

# Developmental Psychology

## **“He Did Girls’ Things!” Hong Kong and Canadian Children’s Reasoning About Moral Judgments of Peers’ Gendered Behaviors**

Karen Man Wa Kwan, Sylvia Yun Shi, Laura N. MacMullin, A. Natisha Nabbijohn, Diana E. Peragine, Doug P. VanderLaan, and Wang Ivy Wong

Online First Publication, March 14, 2024. <https://dx.doi.org/10.1037/dev0001698>

### CITATION

Kwan, K. M. W., Shi, S. Y., MacMullin, L. N., Nabbijohn, A. N., Peragine, D. E., VanderLaan, D. P., & Wong, W. I. (2024, March 14). “He Did Girls’ Things!” Hong Kong and Canadian Children’s Reasoning About Moral Judgments of Peers’ Gendered Behaviors. *Developmental Psychology*. Advance online publication. <https://dx.doi.org/10.1037/dev0001698>

# “He Did Girls’ Things!” Hong Kong and Canadian Children’s Reasoning About Moral Judgments of Peers’ Gendered Behaviors

Karen Man Wa Kwan<sup>1, 2</sup>, Sylvia Yun Shi<sup>2, 3</sup>, Laura N. MacMullin<sup>4</sup>, A. Natisha Nabbijohn<sup>4</sup>,  
Diana E. Peragine<sup>4</sup>, Doug P. VanderLaan<sup>4</sup>, and Wang Ivy Wong<sup>2, 3</sup>

<sup>1</sup>Department of Applied Social Sciences, The Hong Kong Polytechnic University

<sup>2</sup>Department of Psychology, University of Hong Kong

<sup>3</sup>Gender Studies Programme and Department of Psychology, The Chinese University of Hong Kong


<sup>4</sup>Department of Psychology, University of Toronto Mississauga

Children show less positivity toward gender-nonconforming (GN) than gender-conforming (GC) peers. Yet, little is known about children’s reasoning about peers of varying gender expressions, including age-, gender-, and culture-related influences. We investigated how children aged 4- to 5- and 8- to 9-years-old in Hong Kong and Canada ( $N = 678$ ) reason about their moral judgments of GC and GN peers. After viewing vignettes describing GC and GN boys and girls, we asked children whether each target peer’s behavior was right or wrong and why they thought so. We coded children’s reasoning using a new coding scheme developed via inductive content analysis. Overall, children’s most commonly used reasoning styles were global standard, personal choice, gender stereotypes, “don’t know,” and others’ welfare. Children used more gender stereotype-related reasoning when they were older and from Hong Kong, appraising the GN boy, or when they perceived the target’s behavior as wrong. In contrast, children reasoned based on personal choice more when they were from Canada or when they perceived the target’s behavior as right. These findings inform how age-, gender-, and culture-related factors are associated with children’s reasoning about the acceptability or appropriateness of varying kinds of childhood gendered behavior. They provide insights regarding children’s appraisals of different gender expressions by illuminating not only how they view GC and GN peers but also, from their own perspectives, why they do so. These insights have implications for strategies aimed at decreasing gender-related biases and increasing children’s acceptance of gender diversity.

## Public Significance Statement

This study extended prior literature by comparing Hong Kong and Canadian children’s reasoning about moral judgments of peers’ gendered behaviors using a sample much larger than those reported in prior research. Apart from testing important variables including participant gender, age, and target gender and target gender expression, we gained valuable insights into cultural variations. Our findings can inform future work focused on improving children’s appraisals of peer gender nonconformity via strategies that target reasoning styles.

Karen Man Wa Kwan  <https://orcid.org/0000-0003-0214-3390>

Wang Ivy Wong  <https://orcid.org/0000-0002-9428-2832>

Laura N. MacMullin is now at the Department of Psychology, Middlebury College. A. Natisha Nabbijohn is now at the Department of Psychology, University of Guelph.

This research was funded by a Hong Kong Postgraduate Scholarship for PhD study to Karen Man Wa Kwan, a Seed Fund for Basic Research of the University of Hong Kong, a Hong Kong Research Grants Council Early Career Scheme 27402914, a General Research Fund 17600719 to Wang Ivy Wong, Social Sciences and Humanities Research Council (SSHRC) of Canada Graduate Scholarship-Master’s Awards to A. Natisha Nabbijohn and Laura N. MacMullin, and a University of Toronto Connaught New Researcher Award and a SSHRC of Canada Insight Development Grant 4301600445 to Doug P. VanderLaan. The authors thank Nicole Raquinio for the artistic production of the stimuli, Alanna Santarossa for participation in study and stimuli design, and Sui Ping Yeung and Haley James for assistance in data collection. All authors declare that they have no conflicts of interest. The data that support the findings of this study are available from the corresponding authors upon reasonable request. Ethics approval was obtained from the institutional ethics committees of the University of Hong Kong and the University of Toronto. The authors own the copyright of the stimuli design.

Karen Man Wa Kwan served as lead for conceptualization, data curation,

formal analysis, investigation, methodology, project administration, visualization, writing—original draft, and writing—review and editing and served in a supporting role for funding acquisition and resources. Sylvia Yun Shi served as lead for formal analysis and served in a supporting role for data curation, methodology, visualization, writing—original draft, and writing—review and editing. Laura N. MacMullin served in a supporting role for data curation, investigation, and writing—review and editing. A. Natisha Nabbijohn served in a supporting role for data curation, investigation, methodology, and writing—review and editing. Diana E. Peragine served in a supporting role for methodology and writing—review and editing. Doug P. VanderLaan served as lead for conceptualization, funding acquisition, investigation, methodology, project administration, resources, supervision, and writing—review and editing and served in a supporting role for data curation, formal analysis, visualization, and writing—original draft. Wang Ivy Wong served as lead for conceptualization, funding acquisition, investigation, methodology, project administration, resources, supervision, and writing—review and editing and served in a supporting role for data curation, formal analysis, visualization, and writing—original draft.

Correspondence concerning this article should be addressed to Wang Ivy Wong, Gender Studies Programme and Department of Psychology, The Chinese University of Hong Kong, Hong Kong SAR, The People’s Republic of China, or Doug P. VanderLaan, Department of Psychology, University of Toronto, Mississauga, ON, Canada. Email: [iwwong@cuhk.edu.hk](mailto:iwwong@cuhk.edu.hk) or [doug.vanderlaan@utoronto.ca](mailto:doug.vanderlaan@utoronto.ca)

*Keywords:* gender nonconformity, gender stereotypes, reasoning style, gender role attitudes, peer relations

*Supplemental materials:* <https://doi.org/10.1037/dev0001698.supp>

Gender expression refers to aspects of an individual's presentation (e.g., appearance, behaviors, and traits) that are related to the roles of being boys/men or girls/women (American Psychological Association, 2015). Children have varying appraisals toward peers of different gender expressions (Kwan et al., 2020; Nabbijohn et al., 2020). Gender expressions that do not follow the expected gender roles of a given culture are often considered as instances of gender nonconformity (Spivey et al., 2018). Prior research suggested that children were less positive toward gender nonconformity (e.g., less willing to befriend gender-nonconforming [GN] peers and more likely to imitate gender-conforming [GC] than GN peers; Blakemore, 2003; Kwan et al., 2020; Levy et al., 1995; Nabbijohn et al., 2020; Y. Wang et al., 2022; Zucker et al., 1995). As poor peer relations including stigmatization by peers have been found to be a robust correlate of poorer psychological well-being in children (Cohen-Kettenis et al., 2003; MacMullin et al., 2021; Roberts et al., 2013), it is important to understand not just how, but also why children appraise peers with varying gender expressions differently. Children's reasoning styles when explaining their appraisals may potentially reflect their underlying beliefs (Witenberg, 2007). Common reasoning styles children used to explain their appraisals of different gender expressions include personal choice (i.e., appeal to individual preferences), gender stereotypes, others' welfare, and global descriptions of behavior as good or bad (Conry-Murray, 2013; Conry-Murray & Turiel, 2012; Killen et al., 2001, 2013; Smetana, 1986; Theimer et al., 2001).

Prior studies suggested that the reasoning styles used were associated with levels of bias against different gender expressions (Killen et al., 2013). For example, when children reason in support of a certain gender expression, they are more likely to cite personal choice. More than half of the children aged 5-, 6-, 8-, and 9-years-old who supported another child's GN preference (e.g., a boy's preference for a pink bike; Conry-Murray, 2013) and children aged 6- and 8-years-old who believed that a boy preferring a feminine activity should be selected for the activity used personal choice as their justification (Conry-Murray & Turiel, 2012). Gender stereotype-related reasoning is often adopted when children favor gender-stereotypical situations or GC peers. For example, preschoolers who excluded an other-gender peer engaging in a GN activity often rationalized their choice using gender stereotypes (e.g., dolls are for girls, not boys; Killen et al., 2001; Theimer et al., 2001). Other reasoning styles may be less directly related to gender but are nevertheless important, as they are among the most common reasoning styles children use. According to social domain theory (Smetana, 2006; Turiel, 1983), moral reasons (e.g., fairness), as well as social conventional reasons (e.g., group norms) and personal reasons (e.g., personal goals), are fundamental concerns when children interpret social issues. When justifying their (dis)approval of moral transgressions in general, children used reasoning such as fairness, equality, justice, rights, others' welfare (e.g., the described behavior should not bother others), and global descriptions of bad or good (e.g., stealing is bad; Elenbaas, 2019; Smetana, 1986; Stoddart & Turiel, 1985). Little is known about developmental and gender differences in reasoning when children judge peer gender expressions. Thus, we investigated how children aged

4- to 5- and 8- to 9-years-old in Hong Kong and Canada ( $N = 678$ ) reason about their moral judgments of GC and GN peers.

## Developmental and Gender Differences in Children's Reasoning About Gender Norms

According to gender schema theory (Martin et al., 2002; Martin & Halverson Jr, 1981; Martin & Ruble, 2004), once children acquire gender identity around 2–3 years of age, gender becomes a salient characteristic. They are motivated to actively search for gender-related information from the environment. Gender stereotypes are formed as a result, which then influence cognitive processes and behaviors. Through operant conditioning such as reinforcement of GC behaviors, as well as observational learning of models exhibiting GC behaviors (Bandura & Bussey, 2004; Bussey & Bandura, 1999; Mischel, 2015), gender socialization occurs and children learn the expected gender norms over time, as socialization experience accumulates. Their gender schema may initially be rigid and then become more complex with age due to improved cognitive ability and increasing experience with diverse and counter examples of gender norms (Halim, 2016; Lobel et al., 2000). Thus, with advancing cognitive abilities, older children also are more flexible with gender norms, believing that it is possible for people to engage in GN activities (Trautner et al., 2005). However, less rigid views on whether others adopt GN roles do not necessitate that children become more accepting of GN behavior. Rather, prior studies showed that children's appraisals of gender nonconformity are less positive in middle than in early childhood (Blakemore, 2003; Carter & McCloskey, 1984; Kwan et al., 2020; Nabbijohn et al., 2020; but see Stoddart & Turiel, 1985).

Understanding the paradoxical development of greater recognition of, and greater disapproval of, gender nonconformity requires studies on the rationales children use when evaluating GN, in contrast to GC, behaviors. Recent research provides some insights. Miller et al. (2009) found that older children in fourth and fifth grades generated more gender-stereotypical statements to describe boys and girls than did preschoolers. This greater knowledge may contribute to greater reliance on gender stereotypes for judging gendered behaviors. Children may also expand their rationales for assessing the appropriateness of gendered behaviors, drawing on increasingly more abstract concepts. For example, when asked why they thought GN activities were wrong or not, kindergartners' reasoning emphasized the importance of maintaining physical appearance that conforms to one's gender identity, whereas children in middle childhood emphasized personal choice and social expectations (i.e., gender roles can be changed according to different social expectations; Stoddart & Turiel, 1985). Consequently, with age, children may be prone to either utilize increasingly cognitively sophisticated rationales to justify approval of gender nonconformity, or utilize their expanding knowledge of gender stereotypes to rationalize disapproval of gender nonconformity.

Apart from age, children's own gender appears to moderate their reasoning about gender expressions. Leaper (1994) proposed that higher status members tend to be avoidant of characteristics of

lower status members; therefore, boys avoid femininity because it is perceived as being associated with lower social status than masculinity. Boys are also subject to stronger socialization pressure than girls to conform to gender norms. For example, adults and children often show more acceptance of gender nonconformity in girls than boys (Kane, 2006; Roberts et al., 2013; Thomas & Blakemore, 2013). Consistent with these ideas, boys are usually more avoidant of GN behaviors and more accepting of gender stereotypes than girls (Wallien et al., 2010; Wood et al., 2022) and GN boys are perceived more negatively than GN girls (Braun & Davidson, 2017; Kwan et al., 2020; Nabbijohn et al., 2020; Qian et al., 2021). Although these studies did not directly investigate gender differences in reasoning, they point to the heavier emphasis for boys and by boys to adhere to gender norms, thus suggesting that boys may be more likely than girls to use gender stereotype-related reasoning, and this reasoning style may be used more often when evaluating GN boys than GN girls.

### Culture and Children's Reasoning About Gender Norms

Because gender socialization practices are embedded within larger cultural practices and values, developmental changes in children's reasoning styles when explaining their judgment about gender norms need to be studied within diverse communities and through global studies. Existing research in the West (usually Western Europe and North America) has shown that children as young as 4 years old to middle childhood give less positive appraisals toward gender nonconformity than gender conformity (Blakemore, 2003; Levy et al., 1995; Nabbijohn et al., 2020; Zucker et al., 1995). Consistent with these studies, Chinese children between 4 and 9 years old perceived GN peers less positively and preferred them less than GC peers (Kwan et al., 2020; Y. Wang et al., 2022). Although both Chinese and Western children are less positive toward gender nonconformity than gender conformity, whether children across cultures reason similarly about their appraisals is largely unknown.

Although other studies have investigated judgments about gender norms with a cultural perspective, we are aware of only two publications, Conry-Murray et al.'s (2015, 2020) research on U.S. and Korean children, that investigated reasoning styles about gender norms across more than one cultural sample. Specifically, Conry-Murray et al. (2015,  $N = 144$  children; 2020,  $N = 151$  children and 49 adults) examined U.S. and Korean children's judgment and reasoning about gender norms in helping versus nonhelping and public versus private contexts. Korean children were less accepting of gender norm violations and used more gender-related reasoning than U.S. children when the context did not involve helping others, reflecting different cultural emphasis on gender stereotypes. There was little cultural difference when a helping goal was involved, reflecting general prominence of moral concerns or concerns about others' welfare in children. In contrast, the U.S. children used more gender-related reasoning than Korean children in private, though not in public contexts. These provided initial evidence that culture may influence children's use of reasoning styles when explaining judgments about gender norms.

There are reasons to expect that Chinese children's reasoning about gender nonconformity may differ from their Western peers. One difference may be that Chinese children reason based on gender stereotypes more often. For example, the Taoist concept of yin and yang in Chinese culture, which suggests that men and women play different roles to maintain a balance in nature (R. R. Wang, 2005),

might dichotomize the two genders and rationalize adherence to gender norms. Supporting this notion, Chinese women scored higher than U.S. women on benevolent sexism. This scale includes a gender complementary dimension (i.e., stereotypic roles for women), which suggested Chinese women might have endorsed gender complementary relationships more strongly than U.S. women (Z. Chen et al., 2009). Several other findings suggest that gender stereotypes are more visible and endorsed more in Chinese than typical Western societies included in prior research. For example, Chinese students described women using more gender-stereotypical characteristics (e.g., "warm"), whereas U.S. students described women using more liberal characteristics (e.g., "competent"; Lii & Wong, 1982). Also, Chinese magazine advertisements showed less gender egalitarian content than U.S. advertisements (Zhang & Cartwright, 2009). In children's appraisals, Hong Kong children were less positive toward gender nonconformity than Canadian children (Kwan et al., 2020; MacMullin et al., 2020; Nabbijohn et al., 2020). In Conry-Murray et al. (2015), children in Korea (a culture that shares many cultural elements with China, such as Confucianism) also employed more gender-related reasoning than U.S. children in nonhelping contexts. Thus, it is possible that Chinese children reason based on gender stereotypes more than children from Western cultures.

Specific cultural values may also lead Chinese children to reason more based on group harmony or others' welfare and less on personal choice. For example, Chinese culture places more emphasis on relationships with others (Chao, 1995; X. Chen et al., 2003), whereas Canadian culture places more emphasis on individual autonomy (Y.-R. Chen et al., 1998). When examining parenting goals, Chinese mothers emphasized maintaining good relationships with others (Chao, 1995), whereas Canadian mothers emphasized the development of autonomy (Liu et al., 2005); when examining teaching styles, Chinese teachers tended to adopt a controlling style rather than an autonomy-supportive style (Zhou et al., 2012); when examining students' learning styles, Asian students (including Hong Kong) had a stronger tendency to adopt collaborative approaches in learning than Australian students (Ramburuth & McCormick, 2001; Tang, 1996). These patterns suggest Chinese children may reason more based on others' welfare, whereas children from Western countries (e.g., Canada) may reason more based on personal choice than their cultural counterparts.

### The Present Study

Prior findings from the data set employed in this study (Kwan et al., 2020; MacMullin et al., 2020; Nabbijohn et al., 2020) found that children in Hong Kong and Canada were less positive toward GN peers in various respects (e.g., friendship preference, perceived popularity). Interestingly, Chinese, but not Canadian, children judged GN peers' behavior as more wrong than GC peers' behavior. This study builds on earlier analyses of this data set by focusing on children's reasoning about their moral judgments (i.e., judging the correctness of peers' behaviors). Moral judgments are important because they capture children's cognitive bias and are often investigated in studies of children's gender-related appraisals of peers. We included children aged 4–5 and 8–9 years old for several reasons. First, peer gender segregation increases from early childhood and peaks in middle childhood (Leaper, 2022; Mehta & Strough, 2009), and children's behaviors tend to be more gender-typed in same-gender peer groups, partly through modeling and peer scrutinization of GN behaviors

(Kowalski, 2007; Martin & Fabes, 2001). Second, developmental differences in children's judgment and reasoning about gender norms have been reported in this age range (although with mixed findings; Conry-Murray, 2013; Conry-Murray et al., 2015, 2020; Conry-Murray & Turiel, 2012; Killen & Stangor, 2001), making it important to understand children's explanations during this developmental period. Third, as prior studies examining children's gender-related reasoning usually involved children aged 4 years old or above (Conry-Murray, 2015, 2017; Killen et al., 2001; Theimer et al., 2001), it is expected that children above this age have the necessary verbal ability to explain their reasoning. Thus, by building on earlier analyses of our large data set ( $N = 678$ ) of Hong Kong and Canadian children aged 4–5 and 8–9 years old, the present study can contribute to our understanding of children's reasoning styles about their moral judgments (i.e., judging the correctness of GN and GC peers' behaviors) reflecting their appraisals of different gender expressions, as well as the influence of various important variables including age and gender, in two samples from different geographical regions.

The first set of hypotheses concerned the overall usage of reasoning styles. According to the gender schema theory, gender stereotypes develop from an early age and often affect judgments (Killen et al., 2001; Martin, 1989; Theimer et al., 2001). Also, prior empirical findings based on social domain theory have shown that moral concerns, social convention and approval, and personal choice are important in children's judgments about gender norms (Smetana, 2006; Turiel, 1983). Thus, our first hypothesis was that gender stereotype-related reasoning, as well as reasoning reflecting moral concerns (e.g., others' welfare), personal choice, and global standards, would be among the most commonly used reasoning styles (H1a). However, the frequency of the various reasoning styles may be moderated by whether the participants perceived the target's behavior as right or wrong. Thus, it was hypothesized that children who believed the target's behavior was wrong would reason based on gender stereotypes more than children who believed the target's behavior was right (Killen et al., 2001; Theimer et al., 2001) (H1b). In contrast, children who believed that the target was right would reason based on personal choice more than children who believed that the target was wrong (Conry-Murray, 2013; Conry-Murray & Turiel, 2012) (H1c).

The second set of hypotheses concerned children's age. Gender segregation peaks in middle childhood with children more intensively scrutinizing peers' gender nonconformity (Leaper, 2022; Mehta & Strough, 2009), and children's knowledge of gender stereotypes increases with age (Blakemore, 2003; Miller et al., 2009; Trautner et al., 2005). Both factors suggest gender-stereotypical reasoning would be more common among older children. Also, older children or young adults emphasized personal choice more than younger children (Conry-Murray et al., 2020; Stoddart & Turiel, 1985). Thus, it was hypothesized that older children would reason based on gender stereotypes (H2a) and personal choice (H2b) more than younger children.

The third set of hypotheses concerned gender. Based on status differences between masculinity and femininity and heavier sanctions for boys' gender nonconformity, it was hypothesized that the gender of the participant (Leaper, 1994; Wood et al., 2022) and the target being appraised (Braun & Davidson, 2017; Killen et al., 2001; Theimer et al., 2001; Wallien et al., 2010) would be related to children's reasoning. Specifically, it was hypothesized that boys would reason based on gender stereotypes more than girls (H3a) and children would reason based on gender stereotypes more when appraising GN children, particularly GN boys, than GC children (H3b).

The fourth set of hypotheses concerned cultural comparisons. Limited literature has compared gender-related reasoning across different cultures simultaneously (except Conry-Murray et al., 2015, 2020). Our study contributed in this regard by comparing Chinese and Canadian cultures with a sample much larger than those reported in prior research. Chinese culture emphasizes gender stereotypes more strongly (Z. Chen et al., 2009; Lii & Wong, 1982; Zhang & Cartwright, 2009). Also, participants in this study from Hong Kong were found to be less positive toward gender nonconformity than Canadian children (Kwan et al., 2020; MacMullin et al., 2020; Nabbijohn et al., 2020). Thus, it was hypothesized that Hong Kong children would reason based on gender stereotypes more than Canadian children (H4a). Further, based on the relative emphasis placed on interpersonal relationships or personal interest and the various differences in socialization practices between these cultures (Chao, 1995; Liu et al., 2005; Ramburuth & McCormick, 2001; Tang, 1996; Zhou et al., 2012), it was hypothesized that Hong Kong, compared with Canadian, children would reason more based on others' welfare (H4b) and Canadian, compared with Hong Kong, children would reason more based on personal choice (H4c).

## Method

### Participants

This study was based on data from our larger project examining gender-related peer appraisals in Hong Kong and Canadian children (Kwan et al., 2020; MacMullin et al., 2020; Nabbijohn et al., 2020). Hong Kong Chinese children were recruited through kindergartens, primary schools, and education centers, and through advertisements online and on campus. Canadian children were recruited from a community participant pool at a large university in Mississauga. Across the Hong Kong and Canadian samples, 26 participants were excluded because they reported having special educational needs, did not meet the age requirement, did not comply with the study procedures, and/or failed attention check questions during the experiment. Also, one Canadian participant was identified by their parent as gender creative. We did not have enough participants to form a group of nonbinary children for analysis and, thus, this participant was excluded. Our sample was part of a larger project and the sample size was determined to provide conventional levels of statistical power in relation to comparisons performed in our earlier papers (Kwan et al., 2020; MacMullin et al., 2020; Nabbijohn et al., 2020). The Hong Kong sample consisted of 320 participants, of which 109 children (55 boys and 54 girls) were in the younger age group (4–5 years old;  $M_{\text{age}} = 4.51$ ,  $SD = 0.50$ ) and 211 children (105 boys and 106 girls) were in the older age group (8–9 years old;  $M_{\text{age}} = 8.50$ ,  $SD = 0.50$ ). The Canadian sample consisted of 358 participants, of which 176 children (91 boys and 85 girls) were in the younger age group (4–5 years old;  $M_{\text{age}} = 4.55$ ,  $SD = 0.50$ ) and 182 children (93 boys and 89 girls) were in the older age group (8–9 years old;  $M_{\text{age}} = 8.53$ ,  $SD = 0.50$ ). For demographic characteristics of our sample, please refer to Tables 1 and 2.

### Procedures

The study procedures were the same in Hong Kong and Canada. Each child was accompanied by a parent/guardian to the laboratory and both provided consent or verbal assent. Ethics approval was obtained from the institutional ethics committees. The parent/guardian

**Table 1**  
*Demographic Characteristics of Hong Kong Participants*

Variable	Age 4, 5		Age 8, 9			
	Range	N	M (SD) or frequency	M (SD) or frequency		
(a) Education level of the participating parent	NA	(a) 109	Less than high school: (a) 10 (9.2%), (b) 9 (8.3%); High school: (a) 19 (17.4%), (b) 15 (13.9%); Diploma/Certificate: (a) 21 (19.3%), (b) 17 (15.7%); Associate degree: (a) 5 (4.6%), (b) 5 (4.6%); University, bachelor's degree: (a) 35 (32.1%), (b) 44 (40.7%); University, master's degree: (a) 17 (15.6%), (b) 18 (16.7%); University, doctoral degree: (a) 2 (1.8%), (b) 0 (0%)	NA	(a) 211 (b) 207	Less than high school: (a) 35 (16.6%), (b) 38 (18.4%); High school: (a) 40 (19%), (b) 41 (19.8%); Diploma/certificate: (a) 35 (16.6%), (b) 15 (7.2%); Associate degree: (a) 6 (2.8%), (b) 6 (2.9%); University, bachelor's degree: (a) 59 (28%), (b) 64 (30.9%); University, master's degree: (a) 32 (15.2%), (b) 39 (18.8%); University, doctoral degree: (a) 4 (1.9%), (b) 4 (1.9%)
(b) Education level of the participating parent's partner	NA	108	<\$2,000: 0 (0%); \$2,000–\$3,999: 0 (0%); \$4,000–\$5,999: 0 (0%); \$6,000–\$7,999: 0 (0%); \$8,000–\$9,999: 0 (0%); \$10,000–\$14,999: 1 (0.9%); \$15,000–\$19,999: 5 (4.6%); \$20,000–\$24,999: 5 (4.6%); \$25,000–\$29,999: 9 (8.3%); \$30,000–\$39,999: 21 (19.4%); \$40,000–\$59,999: 28 (25.9%); \$60,000–\$79,999: 17 (15.7%); \$80,000–\$99,999: 7 (6.5%); ≥ \$100,000: 15 (13.9%)	NA	210	<\$2,000: 3 (1.4%); \$2,000–\$3,999: 2 (1%); \$4,000–\$5,999: 0 (0%); \$6,000–\$7,999: 0 (0%); \$8,000–\$9,999: 2 (1%); \$10,000–\$14,999: 10 (4.8%); \$15,000–\$19,999: 14 (6.7%); \$20,000–\$24,999: 10 (4.8%); \$25,000–\$29,999: 17 (8.1%); \$30,000–\$39,999: 31 (14.8%); \$40,000–\$59,999: 58 (27.6%); \$60,000–\$79,999: 27 (12.9%); \$80,000–\$99,999: 16 (7.6%); ≥ \$100,000: 20 (9.5%)
Family religious level <sup>a</sup>	0–4	109	97 (1.34)	0–5	210	1.14 (1.34)
Ethnicity	NA	109	Single-ethnic Chinese: 106 (97.2%); Multiethnic Chinese: 3 (2.8%)	NA	211	Single-ethnic Chinese: 208 (98.6%); Multiethnic Chinese: 3 (1.4%)
Participating parent/guardian gender <sup>b</sup>	NA	109	Female: 97 (89%); Male: 12 (11%)	NA	211	Female: 180 (85.3%); Male: 31 (14.7%)
Family religion	NA	109	No religion: 68 (62.4%); Roman Catholic: 17 (15.6%); Protestant: 1 (0.9%); Christian: 21 (19.3%); Buddhist: 2 (1.8%)	NA	211	No religion: 111 (52.6%); Roman Catholic: 23 (10.9%); Protestant: 5 (2.4%); Christian: 56 (26.5%); Buddhist: 13 (6.2%); Hindu: 2 (0.9%); Other religion: 1 (0.5%)

Note. Foreducation level and income, different response categories were used for the Hong Kong and Canadian samples because these categories were more reasonable and more commonly used in surveys of each region. HKD = Hong Kong dollar; NA = not applicable.

<sup>a</sup> Family religious level: 5 = extremely; 4 = very; 3 = moderately; 2 = slightly; 1 = not at all; 0 = no religion. <sup>b</sup> Gender refers to the birth-assigned gender.

**Table 2**  
*Demographic Characteristics of Canadian Participants*

Variable	Age 4, 5			Age 8, 9		
	Range	N	M (SD) or frequency	Range	N	M (SD) or frequency
(a) Education level of the participating parent	NA	(a) 167	Less than high school: (a) 0 (0%); (b) 3 (1.9%);	NA	(a) 171	Less than high school: (a) 0 (0%); (b) 1 (0.6%);
(b) Education level of the participating parent's partner		(b) 161	Some high school: (a) 0 (0%); (b) 4 (2.5%); High school diploma or equivalent: (a) 4 (2.4%), (b) 17 (10.6%); College diploma or trade school certificate: (a) 35 (21%), (b) 39 (24.2%); University, bachelor's degree: (a) 76 (45.5%), (b) 55 (34.2%); University, master's degree: (a) 36 (21.6%), (b) 33 (20.5%); University, postgraduate degree: (a) 16 (9.6%), (b) 10 (6.2%)		(b) 169	Some high school: (a) 0 (0%); (b) 9 (5.3%); High school diploma or equivalent: (a) 8 (4.7%), (b) 15 (8.9%); College diploma or trade school certificate: (a) 35 (20.5%), (b) 48 (28.4%); University, bachelor's degree: (a) 77 (45%), (b) 58 (34.3%); University, master's degree: (a) 36 (21.1%), (b) 28 (16.6%); University, postgraduate degree: (a) 15 (8.8%), (b) 10 (5.9%)
Annual household income (in CAD)	NA	166	< \$23,999: 1 (0.6%); \$24,000–\$49,999: 11 (6.6%); \$50,000–\$79,999: 24 (14.5%); \$80,000–\$124,999: 48 (28.9%); ≥ \$125,000: 82 (49.4%)	NA	170	< \$23,999: 4 (2.4%); \$24,000–\$49,999: 4 (2.4%); \$50,000–\$79,999: 15 (8.8%); \$80,000–\$124,999: 67 (39.4%); ≥ \$125,000: 80 (47.1%)
Family religious level <sup>a</sup>	1–5	175	3.50 (1.07)	1–5	181	3.50 (1.10)
Ethnicity	NA	175	European origins: 73 (41.7%); Non-European origins: 57 (32.6%); Multirethnic origins: 45 (25.7%)	NA	181	European origins: 65 (35.9%); Non-European origins: 63 (34.8%); Multirethnic origins: 53 (29.3%)
Participating parent gender <sup>b</sup>	NA	175	Female: 148 (84.6%); Male: 27 (15.4%)	NA	181	Female: 149 (82.3%); Male: 32 (17.7%)
Family religion	NA	175	No religion: 29 (16.6%); Roman Catholic: 47 (26.9%); Protestant: 6 (3.4%); Christian: 24 (13.7%); Muslim: 12 (6.9%); Buddhist: 2 (1.1%); Hindu: 8 (4.6%); Sikh: 3 (1.7%); Multiple religions: 40 (22.9%); Other religion: 4 (2.3%)	NA	181	No religion: 38 (21%); Roman Catholic: 53 (29.3%); Protestant: 6 (3.3%); Christian: 24 (13.3%); Muslim: 11 (6.1%); Jewish: 3 (1.7%); Hindu: 9 (5%); Sikh: 2 (1.1%); Multiple religions: 33 (18.2%); Other religion: 2 (1.1%)

*Note.* For education level and income, different response categories were used for the Hong Kong and Canadian samples because these categories were more reasonable and more commonly used in surveys of each region. CAD = Canadian dollar; NA = not applicable.

<sup>a</sup>Family religious level: 5 = extremely; 4 = very; 3 = moderately; 2 = slightly; 1 = not at all. <sup>b</sup> Gender refers to the birth-assigned gender.

completed a demographic questionnaire and other questionnaires as part of a larger study. At the same time, the child completed the study with a GC female researcher. Children viewed one of two slideshows depending on whether they were assigned to the intervention or control condition (for detailed procedures and results relating to the conditions, please refer to the online supplemental material A). The intervention was designed to influence the appraisals but not the reasoning styles, and supplementary analyses confirmed that assignment to the intervention versus control condition had no effect on the reasoning styles. Thus, condition was only included as a control variable in this report. After viewing the intervention or control condition slides, the test phase of the procedure commenced. In the test phase, children viewed four vignettes, each describing one of four targets (i.e., GC boy, GN boy, GC girl, and GN girl) in a random order, followed by an interview on their appraisals and reasoning. Each vignette was 75 s long and consisted of five illustrations lasting 15 s each. Each illustration was displayed with prerecorded audio narratives. The illustrations introduced the target's name and grade (the same grade as the participant), the target's preferences for toys, activities, clothing and hairstyle, and gender of playmates. To control for the possible racial bias, the target's skin tone was matched to the participant's complexion of White, East Asian, Brown, and Black individuals. Samples of slideshows in the control and intervention conditions, as well as the vignettes and associated scripts, are provided in the online supplemental material B.

### Reasoning About Moral Judgments

To assess moral judgments, the Canadian version asked the child participants "Was what (name of the target) was doing in the story wrong?" for each target in the vignettes. Children answered on a 3-point scale of *no* (*not wrong*), *don't know*, and *yes* (*wrong*) with corresponding emojis (i.e., smiling face, neutral face, and frowning face) as illustrations. As a matter of translation that considered what sounded more natural in daily conversations, we phrased the question in the Hong Kong version as "Was what (name of the target) was doing in the story right?" Children could respond with one of three options, *yes* (*right*), *don't know*, or *no* (*not right*). After children answered this appraisal question, we asked a follow-up reasoning question: "Why do you think what (name of the target) was doing in the story is wrong (not right) or not wrong (right)?"

### Coding Scheme Development and Reliability

We first pilot-tested the coding scheme of a prior study on children's reasoning in justifying correctness of behaviors (Smetana, 1986) on 80 participants. However, this existing coding scheme did not fit our data well. Many responses could not be unambiguously assigned to a coding category, resulting in low interrater reliability. We thus developed, taking reference of the existing scheme of (Smetana, 1986), a new coding scheme to better fit our data. We developed the preliminary coding schemes for this study using inductive content analysis, a systematic method to describe the meaning of and identifying themes in qualitative data (Elo & Kyngäs, 2008). It is a widely used research technique in which direct information can be grounded on actual data (Hsieh & Shannon, 2005). We proposed coding categories and definitions after screening the actual data, conducted the initial coding and returned to examine the data repeatedly to develop appropriate categories that closely reflected the data. One Hong Kong researcher and two

Canadian researchers conducted the coding of the first 40 participants of the Canadian sample independently using the preliminary coding schemes. The reliabilities were low, with  $\kappa$  values of most codes falling below the conventional acceptable value of .80 (McHugh, 2012). The discrepancies among the three coders were discussed and some categories were revised to better classify the responses. The revised coding scheme was applied for the next 40 participants of the Canadian sample following the previously described coding process. The reliabilities improved and the  $\kappa$  values of most codes were around .80. After discussion of the initial coding results of the first 80 participants mentioned above, the coding scheme was finalized.

In the final stage, two Hong Kong researchers fluent in both Chinese and English coded all the data independently. Coding for each response from a child was either 0 (*did not use the reasoning style*) or 1 (*did use the reasoning style*) in each reasoning style category. Multiple coding of reasoning styles was allowed for each response such that one or more reasoning style categories could be coded as "1." For moral judgments of the four targets, the  $\kappa$  values of the Hong Kong data ranged from .88 to .92, whereas those for the Canadian data ranged from .86 to .91. Thus, all  $\kappa$  values were good (i.e., >.80). When there were discrepancies between the two coders, a third researcher fluent in both Chinese and English made the final judgment. Table 3 shows the finalized coding scheme.

### Translation

Materials such as script of vignettes, questionnaires, and coding scheme were originally in English. To prepare a Chinese version of materials for Hong Kong children, Karen Man Wa Kwan, Wang Ivy Wong and an assistant with a translation degree translated the materials from English to Chinese. All of these researchers are from Hong Kong, proficient in both Chinese and English, and are familiar with Hong Kong culture.

### Statistical Analyses

We first investigated the frequency of each reasoning style in order to identify the most frequently used reasoning styles. This addressed our first hypothesis that gender stereotype-related reasoning, as well as concern for others' welfare, global standards, and personal choice would be among the top reasoning styles. Because the coding scheme included many different reasoning styles (i.e., 18), we focused on the five most frequently used reasoning styles. Then, for the remaining analyses, we considered only the top five for parsimony.

To discern whether these reasoning styles depended on participants' perception of the correctness of targets, we included participant's judgment (i.e., participant's response of whether the target is right or wrong) as a predictor in the analyses. This addressed the rest of our first set of hypotheses, which predicted that children used different kinds of reasoning styles depending on whether they perceived the targets as right or wrong.

To examine whether and how reasoning style was related to participant age group, gender, and culture, we included these variables and their interaction terms as predictors in the analyses. This addressed the second, third, and fourth sets of hypotheses.

As each participant answered the same question for each of the four targets, making it a repeated measures design, we used generalized linear mixed-effects models to account for both within- and between-person variability (Bonate, 2006). The fixed effects

**Table 3**  
*Coding Scheme and Percentage of Children Using Each Reasoning Style*

Percentage of children using this reasoning style (%)	Category	Description	Example(s)
23.4	Global standard	Behaving in accordance with existing social rules/norms (excluding gender norms) Global or generic descriptions of an act as good/bad, right/wrong	What he/she does is normal; he/she is good in general
21.2	Personal choice	Appeal to individual preferences/liking	It is his/her own choice; he/she is doing something that he/she likes
12.2	Gender stereotypes	References to gender roles or gender stereotype-related reasoning	He did girls' things/she did boys' things
12.2	Don't know	Participants cannot think of any reason for the target's correctness/wrongness	I don't know
10.8	Others' welfare	Appeal to the interests of persons other than the target	He/she did not upset/hurt/disturb others
6.6	Mere repetition	Mere repetition of what the target is doing without making any comment	She puts on mom's old dress
3.8	Perceived emotion	References to the target's emotion(s)	He/she is happy
2.8	Prudential reasons	References to nonsocial positive/negative consequences to the target, e.g., personal comfort, health	What he/she is doing is dangerous
2.6	Self-reference	References to the self, e.g., explicitly express that they like the target because the target does something the participants like	He/she is playing something that I like; I am a boy, so I support the boy
2.1	Authority's approval	References to the approval/disapproval or positive/negative reaction(s) (e.g., rewards/punishment) of specific authority figures (e.g., parents, teachers)	Mom/dad doesn't allow me to do so
1.4	Other-reference	References to other people or characters in the vignette	His name is the same as my best friend's name
1	No answer	Say nothing	
0.2	Peer approval	References to the positive/negative responses of peers	Other children won't play with him/her
0.2	Superficial environment element	References to the environment's superficial elements	The car is not of the same color
0.1	Superficial element (not about subjective liking)	Merely mention the target's superficial elements (e.g., appearance, name)	She is beautiful
0.1	Gender of the target	Perceive the target as right/wrong simply because of the gender of the target	He is wrong because he is a boy
0.1	Self-reference (without reason)	References to the self without providing reason(s)	I like him/her
9.2	Others	Any other reason(s)	He/she is different/odd

included between-person variables (i.e., participant's judgment, age group, gender, culture, and condition), within-person variables (i.e., target gender and target gender expression), and several interaction terms including interactions pertaining to gender (i.e., Target Gender  $\times$  Target Gender Expression), interactions to further control for the possible influence of condition (i.e., Culture  $\times$  Condition and Condition  $\times$  Age Group), interactions to explore the possible accumulated socialization effects with age (i.e., Target Gender Expression  $\times$  Age Group, Target Gender  $\times$  Age Group, Gender  $\times$  Age group, Culture  $\times$  Age Group), and interactions to explore the possible accumulated socialization effects by culture with age (i.e., Culture  $\times$  Target Gender Expression  $\times$  Age Group, Culture  $\times$  Target Gender  $\times$  Age Group, Culture  $\times$  Gender  $\times$  Age Group). In the post hoc comparisons following up on the significant findings involving participants' judgment, we only compared children who responded that the target's behavior was either right or wrong (i.e., children who responded "don't know" were not included in the comparisons), which was our main interest. The random effect was participant. The dependent variable in each model was the use of each of the top five reasoning styles. All means presented were adjusted means. As 678 participants answered the question for the four targets (i.e., each participant gave four responses), 2,712 responses were analyzed in each model.

For the purpose of multiple testing correction, we adopted a stringent alpha (i.e., critical  $\alpha$  of .001) for all the analyses. Table 4 summarizes the results of hypothesis testing and Table 5 summarizes the significant findings without a hypothesis. Because none of the demographic variables correlated with the dependent variables for all four target peers (i.e., GC boy, GN boy, GC girl, and GN girl), they were not entered as covariates. Table S1 in the online supplemental material C shows the correlational analysis of the demographic and dependent variables. In the Results section, we first report the frequency of the top five reasoning styles, and then organize the rest of the findings by these reasoning styles.

Other aspects of our larger study explored the appraisals and intervention effect aimed at influencing children's appraisals of GN peers. Results on developmental patterns of appraisals of peers and the intervention effects on such appraisals were detailed in our prior reports (Kwan et al., 2020; MacMullin et al., 2020; Nabbijohn et al., 2020). As the effect of the intervention on reasoning styles was not a focus of the current study and we did not find the reasoning styles to differ across conditions, any results pertaining to reasoning styles across conditions are presented in the online supplemental material A. The data that support the findings of this study are available from the corresponding authors upon reasonable request.

**Table 4**  
Summary of Hypothesis Testing Results

Variable	Hypothesis	Result
Frequency	H1a: Gender stereotype-related reasoning, as well as reasoning reflecting moral concerns (e.g., others' welfare), personal choice, and global standards would be among the most commonly used reasoning styles	Supported
Correctness of targets	H1b: Children who believed that the target was wrong would reason based on gender stereotypes more than children who believed that the target was right	Supported
	H1c: Children who believed that the target was right would reason based on personal choice more than children who believed that the target was wrong	Supported
Age	H2a: Older children would reason based on gender stereotypes more than younger children	Supported
	H2b: Older children would reason based on personal choice more than younger children	Supported
Gender	H3a: Boys would reason based on gender stereotypes more than girls	Not Supported
	H3b: Children would reason based on gender stereotypes more when appraising GN children, particularly GN boys, than GC children	Supported
Culture	H4a: Hong Kong children would reason based on gender stereotypes more than Canadian children	Supported in older children
	H4b: Hong Kong children would reason based on others' welfare more than Canadian children	Supported
	H4c: Canadian children would reason based on personal choice more than Hong Kong children	Supported

## Results

In the combined sample, the top five reasoning styles were (a) global standard (23.4%; e.g., "what he/she does is normal"), (b) personal choice (21.2%; e.g., "it is his/her own choice"), (c) gender stereotypes (12.2%; e.g., "he did girls' things"), (d) "don't know" (12.2%; e.g., "I don't know"), and (e) others' welfare (10.8%; e.g., "he/she did not upset others"). Table 3 shows the percentages of usage of all reasoning styles in the combined sample. These findings supported H1a that gender stereotype-related reasoning, global standard, others' welfare, and personal choice would be commonly related to children's appraisals.

In Figure 1, children's top five reasoning styles were compared across participants' different judgments of the targets' correctness (1a), participant age group (1b), participant gender (1c), target gender and gender expression (1d), and culture (1e). Figure 2 shows the follow-up for the significant interaction between culture and age for the gender stereotypes reasoning style. Tables 4 and 5 summarize all the hypothesis testing results. For condition, please refer to Figure S1 in the online supplemental material A.

Table S2 in the online supplemental material C shows percentages of usage in children for all reasoning styles by judgment of correctness, age group, child's gender, target's gender, culture, and condition. Table S3 in the online supplemental material C shows the correlation matrix among the top five reasoning styles. Table S4 in the online supplemental material C shows the  $F$  value,  $df$ ,  $p$  value, and post hoc test (if any) for each fixed effect.

### Global Standard

There was a main effect of participant's judgment,  $F(2, 2690) = 6.64, p < .001$ . Children who responded that the target's behavior was right were more likely to use global standard than children who

responded that the target's behavior was wrong,  $t(590) = -3.31, p < .001, OR = 2.19, M_{\text{right}} = 0.22, M_{\text{wrong}} = 0.11$ . There was no main effect of participant age, participant gender, target gender, target gender expression, culture, and no interaction.

### Personal Choice

There was a main effect of participant's judgment,  $F(2, 2690) = 21.46, p < .001$ . Children who responded that the target's behavior was right were more likely to use personal choice than children who responded that the target's behavior was wrong,  $t(565) = -4.80, p < .001, OR = 7.81, M_{\text{right}} = 0.12, M_{\text{wrong}} = 0.02$  (supporting H1c). Also, there was a main effect of age group. Older children were more likely to use personal choice than younger children,  $F(1, 2690) = 53.53, p < .001, OR = 16.13, M_{\text{younger}} = 0.01, M_{\text{older}} = 0.09$  (supporting H2b). Moreover, there was a main effect of culture. Canadian children were more likely to use personal choice than Hong Kong children,  $F(1, 2690) = 51.32, p < .001, OR = 6.10, M_{\text{Hong Kong}} = 0.01, M_{\text{Canadian}} = 0.09$  (supporting H4c). There was no main effect of participant gender, target gender, target gender expression, and no interaction.

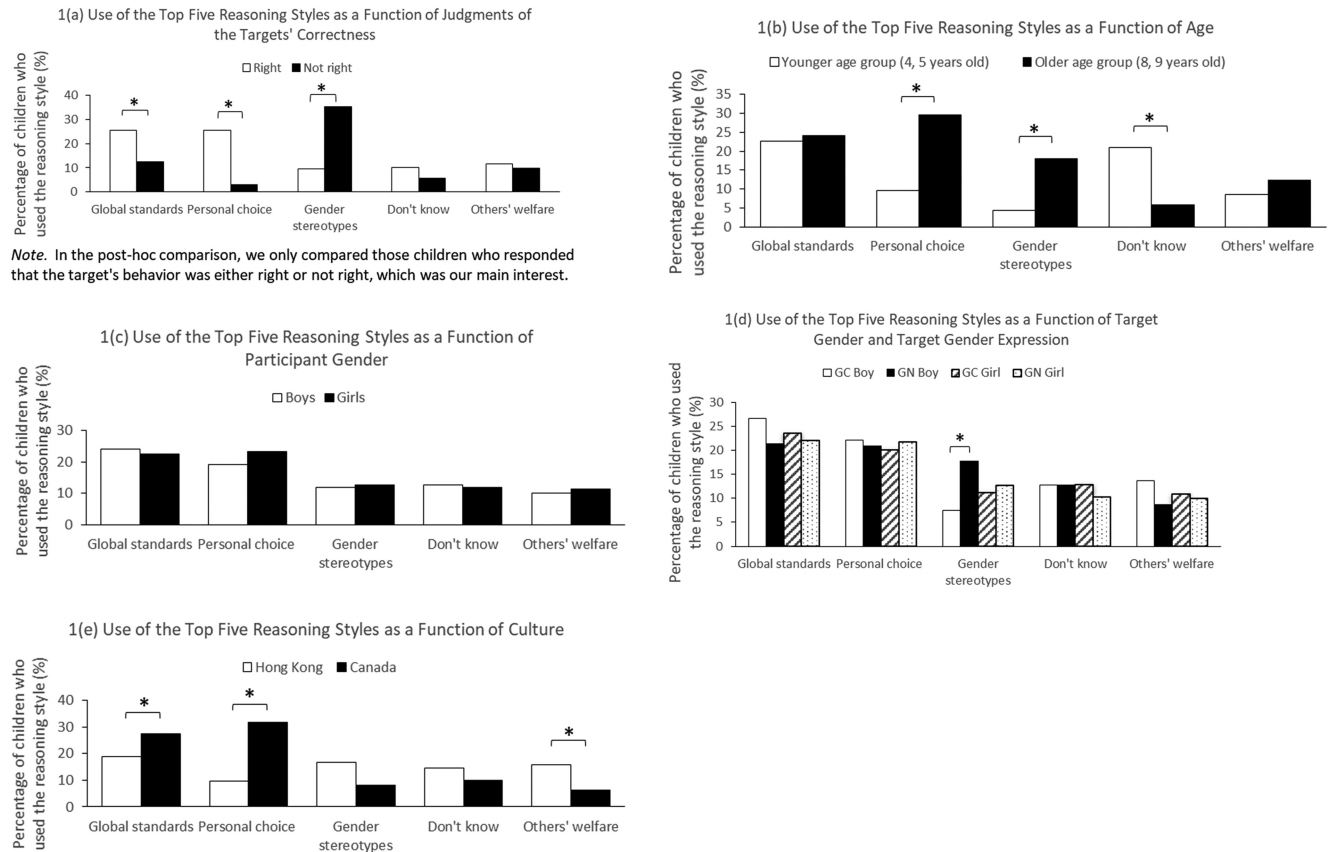
### Gender Stereotypes

There was a main effect of participant's judgment,  $F(2, 2690) = 20.63, p < .001$ . Children who responded that the target's behavior was wrong were more likely to use gender stereotype-related reasoning than children who responded that the target's behavior was right,  $t(298) = 6.13, p < .001, OR = 4.43, M_{\text{right}} = 0.04, M_{\text{wrong}} = 0.14$  (supporting H1b). Also, there was a main effect of age group. Older children were more likely to use gender stereotype-related reasoning than younger children,  $F(1, 2690) = 43.38, p < .001, OR = 8.20, M_{\text{younger}} = 0.02, M_{\text{older}} = 0.15$  (supporting H2a). Further, there was an interaction between culture and age group,  $F(1, 2690) = 10.29, p < .001$ . Post hoc analyses were conducted within each age group. For older children, Hong Kong children were more likely to use gender stereotype-related reasoning than Canadian children,  $F(1, 1564) = 14.56, p < .001, OR = 2.44, M_{\text{Hong Kong}} = 0.24, M_{\text{Canada}} = 0.11$ . No significant cultural comparison was found for the younger age group,  $F(1, 1132) = .97, ns$ . Additional post hoc analyses were conducted within each culture. Older children were

**Table 5**  
Summary of Findings Without a Hypothesis

Variable	Finding
Correctness of targets	Children who believed that the target was right reason based on a global standard more than children who believe that the target was wrong
Age	Younger children reasoned based on "don't know" more than older children

**Figure 1**  
**Children's Use of the Top Five Reasoning Styles in Moral Judgments**



**Note.** (a) Use of the top five reasoning styles as a function of judgments of the targets' correctness. In the post hoc comparison, we only compared those children who responded that the target's behavior was either right or not right, which was our main interest. (b) Use of the top five reasoning styles as a function of age. (c) Use of the top five reasoning styles as a function of participant gender. (d) Use of the top five reasoning styles as a function of target gender and target gender expression. (e) Use of the top five reasoning styles as a function of culture.

\*  $p < .001$ .

more likely to use gender stereotype-related reasoning than younger children in both Hong Kong,  $F(1, 1272) = 45.86, p < .001, OR = 16.13$  and Canada,  $F(1, 1424) = 12.64, p < .001, OR = 3.28$ . Moreover, there was an interaction between target gender expression and target gender,  $F(1, 2690) = 13.75, p < .001$ . Post hoc analyses were conducted within each target gender. When appraising target boys, children were more likely to use gender stereotype-related reasoning when evaluating the GN boy than the GC boy,  $F(1, 1348) = 18.90, p < .001, OR = 2.38, M_{\text{target GC boy}} = 0.07, M_{\text{target GN boy}} = 0.14$  (supporting H3b). No significant difference was found for target girls. There was no main effect of participant gender, target gender, or culture.

### Don't Know

There was a main effect of participant's judgment,  $F(2, 2690) = 36.79, p < .001$ ; but, when comparing only those children who responded the target's behavior was right or wrong, no significant difference was found. Also, there was a main effect of age group. Younger children were more likely to use "don't know" than older children,  $F(1, 2690) = 19.27, p < .001, OR =$

$5.28, M_{\text{younger}} = 0.11, M_{\text{older}} = 0.03$ . There was no main effect of participant gender, target gender, target gender expression, culture, and no interaction.

### Others' Welfare

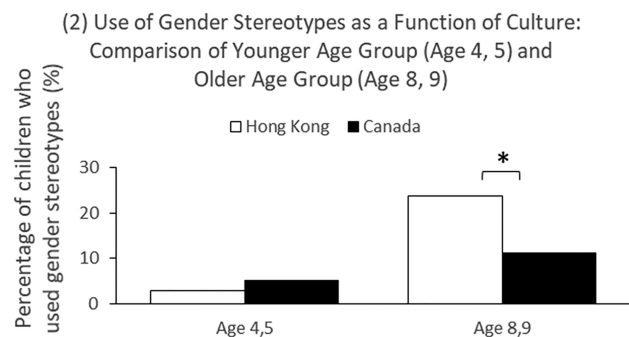
There was a main effect of culture. Hong Kong children were more likely to use others' welfare than Canadian children,  $F(1, 2690) = 27.90, p < .001, OR = 1.33, M_{\text{Hong Kong}} = 0.13, M_{\text{Canadian}} = 0.03$  (supporting H4b). There was no main effect of participant judgment, age, participant gender, target gender, target gender expression, and no interaction.

### Post Hoc Analysis on the Use of "Don't Know" Reasoning Style

"Don't know" was among the top five reasoning styles and younger children responded with "don't know" more than older children did. To examine the possibility that the findings that older children used more personal choice and gender stereotype-related reasoning were not merely an artefact of younger children's greater tendency

**Figure 2**

*Children's Use of the Reasoning Style Gender Stereotypes as a Function of Culture: Comparison of Younger Age Group (Age 4, 5) and Older Age Group (Age 8, 9) in Moral Judgments*



\*  $p < .001$ .

to respond with “don’t know”, the significant main effects of age were examined when excluding those children who used the reasoning style “don’t know”. Results showed that the main effects of age were still present, suggesting that the age differences were not merely the result of younger children’s tendency to use the reasoning style “don’t know” more than older children. For details, please refer to Table S5 in the online supplemental material C.

## Discussion

Prior studies have shown that children appraise peers who violate gender norms less positively than those who conform (Blakemore, 2003; Kwan et al., 2020; Levy et al., 1995; Nabbijohn et al., 2020; Y. Wang et al., 2022; Zucker et al., 1995). However, little is known about how children explain their appraisals, an avenue that may help clarify the root of children’s bias against gender nonconformity and, thus, inform the appropriate target(s) of intervention strategies. This study extended prior literature by examining children’s reasoning styles when appraising peers of varying gender and gender expressions using a large sample and examining reasoning in moral judgments, a core aspect of appraisal examined in studies of children’s appraisals of peers. Apart from testing important variables in previous studies, including participant gender, age, and target gender and target gender expression, we gained valuable insights into cultural variations by examining children’s reasoning across two cultures. Of note, the present study employed a much larger child sample ( $N = 678$ ) than other available studies on reasoning about gender norms across different cultural samples (Conry-Murray et al., 2015,  $N = 144$  children; Conry-Murray et al., 2020;  $N = 151$  children). Importantly, our findings buttress prior work indicating gender stereotypes are a salient consideration in children’s judgments (Martin & Halverson, 1981; Martin & Ruble, 2004; Martin et al., 2002), which may inform the basis of children’s appraisals of GC and GN peers. Overall, when children were older and from Hong Kong, appraising the GN boy, or perceived the target as behaving in a manner that was wrong, they were more likely to reason based on gender stereotypes. On the contrary, when children were from Canada, or perceived the target as right, they were more likely to reason based on personal choice.

As hypothesized (H1a), gender stereotype-related reasoning was one of the top five reasoning styles. Apart from gender stereotype-related

reasoning, children’s reasoning often cited global standard, personal choice, others’ welfare, and “don’t know”. The extent to which children emphasized these elements depended on various factors, which we detail in the subsections below.

## Moral Judgment

Children’s reasoning styles were related to whether they perceived the target’s behavior as right or wrong. Children who responded that the target was wrong reasoned based on gender stereotypes (supporting H1b) more than children who responded that the target was right. On the other hand, children who responded that the target was right reasoned based on personal choice (supporting H1c) and global standard more than children who responded that the target was wrong. These findings are consistent with prior research suggesting that children reason based on gender stereotypes when they support gender-stereotypical expression only (Killen et al., 2001; Theimer et al., 2001), and on personal choice when they are positive toward different gender expressions (Conry-Murray, 2013; Conry-Murray & Turiel, 2012). We further showed that global standard, which refers to judging behaviors in accordance with existing social conventions, is another reasoning style related to positivity toward varying gender expressions. Overall, when children focus on gender norms, they were more likely to believe that gender nonconformity was wrong. In contrast, when children explained that gender expression was a personal choice or believed that a certain gender expression is acceptable by social conventions, they more often perceived those gender expressions as right. However, it is also possible that global standard and personal choice happened to be the most parsimonious explanations for children to give to explain their positive perception of a peer’s behavior. Experimental studies are needed to address these possibilities in the future.

## Age Differences

Older children (i.e., 8- to 9-years-old) reasoned based on gender stereotypes more than younger children (i.e., 4- to 5-years-old) (supporting H2a). Social-cognitive theories propose that gender socialization occurs throughout children’s development (Martin et al., 2002). Prior studies suggested that older children were less positive toward gender nonconformity than younger children (Blakemore, 2003; Carter & McCloskey, 1984; Kwan et al., 2020; Nabbijohn et al., 2020). The decrease in positivity may be related to the increase in gender segregation from early childhood to middle childhood (Leaper, 2022; Mehta & Strough, 2009), and an increase in gender-typed behaviors among same-gender peers and peer scrutiny of GN behaviors (Kowalski, 2007; Martin & Fabes, 2001). Our findings further suggest that reasoning based on gender stereotypes might be associated with older children’s lesser positivity toward peers’ GN behaviors.

Our results also showed that older children reasoned based on personal choice more than younger children (supporting H2b). Similar to our findings, a prior study showed that older children aged 6 and 8 years old (but not younger children aged 4 years old) were more likely to reason based on personal choice than other reasoning styles (Conry-Murray & Turiel, 2012). Based on the gender schema theory (Martin & Halverson, 1981; Martin & Ruble, 2004; Martin et al., 2002), children’s gender schemas become more complex as they encounter and assimilate more diverse examples (e.g., both GC and GN examples) over time. The more refined schemas may

enhance cognitive flexibility, the tendency to assign gender-stereotypical attributes to both genders rather than exclusively to one gender, which also increases with age (Levy et al., 1995; Trautner et al., 2005). Increased cognitive flexibility might contribute to the greater use of the personal choice reasoning style in older children. Another possible explanation is the development of theory of mind (ToM), which consolidates throughout early childhood (Wellman et al., 2001). Increased ToM with age may help older children perceive the perspectives of peers with varying gender expressions and show greater understanding of different behavioral choices (e.g., GC and GN behaviors). Consistent with this notion, prior research showed that children who have developed ToM were more supportive of challenging gender stereotypes than those who have not developed ToM (Mulvey et al., 2016).

Of note, that older children reasoned more based on gender stereotypes but also respected personal choice more appears contradictory; however, increasing use of personal choice may suggest increased cognitive flexibility (Trautner et al., 2005) and/or ToM, which may not necessarily mean increased positivity toward gender nonconformity (Carter & McCloskey, 1984; Kwan et al., 2020; Nabbijohn et al., 2020). Cognitive flexibility is usually assessed by asking children about the ability or possibility for both boys and girls to perform a certain activity (e.g., boys can also play with dolls). On the other hand, positivity toward gender nonconformity can be assessed by asking children questions such as their liking of or intention to befriend the peers (e.g., I want to play with or befriend GN peers; Levy et al., 1995). It is possible that increased knowledge of the possibility for boys and girls to engage in the same activities may not necessarily translate to more positive attitude, at least in 4- to 9-year-old children.

## Gender and Gender Expression

Children reasoned based on gender stereotypes more when judging the correctness of the GN boy than the GC boy (supporting H3b), but no such difference was found between the GN girl and the GC girl. However, the hypothesis (H3a) that boys reasoned based on gender stereotypes more than girls was not supported. Thus, the use of gender stereotype-related reasoning differed by target gender, but not by participant gender.

Breaking gender norms is more socially costly for boys than for girls (Leaper, 1994). This is consistent with the belief that boys should not display feminine behaviors (Braun & Davidson, 2017; Wallien et al., 2010) and stricter gender boundaries tend to be imposed on boys than girls by peers (Witt, 2000) and parents (Kane, 2006). Our findings further suggest children used more gender stereotype-related reasoning for GN boys than GC boys. This might contribute to children's lesser positivity toward GN boys. Also, prior studies suggested that both boys and girls were less positive toward GN boys than GN girls (Kwan et al., 2020; Nabbijohn et al., 2020; Qian et al., 2021). As speculated, if lesser positivity toward GN boys were related to more usage of gender stereotype-related reasoning, it is likely that both boys and girls would use more gender stereotype-related reasoning when appraising GN boys, which might explain why no participant gender difference was found.

## Culture

Our findings showed that Hong Kong children reasoned based on gender stereotypes more than Canadian children, but only among the

older age group of children (partially supporting H4a). According to social-cognitive theories, children are socialized to learn gender norms throughout their development (Martin et al., 2002). Influenced by socialization agents over time, cultural differences might appear later in children's development due to more accumulated socialization effects in older children than younger children. Certain cultural values in China and Canada may contribute to differences in gender socialization practices and thus differences in reasoning related to gender stereotypes. Chinese culture emphasizes Taoism, such that women (yin) and men (yang) play their own role as nature dictates them to maintain a balance (R. R. Wang, 2005). Also, gender stereotypes are more polarized in Chinese culture (Z. Chen et al., 2009; Lii & Wong, 1982; Zhang & Cartwright, 2009) and Chinese culture emphasizes conforming to norms (Chao, 1995; X. Chen et al., 2003). Consistent with these cultural differences, Hong Kong children were less positive toward gender nonconformity than Canadian children (Kwan et al., 2020; MacMullin et al., 2020; Nabbijohn et al., 2020). The parallel difference in gender stereotype-related reasoning reported here suggests that such reasoning might contribute to Hong Kong children's lesser positivity toward GN peers. In line with our findings, Conry-Murray et al. (2015) compared children in the United States and Korea (a culture which shares many cultural elements with China, such as Confucianism) and found some evidence that Korean children used more gender-related reasoning than U.S. children. That said, another study found that Thai children were not biased against GN peers although Thai culture has been characterized as showing strong gender stereotyping (Y. Wang et al., 2022). Therefore, future work is required to further examine the relationship between cultural emphasis on gender stereotypes and the development of gender bias.

Our findings also showed that Hong Kong children reasoned more based on others' welfare (supporting H4b) and Canadian children reasoned more based on personal choice (supporting H4c). Chinese culture emphasizes relationship building (Chao, 1995; X. Chen et al., 2003), whereas Canadian culture is associated with less in-group favoritism and greater focus on individual interest (Y.-R. Chen et al., 1998). In Chinese culture, harmonious social relationships are especially encouraged by parents (Chan et al., 2009; Liu et al., 2005), controlling teaching styles are usually preferred by teachers (Zhou et al., 2012), and collaborative learning is usually favored by students (Ramburuth & McCormick, 2001; Tang, 1996). Cultural differences in values and practices such as these perhaps explain the cultural differences found here regarding concern for others' welfare and personal choice reasoning styles.

## Implications

In light of the age-, gender-, and cultural-related differences in how children explained their appraisals of peers' gender expression and the relationships these explanations had with judging peers positively or negatively, guiding children to consider certain reasoning styles may be relevant to intervention strategies. For example, guiding children to focus on alternative reasoning styles (e.g., personal choice) instead of gender labels and stereotypes may lead to more positive appraisals of gender nonconformity. This element can potentially be incorporated into prior interventions aimed at reducing sexism by teaching children to identify and confront sexist remarks (Lamb et al., 2009; Pahlke et al., 2014). Also, our study found that there was an increasing use of gender stereotype-related

reasoning with age. Prior study examining children aged 4–8 years old found that less gender stereotypes are related to the preference for playing with GN peers, especially for younger children (Sims et al., 2022). This suggested the importance to teach children about issues associated with gender stereotypes from a young age in order to prepare them to recognize and confront gender stereotypes. Regarding cultural differences, our findings suggest that Chinese children tend to emphasize others' welfare. Interventions for Chinese children may thus focus on relationship building with GN peers, and how diversity and inclusion contribute to group harmony. On the other hand, Canadian children tend to emphasize personal choice. Interventions for Canadian children may focus on individuals' personal freedom to choose their own preferred behaviors. Tailoring interventions based on specific cultural values can potentially enhance the effectiveness of the interventions in reducing bias against gender nonconformity.

### Limitations and Future Directions

Prior studies have shown the consistency of gender stereotypes across Chinese and Western cultures in general (Wong & VanderLaan, 2020; Yu et al., 2010), such as the concepts of masculinity and femininity (Cheung, 1996), toy play (Qian et al., 2023; Todd et al., 2018), and color preferences (Yeung & Wong, 2018). Despite the evidence supporting the similarity in gender socialization in Chinese and Western cultures, we acknowledge that there are limitations in making cross-cultural comparisons between Hong Kong and Canadian children in our study. First, we did not include a direct measure of culture or the gender socialization and social-cognitive processes; including these measures in future research will help reveal the source of heterogeneity that exists within and across cultural contexts. Second, the hour-long face-to-face testing procedures precluded recruitment from a population-based sample. Our sample may thus not be ideal in representing the general population in Hong Kong and Canada. However, the Hong Kong sample was recruited from different schools and districts to increase the likelihood that the families would resemble typical families with young children. The Canadian sample had average household income (CAD \$80,000–124,999) comparable to the general local population in Mississauga (Statistics Canada, 2016). The large sample ( $N = 678$ ) and inclusion of two cultures in fact make this study more representative than most studies of its kind. In addition, we found that no demographic variable considered here correlated with the dependent variables—although we acknowledge the limitation that other demographic variables not considered may have influenced the results. Third, we acknowledge the limitation about the stimuli used in the vignette. One item, that is, tackle football, is more common in Western culture than Chinese culture. However, it is worth noting that various ball games including basketball and football are male dominated in Hong Kong as well (Leisure and Cultural Services Department, 2009). Tackle football is a rough and tumble type of game, which is likely to be perceived as a male dominated activity by both Canadian and Chinese participants.

Moreover, this research was conducted in controlled laboratory settings. To increase ecological validity, future research can consider naturalistic settings (e.g., classrooms). Another limitation was that younger children's verbal skills may have limited their articulation of their reasoning. This may have contributed to the higher frequency of "don't know" responses in younger children. That said, post hoc

analyses showed that age differences reported here were not due to younger children's greater tendency to respond with "don't know" than older children. Further, the large majority of children were able to explain their appraisals with a codable reasoning style. In any case, future research might control for children's verbal skills.

### Conclusion

We extended previous studies (Kwan et al., 2020; MacMullin et al., 2020; Nabbijohn et al., 2020) on children's appraisals of peers gender expression by examining the reasoning styles children used to justify their moral judgments. Of note, we employed a large cross-cultural sample, which is rare for this literature. Children used various reasoning styles, with gender stereotype-related reasoning being one of the most commonly used styles. Gender stereotype-related reasoning was associated with less favorable moral judgments of peers' gender-related behaviors, whereas personal choice was associated with more favorable ones. Thus, these reasoning styles may serve as important aspects of children's thinking to target in any interventions aimed at increasing children's positivity toward peers of varying gender expressions. Other important variables related to children's reasoning styles were age, gender of peers, and culture. Older children reasoned more based on gender stereotypes and personal choice than younger children; children reasoned more based on gender stereotypes when appraising GN children, particularly GN boys; and Hong Kong children reasoned more based on others' welfare and less based on personal choice than Canadian children. In the older age group, Hong Kong children also reasoned more based on gender stereotypes than Canadian children. This set of findings on age, gender, and culture lend support to social-cognitive and gender schema perspectives that help guide understanding of the processes that influence children's thinking about peer gender (non)conformity. Importantly, the present study provides insights into the reasoning styles underpinning children's appraisals of varying peer gender expressions. This knowledge can inform future work focused on improving children's appraisals of peer gender nonconformity via strategies that target reasoning styles.

### References

- American Psychological Association. (2015). Guidelines for psychological practice with transgender and gender nonconforming people. *American Psychologist, 70*(9), 832–864. <https://doi.org/10.1037/a0039906>
- Bandura, A., & Bussey, K. (2004). On broadening the cognitive, motivational, and sociostructural scope of theorizing about gender development and functioning: Comment on Martin, Ruble, and Szkrybalo (2002). *Psychological Bulletin, 130*(5), 691–701. <https://doi.org/10.1037/0033-2909.130.5.691>
- Blakemore, J. E. O. (2003). Children's beliefs about violating gender norms: Boys shouldn't look like girls, and girls shouldn't act like boys. *Sex Roles, 48*(9–10), 411–419. <https://doi.org/10.1023/A:1023574427720>
- Bonate, P. L. (2006). Linear mixed effects models. In P. L. Bonate (Ed.), *Pharmacokinetic–pharmacodynamic modeling and simulation* (pp. 181–204). Springer.
- Braun, S. S., & Davidson, A. J. (2017). Gender (non)conformity in middle childhood: A mixed methods approach to understanding gender-typed behavior, friendship, and peer preference. *Sex Roles, 77*(1–2), 16–29. <https://doi.org/10.1007/s11199-016-0693-z>
- Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review, 106*(4), 676–713. <https://doi.org/10.1037/0033-295X.106.4.676>

- 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>
- Chan, S. M., Bowes, J., & Wyver, S. (2009). Chinese parenting in Hong Kong: Links among goals, beliefs and styles. *Early Child Development and Care*, 179(7), 849–862. <https://doi.org/10.1080/03004430701536525>
- Chao, R. K. (1995). Chinese And European American cultural models of the self reflected in mothers' childrearing beliefs. *Ethos*, 23(3), 328–354. <https://doi.org/10.1525/eth.1995.23.3.02a00030>
- Chen, X., Rubin, K., Liu, M., Chen, H., Wang, L., Li, D., Gao, X., Cen, G., Gu, H., & Li, B. (2003). Compliance in Chinese and Canadian toddlers: A cross-cultural study. *International Journal of Behavioral Development*, 27(5), 428–436. <https://doi.org/10.1080/01650250344000046>
- Chen, Y.-R., Brockner, J., & Katz, T. (1998). Toward an explanation of cultural differences in in-group favoritism: The role of individual versus collective primacy. *Journal of Personality and Social Psychology*, 75(6), 1490–1502. <https://doi.org/10.1037/0022-3514.75.6.1490>
- Chen, Z., Fiske, S. T., & Lee, T. L. (2009). Ambivalent sexism and power-related gender-role ideology in marriage. *Sex Roles*, 60(11–12), 765–778. <https://doi.org/10.1007/s11199-009-9585-9>
- Cheung, F. M. (1996). Gender role development. In S. Lau (Ed.), *Growing up the Chinese way: Chinese child and adolescent development* (pp. 45–67). The Chinese University Press.
- Cohen-Kettenis, P. T., Owen, A., Kaijser, V. G., Bradley, S. J., & Zucker, K. J. (2003). Demographic characteristics, social competence, and behavior problems in children with gender identity disorder: A cross-national, cross-clinic comparative analysis. *Journal of Abnormal Child Psychology*, 31(1), 41–53. <https://doi.org/10.1023/A:1021769215342>
- Conry-Murray, C. (2013). Children's reasoning about gender-atypical preferences in different settings. *Journal of Experimental Child Psychology*, 115(1), 210–217. <https://doi.org/10.1016/j.jecp.2012.09.007>
- Conry-Murray, C. (2015). Children's judgments of inequitable distributions that conform to gender norms. *Merrill-Palmer Quarterly*, 61(3), 319–344. <https://doi.org/10.13110/merrpalmqur1982.61.3.0319>
- Conry-Murray, C. (2017). Children's distributive justice: The role of gender norms in different settings. *European Journal of Developmental Psychology*, 14(1), 62–74. <https://doi.org/10.1080/17405629.2016.1152176>
- Conry-Murray, C., Kim, J. M., & Turiel, E. (2015). Judgments of gender norm violations in children from the United States and Korea. *Cognitive Development*, 35, 122–136. <https://doi.org/10.1016/j.cogdev.2015.04.002>
- Conry-Murray, C., Kim, J. M., & Turiel, E. (2020). Culture and children's reasoning about preferences and gender norms. *Journal of Experimental Child Psychology*, 196, Article 104861. <https://doi.org/10.1016/j.jecp.2020.104861>
- Conry-Murray, C., & Turiel, E. (2012). Jimmy's baby doll and Jenny's truck: Young children's reasoning about gender norms. *Child Development*, 83(1), 146–158. <https://doi.org/10.1111/j.1467-8624.2011.01696.x>
- Elenbaas, L. (2019). Against unfairness: Young children's judgments about merit, equity, and equality. *Journal of Experimental Child Psychology*, 186, 73–82. <https://doi.org/10.1016/j.jecp.2019.05.009>
- Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107–115. <https://doi.org/10.1111/j.1365-2648.2007.04569.x>
- Halim, M. L. D. (2016). Princesses and superheroes: Social-cognitive influences on early gender rigidity. *Child Development Perspectives*, 10(3), 155–160. <https://doi.org/10.1111/cdep.12176>
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. <https://doi.org/10.1177/1049732305276687>
- Kane, E. W. (2006). “No way my boys are going to be like that!” Parents' responses to children's gender nonconformity. *Gender & Society*, 20(2), 149–176. <https://doi.org/10.1177/0891243205284276>
- Killen, M., Mulvey, K. L., & Hitti, A. (2013). Social exclusion in childhood: A developmental intergroup perspective. *Child Development*, 84(3), 772–790. <https://doi.org/10.1111/cdev.12012>
- Killen, M., Pisacane, K., Lee-Kim, J., & Ardila-Rey, A. (2001). Fairness or stereotypes? Young children's priorities when evaluating group exclusion and inclusion. *Developmental Psychology*, 37(5), 587–596. <https://doi.org/10.1037/0012-1649.37.5.587>
- Killen, M., & Stangor, C. (2001). Children's social reasoning about inclusion and exclusion in gender and race peer group contexts. *Child Development*, 72(1), 174–186. <https://doi.org/10.1111/1467-8624.00272>
- Kowalski, K. (2007). The development of social identity and intergroup attitudes in young children. In O. Saracho & B. Spodek (Eds.), *Contemporary perspectives on social learning in early childhood education* (pp. 51–84). IAP.
- Kwan, K. M. W., Shi, S. Y., Nabbijohn, A. N., MacMullin, L. N., VanderLaan, D. P., & Wong, W. I. (2020). Children's appraisals of gender nonconformity: Developmental pattern and intervention. *Child Development*, 91(4), 780–798. <https://doi.org/10.1111/cdev.13316>
- Lamb, L. M., Bigler, R. S., Liben, L. S., & Green, V. A. (2009). Teaching children to confront peers' sexist remarks: Implications for theories of gender development and educational practice. *Sex Roles*, 61(5–6), 361–382. <https://doi.org/10.1007/s11199-009-9634-4>
- Leaper, C. (1994). Exploring the consequences of gender segregation on social relationships. *New Directions for Child and Adolescent Development*, 1994(65), 67–86. <https://doi.org/10.1002/cd.23219946507>
- Leaper, C. (2022). Origins and consequences of childhood gender segregation: Toward an integrative developmental systems model. In D. P. VanderLaan & W. I. Wong (Eds.), *Gender and sexuality development* (pp. 159–205). Springer. [https://doi.org/10.1007/978-3-030-84273-4\\_6](https://doi.org/10.1007/978-3-030-84273-4_6)
- Leisure and Cultural Services Department. (2009). *Consultancy study on sport for all—Participation patterns of Hong Kong people in physical activities*. <https://www.lcsd.gov.hk/en/sportforall/index.html>
- Levy, G. D., Taylor, M. G., & Gelman, S. A. (1995). Traditional and evaluative aspects of flexibility in gender roles, social conventions, moral rules, and physical laws. *Child Development*, 66(2), 515–531. <https://doi.org/10.2307/1131594>
- Lii, S.-Y., & Wong, S.-Y. (1982). A cross-cultural study on sex-role stereotypes and social desirability. *Sex Roles*, 8(5), 481–491. <https://doi.org/10.1007/BF00287714>
- Liu, M., Chen, X., Rubin, K. H., Zheng, S., Cui, L., Li, D., Chen, H., & Wang, L. (2005). Autonomy- vs. connectedness-oriented parenting behaviours in Chinese and Canadian mothers. *International Journal of Behavioral Development*, 29(6), 489–495. <https://doi.org/10.1177/01650250500147063>
- Lobel, T. E., Bar-David, E., Gruber, R., Lau, S., & Bar-Tal, Y. (2000). Gender schema and social judgments: A developmental study of children from Hong Kong. *Sex Roles*, 43(1/2), 19–42. <https://doi.org/10.1023/A:1007035611440>
- MacMullin, L. N., Bokeloh, L. M., Nabbijohn, A. N., Santarossa, A., van der Miesen, A. I., Peragine, D. E., & VanderLaan, D. P. (2021). Examining the relation between gender nonconformity and psychological well-being in children: The roles of peers and parents. *Archives of Sexual Behavior*, 50(3), 823–841. <https://doi.org/10.1007/s10508-020-01832-6>
- MacMullin, L. N., Nabbijohn, A. N., Kwan, K. M. W., Santarossa, A., Peragine, D. E., James, H. J., Wong, W. I., & VanderLaan, D. P. (2020). Testing an intergroup relations intervention strategy to improve children's appraisals of gender-nonconforming peers. *The Canadian Journal of Human Sexuality*, 29(2), 167–181. <https://doi.org/10.3138/cjhs.2020-0001>
- Martin, C. L. (1989). Children's use of gender-related information in making social judgments. *Developmental Psychology*, 25(1), 80–88. <https://doi.org/10.1037/0012-1649.25.1.80>
- Martin, C. L., & Fabes, R. A. (2001). The stability and consequences of young children's same-sex peer interactions. *Developmental Psychology*, 37(3), 431–446. <https://doi.org/10.1037/0012-1649.37.3.431>

- Martin, C. L., & Halverson, Jr, C. F. (1981). A schematic processing model of sex typing and stereotyping in children. *Child Development*, 52(4), 1119–1134. <https://doi.org/10.2307/1129498>
- Martin, C. L., & Ruble, D. (2004). Children's search for gender cues: Cognitive perspectives on gender development. *Current Directions in Psychological Science*, 13(2), 67–70. <https://doi.org/10.1111/j.0963-7214.2004.00276.x>
- Martin, C. L., Ruble, D. N., & Szkrybalo, J. (2002). Cognitive theories of early gender development. *Psychological Bulletin*, 128(6), 903–933. <https://doi.org/10.1037/0033-2909.128.6.903>
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282. <https://doi.org/10.11613/BM.2012.031>
- Mehta, C. M., & Strough, J. (2009). Sex segregation in friendships and normative contexts across the life span. *Developmental Review*, 29(3), 201–220. <https://doi.org/10.1016/j.dr.2009.06.001>
- Miller, C. F., Lurye, L. E., Zosuls, K. M., & Ruble, D. N. (2009). Accessibility of gender stereotype domains: Developmental and gender differences in children. *Sex Roles*, 60(11–12), 870–881. <https://doi.org/10.1007/s11199-009-9584-x>
- Mischel, W. (2015). A social learning view of sex differences in behavior. In V. Burr (Ed.), *Gender and psychology* (pp. 108–129). Routledge/Taylor & Francis Group.
- Mulvey, K. L., Rizzo, M. T., & Killen, M. (2016). Challenging gender stereotypes: Theory of mind and peer group dynamics. *Developmental Science*, 19(6), 999–1010. <https://doi.org/10.1111/desc.12345>
- Nabbijohn, A. N., MacMullin, L. N., Kwan, K. M. W., Santarossa, A., Peragine, D. E., Wong, W. I., & VanderLaan, D. P. (2020). Children's bias in appraisals of gender-variant peers. *Journal of Experimental Child Psychology*, 196, Article 104865. <https://doi.org/10.1016/j.jecp.2020.104865>
- Pahlke, E., Bigler, R. S., & Martin, C. L. (2014). Can fostering children's ability to challenge sexism improve critical analysis, internalization, and enactment of inclusive, egalitarian peer relationships? *Journal of Social Issues*, 70(1), 115–133. <https://doi.org/10.1111/josi.12050>
- Qian, M., Wang, Y., Wong, W. I., Fu, G., Zuo, B., & VanderLaan, D. P. (2021). The effects of race, gender, and gender-typed behavior on children's friendship appraisals. *Archives of Sexual Behavior*, 50(3), 807–820. <https://doi.org/10.1007/s10508-020-01825-5>
- Qian, M., Wong, W. I., Nabbijohn, A. N., Wang, Y., MacMullin, L. N., James, H. J., Fu, G., Zuo, B., & VanderLaan, D. P. (2023). Children's implicit gender-toy association development varies across cultures. *Developmental Psychology*, 59(12), 2287–2295. <https://doi.org/10.1037/dev0001590>
- Ramburuth, P., & McCormick, J. (2001). Learning diversity in higher education: A comparative study of Asian international and Australian students. *Higher Education*, 42(3), 333–350. <https://doi.org/10.1023/A:1017982716482>
- Roberts, A. L., Rosario, M., Slopen, N., Calzo, J. P., & Austin, S. B. (2013). Childhood gender nonconformity, bullying victimization, and depressive symptoms across adolescence and early adulthood: An 11-year longitudinal study. *Journal of the American Academy of Child & Adolescent Psychiatry*, 52(2), 143–152. <https://doi.org/10.1016/j.jaac.2012.11.006>
- Sims, R. N., Rizzo, M. T., Mulvey, K. L., & Killen, M. (2022). Desire to play with counterstereotypical peers is related to gender stereotypes and playmate experiences. *Developmental Psychology*, 58(3), 510–521. <https://doi.org/10.1037/dev0001269>
- Smetana, J. G. (1986). Preschool children's conceptions of sex-role transgressions. *Child Development*, 57(4), 862–871. <https://doi.org/10.2307/1130363>
- Smetana, J. G. (2006). Social-cognitive domain theory: Consistencies and variations in children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 119–154). Erlbaum.
- Spivey, L. A., Huebner, D. M., & Diamond, L. M. (2018). Parent responses to childhood gender nonconformity: Effects of parent and child characteristics. *Psychology of Sexual Orientation and Gender Diversity*, 5(3), 360–370. <https://doi.org/10.1037/sgd0000279>
- Statistics Canada. (2016). *Mississauga, CY [Census subdivision], Ontario and Canada [Country] (table)*. 2016 Census. <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CSD&Code1=3521005&Geo2=PR&Code2=01&SearchText=mississauga&SearchType=Begins&SearchPR=01&B1=All&TABID=1&type=0>
- Stoddart, T., & Turiel, E. (1985). Children's concepts of cross-gender activities. *Child Development*, 56(5), 1241–1252. <https://doi.org/10.2307/1130239>
- Tang, C. (1996). Collaborative learning: The latent dimension in Chinese students' learning. In D. Watkins & J. Biggs (Eds.), *The Chinese learner: Cultural, psychological and contextual influences* (pp. 183–204). Oxford University Press.
- Theimer, C. E., Killen, M., & Stangor, C. (2001). Young children's evaluations of exclusion in gender-stereotypic peer contexts. *Developmental Psychology*, 37(1), 18–27. <https://doi.org/10.1037/0012-1649.37.1.18>
- Thomas, R. N., & Blakemore, J. E. O. (2013). Adults' attitudes about gender nonconformity in childhood. *Archives of Sexual Behavior*, 42(3), 399–412. <https://doi.org/10.1007/s10508-012-0023-7>
- Todd, B. K., Fischer, R. A., Di Costa, S., Roestorf, A., Harbour, K., Hardiman, P., & Barry, J. A. (2018). Sex differences in children's toy preferences: A systematic review, meta-regression, and meta-analysis. *Infant and Child Development*, 27(2), Article e2064. <https://doi.org/10.1002/icd.2064>
- Trautner, H. M., Ruble, D. N., Cyphers, L., Kirsten, B., Behrendt, R., & Hartmann, P. (2005). Rigidity and flexibility of gender stereotypes in childhood: Developmental or differential? *Infant and Child Development*, 14(4), 365–381. <https://doi.org/10.1002/icd.399>
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.
- Wallien, M. S., Veenstra, R., Kreukels, B. P., & Cohen-Kettenis, P. T. (2010). Peer group status of gender dysphoric children: A sociometric study. *Archives of Sexual Behavior*, 39(2), 553–560. <https://doi.org/10.1007/s10508-009-9517-3>
- Wang, R. R. (2005). Dong Zhongshu's transformation of “Yin-Yang” theory and contesting of gender identity. *Philosophy East and West*, 55(2), 209–231. <https://doi.org/10.1353/pew.2005.0013>
- Wang, Y., Qian, M., Nabbijohn, A. N., Wen, F., Fu, G., Zuo, B., & VanderLaan, D. P. (2022). Culture influences the development of children's gender-related peer preferences: Evidence from China and Thailand. *Developmental Science*, 25(4), Article e13221. <https://doi.org/10.1111/desc.13221>
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72(3), 655–684. <https://doi.org/10.1111/1467-8624.00304>
- Witenberg, R. T. (2007). The moral dimension of children's and adolescents' conceptualisation of tolerance to human diversity. *Journal of Moral Education*, 36(4), 433–451. <https://doi.org/10.1080/03057240701688002>
- Witt, S. D. (2000). The influence of peers on children's socialization to gender roles. *Early Child Development and Care*, 162(1), 1–7. <https://doi.org/10.1080/0300443001620101>
- Wong, W. I., & VanderLaan, D. P. (2020). Sex differences in early life: A cross-cultural perspective. In F. M. Cheung & D. F. Halpern (Eds.), *Cambridge international handbook on psychology of women* (pp. 83–95). Cambridge University Press.
- Wood, L. A., Hutchison, J., Aitken, M., & Cunningham, S. J. (2022). Gender stereotypes in UK children and adolescents: Changing patterns of knowledge and endorsement. *British Journal of Social Psychology*, 61(3), 768–789. <https://doi.org/10.1111/bjso.12510>
- Yeung, S. P., & Wong, W. I. (2018). Gender labels on gender-neutral colors: Do they affect children's color preferences and play performance? *Sex Roles*, 79(5–6), 260–272. <https://doi.org/10.1007/s11199-017-0875-3>

- Yu, L., Winter, S., & Xie, D. (2010). The child play behavior and activity questionnaire: A parent-report measure of childhood gender-related behavior in China. *Archives of Sexual Behavior, 39*(3), 807–815. <https://doi.org/10.1007/s10508-008-9403-4>
- Zhang, L., & Cartwright, D. (2009). A comparison of gender role portrayals in magazine advertising: The United States, China and Thailand. *Management Research News, 32*(7), 683–700. <https://doi.org/10.1108/01409170910965279>
- Zhou, N., Lam, S.-F., & Chan, K. C. (2012). The Chinese classroom paradox: A cross-cultural comparison of teacher controlling behaviors. *Journal of Educational Psychology, 104*(4), 1162–1174. <https://doi.org/10.1037/a0027609>
- Zucker, K. J., Wilson-Smith, D. N., Kurita, J. A., & Stern, A. (1995). Children's appraisals of sex-typed behavior in their peers. *Sex Roles, 33*(11–12), 703–725. <https://doi.org/10.1007/BF01544775>

Received November 8, 2022

Revision received November 15, 2023

Accepted November 19, 2023 ■