Children's Physical Gender Schemas: Acquisition and Features

Darryl B. Hill[1] · Harriet R. Tenenbaum[2]

Abstract
Studies on how physical gender schemas develop in children have traditionally utilized forced-choice and close-ended tasks, finding that the ability to make gender-related distinctions develops in the first years of a child's life. To reduce demand characteristics that reinforce gender binaries in children's models of gender, we relied on open-ended discourse analysis to study children's physical gender schemas. We focused on whether children's ability to ask questions that distinguish gender groups was greater in older than younger children. Participants were 44 3–4-year-olds, 35 5–6-year-olds, and 23 7–8-year-old children in the U.K. who were led through a guessing game to elicit gender-related beliefs and compare their beliefs about gender to their beliefs about other entities such as living things. When asking questions to distinguish gender binary groups, older children judging gendered individuals were more likely to ask questions that stereotypically distinguished the gender groups than younger children. Older children were also more likely to focus on individuals' biological properties, clothing, and hair length than were younger children. Thus, the development of a child's understanding of physical gender schemas gender is discrete, developing gradually at least until the age of 8.

Keywords Gender attributions · Gender stereotypes · Physical gender schema · Children · Cognitive development · Questions

Gender is a macrosocial quality, a broad social structure, often visibly coded in our society to enable clear gender attributions. Gender identity attributions and rules become cognitive structures that organize information related to gender (see Leaper, 2015, for a review of cognitive theories), and the resulting gender schema and stereotypes help simplify a wide array of perceptions that people easily apply to self and others. Yet, rigid and binary gender schemas may deter children from learning about activities not deemed appropriate for their gender (Leaper & Bigler, 2018; Weisgram et al., 2014). As a result, gender schemas may restrict interest in future occupations that do not match children's gender self-schemas (Weisgram et al., 2010). In addition, gender schemas may

also cause children to play with those children they identify as the same gender, limiting friendship opportunities (Martin et al., 2017b). Unfortunately, however, we lack knowledge of how children determine another person's gender at a young age, and how these cues develop across childhood. Without this knowledge, it is not possible to design interventions to help children understand that gender schemas should not limit their friendship opportunities, activities, and occupations. The present study used a novel methodology to identify the physical gender schemas that 3- to 8-year-old children use to identify other people's gender.

Gender Schema Theory

Gender schema theory (Leaper, 2015; Martin & Ruble, 2010) predicts that children acquire their gender schemas gradually rather than all at once. Children appropriate simpler aspects of the schema first and then comprehend more complicated notions (Tenenbaum et al., 2010). Obvious and simple physical characteristics tend to be the first cues used by children in the absence of other gendering cues, at least for girl targets (Miller et al., 2009). That is, young children begin with

✉ Harriet R. Tenenbaum
  h.tenenbaum@surrey.ac.uk

  Darryl B. Hill
  darryl.hill@csi.cuny.edu

[1] Department of Psychology, College of Staten Island and Graduate Center, City University of New York, New York, USA

[2] Department of Psychology, University of Surrey, Guildford, England

simple and traditional theories of gender (e.g., boys have short hair). At the same time, children also invoke unconventional notions (e.g., girls do not wear glasses) (Tenenbaum et al., 2010). Yet, rather than simply assimilating traditional gender stereotypes, as they age, children develop more complex schemas, such as the possibility that someone might show mixed signs of gender (Tenenbaum et al., 2010). With increasing age, children also incorporate additional dimensions such as occupations, activities, and traits (Signorella & Frieze, 2008) into their schemas. Although some of the information used for gender identity attributions are gendered role behaviors and activities (e.g., Berndt & Heller, 1986; Biernat, 1991; Miller et al., 2009), physical schemas come first (Miller et al., 2009). Children's focus on the physical may explain young children's rigid adherence to a gender-stereotypical appearance (Halim et al., 2018). Thus, physical gender schemas are important because judgements of a child's gender are associated with other inferences such as what kinds of activities they might enjoy and traits they might have (Kaiser, 1997). For these reasons, it is important to examine the content of children's physical gender schemas to understand how and when they develop.

Physical Gender Schemas

Some researchers assert that the physical gender schema exists as early as 8 months with children of this age detecting differences in hairstyle, clothing, and body build (Blakemore, 2003). Other studies report that 3-year-old children can label others and themselves as boys or girls using cues like clothing, a skill that develops with age up to 7 or 8 years (Campbell et al., 2004; Martin & Little, 1990). Yet, it is not known what happens between 8 months and 8 years with children's models of gender.

Another limitation with previous research is that most past work relies on closed-ended tasks (e.g., Blakemore, 2003; Campbell et al., 2004). When most children are above chance in following gender norms, researchers credit children of a particular age with having appropriated adult gender schemas. However, some children may not strictly follow these schemas. For example, open-ended tasks allow children to express non-traditional ideas about gender identity. In open-ended tasks, some children follow idiosyncratic gender reasoning strategies until at least 5 to 6 years old when they begin to invoke conventional physical gender schemas more than unconventional ones (Martin & Ruble, 2004; Tenenbaum et al., 2010). Thus, open-ended tasks are preferable to close-ended tasks for understanding children's emic view of gender. To assess children's opinion and knowledge about domains other than gender, open-ended methods, such as guessing games, are used to understand the range of beliefs children may hold (Coenen et al., 2019),

thus avoiding demand characteristics that overly dichotomize children's ideas of gender.

Early Work on Children's Gendering of Others

One of the first studies on the development of children's physical gender schemas used open-ended methods and found that young children had idiosyncratic constructions of gender, which they gradually tested, eventually understanding how their cultural community comprehended gender identity, with most adopting conventional gender reasoning after the pre-school age of 5–6 years (Kessler & McKenna, 1978). This ethnographic study examined the physical factors children invoked to differentiate girls and boys (Kessler & McKenna, 1978). The researchers asked a gender-balanced sample of 10 pre-school (ages 3.5 to 4.5 years), 10 kindergarten (ages 5 to 6 years), and 10 third-grade children (ages 8 to 9) to consider drawings of boys and girls, make a gender attribution, and then explain it. Participants were also asked to guess the gender of a person in 10 guesses. Only a quarter of the respondents asked questions about genitals, with most relying on gender roles and secondary sexual characteristics to assess gender. In a third variation with the same sample, researchers presented young children with simple outline figure drawings of undressed children, with varying content: some had explicit genitals, others not, and many had conflicting cues for gender (e.g., a penis with a figure wearing barrettes in the hair). They asked simply for the participant to guess the gender. In the case just mentioned, if a figure had a penis, children often claimed this figure was male, even when they had no body hair, a curvy body with breasts, and long hair. In contrast to their previous finding, genitals were the primary basis of their gender attributions. However, there are a few limitations of this study. First, it is rare to be classed into a gender category by ones' genitals to a stranger. Second, this study was a small ethnographic sample in which age group differences were not clearly delineated. At the same time, the conclusions were clear: although some children have not yet learned the rules of identifying another's gender, both primary and secondary sex characteristics were used by many children to identify another's gender.

Physical Gender Schemas Using Forced-Choice Methods

Many studies have used forced-choice methods to examine how children identify gender. Some of these studies show that perception of the body is a quality children use to attribute gender. Johnson et al. (2010) tested the ability of 4, 5, and 6-year-olds to use the relative proportion of waist to hips (i.e., curviness) to make gender attributions.

Four-year-olds did not use "curviness" to categorize sex, but the older children showed a tendency increasing with age to categorize those with less curvy figures as men. The authors argue between ages of 4 and 6 years, children acquire the ability to assign gender to bodies. However, whether children are aware of these body cues and able to articulate them is unknown.

In addition to bodies, children also rely on clothing to assign gender identity. Indeed, Kaiser (1997) summarized the growing consensus that children learn by ages 2–3 years which kinds of clothing are appropriate for their own gender. Children label clothing that includes light bright colors, with fine lace or other adornments, and dresses as feminine; and dark colors, pants, and athletic as masculine. When given a forced-choice task to identify clothing as belonging to girls or boys, most 3- to 4-year-olds conformed to gender stereotypes (Blakemore, 2003; Martin & Little, 1990).

Color has also repeatedly been found to be a central indicator of gender. For example, toys, clothes, books, even bedroom furniture are color-coded for gender: boys' possessions are colored darker, and girls' have light pastel colors, and gender-neutral objects are colored darker to appeal to boys who shun pink (Auster & Mansbach, 2012; Berry & Wilkins, 2017; LoBue & DeLoache, 2011; Weisgram et al., 2014). More recently, researchers asserted that by the age of 5 years, children use more gender-stereotypical colors for gendered targets (Navarro et al., 2014), and the gendering of clothing color may prime perceivers for a masculine or feminine read (Cunningham & Macrae, 2011). These studies suggest that clothing and color are important to gender attributions, and that children acquire the ability for gender attributions early and gradually.

Faces may also play a role. Research on infant perception of gender indicates infants can gender faces by 3 months, tend to show a preference for female faces, and expect to hear a match between gendered faces, voices, and bodies within the first year (see Hock et al., 2015 for a review). Hock et al. (2015) showed 3.5 month and 5.5-month-old infants pictures of various faces and bodies, some with gender congruent bodies and faces, and some incongruent, and found that many 5.5-month infants detected gender incongruent bodies and faces, while those at 3.5 months did not. These researchers surmised that gender attributions based on bodies and faces emerged between these two time periods for most, but this was a gradual process. Moreover, it is unclear exactly what about faces were being used by children to attribute gender. It seems reasonable to assume that eyes, noses, mouths, and face shape were not what infants were using—rather they were probably cuing on hair length.

Physical Gender Schemas Using Open-Ended Methods

The above studies on physical gender schema found that many 6-year-olds use color, clothing, bodies, and infants less than one-year-old can use faces to gender others. Yet some do not appropriate this information, and others decide that the rules sometimes are not applicable. However, forced-choice methods limit what children can report. Tenenbaum et al. (2010) conducted two studies on physical gender schemas. In the first study, children drew a picture of a boy and a girl, and then made gender identity attributions about their own and other participants' drawings and explained their attributions. The researchers categorized participant responses as conventional (consistent with gender stereotypes) and non-conventional (such as reversed, idiosyncratic, or flexible gender reasoning). They found a developmental effect, with older children (aged 5 and above) providing more conventional schemas and younger children offering more unconventional gender reasoning. Despite the age-related increases in conventional reasoning, some children past 6 years of age still proffered unconventional gender explanations, likely due to a growing awareness of flexible gender roles or individual differences in gender presentation.

Gender reasoning might also work backwards: if someone enjoys a gender feminine stereotypical activity, or have biological properties like being able to lactate, then they must be female and have feminine traits. It seems that the qualities children use to determine gender depend on the gender of the person being perceived. Miller et al. (2009) found that children 3 to 10 years, a wide range, rely more on appearance when judging the gender of girls, and boys rely on activity and trait-based domains first, then appearance. In terms of physical appearance, clothing (pants versus dresses) and length of hair were key to assessing physical gender schema in Miller et al.'s study.

Perhaps children build varying models of gender schema, some rich and complex, and some children rely on more simple binary models. Halim et al. (2014) confirmed that for 3- to 6-year-old girls, rigidity in stereotypes about gender appearance peaks, followed by a period of flexibility in these stereotypes, yet the 6-year-old boys in the study had yet to demonstrate similar flexibility. Tenenbaum et al. (2010) contradicted this finding when they found no such delayed effect for boys.

Summary and Critique

Although much research has been conducted on children's early physical gender schemas, more research is needed for many reasons. First, many of these studies were conducted 30–40 years ago in the United States, so it is important to

examine whether gender stereotypes from the past generalize to contemporary views of gender identity elsewhere. Moreover, the studies using forced choice methods reported that children learn gender schemas quite early in life. Past studies have also taken a piece-meal approach, examining one aspect of physical gender schema at a time. No study has examined appearance, activities, and biological properties simultaneously in the role of physical gender schemas, so it remains unknown which categories children use most when differentiating genders. Furthermore, much of the past research has relied on recognition-based tasks (i.e., forced choice answers; for an exception see Miller et al., 2009), which may have constrained unconventional gender attributions and reasoning, making it seem as if children are more conventional than they really are. A more open-ended approach might reveal unexpected ways in which young people understand gender, especially when they rely on unconventional or flexible gendered reasoning. Indeed, as a result of past work with a focus on closed-ended methodologies, little is known about the content of children's physical gender schemas or the ages at which children rely on different categories (e.g., color, hair length) in their schemas. Significantly, if society wants children to become more gender flexible about others' gender categories, we need to know the basis for their stereotyped judgments in the first place.

A further concern is that studies examining children's gender schematic judgements only ask participants to make judgements about gendered entities. As a result, there is no comparison for these judgements. Not until the age of 5 years are children able to distinguish animate from inanimate objects in categorization tasks above chance (Wright et al., 2015). It is unknown, thus, whether the basis for gender attribution is unique to decisions about gender or the categorization of other entities. If children rely more on superficial features about all categorization tasks, society may need to intervene more generally to reduce stereotyping in general rather than intervening to reduce gender stereotyping in and of itself.

The Present Study

The goals of the present study were twofold. The first was to document how children's understanding of the physical gender schema enables them to attribute others' gender to understand the content of children's gender schemas. The second goal was to track the emergence of these attributions over childhood. We used a novel methodology, a guessing game, to elicit discourse from children that would reveal which elements of the stimuli guide gender attributions. Children ages 3 to 8 were asked to play a game in which they needed to distinguish between pairs of entities (e.g.,

a mother and a father; a cat and a rose) by asking yes/no questions. Children were asked about three gendered comparisons and three non-gendered comparisons. This method reduced demand characteristics and provided comparison controls to better understand their gender attributions, maximizing the freedom for respondents to provide unconventional gendered reasoning.

We advanced two main hypotheses. First, based on Tenenbaum et al. (2010) in which children older than 5 years proffered more conventional ideas than younger children, we predicted an interaction effect. We expected children older than the age of 5 years to ask more distinguishing (operationalized by being able to differentiate between the entities) than non-distinguishing questions on comparisons that involved gendered entities, whereas we expected children younger than 5 years to ask the same numbers of the types of distinguishing and non-distinguishing questions. To rule out a confound that asking distinguishing questions might be too difficult for the children (e.g., Ruggeri et al., 2021), we also examined children's ability to ask distinguishing versus non-distinguishing questions involving other living things. Not until the age of 5 years are children able to distinguish animate from inanimate objects in categorisation tasks above chance (Wright et al., 2015).

Second, based on the pioneering work by Kessler and McKenna (1978) demonstrating that older children have more conventional gender schemas, we expected that children above the age of 5 years would use physical dimensions such as hair length, color, clothing, and biological aspects of gender in their distinguishing questions about gender more often than other dimensions.

Method

Participants

The final sample included 44 3–4-year-olds, 35 5–6-year-olds, and 23 7–8-year-old children (54% identified as boys and 46% identified as girls) from four state schools in London, United Kingdom, which included nursery children and two private nurseries. The catchment areas of all the schools are ethnically diverse ranging from 20 to 60% ethnic minority. Ethnic minority children were typically Asian British (e.g., Pakistani, Sri Lankan), Caribbean, and African British, and mixed-race backgrounds. Eighteen percent to 40% of students have English as an additional language. Other than English, the top languages spoken in these neighborhoods are Tamil, Polish, Arabic, and German. The catchment areas included low-income families living in council estates and middle-income families based on the English government's indices of multiple deprivation. Five participants with incomplete data were removed from the analyses.

Anonymized data are available on the OSF website: https://osf.io/8jm6n/.

Procedure

This study was reviewed and approved by the Institutional Review Boards at the home institutions for each author. Email messages were sent to principals of state schools and nurseries in London. Principals who agreed to participate were sent parental information and consent forms to be distributed in the grades of interest. Parental information sheets explained that the research focused on children's understanding of gender and researchers would ask children a series of questions that would be audio-recorded.

After parents provided returned consent forms with written consent for children's participation and children provided assent, children were led through a guessing game inspired by Chouinard et al. (2007) to elicit gender-related beliefs and compare the beliefs about gender to beliefs about other non-gendered entities such as objects. A research assistant unaware of the hypotheses read the child instructions and started with a training trial wherein the child was shown a picture of a strawberry and a scooter. Then, the child hid one of the objects behind a paper curtain. The researcher modeled how to ask questions to differentiate what was hidden to scaffold the child (e.g., "Can I eat it?" would successfully lead to discovering that a strawberry was hidden). Next, the researcher showed children a picture of a teddy bear and a teacup. The researcher hid one picture behind a paper cup. Children were told to ask yes/no questions to help them figure out what was hidden. Children were provided feedback for the trial.

After children successfully completed the trial, they were told that they were going to play a guessing game in which they were not shown pictures. Instead, the experimenter would name two things and the child had to ask questions to figure out which of two things was the target of the game. The non-gender related pairs consisted of a ball and a cat; a rose and a cat; and a baby and a cat, in counterbalanced order. Each gendered pair was only presented once, presented in between the non-gender related stimuli. The gendered pairs included a mother and father; boy and a girl; and a man and a woman. In this way, the non-gendered entities formed a control comparison group. We did not use pictures for the items because we did not want to introduce bias for the gender pairs. All items were presented in a counter-balanced randomized order for all participants. Children were allowed to ask as many questions as they wanted. To keep children interested and to keep the procedure as consistent as possible for all children, we always answered yes to their questions whether it distinguished the entities or not, even though this might appear as if the researcher was endorsing traditional gender

beliefs (see Table 1 for examples). We were only interested in the types of questions children asked. Responses were audio-recorded and transcribed verbatim. Each trial was completed when children successfully guessed the entity. Children were not instructed to guess the entity with as few questions as possible. There was no debrief.

The two authors read the transcripts and coded children's questions. Both experimenters coded a third of all cases. Coding and inter-rater reliability took part in two phases. First, researchers determined whether the question would distinguish the entities (distinguishing or non-distinguishing questions). Second, the researchers coded the questions into categories (Table 2 displays these categories) using conventional gendered assumptions. Each question was categorized into at least one group; some were coded into multiple categories. For example, "Does it wear trousers?" was coded as "clothing" but not "biological property" because an inanimate doll could wear trousers. Once their level of agreement was acceptable ($\kappa = .79$ for distinguishing question/non-distinguishing question and $\kappa = .86$ for categories), the second author coded the remaining transcripts.

Children's questions were also judged as to whether their questions distinguished the hidden entity. For a question to be coded as distinguishing, the question would lead to a correct determination of which entity was hidden (e.g., "Is it prickly?" would distinguish a rose from a cat). For the gendered entities, coders relied on conventional gender stereotypes to make this determination (e.g., "Does it have long hair?" was considered distinguishing because contemporary women in the United Kingdom stereotypically have long hair). We compared all the different kinds of questions and examined which categories of questions they used most for making guesses about gendered entities (man vs. woman, boy vs. girl, mom vs. dad) compared to non-gendered entities (cat vs. ball, baby vs. cat, cat vs. rose). We summed the frequency of questions and distinguishing questions for each category for the gendered stimuli and the non-gendered stimuli separately.

Results

Analysis Plan

To assess part of the second hypothesis, children's questions were tabulated for the categories listed in Table 2 (e.g., biological properties, hair length) counting the simple frequency they were used. Table 3 lists the frequency of total questions and distinguishing questions for the three gender comparisons (i.e., mother and father, boy and a girl, a man and a woman). In general, children asked most questions about clothing and hair length. However, the most typical distinguishing category of question for all gender stimuli

Table 1 Participant Gender, Age, Excerpts, and Codes From Guessing Game

Participant Gender and Age	Excerpt	Attribute Code	Distinguish Genders
Girl aged 7 years	Interviewer: If I say I had either a cat or a ball, and I'm going to hide one there, what would you ask me? Child: Is it fluffy? Interviewer: It is fluffy Child: Cat Interviewer: Right, excellent, good question, that helped you	Physical attribute	Yes
Girl aged 7 years	Interviewer: Okay, what if I hide either a boy or a girl there, what question would you ask me? Child:Does it have long hair or short hair? Child:Girl Interviewer: Excellent	Hair length	Yes
Boy aged 6 years	Interviewer: What if I hid a cat or a rose what question would you ask me? Child: Erm if it's a rose, erm, I would ask is it prickly or? Interviewer: It is prickly Child: Or I would ask has it got petals? Interviewer: Yes to both Child: A rose	Physical attribute	Yes
Boy aged 6 years	Interviewer: What if it's a mummy or a daddy? Child: Erm Interviewer: Mummy or a daddy what question would you ask me? Child: Daddy Interviewer: Yeah, what would you ask me? Child: I would ask, erm, does he sometimes wear ties? Interviewer: Does he, yes he does Child:Daddy Interviewer: Yes	Clothing	Yes
Girl aged 4 years	Interviewer: Ok, I have hidden a man or a woman. What question could you ask me? Child: A woman Interviewer: What would you ask me before to find out which it is? Child: Women have curly hair Interviewer: Can you ask a question? Child: Does she have curly hair? Interviewer: Yes it does Child: A woman Interviewer: Yes	Hair	No
Girl aged 4 years	Interviewer: Now I am going to hide either a cat or a ball Interviewer: What could you ask me? Child: A ball Interviewer: Ok, what would you ask me about the ball before guessing? Child: It's bouncy Interviewer: Can you ask a question? Child: Is it bouncy? Interviewer: It does bounce Child: It's a ball	Physical attribute	Yes
Girl aged 7 years	Interviewer: Okay, the next one, mommy and daddy Interviewer: Okay, I'm hiding one behind the curtain Child: Um, does it usually go off to work, while the other parent looks after the children? Interviewer: Usually Child: Uh, I don't know. Is it uh, is it usually at home more on the weekends or on weekdays? Interviewer: I can only say yes or no Child: Is this person usually at home on weekends? Interviewer: Yes, weekends Child: It is the dad	Activity	Yes

Table 2 Coding Categories, Definitions, and Examples

Category	Definition	Example
Action	Refers to an action	Does it bounce? Can I kick it?
Activity	Refers to an activity	Is it playing? Does it take care of children?
Biological property	Refers to a biological property of living things	Does it grow? Does it give birth?
Body Parts	Refers to body parts	Does it have lips?
Clothing	Refers to clothing or any other possession	Does it have a hat? Does it wear trousers?
Color	Refers to color	Is it blue?
Hair	Refers to aspects of hair other than length	Does it have fringe?
Hair Length	Refers to hair length	Does it have short hair?
Physical Attributes	Refers to other physical cues	Is it fluffy? Does it have petals?
Size	Refers to size	Is it big?
Sound	Refers to sound	Does it meow? Does it squeak?

was hair length. There were no effects of the child's own gender identity, so we did not include this factor in our analyses.

To examine whether the 5- to 8-year-old age groups asked more distinguishing (operationalized by being able to differentiate between the entities) than non-distinguishing questions involving gendered entities (Hypothesis 1), we conducted a 2 (distinguishing, non-distinguishing questions) × 3 (Age Groups 3–4, 5–6, 7–8) mixed-design ANOVA. To rule out the possibility that asking distinguishing questions might be too difficult for the children, we also conducted a 2 (distinguishing, non-distinguishing questions) × 3 (Age Groups 3–4, 5–6, 7–8) mixed-design ANOVA on the living things entities. With a medium to large effect ($f = .25$) based on Tenenbaum et al. (2010), an alpha of .05, and power of .80, G*Power returned a sample size of 105 for a

Table 3 Total Number of Questions and Distinguishing Questions by Category for the Gender Comparisons

Category	Total Questions		Distinguishing Questions	
	M	*SD*	*M*	*SD*
Hair Length	.97$_a$	1.16	.95$_a$	1.15
Clothes	1.20$_a$	1.79	.63$_{a,b}$	1.09
Activity	.87$_{a,c}$	1.60	.41$_{b,c,g}$	.97
Physical Characteristic	.34$_{b,c,d}$	.79	.24$_{b,c,d}$	.66
Biological Feature	.25$_{b,e}$	.53	.20$_{c,e}$	.42
Action	.22$_{b,e}$	.68	.11$_{c,e,f}$	.56
Color	.25$_{b,e}$	.62	.10$_{d,e}$	.36
Hair Other than Length	.39$_{b,c}$	.65	.10$_{d,e,g}$	.33
Body	1.12$_{a,b}$	3.00	.10$_{d,e,g}$	.42
Size	.07$_e$	.32	.06$_{d,f}$	.31
Sound	.05$_e$	.22	.04$_{d,e}$	.20

The numbers of questions asked about the three gender comparisons by category are listed above. Categories in rows with the same subscript are not significantly different from each other

2 × 3 mixed-design ANOVA model (Faul et al., 2007). To examine whether children in the 5- to 8-year-old age groups used more physical dimensions such as hair, color, clothing, and biological aspects of gender than other categories we conducted a 3 (Age Groups 3–4, 5–6, 7–8) × 11 (category of questions) mixed-design ANOVA model. With a medium to large effect ($f = .25$) based on Tenenbaum et al. (2010), an alpha of .05, and power of .80, G*Power returned a sample size of 62 for a 2 × 3 mixed-design ANOVA model (Faul et al., 2007). In both analyses, age was a between-subjects factor and questions were within-subjects factors. Where we found statistically significant findings, we followed up with post-hoc analyses.

Types of Stimuli and Age Group Analyses

To understand the total number of questions that children asked about the two types of stimuli at the different age groups, we conducted a 2 (Type of Stimuli: Gender, Living Things) × 3 (Age Groups) mixed-design ANOVA on the total number of questions children asked. Type of stimuli served as a within-subjects factor and age served as a between-subjects factor. There was no main effect of Type of Stimuli, $F(1, 99) = .02$, $p = .89$, and Type did not interact with Age Group, $F(1, 99) = 1.16$, $p = .32$. There was a main effect of Age Group, $F(2, 99) = 8.81$, $p < .001$, $_p\eta^2 = .15$. To tease apart this effect, we conducted three ANOVA models to examine differences between the age groups. We controlled for the three tests by using an alpha of .01 (.05 divided by 3). Children in the 7- to 8-year-old age group asked significantly more questions ($M = 15.30$, $SD = 5.58$) than did children in the 5- to 6-year-old age group ($M = 11.34$, $SD = 5.10$), $F(1, 56) = 7.77$, $p = .007$, $_p\eta^2 = .12$, and children in the 3- to 4-year-old age group, ($M = 9.16$, $SD = 6.16$), $F(1, 65) = 16.00$, $p < .0001$, $_p\eta^2 = .20$. Children in the 3- to 4-year-old age group did not differ from children in the 5- to

6-year-old age group in the total number of questions asked, $F(1, 77) = 2.84$, $p = .10$.

Distinguishing Questions about Gender and Age Groups

Question Frequency

To examine our hypothesis that children aged 5 years and older would use more distinguishing questions compared to non-distinguishing questions about the gender stimuli, we conducted a 2 (Type of Question: Distinguishing, Non-Distinguishing) × 3 (Age Group) mixed-design ANOVA on the number of distinguishing versus non-distinguishing questions children asked about the gender stimuli. Type of questions served as a within-subjects factor and age served as a between-subjects factor. In partial support of our first hypothesis, the ANOVA revealed a statistically significant interaction between Age and Type of Questions, $F(1, 99) = 11.99$, $p < .0001$, $_p\eta^2 = .19$. To tease apart this effect, we tested three ANOVA models to examine where there were differences in the age groups. We controlled for the multiple comparisons by using an alpha of .01 (.05 divided by three). Children in the 7- to 8-year-old age group asked more distinguishing questions ($M = 5.74$, $SD = 2.73$) than non-distinguishing questions ($M = 1.74$, $SD = 1.74$), $F(1, 22) = 30.44$, $p < .0001$, $_p\eta^2 = .58$. In contrast to the hypothesis, children in the 5- to 6-year-old age group did not ask more distinguishing questions ($M = 2.63$, $SD = 1.77$) than non-distinguishing questions ($M = 2.97$, $SD = 3.30$), $F(1, 34) = .20$, $p = .66$. Similarly, children in the 3- to 4-year-old age group did not ask more distinguishing questions ($M = 1.66$, $SD = 1.92$) than non-distinguishing questions ($M = 3.23$, $SD = 4.03$), $F(1, 43) = 4.53$, $p = .04$, $_p\eta^2 = .10$. There was no main effect for Type of Question, $F(1, 99) = 2.29$, $p = .13$.

Question Proportions

To make sure that older children's use of distinguishing questions about gender was not simply that they asked more questions generally, we conducted a between-subjects ANOVA model with 3 levels (3–4, 5–6, 7–8) using the proportion of distinguishing questions about gender divided by the total number of questions asked when presented with the gender stimuli. The ANOVA revealed a statistically significant interaction effect of Age on the proportion of distinguishing questions, $F(2, 99) = 9.09$, $p < .001$, $_p\eta^2 = .16$. To tease apart this effect, we tested three follow-up ANOVA models to examine if there were differences in the age groups. We controlled for the multiple comparisons by using an alpha of .01 (.05 divided by three). Children in the 7- to 8-year-old age group asked a higher proportion of distinguishing questions ($M = .78$, $SD = .19$) than did children

in the 5- to 6-year-old age group ($M = .56$, $SD = .35$), $F(1, 56) = 7.54$, $p = .008$, $_p\eta^2 = .12$ or children in the 3- to 4-year-old age group ($M = .40$, $SD = .40$), $F(1, 65) = 18.18$, $p < .001$, $_p\eta^2 = .22$. In contrast, there was no difference between the 3- to 4-year-old age group and the 5- to- 6-year-old age group, $F(1, 77) = 3.43$, $p = .07$.

Distinguishing Questions about Other Domains

To tease apart the possible confound around young children being unable to ask distinguishing questions, we conducted a 2 (Type of Question: Distinguishing, Non-Distinguishing) × 3 (Age Group) mixed-design ANOVA on the number of questions children asked about the comparison (living things) condition. First, there was a main effect of Type of Question with children of all ages asking more distinguishing questions ($M = 4.09$, $SD = 2.76$) than non-distinguishing questions ($M = 1.49$, $SD = 1.73$), $F(1, 99) = 92.95$, $p < .0001$, $_p\eta^2 = .48$. This effect was qualified by a statistically significant Type of Question x Age Group interaction effect, $F(1, 99) = 11.20$, $p < .0001$, $_p\eta^2 = .18$. To tease apart this effect, we tested three ANOVA models to examine where there were differences in the age groups. We controlled for the multiple comparisons by using an alpha of .01 (.05 divided by three). Children in the 7- to 8-year-old age group asked more distinguishing questions ($M = 6.48$, $SD = 2.92$) than non-distinguishing questions ($M = 1.35$, $SD = 1.58$), $F(1, 22) = 49.96$, $p < .0001$, $_p\eta^2 = .69$. Second, children in the 5- to 6-year-old age group asked more distinguishing questions ($M = 4.03$, $SD = 2.16$) than non-distinguishing questions ($M = 1.71$, $SD = 1.98$), $F(1, 34) = 24.00$, $p < .0001$, $_p\eta^2 = .41$. Finally, children in the 3- to 4-year-old age group asked more distinguishing questions ($M = 2.89$, $SD = 2.30$) than non-distinguishing questions ($M = 1.39$, $SD = 1.62$), $F(1, 43) = 11.60$, $p = .001$, $_p\eta^2 = .21$. Although children in all age groups were able to ask more distinguishing than non-distinguishing questions, the effect size was greater in older than younger children. When we looked at the proportion of distinguishing questions asked about the comparison condition, there was no differences based on Age Group, $F(2, 99) = 3.09$, $p = .05$. Thus, we ruled out the confound that age was related to being able to ask distinguishing questions.

We also examined whether children asked more distinguishing questions about gender than the living things and whether the effect varied across age group by conducting a 2 (Distinguishing Questions: Gender, Living) × 3 (Age Group) ANOVA. Children asked more distinguishing questions about living things ($M = 4.09$, $SD = 2.92$) than gender ($M = 2.90$, $SD = 2.60$), $F(1, 99) = 24.74$, $p < .001$, $_p\eta^2 = .20$. This finding suggests that children have more conventional schemas (or adult-like knowledge) about living things than

gender at these ages. There was no interaction with Age Group, $F(2, 99) = .65$, $p = .53$.

Physical Category by Age Group

To examine if the questions asked by the two older age groups involved more physical dimensions such as hair, color, clothing, and biological aspects of gender, in their distinguishing questions, than other categories we conducted a 3 (Age Group) × 11 (Category) mixed-design ANOVA on the type of questions children asked for the gender stimuli. There was a main effect of Category, $F(5.46, 540.73) = 29.29$, $p < .0001$, $_p\eta^2 = .23$, which was subsumed by an Age Group x Category interaction effect, $F(10.92, 540.73) = 7.84$, $p < .0001$, $_p\eta^2 = .14$. The main effect of category results may be found below. We also tested 11 ANOVA models to examine which categories of distinguishing questions were most prevalent at the different ages. We used a protected alpha of .004 for these analyses. Table 4 displays these means.

We expected older children (aged 5 to 8 years) to ask more questions invoking the categories of clothing, clothing color, hair length, and biological features than other categories. Using the protected alpha, there were no age differences based on action, $F(1,99) = 3.57$, $p = .07$, activity, $F(1,99) = 3.17$, $p = .03$, physical characteristics, $F(1,99) = .79$, $p = .46$, body, $F(1,99) = .65$, $p = .52$, color, $F(1,99) = 1.67$, $p = .19$, hair other than length, $F(1,99) = 3.82$, $p = .03$, sound, $F(1,99) = 1.00$, $p = .37$, and size, $F(1,99) = 1.28$, $p = .28$.

Partially supporting the second hypothesis, there was an effect of Age Group on biological characteristic questions that distinguished the gender stimuli, $F(1, 99) = 15.13$, $p < .0001$, $_p\eta^2 = .23$. Using a p-value of .01, follow-up tests indicated that 7- to 8-year-old children used more distinguishing questions related to biological properties than did 5- to 6-year-old children, $F(1, 56) = 11.61$, $p = .001$, $_p\eta^2 = .17$, and 3- to 4-year-old children, $F(1, 65) = 27.74$, $p = .001$, $_p\eta^2 = .30$. The 3- to 4-year-old group did not differ from the 5- to 6-year-old group, $F(1, 77) = 3.00$, $p = .13$.

Second, there was an effect of age on clothing questions that distinguished the genders, $F(1, 99) = 20.29$, $p < .0001$, $_p\eta^2 = .29$. Using a p-value of .01, follow-up tests indicated that 7- to 8-year-old children used more distinguishing questions related to clothing than did 5- to 6-year-old children, $F(1, 56) = 9.34$, $p = .001$, $_p\eta^2 = .17$, and 3- to 4-year-old children, $F(1, 65) = 37.01$, $p < .0001$, $_p\eta^2 = .36$. The 3- to 4-year-old group used fewer clothing questions than the 5- to 6-year-old group, $F(1, 77) = 14.36$, $p = .0001$, $_p\eta^2 = .16$.

Finally, there was an effect of age on hair length questions that distinguished the genders, $F(1, 99) = 17.80$, $p < .0001$, $_p\eta^2 = .27$. Using a p-value of .01, follow-up tests indicated that 7- to 8-year-old children used more distinguishing questions related to hair length than did 5- to 6-year-old children, $F(1, 56) = 11.95$, $p = .001$, $_p\eta^2 = .18$, and 3- to 4-year-old children, $F(1, 65) = 39.52$, $p < .0001$, $_p\eta^2 = .38$. The 3- to 4-year-old group did not differ from the 5- to 6-year-old group, $F(1, 77) = 2.30$, $p = .0001$, $_p\eta^2 = .13$.

We also examined whether children in the different age groups relied on different categories of distinguishing questions. Table 4 displays these means. We used a protected alpha of .0001 for follow-up tests. Although there was a statistically significant effect, $F(5, 206) = 3.58$, $p = .005$, $_p\eta^2 = .08$ in the 3- to 4-year-old group, no categories were used differently at a statistically significant level. In contrast, for 5- to 6-year-olds, $F(4, 304) = 9.09$, $p < .0001$, $_p\eta^2 = .21$, and 7- to 8-year-olds, $F(3, 74) = 16.23$, $p < .0001$, $_p\eta^2 = .42$, hair length and clothes were used significantly more frequently than the other categories.

Table 4 Category Use of Distinguishing Questions about Gender in Different Age Groups

Category	3 to 4 Years Old		5 to 6 Years Old		7 to 8 Years Old	
	M	SD	M	SD	M	SD
Hair Length	.43$_a$	.85	.94$_a$	1.08	1.97$_a$	1.11
Clothes	.09$_a$	.47	.66$_{a,c}$	.84	1.61$_{a,b}$	1.53
Activity	.45$_a$	.87	.11$_{b,c}$	.40	.78$_{a,b,d}$	1.51
Biological Feature	.05$_a$	.21	.14$_{b,c}$	.36	.57$_{b,d}$	.59
Hair Other than Length	.05$_a$	.21	.06$_b$	.24	.26$_{b,d}$	.54
Color	.07$_a$	.33	.06$_b$	.34	.22$_{c,d}$	.42
Physical Characteristic	.20$_a$	.70	.34$_{a,b,c}$	.76	.13$_{c,d}$	.34
Size	.07$_a$	.33	.00$_b$	.00	.13$_{c,d}$	.46
Action	.02$_a$	.15	.26$_{a,b,c}$	.92	.04$_{c,d}$	.21
Body	.16$_a$	.81	.03$_b$	.17	.04$_{c,d}$	.21
Sound	.07$_a$	.25	.03$_b$	.17	.00$_{c,d}$	.00

Categories in rows with the same subscript are not significantly different from each other

Discussion

Our study examined the content of children's physical gender schemas by examining the types of questions children asked to distinguish women from men and girls from boys. Children's questions about gender included a range of categories, including hair length, clothes, activities, physical features, biological features, action, color, other characteristics of hair, body, size, and sound. We found that, as predicted, children older than 6 years offered more questions and more distinguishing questions across stimuli. There was a marked increase in each age group. However, there was also an increase in invoking clothing and hair.

In terms of developmental increases in the ability to ask questions, children between the ages of 3 and 6 asked a greater number of informative questions that distinguished the living thing categories than questions that did not distinguish categories, which is consistent with recent work (e.g., Ruggeri et al., 2021). This difference was even greater in 7- to 8-year-old children suggesting the developmental progress in this ability is not yet completed at age 6 years. Our findings suggest that children further increase in gaining knowledge about social categories like gender even after the age of 6 years. Tenenbaum et al. (2010) also found that children's gender schemas continued to evolve through to 8 years. Although Ruggeri et al. (2021) suggested that children have a good ability to ask questions that distinguish different types of categories by 6 years, our work suggests that this ability continues to develop even past this age.

Children were also able to ask more distinguishing questions about living things than they were about gender even though past work reported that these schemas develop at roughly the same ages (Leaper, 2015; Margett-Jordan et al., 2017). Research had reported that between 3 and 4 years of age, children become more similar to adults in distinguishing animals and plants from mobile and immobile artefacts (Margett & Witherington, 2011). Children also begin to restrict the biological properties of growing, eating, and drinking to living things (Margett-Jordan et al., 2017). A majority of children in the Margett-Jordan et al. (2017) study did not report that plants were alive. However, in the present study, children were able to ask more distinguishing questions overall about living things than about gender. These findings tentatively suggest that social categories, such as gender, may not rest on invariant principles compared to more biologically-oriented categories, such as living things. Gender binary categories are salient to children because of the way that society organizes gender and not because of any underlying principles (Hyde et al., 2019). These findings suggest that children appropriate social categories with difficulty.

The category of children's questions also differed by age group. We had predicted that older children would rely on clothing, hair length, color, and biological properties more than younger children would because these are conventional cues that children use when they are older (Auster & Mansbach, 2012; Blakemore, 2003). Although older children did not ask more distinguishing questions that relied on color to distinguish gender groups, they did base more distinguishing questions on clothing, hair length, and biological properties (e.g., giving birth). These findings suggest that each age group increased in their conventional understanding of gender. When we examined children's questions separately by age group, we found that children aged 5 and older were most likely to use hair length and clothing compared to other categories. These older groups rarely used sound (e.g., high pitched voice) or body (e.g., having breasts) in their guesses. Older children may not have asked about body parts because they may have believed that such questions were inappropriate even though these questions would distinguish gender groups. In contrast, with the corrected alpha level, there was no difference in the categories used by children in the 3- to 4-year-old age group. This finding may support previous work suggesting that with age, children become more similar to each other and less idiosyncratic in the particular physical cues they use to gender others (Tenenbaum et al., 2010). Although children may take until 10 years to develop their physical gender schemas (Ruble et al., 2006), children showed rapid increases across the different age groups. Thus, gender schemas seem to become more conventional in successive age groups.

In terms of the content of these schemas, past work has suggested that children are adept at using hairstyle, clothing, body build (Blakemore, 2003), and color (Auster & Mansbach, 2012) to distinguish women and men on a physical level. In the present study, children relied on hair length followed by clothes, activities, and physical activities. Surprisingly, clothing color was not one of the categories children invoked the most to distinguish the categories. Using a novel methodology, we were able to gain a more in-depth understanding of the range of categories children used than when simply asking closed-ended questions, which tends to make it seem as if children's gender reasoning is more conventional than they really are. Our work suggests that when using open-ended responses, children also rely on other categories, such as sound (e.g., deep voice), although rarely. Thus, the content of gender schemas may include more categories and be richer than one would expect.

Limitations and Future Directions

A major limitation of this study is that it asked respondents to make binary gender attributions (Hyde et al., 2019). A future direction would be to ask questions in a more

open-ended manner rather than relying on binary gender categories. Another direction would be to ask children with non-binary gender identities about their physical gender schemas, thus tracking how children's gender identity, as opposed to their sex, influences gender schema development. Although gender non-conforming children may be as essentialist about gender as cisgender children (Gülgöz et al., 2019), the content of their physical gender schemas may vary. Finally, how similar children see their own gender and the gender of others (Andrews et al., 2016) may also influence their gender schemas, so future research should include additional measures of children's identity.

Another limitation of this study is the reliance on questions from children younger than 6 years. Indeed, Ruggeri et al. (2021) argued that young children often have difficulty asking distinguishing questions when playing games such as 20 questions with novel characters. Other work, however, has found that children as young as 3 years (Callanan & Oakes, 1992; Chouinard et al., 2007) are able to ask intelligent questions in everyday contexts. Our study used children's question-asking to understand their thinking about gender, and as such children's ability to ask questions may be a limitation of this research. Additionally, children's skill at playing 20 questions was not assessed, which could be used as a control in future research. Related to using question is that children's oral language skills may have also inhibited their ability to ask questions more generally.

Moreover, children may have answered differently if they were told that the goal was to ask as few questions as possible. Future work could investigate how changes, such as limiting the rules of the game, may influence children's decisions. We did not ask children to guess in as few questions as possible because we wanted to uncover the full range of their gender schemas. However, playing the game in such a way may have uncovered what they assessed as most central to gender. Additionally, we did not want to bias children's answers by showing them pictures or providing accurate feedback, so this may have influenced how children asked future questions. Whether children provided conventional or non-conventional answers, we did not correct them. Thus, children may have interpreted our answers as confirming their prior beliefs. Both these decisions may be seen as limitations in the current study.

From sociocultural theory, children's folk theories are developed through everyday activities in which children participate actively to create opportunities for the practice of gendered behaviors (Rogoff et al., 2018). For this reason, children's experiences with peer groups (Martin et al., 2017b), parents (Leman & Tenenbaum, 2014), and even schools (Spinner et al., 2021) all influence the content of children's schemas. Future research should look at how everyday experiences, such as children's choice of playmates influence the content of physical gender schemas.

Children's everyday experiences also occur within their socio-demographic background. A limitation of the present study is that we did not consider these variables. We do know, however, that these children's schools had a range of students from different backgrounds. Thus, this study provides a snapshot of children living in a highly diverse multilingual urban area. Children in other regions of the U.K., or elsewhere, may have responded differently, which future research should examine.

Practice Implications

Children's adherence to gender schemas can be detrimental for their toy choices, career aspirations, friendship decisions, and self-esteem (Leaper, 2015). Indeed, these gender stereotypes can serve as a gatekeeper, or barrier, to children's learning, future occupational decisions, and peer interactions (Martin et al., 2017a; Weisgram et al., 2010). To challenge these schemas, it is important to understand the content of these schemas. For example, this study indicates that although color has been found to be central to children's schemas in past research (Berry & Wilkins, 2017), children may be more likely to rely on hair length and clothing to make gender attributions than color. Our findings, thus, suggest that interventions should initially focus on challenging hair length and clothing as gatekeepers of gender conformity rather than color. This work also has the potential to guide future school-based interventions. For example, the linear increase in children's invoking clothes may be related to the greater use of gender-stereotyped uniforms in primary schools. Thus, interventions will need to carefully consider the social context that surrounds children to move away from gender binary categories (Hyde et al., 2019).

Conclusion

In sum, our study found that children use a variety of categories to make sense of physical gender attributions. Nevertheless, children become more conventional in their schemas as they age suggesting that physical gender schemas are certainly not complete at the age of 5 years when children become more flexible (Trautner et al., 2005). Thus, children's gender flexibility coincides with a time when they gain increasing conventional knowledge. Additionally, our work suggests that the development of conventional schemas is a slow process that continues to undergo revision until at least the age of 8 years, a process that may include an initial flexibility for some children, and for other a reversion to conventional schemas.

Declarations

References

Andrews, N. C., Martin, C. L., & Gallagher, A. M. (2016). On the association between self-reported own-and other-gender similarity and the use of physical and relational aggression in sixth grade children. *Archives of Sexual Behavior, 45*, 1817–1826. https://doi.org/10.1007/s10508-015-0685-z

Auster, C., & Mansbach, C. (2012). The gender marketing of toys: An analysis of color and type of toy on the Disney store website. *Sex Roles, 67*, 375–388. https://doi.org/10.1007/s11199-012-0177-8

Berndt, T. J., & Heller, K. A. (1986). Gender stereotypes and social inferences: A developmental study. *Journal of Personality and Social Psychology, 50*, 889–898. https://doi.org/10.1037/0022-3514.50.5.889

Berry, T., & Wilkins, J. (2017). The gendered portrayal of inanimate characters in children's books. *Journal of Children's Literature, 43*(2), 4–15.

Biernat, M. (1991). Gender stereotypes and the relationship between masculinity and femininity: A developmental analysis. *Journal of Personality and Social Psychology, 61*, 351–365. https://doi.org/10.1037/0022-3514.61.3.351

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*, 411–419. https://doi.org/10.1023/A:1023574427720

Callanan, M. A., & Oakes, L. M. (1992). Preschoolers' questions and parents' explanations: Causal thinking in everyday activity. *Cognitive Development, 7*, 213–233. https://doi.org/10.3389/Ffpsyg.2020.586819

Campbell, A., Shirley, L., & Candy, J. (2004). A longitudinal study of gender-related cognition and behaviour. *Developmental Science, 7*(1), 1–9. https://doi.org/10.1111/j.1467-7687.2004.00316.x

Chouinard, M. M., Harris, P. L., & Maratsos, M. P. (2007). Children's questions: A mechanism for cognitive development. *Monographs of the Society for Research in Child Development*, i-129. https://doi.org/10.1111/j.1540-5834.2007.00412.x

Coenen, A., Nelson, J. D., & Gureckis, T. M. (2019). Asking the right questions about the psychology of human inquiry: Nine open challenges. *Psychonomic Bulletin & Review, 26*, 1548–1587. https://doi.org/10.3758/s13423-018-1470-5

Cunningham, S. J., & Macrae, C. N. (2011). The colour of gender stereotyping. *British Journal of Psychology, 102*, 598–614. https://doi.org/10.1111/j.2044-8295.2011.02023.x

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. https://doi.org/10.3758/BF03193146

Gülgöz, S., DeMeules, M., Gelman, S. A., & Olson, K. R. (2019). Gender essentialism in transgender and cisgender children. *PloS One, 14*(11), Article e0224321. https://doi.org/10.1371/journal.pone.0224321

Halim, M. L., Gutierrez, B. C., Bryant, D. N., Arredondo, M., & Takesako, K. (2018). Gender is what you look like: Emerging gender identities in young children and preoccupation with appearance. *Self and Identity, 17*, 455–466. https://doi.org/10.1080/15298868.2017.1412344

Halim, M. L., Ruble, D. N., Tamis-LeMonda, C. S., Zosuls, K. M., Lurye, L. E., & Greulich, F. K. (2014). Pink frilly dresses and the avoidance of all things girly: Children's appearance rigidity and cognitive theories of gender development. *Developmental Psychology, 50*, 1091–1101. https://doi.org/10.1037/a0034906

Hock, A., Kangas, A., Zieber, N., & Bhatt, R. S. (2015). The development of sex category representation in infancy: Matching of faces and bodies. *Developmental Psychology, 51*, 346–352. https://doi.org/10.1037/a0038743

Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & van Anders, S. M. (2019). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist, 74*(2), 171. https://doi.org/10.1037/amp0000307

Johnson, K. L., Lurye, L. E., & Tassinary, L. G. (2010). Sex categorization among preschool children: Increasing utilization of sexually dimorphic cues. *Child Development, 81*, 1346–1355. https://doi.org/10.1111/j.1467-8624.2010.01476.x

Kaiser, S. (1997). *The social psychology of clothing: Symbolic appearance in context* (2nd ed. revised). Fairchild, New York.

Kessler, S. J., & McKenna, W. (1978). *Gender: an ethnomethodological approach*. NewYork: Wiley.

Leaper, C. (2015). Gender and social-cognitive development. In R. M. Lerner (Series Ed.), L. S. Liben & U. Muller (Vol. Eds.), *Handbook of child psychology and developmental science* (7th ed.), Vol. 2: Cognitive processes (pp. 806–853). Wiley.

Leaper, C., & Bigler, R. S. (2018). Societal causes and consequences of gender-typing of children's toys. In E. S. Weisgram & L. M. Dinella (Eds.), *Gender-typing of children's toys: How early play experiences impact development*. American Psychological Association. https://doi.org/10.1037/0000077-013

Leman, P. J. & Tenenbaum, H. R. (2014). Introduction. In P. J. Leman & H. R. Tenenbaum (Eds.), *Gender development* (pp. xv–xxiii). Psychology Press.

LoBue, V., & DeLoache, J. S. (2011). Pretty in pink: The early development of gender-stereotyped colour preferences. *British Journal of Developmental Psychology, 29*, 656–667. https://doi.org/10.1111/j.2044-835X.2011.02027.x

Margett-Jordan, T., Falcon, R. G., & Witherington, D. C. (2017). The development of preschoolers' living kinds concept: A longitudinal study. *Child Development, 88*, 1350–1367. https://doi.org/10.1111/cdev.12709

Margett, T. E., & Witherington, D. C. (2011). The nature of preschoolers' concept of living and artificial objects. *Child*

Development, 82, 2067–2082. https://doi.org/10.1111/j.1467-8624.2011.01661.x

Martin, C. L., Andrews, N. C., England, D. E., Zosuls, K., & Ruble, D. N. (2017a). A dual identity approach for conceptualizing and measuring children's gender identity. *Child Development, 88*, 167–182. https://doi.org/10.1111/cdev.12568

Martin, C. L., Fabes, R. A., Hanish, L. D., Gaertner, B., Miller, C. F., Foster, S., & Updegraff, K. A. (2017b). Using an intergroup contact approach to improve gender relationships: a case study of a classroom-based intervention. In A. Rutland, D. Nesdale, & C. Spears Brown (Eds.), *The Wiley Handbook of Group Processes in Children and Adolescents* (pp. 435–454). Wiley.

Martin, C. L., & Little, J. K. (1990). The relation of gender understanding to children's sex-typed preferences and gender stereotypes. *Child Development, 61*, 1427–1439. https://doi.org/10.2307/1130753

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. (2010). Patterns of gender development. *Annual Review of Psychology*, *61*, 353–81. https://doi.org/10.1146/annurev.psych.093008.100511

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*, 870–881. https://doi.org/10.1007/s11199-009-9584-x

Navarro, R., Martínez, V., Yubero, S., & Larrañaga, E. (2014). Impact of gender and the stereotyped nature of illustrations on choice of color: Replica of the study by Karniol (2011) in a Spanish sample. *Gender Issues, 31*, 142–162. https://doi.org/10.1007/s12147-014-9122-1

Rogoff, B., Dahl, A., & Callanan, M. (2018). The importance of understanding children's lived experience. *Developmental Review*, *50*, 5–15. https://doi.org/10.1016/j.dr.2018.05.006

Ruble, D. N., Martin, C. L., & Berenbaum, S. (2006). Gender development. In N. Eisenberg (Ed.), *Handbook of child psychology*: *Social, emotional and personality development* (6th ed., Vol. 3, pp. 858–932). Wiley.

Ruggeri, A., Walker, C. M., Lombrozo, T., & Gopnik, A. (2021). How to help young children ask better questions?. *Frontiers in Psychology, 11*, 2908. https://doi.org/10.3389/fpsyg.2020.586819

Signorella, M. L., & Frieze, I. H. (2008). Interrelations of gender schemas in children and adolescents: Attitudes, preferences, and self-perceptions. *Social Behavior and Personality: An International Journal*, *36*(7), 941–954. https://doi.org/10.2224/sbp.2008.36.7.941

Spinner, L., Tenenbaum, H. R., Cameron, L., Wallinheimo, A. (2021). A school-based intervention to reduce gender-stereotyping. *School Psychology International*, *42*, 422–449. https://doi.org/10.1177/2F01430343211009944

Tenenbaum, H. R., Hill, D. B., Joseph, N., & Roche, E. (2010). 'It's a boy because he's painting a picture': Age differences in children's conventional and unconventional gender schemas. *British Journal of Psychology, 101*, 137–154. https://doi.org/10.1348/000712609X433122

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: An International Journal of Research and Practice, 14*(4), 365–381. https://doi.org/10.1002/icd.399

Weisgram, E. S., Bigler, R. S., & Liben, L. S. (2010). Gender, values, and occupational interests among children, adolescents, and adults. *Child Development, 81*, 778–796. https://doi.org/10.1111/j.1467-8624.2010.01433.x

Weisgram, E. S., Fulcher, M., & Dinella, L. M. (2014). Pink gives girls permission: Exploring the roles of explicit gender labels and gender-typed colors on preschool children's toy preferences. *Journal of Applied Developmental Psychology, 35*, 401–409. https://doi.org/10.1016/j.appdev.2014.06.004

Wright, K., Poulin-Dubois, D., & Kelley, E. (2015). The animate–inanimate distinction in preschool children. *British Journal of Developmental Psychology, 33*, 73–91. https://doi.org/10.1111/bjdp.12068