviors and show group-based discrimination in their behavioral intentions to affiliate with peers. These findings advance our understanding of structural reasoning in children and could potentially shed light on future intervention efforts to combat stereotyping and prejudice against non-typical members.

NOTES

- ¹ Note that following Vasilyeva et al. (2018), all participants answered questions about girls' behaviors.
- ² Results indicate that they still did not clearly differentiate these two explanations (see supplemental materials for detailed results).
- ³ There was an unanticipated gender effect, with boys giving overall lower evaluations than girls ($M_{\text{girl}} = 3.11$, $SD_{\text{girl}} = 1.08$; $M_{\text{boy}} = 2.88$, $SD_{\text{boy}} = 1.20$): $B = -0.40$, 95% CI $[-0.78, -0.03]$, $p = 0.04$, $R^2 = 0.04$, though this finding did not interact with other reported findings and so will not be a point of emphasis here.
- ⁴ We note that in the pre-registration we mistakenly pre-registered a mixed-model with a random intercept for participants, but because there is only one score per participant we actually fit a standard linear model.
- ⁵ Our preregistered sample size was $n = 96$, with $n = 48$ in each age group (same as Study 1); however, due to an unexpected experimenter error in the counterbalancing procedure, we collected a slightly unbalanced sample.

ACKNOWLEDGMENTS

We thank Nadya Vasilyeva for sharing their materials with us, Shaneka Crossman, Jessica Griggs, Laura Rodriguez, Ilayda Orhan, Sophie Arnold, Haley Hegefeld, and Auguste White for help with data collection, and Ilayda Orhan and Haley Hegefeld for coding.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no competing interests.

DATA AVAILABILITY STATEMENT

All study materials, data files, and analysis code that replicates all results in the manuscript and supplemental materials are also openly

shared and are available on an anonymous link at: https://osf.io/6kx4e/?view_only=33413fde2df346328f32e0bc5e72d11e.

REFERENCES

- Amsterlaw, J. (2006). Children's beliefs about everyday reasoning. *Child Development*, 77(2), 443–464. <https://doi.org/10.1111/j.1467-8624.2006.00881.x>
- Bonilla-Silva, E. (1997). Rethinking racism: Toward a structural interpretation. *American Sociological Review*, 62(3), 465–480. <https://doi.org/10.2307/2657316>
- Carter, D. B., & McCloskey, L. A. (1984). Peers and the maintenance of sex-typed behavior: The development of children's conceptions of cross-gender behavior in their peers. *Social Cognition*, 2(4), 294–314. <https://doi.org/10.1521/soco.1984.2.4.294>
- Chen, J. M., & Ratliff, K. A. (2018). Psychological essentialism predicts inter-group bias. *Social Cognition*, 36(3), 301–323. <https://doi.org/10.1521/soco.2018.36.3.301>
- Cimpian, A., & Salomon, E. (2014). The inference 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/S0140525113002197>
- Costanzo, P. R., & Shaw, M. E. (1966). Conformity as a function of age level. *Child Development*, 37(4), 967–975. JSTOR. <https://doi.org/10.2307/1126618>
- Diemer, M. A., Rapa, L. J., Voight, A. M., & McWhirter, E. H. (2016). Critical consciousness: A developmental approach to addressing marginalization and oppression. *Child Development Perspectives*, 10(4), 216–221. <https://doi.org/10.1111/cdep.12193>
- Dunlea, J., & Heiphetz, L. (in press). Language shapes children's attitudes: Consequences of internal, behavioral, and societal information in punitive and non-punitive contexts. *Journal of Experimental Psychology: General*. <https://philpapers.org/rec/DUNLSC>
- Dunlea, J. P., & Heiphetz, L. (2020). Children's and adults' understanding of punishment and the criminal justice system. *Journal of Experimental Social Psychology*, 87, 103913. <https://doi.org/10.1016/j.jesp.2019.103913>
- Eidelman, S., & Crandall, C. S. (2014). Chapter Two—The intuitive traditionalist: How biases for existence and longevity promote the status quo. In (J. M. Olson & M. P. Zanna Eds.), *Advances in experimental social psychology*, (Vol., 50, pp. 53–104). Academic Press. <https://doi.org/10.1016/B978-0-12-800284-1.00002-3>
- Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences*, 8(9), 404–409. <https://doi.org/10.1016/j.tics.2004.07.001>
- Godfrey, E. B., Burson, E. L., Yanisch, T. M., Hughes, D., & Way, N. (2019). A bitter pill to swallow? Patterns of critical consciousness and socioemotional and academic well-being in early adolescence. *Developmental Psychology*, 55(3), 525–537. <https://doi.org/10.1037/dev0000558>
- Goldring, M. R., & Heiphetz, L. (2020). Sensitivity to ingroup and outgroup norms in the association between commonality and morality. *Journal of Experimental Social Psychology*, 91, 104025. <https://doi.org/10.1016/j.jesp.2020.104025>
- Gülgöz, S., Alonso, D. J., Olson, K. R., & Gelman, S. A. (2021). Transgender and cisgender children's essentialist beliefs about sex and gender identity. *Developmental Science*, 1–14. <https://doi.org/10.1111/desc.13115>
- Gülgöz, S., DeMeules, M., Gelman, S. A., & Olson, K. R. (2019). Gender essentialism in transgender and cisgender children. *Plos One*, 14(11), e0224321. <https://doi.org/10.1371/journal.pone.0224321>
- Gülgöz, S., Gomez, E. M., DeMeules, M. R., & Olson, K. R. (2018). Children's evaluation and categorization of transgender children. *Journal of Cognition and Development*, 19(4), 325–344. <https://doi.org/10.1080/15248372.2018.1498338>
- Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology*, 39(1), 113–127. <https://doi.org/10.1348/014466600164363>

- Haslanger, S. (2016). What is a (social) structural explanation? *Philosophical Studies*, 173(1), 113–130. <https://doi.org/10.1007/s11098-014-0434-5>
- Haun, D. B. M., & Tomasello, M. (2011). Conformity to peer pressure in preschool children. *Child Development*, 82(6), 1759–1767. <https://doi.org/10.1111/j.1467-8624.2011.01666.x>
- 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>
- Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, 25(6), 881–919. <https://doi.org/10.1111/j.1467-9221.2004.00402.x>
- Juvan, E., & Dolnicar, S. (2014). The attitude-behaviour gap in sustainable tourism. *Annals of Tourism Research*, 48, 76–95. <https://www.cabdirect.org/cabdirect/abstract/20143369382>
- Killen, M. (2007). Children's social and moral reasoning about exclusion. *Current Directions in Psychological Science*, 16(1), 32–36. <http://journals.sagepub.com/doi/abs/10.1111/j.1467-8721.2007.00470.x>
- Kollmuss, A., & Agyeman, J. (2010). Mind the Gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental Education Research*, 8(3), 239–260. <https://doi.org/10.1080/13504620220145401>
- Kraus, M. W., & Park, J. W. (2017). The structural dynamics of social class. *Current Opinion in Psychology*, 18, 55–60. <https://doi.org/10.1016/j.copsyc.2017.07.029>
- Legare, C. H., Evans, E. M., Rosengren, K. S., & Harris, P. L. (2012). The coexistence of natural and supernatural explanations across cultures and development. *Child Development*, 83(3), 779–793. <https://doi.org/10.1111/j.1467-8624.2012.01743.x>
- Merolla, D. M., & Jackson, O. (2019). Structural racism as the fundamental cause of the academic achievement gap. *Sociology Compass*, 13(6), e12696. <https://doi.org/10.1111/soc4.12696>
- Nyhout, A., & Ganea, P. A. (2019). Mature counterfactual reasoning in 4- and 5-year-olds. *Cognition*, 183, 57–66. <https://doi.org/10.1016/j.cognition.2018.10.027>
- Peretz-Lange, R. (2021). Why does social essentialism sometimes promote, and other times mitigate, prejudice development? A causal discounting perspective. *Cognitive Development*, 59, 101085. <https://doi.org/10.1016/j.cogdev.2021.101085>
- Peretz-Lange, R., Perry, J., & Muentener, P. (2021). Developmental shifts toward structural explanations and interventions for social status disparities. *Cognitive Development*, 58, 101042. <https://doi.org/10.1016/j.cogdev.2021.101042>
- Prasada, S., & Dillingham, E. M. (2009). Representation of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science*, 33(3), 401–448. <https://doi.org/10.1111/j.1551-6709.2009.01018.x>
- Prentice, D. A., & Miller, D. T. (2007). Psychological essentialism of human categories. *Current Directions in Psychological Science*, 16(4), 202–206. <https://doi.org/10.1111/j.1467-8721.2007.00504.x>
- Rafetseder, E., & Perner, J. (2014). Counterfactual reasoning: Sharpening conceptual distinctions in developmental studies. *Child Development Perspectives*, 8(1), 54–58. <https://doi.org/10.1111/cdep.12061>
- Rafetseder, E., Schwitalla, M., & Perner, J. (2013). Counterfactual reasoning: From childhood to adulthood. *Journal of Experimental Child Psychology*, 114(3), 389–404. <https://doi.org/10.1016/j.jecp.2012.10.010>
- Rhodes, M., & Mandalaywala, T. M. (2017). The development and developmental consequences of social essentialism. *Wiley Interdisciplinary Reviews: Cognitive Science*, 8(4), e1437. <https://doi.org/10.1002/wcs.1437>
- Ridgeway, C. L., & Smith-Lovin, L. (1999). The gender system and interaction. *Annual Review of Sociology*, 25(1), 191–216. <https://doi.org/10.1146/annurev.soc.25.1.191>
- Rizzo, M. T., Elenbaas, L., & Vanderbilt, K. E. (2018). Do children distinguish between resource inequalities with individual versus structural origins? *Child Development*, 91(2), 439–455. <https://doi.org/10.1111/cdev.13181>
- Rizzo, M. T., & Killen, M. (2020). Children's evaluations of individually and structurally based inequalities: The role of status. *Developmental Psychology*, 56(12), 2223–2235. <https://doi.org/10.1037/dev0001118>
- Roberts, S. O., Gelman, S. A., & Ho, A. K. (2017). So it is, so it shall be: Group regularities license children's prescriptive judgments. *Cognitive Science*, 41(S3), 576–600. <https://doi.org/10.1111/cogs.12443>
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In (L. Berkowitz Ed.), *Advances in Experimental Social Psychology*, (Vol. 10, pp. 173–220). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60357-3](https://doi.org/10.1016/S0065-2601(08)60357-3)
- Ross, L. (2018). From the fundamental attribution error to the truly fundamental attribution error and beyond: My research journey. *Perspectives on Psychological Science*, 13(6), 750–769. <https://doi.org/10.1177/1745691618769855>
- Rubin, Z., & Peplau, L. A. (1975). Who believes in a just world? *Journal of Social Issues*, 31(3), 65–89. <https://doi.org/10.1111/j.1540-4560.1975.tb00997.x>
- Schmidt, M. F. H., & Tomasello, M. (2012). Young children enforce social norms. *Current Directions in Psychological Science*, 21(4), 232–236. <https://doi.org/10.1177/0963721412448659>
- Van Wye, E. M., Wang, M., & Roberts, S. O. (2020). Explanations for norm violations affect preschoolers' judgments of norm violators. *Journal of Experimental Psychology: General*, No Pagination Specified-No Pagination Specified. <https://doi.org/10.1037/xge0000942>
- 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>
- Walker, M. B., & Andrade, M. G. (1996). Conformity in the Asch task as a function of age. *The Journal of Social Psychology*, 136(3), 367–372. <https://doi.org/10.1080/00224545.1996.9714014>
- Watts, R. J., Diemer, M. A., & Voight, A. M. (2011). Critical consciousness: Current status and future directions. *New Directions for Child and Adolescent Development*, 2011(134), 43–57. <https://doi.org/10.1002/cd.310>
- Yang, X., & Dunham, Y. (2019). Minimal but meaningful: Probing the limits of randomly assigned social identities. *Journal of Experimental Child Psychology*, 185, 19–34. <https://doi.org/10.1016/j.jecp.2019.04.013>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Yang, X., Naas, R., & Dunham, Y. (2021). Testing the limits of structural thinking about gender. *Developmental Science*, e13169. <https://doi.org/10.1111/desc.13169>