



Gender Labels on Gender-Neutral Colors: Do they Affect Children's Color Preferences and Play Performance?

Sui Ping Yeung¹ · Wang Ivy Wong¹

© Springer Science+Business Media, LLC, part of Springer Nature 2017

Abstract

Gender-typed color preferences are widely documented, and there has been increasing concern that they affect children's play preferences. However, it is unclear whether such color preferences exist across cultures, how they have emerged, and how gender color-coding affects performance. Chinese preschoolers ($n = 126$) aged 59 to 94 months were tested. First, we assessed their gender-typed color preferences using forced-choice tasks with color cards and pictures of neutral toys in gender-typed colors. Second, we tested if gender labels could affect color preferences by labeling two gender-neutral colors as gender-typed and assessed children's liking for them using a rating task and a forced-choice task with pictures of neutral toys in the labeled colors. Third, we assigned children a tangram puzzle (i.e., a puzzle using geometric pieces) painted either in the gender-appropriate or gender-inappropriate color and measured the number of pieces they completed and their speed. Results showed that Chinese children exhibited the same gender-typed color preferences as Western children did. Moreover, applying gender labels amplified a gender difference in color preferences, thus providing direct and strong evidence for the social-cognitive pathway underlying gender-typed preferences. Finally, color-coding as gender-appropriate or -inappropriate had no impact on performance but the gender labels improved boys' performance. These results add to knowledge on how gender-related information affects children's responses to the social world and suggest that the current gender color divide should be reconsidered.

Keywords Gender labels · Gender color-coding · Color preferences · Play performance · Gender differences

“Gender Revolution,” a special issue of National Geographic Magazine in January 2017, has caught worldwide attention (Goldberg 2017). A striking image is the photo of a transgender nine-year-old girl dressed in pink from head to toe on the cover. Other images show girls and boys surrounded by exclusively pink or blue possessions (Zuckerman 2017). It is easy to observe, for instance in shops and advertisements, that pink is commonly used in a wide range of products targeting girls and blue in products targeting boys. Pink and blue have become gender-typed as symbols of femaleness and maleness, respectively, and appear to be the most gender-typed among different colors in the recent decades (Chiu et al.

2006; Del Giudice 2012). The colors themselves can thus serve as visual gender labels (Wong and Hines 2015a).

The prevalence of gender labels and of gender color-coding (i.e., the use of gender-typed colors in differentiating objects by gender) may affect how children respond to the environment as proposed by gender schema theory (Martin and Halverson 1981). The possibility that such labels affect child development has aroused the concerns of parents, educators, and researchers. Although research has demonstrated a gender difference in children's color preferences and the effects of gender color-coding on children's gender assignment of and preferences for toys (Weisgram et al. 2014; Wong and Hines 2015a, b), these studies only provided a picture of the West and did not address how a gender difference in color preferences emerged. Moreover, there is little research on whether gender color-coding has behavioral consequences such as affecting performance. Therefore, the present study aimed to examine (a) if Chinese children would show gender-typed preferences for pink and blue, (b) if a gender difference in color preferences could be created by merely applying gender labels to the colors, and (c) if the colors, after becoming

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11199-017-0875-3>) contains supplementary material, which is available to authorized users.

✉ Wang Ivy Wong
iw Wong@hku.hk

¹ Department of Psychology, The University of Hong Kong, Pokfulam, Hong Kong, People's Republic of China

gender-typed, would affect children's performance in their play with materials coded in the color labeled as for their own or the other gender. Findings would contribute to revealing the social-cognitive pathway underlying gender-typed color preferences and the potential impacts of gender labels and gender color-coding.

Gender Schema Theory

Gender schema theory (Martin and Halverson 1981) proposed that once children have acquired gender identity, they begin to actively seek out gender-related information from the environment and assimilate the information into their gender schema, which then guides their behavior on what is appropriate or inappropriate for their gender (Fagot and Leinbach 1989; Martin and Ruble 2004; Martin et al. 2002). These standards of gender-appropriateness influence how children interact with their surroundings (Halim and Ruble 2010; Maccoby and Jacklin 1974). An example is that children's involvement in housework could be affected by parents' division of labor, with girls performing domestic chores such as cooking (an act usually performed by mothers) and boys performing maintenance chores like wall-painting (an act usually performed by fathers) (Antill et al. 1996; Basow 1992).

As for the case of colors, information about the gender attribute of colors may teach children that colors are gender-typed. In recent years, the marketing of children's merchandise has been increasingly gender-specific (Cunningham and Macrae 2011). For instance, Disney products, which dominate the children's entertainment industry across the globe, are highly gender-typed and provide strong cues in the gender attribute of colors by using pink pervasively in girl-typical toys such as dolls but bold colors including blue in boy-typical toys such as vehicles (Auster and Mansbach 2012). Another example is LEGO®'s "LEGO Friends" released in 2012. The line is designed for girls, with a lot more pink bricks used relative to traditional LEGO® sets targeting at boys (Black et al. 2016). The use of gender-typed colors in clothing and room décor is also prevalent (Pomerleau et al. 1990; Sweet 2013).

Because different colors are frequently paired with girl-typed or boy-typed objects, and because adults tend to choose products ranging from toys to clothes in these gender-typical colors for children (Kane 2006; Pomerleau et al. 1990), girls and boys have been bombarded with pink and blue, respectively, since their early years. Because of frequent exposure to the color divide, children may acquire information that colors are gender-typed, where pink is for girls and blue is for boys (Paoletti 2012). Once children incorporate the gender attribute of colors into their gender schema, they may regard the socially constructed "gender-typical" colors as appropriate for their

gender and the "gender-atypical" colors as inappropriate for their gender.

Gender-Typed Color Preferences and their Origin

Given that pink and blue have been strongly associated with the genders, gender differences in preferences for pink and blue have been found in studies using pure color stimuli (Chiu et al. 2006; Hurlbert and Ling 2007) or real objects of different colors (LoBue and DeLoache 2011; Picariello et al. 1990) and employing methods such as forced choices (LoBue and DeLoache 2011; Wong and Hines 2015b), observations (Wong and Hines 2015a) or self-report (Cohen 2013; Ellis and Ficek 2001). For example, when children aged 3–6 years-old were presented with toy felt pigs of different colors, girls tended to choose the pink one and boys the one in dark colors (e.g., navy blue) as their favorite (Picariello et al. 1990). When children aged around 2–3 years-old were asked to choose from pairs of identical objects in pink or blue or to play with identical toys in pink or blue, girls preferred pink items more and boys preferred blue items more (Wong and Hines 2015b).

However, these studies were conducted with Western samples. Some research examined the color preferences of Asians such as Chinese, Japanese, and Indonesian. However, they did not focus on children or on gender differences (Chattopadhyay et al. 2010; Saito 1994, 1996). It is unclear whether children from the East share the same gender-typed color preferences as Western children do, with girls preferring pink more and boys preferring blue more. Research with Chinese children could show the prevalence of such gender-typed color preferences in different cultures.

We should note that the origin of gender-typed color preferences is still unclear. Some researchers suggested that they are inborn, originating from differences in cone-contrast sensitivity underlying the visual system evolved from gender role divisions (e.g., hunting vs. fruit-picking) of early humans (Alexander 2003; Hurlbert and Ling 2007). Yet, this proposition has been challenged. A recent study found gender differences in the color preferences of British adults but not Himba adults (a nonindustrialized population), suggesting that gender-typed pink-blue preferences are not universal and are culturally based (Taylor et al. 2013). In addition, gender-typed preferences for pink and blue only appear to emerge when children turn 2 years-old (Jadva et al. 2010; LoBue and DeLoache 2011) and to become stronger as children grow older (Wong and Hines 2015b).

Chiu et al. (2006) provided further insight into the cause of gender-typed color preferences by comparing the color preferences of children with and without gender

identity disorder (GID), who express distress about their sex assigned at birth and identify themselves as the other gender. They found that girls without GID preferred pink more than did boys without GID but such preferences were reversed among children with GID, suggesting that these preferences result from identification with gender norms. Another evidence that the gender-typing of these colors is a cultural product is the finding that society's perception of these colors can be different across time. In the early 1900s, the pink-blue divide was not as rigid as today (Del Giudice 2017). Pink was sometimes viewed as a more masculine color whereas blue was sometimes regarded as a more feminine color (Paoletti 1987).

The non-universality, late emergence of the gender differences, the reversal of gender-typed color preferences among children with and without GID, and the malleability of the colors' gendered nature imply a social-cognitive pathway of gender-typed color preferences. Specifically, verbal gender labels have been shown to affect children's perception of objects' gender attribute and interest. When toys are labeled as for girls or for boys, children like the toys labeled as for their own gender more than the toys labeled as for the other gender (Masters et al. 1979; Weisgram et al. 2014). Although these studies showed that children establish gender-based knowledge when gender labels are applied to concrete materials (i.e., the toys), children may also establish gender-based knowledge on abstract qualities such as shapes and colors (Bem 1981; Leinbach et al. 1997). When gender labels are applied to gender-neutral abstract qualities such as colors, the colors may become gender-typed and their gender attribute may be assimilated into children's gender schema, which may then increase children's liking for the color labeled as for their own gender.

However, the possibility that gender differences for abstract qualities such as colors can emerge through a random labeling mechanism has not been tested directly. All the known studies that have so far been said to support a social-cognitive pathway for the gender difference in pink-blue preferences only provide indirect evidence. They are either based on whether the gender difference is found in certain cultures (e.g., Taylor et al. 2013) or on the age the gender difference is found or not (e.g., LoBue and DeLoache 2011). An experiment that manipulates the social-cognitive factors and that eventually creates a gender difference would offer direct and strong evidence for the social-cognitive pathway of gender-typed color preferences. Therefore, we tested whether applying gender labels to gender-neutral colors would affect girls' and boys' liking for these colors and create a gender difference in their color preferences. This evidence may help explain how, from a social-cognitive perspective, colors such as pink and blue, which might have been originally gender-neutral, have become gender-typed.

Behavioral Consequences

When colors become gender-typed, they can serve as visual gender labels that define the gender-appropriateness of objects. Thus, gender color-coding has been found to affect children's gender assignment of toys. Children aged 3-years-old already understand the gender-typing of pink and blue (Martin et al. 2012; Ruble and Martin 1998) and assign toys to girls or boys based on their colors (Cunningham and Macrae 2011; Weisgram et al. 2014). Color is children's most frequently cited reason when they sort ambiguous or neutral toys by gender (Cherney and Dempsey 2010). Gender color-coding also affects children's own preferences, with children expressing greater verbal interest for toys painted in gender-typical colors (Weisgram et al. 2014) and playing with the gender-atypical toy more when it is painted in a gender-typical color than when it is painted in a gender-atypical color (Wong and Hines 2015a).

Although these studies reveal the effects of gender color-coding on gender assignment of and preferences for objects, it is largely unknown whether gender color-coding would have other behavioral consequences. The current debate on the use of colors to intervene in play preferences is mainly concerned with how to encourage children, girls in particular, to play more with boy-typical toys as a way to improve spatial skills (Casey et al. 2008; Jirout and Newcombe 2015). Some suggest applying gender-typed colors to gender-atypical toys (Black et al. 2016) whereas others suggest removing the color divide altogether in order to avoid creating new gender stereotypes (Cunningham and Macrae 2011). Studies examining the play performance of children in the face of materials coded in gender-appropriate, gender-inappropriate, or gender-neutral colors will help to evaluate the developmental consequences of using colors as an intervention for children's play.

Only two known studies examined the effect of color on play performance. One examined the effect of the color of a masculine construction toy, LEGO® bricks, on children's play performance (Fulcher and Hayes 2017). The authors employed the idea of stereotype threat (i.e., an awareness of being judged adversely based on stereotypes; Spencer et al. 1999) and hypothesized that feminine colors would activate girls' stereotypes about inferior performance on a masculine task and thus girls, but not boys, would perform worse when they receive feminine-colored bricks (i.e., pink/purple) than when they receive masculine-colored bricks (i.e., blue/green). Results did not support this hypothesis; when children were instructed to build certain objects, brick color had no impact on the speed or accuracy of girls' or boys' construction. Another study (Mulvey et al. 2017) assessed engineering aptitude of preschool and primary school children when they were given feminine-colored (i.e., pastel colors) or masculine-colored (i.e., primary colors) engineering materials.

They hypothesized that counter-stereotypic colors would impede performance, especially that of boys, because it is more difficult for boys to act in counter-stereotypic ways. However, their study also found no main or interaction effects of color. These studies suggest that gender color-coding has a minimal effect on children's play performance.

However, it is too early to conclude that color has no impact on performance. Past research consistently showed that the gender-appropriateness of tasks affected how children performed. When a gender-neutral perceptual motor task was labeled as relating to electronics (i.e., stereotypically boy-typical) or to needlework (i.e., stereotypically girl-typical), children performed better when the labels were consistent with their gender (Davies 1986, 1989; Hargreaves et al. 1985). Other research showed that when children were told that an unfamiliar gender-neutral toy game was designed for their own gender, they tended to be more attracted to it and perform better (Montemayor 1977). This enhanced performance was suggested to be caused by an increased interest: when children feel interested, they become more attentive, more persistent, and more motivated to work hard on the task (Hidi 2000; Locke and Latham 1990; Van Yperen 2003). Because gender-typed colors provide cues about the gender-appropriateness of objects and affect interest (Weisgram et al. 2014), and because gender-appropriateness of the tasks can affect performance (Davies 1986), it is possible that colors, after becoming gender-typed, would serve as visual gender labels denoting the gender-appropriateness, alter interest, and as a result affect the performance of both girls and boys. Children engaging in a gender-neutral task (instead of a stereotype-relevant, masculine task as in Fulcher and Hayes 2017, and Mulvey et al. 2017) but assigned task materials coded in the color labeled as for their own gender (i.e., gender-appropriate) may perform better than those playing with materials coded in the color labeled as for the other gender (i.e., gender-inappropriate).

The Present Study

Given that it is unexamined whether gender-typed color preferences exist among children in the East, that the social-cognitive pathway of such preferences is still unclear, and that little is known about the behavioral consequences of gender color-coding, we proposed three sets of hypotheses. First, concerning preferences for pink versus blue, like Western children, we expect that Chinese girls will like pink more and boys will like blue more (Hypothesis 1).

Second, concerning the social-cognitive pathway, when gender labels are applied to gender-neutral colors, children will show a greater liking for the color labeled as for their own gender than for the color labeled as for the other gender and more so than will children in the condition where no

gender labels are applied (Hypothesis 2a). A between-gender difference in the preferences for these colors will also emerge when gender labels are applied, with girls preferring the girl-labeled color more than boys do and boys preferring the boy-labeled color more than girls do (Hypothesis 2b). Third, concerning the impact of gender color-coding, when gender labels are applied to the gender-neutral colors, children, both girls and boys, playing with materials painted in the color labeled as for their own gender will perform better than those playing with materials painted in the color labeled as for the other gender (Hypothesis 3).

Because our study investigated the effects of gender labels and gender color-coding on girls' and boys' cognition and behavior, it was important to study children who were able to identify their own gender and were at the stage of active acquisition of gender-related information (Martin and Halverson 1981). According to cognitive-developmental theory, gender development goes through three stages: gender identity, gender stability, and gender consistency (Kohlberg 1966). Although some research shows that not all children pass through these stages linearly (Cohen-Kettenis and Pfäfflin 2003), reviews have found that most children acquire gender identity at the age of two (Ruble et al. 2007; Zosuls et al. 2009). In addition, when children reach five years of age and before they reach the consistency stage around 7-years-old, they become very rigid in following gender norms (Ruble et al. 2007; Serbin et al. 1993). Therefore, we studied children at preschool years aged around 5–7 years-old.

Method

Participants

We recruited 129 preschoolers from two kindergartens in Hong Kong. Three participants were excluded from analyses: One boy was reported by his parent as having color weakness, one girl had an outlier value with a *z*-score above 3 on the tangram task (i.e., a puzzle using geometric pieces), and one girl withdrew. The remaining 126 participants had normal color vision and no learning difficulties as reported by their parents. All participants were Chinese, aged 59 to 94 months ($M = 67.89$ [5.66 years], $SD = 5.59$). There were 61 boys (48.4%; $M_{age} = 68.10$ months, $SD = 6.28$) and 65 girls (51.6%; $M_{age} = 67.69$ months, $SD = 4.89$). One parent of each participant completed a questionnaire on demographic characteristics. Four parents (3.2%) did not report monthly household income. Of others who did, the income ranged from HKD8,000 to HKD100,000 (i.e., around US\$1025–12,820) with a mean of HKD35,557 (i.e., around US\$4558) and a standard deviation of HKD21,458 (i.e., around US\$2751). Three parents (2.4%) did not report their own and their spouse's education level. Of those who did, 35.8% of fathers

($n = 45$) and 29.4% of mothers ($n = 37$) had tertiary qualifications, 32.5% of fathers ($n = 41$) and 42.9% of mothers ($n = 54$) completed high school, and 29.4% of fathers ($n = 37$) and 25.4% of mothers ($n = 32$) completed junior secondary education or below.

Design and Procedure

The full design was a 2 (Gender: girls vs. boys) \times 2 (Label: yes vs. no) \times 2 (Color-coding: gender-appropriate vs. gender-inappropriate). Participants were randomly assigned to the experimental (label) or the control (no label) group. Only the label group was exposed to gender labels (Girls \times Label: $n = 34$; Girls \times No label: $n = 31$; Boys \times Label: $n = 30$; Boys \times No label: $n = 31$). They were then further randomly assigned to the gender-appropriate or the gender-inappropriate color condition (with $n = 15$ as the smallest cell size). Those in the gender-appropriate condition received play materials (i.e., tangram puzzles that use geometric pieces) in the color labeled as for their gender and those in the gender-inappropriate condition received play materials in the color labeled as for the other gender.

Our study had received ethical approval from an institutional research ethics committee of a local university. All children participated with the written consent of their parent. Children's verbal assent was also obtained prior to the experiment. Children were tested individually in a quiet room in the kindergarten by a female experimenter. The experimenter wore black so as not to provide any color cues.

We first assessed children's preferences for pink versus blue by showing them pink-blue pairs of color cards and pictures of toys. After that, we evaluated children's pre-existing likings for two colors that were found to be gender-neutral in a pilot test, yellow and green, by showing them yellow and green color cards. After this came the manipulation procedure. Only children in the label group were told that yellow is a color for girls and green for boys. The manipulation was checked by asking children to indicate which color is for girls and which is for boys. After the manipulation, children's new preferences for yellow and green were examined by showing them pictures of toys in yellow or green. Before moving on to the tangram task, children in the label group were asked to indicate the gender attribute of yellow and green again so as to ensure that they remembered the gender labels. Then, all children were given a tangram either in yellow or green and had ten minutes to complete as many tangram patterns as they could.

After completing all the testing procedures, children were debriefed. The experimenter explained clearly to the children in the label group that yellow and green are in fact colors for both genders and that both girls and boys can like these two

colors and play with toys in these colors as they wish. All children expressed their understanding.

Materials and Measures

Preferences for Pink Versus Blue

To assess children's pink-blue preferences, we used two forced-choice tasks to increase reliability (Wong and Hines 2015b). In the first task, six pure color cards (sized 12 cm \times 12 cm)—three from the blue collection and three from the pink collection—were employed (see online supplement, Fig. 1s, a–f). Each card displayed one of these colors (hues indicated in brackets): Greenish blue (116), navy blue (158), sky blue (136), purplish pink (207), reddish pink (242), and typical pink (221). These colors all had a saturation level at 240 and a luminance level at 140. Their hues were determined in a pilot test of ten adults (five male, five female), who were shown 86 shades created on Microsoft Power Point, hues ranging from 0 to 255, with 3-point intervals. They indicated the shade they thought was the most representative of each of the above colors and their choices were averaged. The six color cards formed nine pink-blue pairs. Children were shown each pair in random order. The left-right position of the pink and the blue color cards was counterbalanced. Children pointed at the color they liked more. A point was given when they pointed at pink. The total score could thus range from zero to nine.

In the second task, pictures (sized 15 cm \times 15 cm) displaying three pink-blue pairs of gender-neutral play materials were used (see online supplement, Fig. 1s g–h, for an example of a pink-blue pair). In each pair, the materials were identical except that one was pink and one was blue. The sets of materials were balloons, crayons, and star stickers, which were rated or used in previous studies as gender-neutral (Arthur et al. 2009; Blakemore and Centers 2005; Masters et al. 1979; Wong and Hines 2015b).

Children were shown each pair of play materials in random order. The left-right position of the pink and the blue items was counterbalanced. Children pointed at the item they liked more in each pair. A point was given when the pink item was chosen. The total score could thus range from zero to three. Scores for the two tasks were positively correlated, $r(124) = .74$, $p < .001$. The raw score of each task was converted into a standardized score (z -score). The two z -scores were then averaged to form a color composite score. A positive score indicated a greater liking for pink and a negative score indicated a greater liking for blue.

Manipulation of Gender Labeling

In the experiment, gender labels were applied to two gender-neutral colors: Typical yellow (42) and typical green (79). Both colors had a saturation level at 240 and

a luminance level at 140, and their hues were again determined by ten adults with the said pilot procedure. The gender-neutrality of these two colors was pilot tested with ten children (four boys, six girls) aged 4–10 years-old. The ten children sorted 13 colors (i.e., typical yellow, typical green, the six colors used to assess pink-blue preferences, plus typical red [0], reddish orange [8], typical orange [16], yellowish orange [27], and yellowish green [50]) as “for boys,” “for girls,” or “for both boys and girls.” Typical yellow and typical green were perceived as the two most gender-neutral colors, with eight of ten children and nine of ten children sorting them as for both genders, respectively.

Only the label group was exposed to the following manipulation procedure. For the no label group, no such labeling procedure was employed. Given the gender-neutrality of both colors, by the researchers’ arbitrary assignment, yellow was always labeled as “for girls” and green as “for boys.” The 12 cm × 12 cm color cards of yellow and green were shown to the children one by one (see online supplement, Fig. 2s a–b). The experimenter presented the yellow card and said: “In fact, yellow is a color for girls. Many girls at your age told me that they like yellow very much because they think yellow is a symbol of girlhood. I think so, too. Yellow is a color for girls.” The experimenter also presented the green card and said: “In fact, green is a color for boys. Many boys at your age told me that they like green very much because they think green is a symbol of boyhood. I think so, too. Green is a color for boys.” The order of the colors presented was counterbalanced across participants. To check the manipulation, children were asked to indicate which color is for girls and which color is for boys by pointing at the corresponding color card according to what the experimenter had said. If they failed to identify the labels correctly, the labeling procedure was repeated.

Preferences for Yellow Versus Green

Prior to the yellow-green color manipulation, participants’ pre-existing likings for typical yellow and typical green were assessed. Both the label and the no label groups were shown the yellow and the green color cards, one-by-one in random order, and were asked to indicate how much they liked each color by pointing at one of five schematic faces morphing from a frown (1 = *strongly dislike*) to a big smile (5 = *strongly like*). These faces were in black-and-white so as not to distract children from the color stimuli or provide any color cues. This procedure allowed us to control for children’s pre-existing likings for yellow and green in subsequent analyses.

After the test for pre-existing likings and the manipulation, children’s new preferences for yellow and green were assessed with two tasks. In the first task, 21 pictures (sized 15 cm × 15 cm) of gender-neutral play materials

were used. Sixteen of them displayed eight pairs of identical materials in which one was yellow and one was green. The eight pairs of picture illustrations were balloons, crayons, kites, play dough, sand toy sets, slinkies, star stickers, and xylophones (see online supplement, Fig. 2s c–d, for examples of picture illustrations). The remaining five pictures were fillers displaying another five play materials (i.e., cash register, doctor kit, drawing board, karaoke machine, tricycle) in various colors except yellow and green. The fillers were to mask the focus on yellow and green so as to elicit more implicit responses. All of these play materials were rated or used in past studies as gender-neutral (Arthur et al. 2009; Blakemore and Centers 2005; Masters et al. 1979; Wong and Hines 2015b). Children were presented with the pictures one by one in random order. They were asked, “How much do you like it?” and were told to indicate their liking on a 5-point scale by pointing at the corresponding schematic face. Their ratings were added up respectively to generate two scores: Total liking for materials in yellow and total liking for materials in green, each with a minimum of eight points and a maximum of 40 points.

In the second task, the 16 pictures of yellow and green play materials used in the prior task that formed eight yellow-green pairs were administered as a forced-choice task. Each pair was presented in random order and the left-right position of the yellow and the green items was counterbalanced. Children were asked to point at the picture they liked more in each pair. The numbers of chosen yellow and green items were both recorded, each with a minimum of zero points and a maximum of eight points.

To reduce the number of analyses and to better reflect children’s relative preferences for yellow versus green, the score of liking for green was subtracted from the score of liking for yellow to generate a difference score indicating children’s preference for yellow over green for each task, before and after the manipulation. That is, each child had three difference scores: (a) one was before the manipulation indicating their pre-existing liking that would be controlled for in subsequent analyses and the other two were after the manipulation; (b) one from the rating task and (c) the other from the forced-choice task. A positive score indicated a greater preference for yellow over green and a negative score indicated a greater preference for green over yellow. To increase reliability, scores of the two tasks after the manipulation, which were positively correlated, $r(124) = .36$, $p < .001$, were combined for analysis. The raw difference scores were converted into z -scores to make them comparable. They were averaged to form a color composite score for new preference for yellow over green after the manipulation. A positive score indicated a greater liking for yellow and a negative score indicated a greater liking for green.

Tangram Task

The impact of gender color-coding on play performance was assessed with tangram. Tangram is a puzzle comprising seven geometric pieces including a square, a parallelogram and five triangles different in size. Tangram had been pilot tested to be gender-neutral, with ten of ten adults and nine of ten children rating it as for “both genders.” Two sets of 13 cm × 13 cm tangram were used: One was painted in yellow and the other was in green (see online supplement, Fig. 3 s, part a). The colors of the tangrams were in the exact hue, saturation, and luminance level as the colors displayed in the test for pre-existing likings and the labeling procedure.

Children in the label or no label group were further randomly assigned to the gender-appropriate or gender-inappropriate color conditions. In the gender-appropriate color condition, girls were given a yellow tangram and boys a green tangram (i.e., to children from the label group, the tangram color had been labeled as for their own gender; to those from the no label group, the tangram color was gender-neutral because participants had not been exposed to any gender labels). In the gender-inappropriate color condition, girls were given a green tangram and boys a yellow tangram (i.e., the tangram color had been labeled as for the other gender).

The tangram task required children to form ten patterns. Each pattern was pilot tested with ten adults to be gender-neutral, with at least seven respondents indicating it as having no connotation of gender. These adults also rated each pattern’s difficulty level on a 5-point scale (1 = *very easy*; 5 = *very difficult*). Their ratings were averaged to determine the difficulty level of the patterns. The patterns in the order from the lowest to the highest level of difficulty were fir tree, bird, fish, house, teapot, dog, turtle, t-shirt, tree, and whale (downloadable from Tangram Channel: <https://www.tangram-channel.com>). Each was displayed in silhouette in one-to-one size ratio to the tangram and printed on an A3 paper against a light grey background. To fit the ability of preschoolers, in each silhouette, the experimenter always placed one of the largest triangles and the parallelogram at their correct positions (see online supplement, Figs. 3 s b–g for examples of silhouettes). Children only needed to allocate the remaining five pieces of tangram to complete each pattern.

Before the commencement of the task, the labeling manipulation was checked again by asking children in the label group to indicate which color is for girls and which color is for boys so as to ensure that they remembered the gender labels. After that, children in both conditions were given a practice trial. They were taught to rearrange the five separate pieces of tangram into a duck pattern. All five pieces must be used and laid flat next to each other closely without overlap.

When the test trial began, the ten patterns were presented to the children one-by-one in order of increasing difficulty. Children were told to try their best on their own to

complete the patterns by placing the tangram pieces to the corresponding position on the silhouette within 10 min. Once children successfully completed a pattern, the experimenter immediately removed the five tangram pieces from the silhouette of that pattern, gave them back the five tangram pieces together with a silhouette of another pattern, and asked them to continue. When children asked for help, claimed that they completed the shape when they actually did not, or expressed their wish to give up, the experimenter did not provide any hints but instead encouraged them to keep trying. No participant quitted the task midway. The numbers of tangram pieces correctly placed on each presented silhouette during the test were added to generate a total score. Because each pattern needed five pieces to complete and because children could be presented with ten patterns at most, the maximum score was 50. The average time in seconds needed to place each piece to the corresponding position was also recorded (i.e., 600 s divided by the total number of tangram pieces completed). These two variables reflected children’s performance on the tangram task.

Results

Preliminary Analyses

To ensure group comparability, Chi-squared tests and one-way ANOVAs were conducted. The groups did not differ in gender, age, monthly household income, or parental education (all p s > .05), indicating that the group assignments were random and that the groups were comparable. To identify potential covariates to be included in subsequent analyses, we looked at the correlations of the demographic variables with the outcome variables. Monthly household income positively correlated with children’s new preference for yellow over green after the manipulation, $r(124) = .24$, $p = .008$. We also found that father’s education correlated positively with the number of tangram pieces completed, $r(124) = .20$, $p = .029$, and negatively with the average time needed to correctly place each tangram piece, $r(124) = -.20$, $p = .027$. Thus, to prevent confounding, monthly household income was controlled for when analyzing children’s new preference for yellow over green and parental education was controlled for when analyzing their performance on the tangram task.

Because an independent-samples t -test showed that, before manipulation, girls ($M = .86$, $SD = 1.31$) liked yellow more (or green less) than did boys ($M = -.28$, $SD = 1.61$), $t(124) = -4.37$, $p < .001$, $d = .78$, despite that yellow and green were perceived by the children in the pilot test as gender-neutral, we also looked at the correlation of pre-existing liking for yellow over green with the outcome variables. The analyses indicated that pre-existing liking correlated positively with children’s

new preference for yellow over green after the manipulation, $r(124) = .36, p < .001$, and the number of tangram pieces completed, $r(124) = .22, p = .014$, and negatively with the average time needed to complete each piece, $r(124) = -.27, p = .002$. Therefore, for analyses that included children's new preference for yellow over green and their performance on the tangram task, pre-existing liking for yellow over green was statistically controlled.

Hypotheses Testing

Hypothesis 1: Preferences for Pink Versus Blue

A planned independent-samples *t*-test was conducted to examine children's preference for pink over blue. There was a significant difference in girls' and boys' preferences for pink over blue, with girls ($M = .60, SD = .62$) liking pink more (i.e., blue less) than did boys ($M = -.64, SD = .77$), $t(124) = -9.89, p < .001, d = -1.76$.

Hypothesis 2: Preferences for Yellow Versus Green

A 2 (Gender) \times 2 (Label: yes vs. no) ANCOVA was conducted to test for children's new preference for yellow over green after controlling for pre-existing liking and monthly household income. There was a significant main effect of gender, such that girls ($M = .33, SD = .08$) preferred yellow to green more than did boys ($M = -.32, SD = .09$), $F(1,116) = 27.90, p < .001, d = .97$. A significant two-way interaction between gender and labeling was found, $F(1,116) = 14.14, p < .001, d = .70$ (see Fig. 1). Pairwise comparisons showed that within gender, girls in the label group ($M = .55, SD = .11$) had a greater liking for yellow over green than did girls in the no label group ($M = .12, SD = .12$), $F(1,116) = 7.85, p = .006, d = .52$, and that boys in the label group ($M = -.53, SD = .12$) had a greater liking for green over yellow (or a lesser liking for

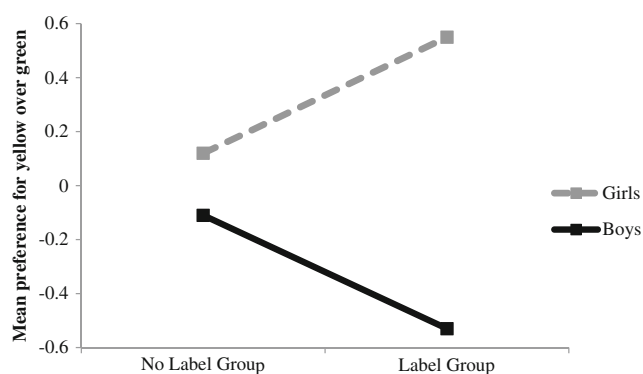


Fig. 1 The significant interaction between gender and labeling for children's new preference for yellow over green after the manipulation, with positive scores indicating a greater liking for yellow (labeled as for girls) over green (labeled as for boys) and negative scores indicating a greater liking for green over yellow

yellow over green) than did boys in the no label group ($M = -.11, SD = .12$), $F(1,116) = 6.40, p = .013, d = .47$. Pairwise comparisons also indicated that in the no label group, girls ($M = .12, SD = .12$) and boys ($M = -.11, SD = .12$) did not differ in their likings for yellow over green, $F(1,116) = 1.70, p = .195, d = .24$, but in the label group, girls ($M = .55, SD = .11$) liked yellow more than did boys ($M = -.53, SD = .12$), or, in other words, boys liked green more than did girls, $F(1,116) = 41.56, p < .001, d = 1.19$. In short, children in the label group, both girls and boys, preferred the color labeled as for their own gender but children in the no label group did not show such preference.

Hypothesis 3: Tangram Performance

We conducted 2 (Gender) \times 2 (Labeling) \times 2 (Color-coding: gender-appropriate vs. gender-inappropriate) ANCOVAs to examine children's performance on the tangram task after controlling for pre-existing color liking and parental education. In terms of the number of pieces completed, there were no main effects of gender, labeling, or color-coding, nor a three-way interaction. The hypothesized interaction between labeling and color-coding was not significant. However, a two-way interaction between gender and labeling was found, $F(1,112) = 4.03, p = .047, d = .38$ (see Fig. 2). Pairwise comparisons showed that within gender, boys in the label group ($M = 31.80, SD = 1.74$) completed more tangram pieces than did boys in the no label group ($M = 26.61, SD = 1.66$), $F(1,112) = 4.85, p = .03, d = .42$, whereas the performance of girls in the label group ($M = 25.51, SD = 1.61$) did not differ from that of girls in the no label group ($M = 26.96, SD = 1.74$), $F(1,112) = .39, p = .534, d = .11$. Pairwise comparisons also showed that in the no label group, the number of tangram pieces boys ($M = 26.61, SD = 1.66$) and girls ($M = 26.96, SD = 1.74$) completed did not differ, $F(1,112) = .021, p = .886$, but in the label group, boys ($M = 31.80, SD = 1.74$) completed more tangram pieces than did girls ($M = 25.51, SD = 1.61$), $F(1,112) = 6.72, p = .011, d = .49$. As for the

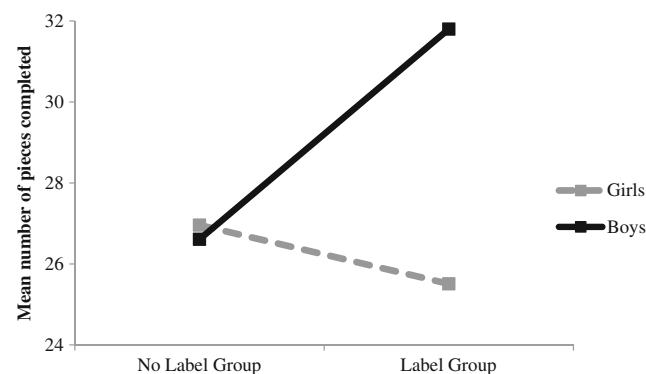


Fig. 2 The significant interaction between gender and labeling for children's performance on the tangram task in terms of the number of pieces completed

average time needed to complete each tangram piece, unlike the results of the number of pieces completed, there were no main effects or two-way or three-way interactions. In sum, labeling magnified the gender difference in the number of pieces completed because boys in the label group completed more pieces than any other group.

Discussion

The present study looked into children's gender-typed color preferences and the effects of gender labels and gender color-coding on both preferences and performance. We demonstrated that, in support of a social-cognitive pathway, randomly applied gender labels could amplify gender differences in preferences for otherwise gender-neutral colors. More importantly, we found that although children's play performance was not affected by whether the color of the play material was gender-appropriate or -inappropriate, exposure to any gender labels enlarged gender differences in performance.

Gender-Typed Color Preferences in Chinese Children

Consistent with our first hypothesis and studies conducted in the West (e.g., Picariello et al. 1990; Wong and Hines 2015b), Chinese children showed gender-typed preferences for pink versus blue. Although one other study found an absence of these gender-typed preferences in a remote non-industrialized culture (Himba; Taylor et al. 2013), these preferences are present in young children in non-Western, industrialized cultures. This finding is not surprising given the high degree of Westernization and the prevalence of gender color-coding typical of Western cultures in Hong Kong (Thomas 1999). Indeed, many gender differences and stereotypes in developed Asian regions resemble those in the West (Chen and Rao 2011; Lee and Collins 2008; Yu et al. 2010). One point to note is that the effect size of the gender difference in pink-blue preferences ($d = -1.76$) is very large. Given that gender differences in other psychological areas are generally smaller than 1 (Hines 2010) and most are smaller than .3 (Hyde 2005), our finding supports the notion that gender-typed liking for pink versus blue is a particularly salient gender difference. We did not include adults, but a few studies have found adult Chinese to show similar gender-typed preferences for pink and blue (Hurlbert and Ling 2007), so it is likely that these gender differences will not disappear in these Chinese children when they grow older.

Social-Cognitive Influences on Gender-Typing

Our second set of hypotheses was also supported. As predicted by gender schema theory (Martin and Halverson 1981) and in accordance with past research on the effects of gender labels

on children's preferences (Masters et al. 1979; Weisgram et al. 2014), we found that both girls and boys in the label group had a greater liking for the color arbitrarily labeled as for their own gender than did children in the no label group, suggesting that, by applying gender labels, not only concrete materials such as toys could become gender-typed, but also abstract qualities such as colors, with children increasing or decreasing their likings for particular colors based on the gender labels available in their social environment.

Moreover, our findings revealed that gender differences could be created merely by applying gender labels. The interaction effect between gender and labeling indicated that, after controlling for children's pre-existing liking for the colors, when the colors were not attached with gender labels, girls' and boys' preferences for the colors did not differ ($d = .24$); but when gender labels were attached to the colors, a corresponding gender difference in the preferences for these two colors emerged. The effect size of this newly developed gender difference ($d = 1.19$) is larger than that of many gender differences in other psychological areas ($d < 1$; Hines 2010), indicating the powerfulness of gender labels in giving rise to gender differences. By manipulating the gender labels in an experimental setting, the present study has provided direct and strong evidence for social-cognitive influences on children's gender-typed color preferences.

Gender Labels, Gender Color-Coding, and Play Performance

Our results did not support the third hypothesis. We examined the potential impact of gender color-coding and expected that when gender labels were applied, children playing with task materials painted in the gender-appropriate color would perform better than those playing with materials painted in the gender-inappropriate color whereas the performance of children in the no label group would not differ regardless of the color of the tangram. However, the non-significant interaction between labeling and color-coding showed that colors, either gender-appropriate or gender-inappropriate, did not improve or impair children's performance. Post-hoc power analyses using G*Power (Erdfelder et al. 1996) suggest that the non-significant results are unlikely to be due to a lack of statistical power because the current study has a power of .80 to detect medium-sized effects ($f = .25$). It is also unlikely that the non-significant effect of gender color-coding is due to the manipulation being too weak to turn the gender-neutral colors into gender-typed or that an effect on performance would have been found if colors that are more gender-typed were manipulated. In fact, the null result concerning color-coding coincides with two studies using colors that are already strongly gender-typed (i.e., pink vs. blue and pastel colors vs. primary colors), which also found no significant effect of the gender-appropriateness of color on the aptitude of brick building

(Fulcher and Hayes 2017) and engineering play (Mulvey et al. 2017). Therefore, color-coding is unlikely to have a substantial impact on children's performance.

Despite the non-significant impact of colors, we found an effect of gender labels on boys' performance. The significant interaction between gender and labeling suggested that exposure to gender labels improved boys' (but not girls') performance on the subsequent tangram task regardless of whether the boys received the tangram painted in the gender-appropriate or gender-inappropriate color. Boys who were exposed to labels completed more pieces than did boys in the no label group and also girls in the label group. Previous studies found that both girls and boys performed better when the task or the game was explicitly and directly labeled as for their own gender than when it was labeled as for the other gender (Davies 1986; Montemayor 1977). Although some studies showed that the effect of gender labels was more apparent in boys' play performance than in girls' (Gold and Berger 1978; Stein et al. 1971), these studies manipulated the labels by directly labeling the task as either gender-appropriate or gender-inappropriate and compared the performance of children encountering different gender labels.

Unlike past research, in the present study, all children in the label group were exposed to the same gender labels (i.e., yellow is for girls and green is for boys) and the only difference they encountered was the color of the task material they received. Although the colors were shown to have minimal effects, our results provided another picture of the effect of gender labels by showing that as long as boys had been exposed to information about gender-appropriateness, their subsequent performance improved. A possible explanation for the boys' enhanced performance is the stereotype boost effect, which refers to the phenomenon that an individual's performance on a stereotype-relevant task enhances when the positively stereotyped group identity is made salient through environmental cues (Armenta 2010; Shih et al. 2002). For example, when Asian women were reminded of their ethnic identity, their performance on the math tests improved; but when their gender identity was made salient, they performed worse (Ambady et al. 2001; Shih et al. 1999; Steele and Aronson 1995).

The tangram task used in the present study was viewed as gender-neutral (i.e., "for both boys and girls") by both children and adults in the pilot test. However, because playing with a tangram involves spatial skills (Lee et al. 2009), which often find a male advantage across countries and ages (Linn and Petersen 1985; Voyer et al. 1995), children may expect that boys are better at playing with a tangram. Based on past studies showing a stronger effect of gender labels on boys than on girls (Gold and Berger 1978), it is possible that the gender cues in our study may have aroused only the boys' awareness of their gender identity and then activated their cognition about male superiority in

spatial ability, which then boosted their confidence and improved their performance on the tangram task.

Limitations and Future Research Directions

Although our study contributed knowledge of the cultural consistency and social-cognitive pathway of gender-typed color preferences and how gender labels and gender color-coding affect performance, there are limitations. First, although revealing a possible social-cognitive mechanism on how children develop gender-typed color preferences, we cannot rule out inborn factors. Second, we picked typical yellow and typical green for manipulation because in the pilot test, the majority of children sorted them as for both genders. Yet, when we looked at the participants' pre-existing likings for these two colors, a gender difference was found, with girls liking yellow more and boys liking green more, although the effect size ($d = .78$) was still much smaller than that of the gender-typed preferences for pink versus blue ($d = 1.76$). It is possible that children in the pilot study inaccurately assumed the gender attribute of yellow and green. It is also possible that these two colors are not defined by society as gender-typed but inborn factors play a part in shaping this gender difference. However, the pre-existing liking for yellow over green was controlled for in subsequent analyses, so a pre-existing gender difference did not confound the effect of the gender labels on children's post-manipulation liking.

Third, our study found that exposure to gender labels improved boys', but not girls', subsequent play performance, thus creating a male advantage in performance. We propose that such enhanced performance may be explained by a stereotype boost effect (i.e., the activation of male identity and of the belief in a male superiority in stereotype-relevant domains; Shih et al. 2002). However, it is still unknown why only boys' performance changed after receiving information about gender-appropriateness, regardless of whether they were given gender-appropriate or gender-inappropriate materials. Future research can further examine this phenomenon and the underlying mechanism. It would also be interesting to take into account individual's gender-role attitudes as well as society's level of gender equality because these factors may affect the way individuals process gender-related information. We also suggest that future research should assess and control for children's spatial abilities and past experience with the toy (i.e., tangram as in our study) which may moderate the manipulation effects on children's performance.

Practice Implications

Our study has practice implications for toymakers and parents. By showing that gender-typed color preferences are prevalent even in an Eastern society and for the first time that gender labels can create gender differences in not only color

preferences but also play performance, we believe that, large-scale, cross-cultural gender-specific marketing, in which the majority of toys targeting girls and boys are coded in different colors, put in different aisles, and labeled as “For Girls” or “For Boys” (Auster and Mansbach 2012) is liable for these gender differences, which can lead to long-term developmental outcomes.

Different types of toys offer different learning experience. For example, playing with girl-typical toys promotes social abilities and playing with boy-typical toys promotes spatial abilities (Blakemore and Centers 2005). However, girls and boys differ in social and spatial skills (Ickes et al. 2000; Voyer et al. 1995), and they tend to engage in gender-typical play and avoid cross-gender activities (Green et al. 2004). Researchers and educators have thus advocated for children to play with both gender-typical and gender-atypical toys so that they can develop a larger variety of skills (Caldera et al. 1989; Cherney and London 2006; Li and Wong 2016; Sprafkin et al. 1983). Some proposed making use of the gender color divide by applying gender-typed colors to gender-atypical toys to encourage cross-gender play (e.g., Black et al. 2016), a strategy which may be effective in increasing preference as previous research indicated (e.g., Weisgram et al. 2014; Wong and Hines 2015a). Incidentally, LEGO®’s sales to girls increased substantially since they launched the LEGO® for Girls line, which, among other marketing strategies, involves a heavy use of pink (Wachman 2012). Despite finding no effect of colors on performance, our findings on the effects of gender labels on performance do not support this reversal or the current pink-blue divide. They suggest that it is the exposure to gender labels or reminder of a gender divide per se, rather than whether a gender-appropriate or -inappropriate version was given to children, that had an impact on play performance. That is, although applying gender-typed colors to cross-gender toys may achieve the aim of attracting children to play more with certain toys (e.g., “getting girls to build”), the trade-off may be an unintended widening of a gender gap in performance. We suggest that toymakers and parents avoid gender-labeling the toys, remove the color divides, and simply adopt a wide range of colors for both boys’ and girls’ toys.

Conclusion

Our study contributes to showing the prevalence of gender-typed preferences for pink versus blue by showing that such preferences were also observed in Chinese children, with girls liking pink more and boys liking blue more. Another significant contribution of our study is that we have provided direct and strong evidence of social-cognitive influences on the development of gender-typed color preferences by demonstrating that a gender difference in color preferences could be created merely by

gender labels. Besides exploring preferences, our study also examined the effects of gender color-coding and gender labels on performance and found that gender color-coding had minimal effect on performance, but having any gender labels could widen the gender gap in play performance. The present study facilitates the understanding of how gender-related information affects children’s development and suggests that the current gender color divide should be reconsidered.

Compliance with Ethical Standards This project was approved by the Psychology Departmental Research Ethics Committee at the University of Hong Kong. All children participated with the written consent of their parent. Children’s verbal assent was also obtained prior to the experiment.

References

- Alexander, G. M. (2003). An evolutionary perspectives of sex-typed toy preferences: Pink, blue and the brain. *Archives of Sexual Behavior*, 32, 7–14. <https://doi.org/10.1023/A:1021833110722>.
- Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science*, 12, 385–390. <https://doi.org/10.1111/1467-9280.00371>.
- Antill, J. K., Goodnow, J. J., Russell, G., & Cotton, S. (1996). The influence of parents and family context on children’s involvement in household tasks. *Sex Roles*, 34(3), 215–236. <https://doi.org/10.1007/BF01544297>.
- Armenta, B. E. (2010). Stereotype boost and stereotype threat effects: The moderating role of ethnic identification. *Cultural Diversity and Ethnic Minority Psychology*, 16(1), 94–98. <https://doi.org/10.1037/a0017564>.
- Arthur, A. E., Bigler, R. S., & Ruble, D. N. (2009). An experimental test of the effects of gender constancy on sex typing. *Journal of Experimental Child Psychology*, 104, 427–446. <https://doi.org/10.1016/j.jecp.2009.08.002>.
- Auster, C. J., & Mansbach, C. S. (2012). The gender marketing of toys: An analysis of color and type of type on the Disney store website. *Sex Roles*, 67, 375–388. <https://doi.org/10.1007/s11199-012-0177-8>.
- Basow, S. A. (1992). *Gender stereotypes and roles* (3rd ed.). Pacific Grove: Brooks/Cole Publishing Company.
- Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review*, 88, 354–364. <https://doi.org/10.1037/0033-295X.88.4.354>.
- Black, R. W., Tomlinson, B., & Korobkova, K. (2016). Play and identity in gendered LEGO franchises. *International Journal of Play*, 5, 64–76. <https://doi.org/10.1080/21594937.2016.1147284>.
- Blakemore, J. E. O., & Centers, R. E. (2005). Characteristics of boys’ and girls’ toys. *Sex Roles*, 53(9–10), 619–633. <https://doi.org/10.1007/s11199-005-7729-0>.
- Caldera, Y. M., Huston, A. C., & O’Brien, M. (1989). Social interactions and play patterns of parents and toddlers with feminine, masculine and neutral toys. *Child Development*, 60, 70–76. <https://doi.org/10.2307/1131072>.
- Casey, B. M., Andrews, N., Schindler, H., Kersh, J. E., Samper, A., & Copley, J. (2008). The development of spatial skills through interventions involving block building activities. *Cognition and Instruction*, 26, 269–309. <https://doi.org/10.1080/07370000802177177>.

- Chattopadhyay, A., Gorn, G. J., & Drake, P. (2010). Differences and similarities in hue preferences between Chinese and Caucasians. In A. Krishna (Ed.), *Sensory marketing: Research on the sensuality of products* (pp. 219–239). New York: Taylor and Francis Group.
- Chen, E. S. L., & Rao, N. (2011). Gender socialization in Chinese kindergartens: Teachers' contributions. *Sex Roles*, *64*, 103–116. <https://doi.org/10.1007/s11199-010-9873-4>.
- Cherney, I. D., & Dempsey, J. (2010). Young children's classification, stereotyping, and play behavior for gender neutral and ambiguous toys. *Journal of Educational Psychology*, *30*(6), 651–669. <https://doi.org/10.1080/01443410.2010.498416>.
- Cherney, I. D., & London, K. (2006). Gender-linked differences in the toys, television shows, computer games, and outdoor activities of 5- to 13-year-old children. *Sex Roles*, *54*, 717–726. <https://doi.org/10.1007/s11199-006-9037-8>.
- Chiu, S. W., Gervan, S., Fairbrother, C., Johnson, L. L., Owen-Anderson, A. F. H., Bradley, S. J., ... Zucker, K. J. (2006). Sex-dimorphic color preference in children with gender identity disorder: A comparison to clinical and community controls. *Sex Roles*, *55*, 385–395. <https://doi.org/10.1007/s11199-006-9089-9>.
- Cohen, P. N. (2013). Children's gender and parents' color preferences. *Archives of Sexual Behavior*, *42*, 393–397. <https://doi.org/10.1007/s10508-012-9951-5>.
- Cohen-Kettenis, P. T., & Pfäfflin, F. (2003). *Transgenderism and intersexuality in childhood and adolescence*. London: Sage.
- 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>.
- Davies, D. R. (1986). Children's performance as a function of sex-typed labels. *British Journal of Social Psychology*, *25*, 173–175. <https://doi.org/10.1111/j.2044-8309.1986.tb00717.x>.
- Davies, D. R. (1989). The effects of gender-typed labels on children's performance. *Current Psychology: Research & Reviews*, *8*, 267–272. <https://doi.org/10.1007/BF02686725>.
- Del Giudice, M. (2012). The 20th century reversal of pink-blue gender coding: A scientific urban legend? [letter to the Editor]. *Archives of Sexual Behavior*, *41*, 1321–1323. <https://doi.org/10.1007/s10508-012-0002-z>.
- Del Giudice, M. (2017). Pink, blue, and gender: An update. *Archives of Sexual Behavior*, *46*, 1555–1563. <https://doi.org/10.1007/s10508-017-1024-3>.
- Ellis, L., & Fieck, C. (2001). Color preferences according to gender and sexual orientation. *Personality and Individual Differences*, *31*, 1375–1379. [https://doi.org/10.1016/S0191-8869\(00\)00231-2](https://doi.org/10.1016/S0191-8869(00)00231-2).
- Erdfelder, E., Faul, F., & Buchner, A. (1996). GPower: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 1–11. <https://doi.org/10.3758/BF03203630>.
- Fagot, B. I., & Leinbach, M. D. (1989). The young child's gender schema: Environmental input, internal organization. *Child Development*, *60*, 663–672. <https://doi.org/10.2307/1130731>.
- Fulcher, M., & Hayes, A. R. (2017). Building a pink dinosaur: The effects of gendered construction toys on girls' and boys' play. *Sex Roles*. Advance online publication. <https://doi.org/10.1007/s11199-017-0806-3>.
- Gold, D., & Berger, C. (1978). Problem solving performance of young boys and girls as a function of task-appropriateness and sex-identity. *Sex Roles*, *4*, 183–195. <https://doi.org/10.1007/BF00287499>.
- Goldberg, S. (2017, January). Why we put a transgender girl on the cover of National Geographic. *National Geographic: Gender Revolution*. Retrieved from <http://www.nationalgeographic.com/magazine/2017/01/editors-note-gender/>.
- Green, V. A., Bigler, R., & Catherwood, D. (2004). The variability and flexibility of gender-typed toy play: A close look at children's behavioral responses to counterstereotypic models. *Sex Roles*, *78*, 371–386. <https://doi.org/10.1023/B:SERS.0000049227.05170.aa>.
- Halim, M. L., & Ruble, D. (2010). Gender identity and stereotyping in early and middle childhood. In J. Chrisler & D. McCreary (Eds.), *Handbook of gender research in psychology* (pp. 495–525). New York: Springer.
- Hargreaves, D. J., Bates, H. M., & Foot, J. M. C. (1985). Sex-typed labeling affects task performance. *British Journal of Social Psychology*, *24*, 153–155. <https://doi.org/10.1111/j.2044-8309.1985.tb00674.x>.
- Hidi, S. (2000). An interest researcher's perspective: The effects of extrinsic and intrinsic factors on motivation. In S. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 309–339). San Diego: Academic Press.
- Hines, M. (2010). Sex-related variation in human behavior and the brain. *Trends in Cognitive Sciences*, *14*, 448–456. <https://doi.org/10.1016/j.tics.2010.07.005>.
- Hurlbert, A. C., & Ling, Y. (2007). Biological components of sex differences in color preference. *Current Biology*, *17*, 623–625. <https://doi.org/10.1016/j.cub.2007.06.022>.
- Hyde, J. (2005). The gender similarities hypothesis. *American Psychologist*, *60*, 581–592. <https://doi.org/10.1037/0003-066X.60.6.581>.
- Ickes, W., Gesn, P. R., & Graham, T. (2000). Gender differences in empathic accuracy: Differential ability or differential motivation? *Personal Relationships*, *7*, 95–109. <https://doi.org/10.1177/0146167201276007>.
- Jadva, V., Hines, M., & Golombok, S. (2010). Infants' preferences for toys, colors and shapes: Sex differences and similarities. *Archives of Sexual Behavior*, *39*, 1261–1273. <https://doi.org/10.1007/s10508-010-9618-z>.
- Jirout, J. J., & Newcombe, N. S. (2015). Building blocks for developing social skills: Evidence from a large, representative U. S. Sample. *Psychological Science*, *26*, 302–310. <https://doi.org/10.1177/0956797614563338>.
- Kane, E. W. (2006). “No way my boys are going to be like that!”: Parents' responses to children's gender nonconformity. *Gender and Society*, *20*, 149–176. <https://doi.org/10.1177/0891243205284276>.
- Kohlberg, L. (1966). A cognitive-developmental analysis of children's sex-role concepts and attitudes. In E. E. Maccoby (Ed.), *The development of sex differences* (pp. 82–173). Stanford: Stanford University Press.
- Lee, J. F. K., & Collins, P. (2008). Gender voices in Hong Kong English textbooks—Some past and current practices. *Sex Roles*, *59*, 127–137. <https://doi.org/10.1007/s11199-008-9414-6>.
- Lee, J., Lee, J. O., & Collins, D. (2009). Enhancing children's spatial sense using tangrams. *Childhood Education*, *86*(2), 92–94. <https://doi.org/10.1080/00094056.2010.10523120>.
- Leinbach, M. D., Hort, B. E., & Fagot, B. I. (1997). Bears are for boys: Metaphorical associations in young children's gender stereotypes. *Cognitive Development*, *12*, 107–130. [https://doi.org/10.1016/S0885-2014\(97\)90032-0](https://doi.org/10.1016/S0885-2014(97)90032-0).
- Li, R. Y. H., & Wong, W. I. (2016). Gender-typed play and social abilities in boys and girls: Are they related? *Sex Roles*, *74*, 399–410. <https://doi.org/10.1007/s11199-016-0580-7>.
- Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, *56*, 1479–1498. <https://doi.org/10.2307/1130467>.
- 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>.
- Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs: Prentice Hall.
- Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford: Stanford University Press.

- Martin, C. L., & Halverson, C. F. (1981). A schematic processing model of sex typing and stereotyping in children. *Child Development, 52*, 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*, 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*, 903–933. <https://doi.org/10.1037/0033-2909.128.6.903>.
- Martin, C. L., DiDonato, M. D., Clary, L., Fabes, R. A., Kreiger, T., Palermo, F., ... Hanish, L. (2012). Preschool children with gender normative and gender non-normative peer preferences: Psychosocial and environmental correlates. *Archives of Sexual Behavior, 41*, 831–847. <https://doi.org/10.1007/s10508-012-9950-6>.
- Masters, J. C., Ford, M. E., Arend, R., Grotevant, H. D., & Clark, L. V. (1979). Modeling and labeling as integrated determinants of children's sex-typed imitative behavior. *Child Development, 50*, 364–371. <https://doi.org/10.2307/1129411>.
- Montemayor, R. (1977). Children's performance in a game and their attraction to it as a function of sex-typed labels. *Child Development, 48*, 152–156. <https://doi.org/10.2307/1127761>.
- Mulvey, K. L., Miller, B., & Rizzardi, V. (2017). Gender and engineering aptitude: Is the color of science, technology, engineering, and math materials related to children's performance? *Journal of Experimental Child Psychology, 160*, 119–126. <https://doi.org/10.1016/j.jecp.2017.03.006>.
- Paoletti, J. B. (1987). Clothing and gender in America: Children's fashions, 1890-1920. *Signs, 13*, 136–143. <https://doi.org/10.1086/494390>.
- Paoletti, J. B. (2012). *Pink and blue: Telling the boys from the girls in America*. Bloomington: Indiana University Press.
- Picariello, M. L., Greenberg, D., & Pillemer, D. B. (1990). Children's sex-related stereotyping of colors. *Child Development, 61*, 1453–1460. <https://doi.org/10.2307/1130755>.
- Pomerleau, A., Bolduc, D., Malcuit, G., & Cossette, L. (1990). Pink or blue: Environmental gender stereotypes in the first two years of life. *Sex Roles, 22*, 359–367. <https://doi.org/10.1007/BF00288339>.
- Ruble, D. N., & Martin, C. (1998). Gender development. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 933–1016). New York: Wiley.
- Ruble, D. N., Taylor, L. J., Cyphers, L., Greulich, F. K., Lurye, L. E., & Shrout, P. E. (2007). Effect of gender constancy on age-related changes in gender beliefs. *Child Development, 78*, 1121–1136. <https://doi.org/10.1111/j.1467-8624.2007.01056.x>.
- Saito, M. (1994). A cross-cultural study on color preference in three Asian cities: Comparison between Tokyo, Taipei and Tianjin. *Japanese Psychological Research, 36*(4), 219–232. <https://doi.org/10.4992/psycholres1954.36.219>.
- Saito, M. (1996). A comparative study of color preferences in Japan, China, and Indonesia, with emphasis on the preference for white. *Perceptual and Motor Skills, 83*, 115–128. <https://doi.org/10.2466/pms.1996.83.1.115>.
- Serbin, L. A., Powlishta, K. K., & Gulko, J. (1993). The development of sex typing in middle childhood. *Monographs of the Society of Research in Child Development, 58*(2), 1–74. <https://doi.org/10.2307/1166118>.
- Shih, M., Ambady, N., Richeson, J. A., Fujita, K., & Gray, H. M. (2002). Stereotype performance boosts: The impact of self-relevance and the manner of stereotype activation. *Journal of Personality and Social Psychology, 83*, 638–647. <https://doi.org/10.1037/0022-3514.83.3.638>.
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80–83. <https://doi.org/10.1111/1467-9280.00111>.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4–28. <https://doi.org/10.1006/jesp.1998.1373>.
- Sprafkin, C., Serbin, L. A., Denier, C., & Connor, J. M. (1983). Sex-differentiated play: Cognitive consequences and early interventions. In M. B. Liss (Ed.), *Social and cognitive skills: Sex roles and child's play* (pp. 167–192). New York: Academic Press.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797–811. <https://doi.org/10.1037/0022-3514.69.5.797>.
- Stein, A., Pohly, S., & Mueller, E. (1971). The influence of masculine, feminine and neutral tasks on children's achievement behavior, expectancies of success and attainments values. *Child Development, 42*, 195–207. <https://doi.org/10.2307/1127075>.
- Sweet, E. V. (2013, August). *Same as it ever was? Gender and children's toys over the 20th century*. Paper presented at the 108th annual American Sociological Association meeting in New York, NY.
- Taylor, C., Clifford, A., & Franklin, A. (2013). Color preferences are not universal. *Journal of Experimental Psychology: General, 142*(4), 1015–1027. <https://doi.org/10.1037/a0030273>.
- Thomas, R. M. (1999). *Human development theories: Windows on culture*. Thousand Oaks: Sage.
- Van Yperen, N. W. (2003). Task interest and actual performance: The moderating effects of assigned and adopted purpose goals. *Journal of Personality and Social Psychology, 85*(6), 1006–1015. <https://doi.org/10.1037/0022-3514.85.6.1006>.
- Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*, 250–270. <https://doi.org/10.1037/0033-2909.117.2.250>.
- Wachman, R. (2012, March). Lego's profits rise as it thinks pink. *The Guardian*. Retrieved from <https://www.theguardian.com/business/2012/mar/01/lego-profits-rise-pink>.
- 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>.
- Wong, W. I., & Hines, M. (2015a). Effects of gender color-coding on toddlers' gender-typical toy play. *Archives of Sexual Behavior, 44*(5), 1233–1242. <https://doi.org/10.1007/s10508-014-0400-5>.
- Wong, W. I., & Hines, M. (2015b). Preferences for pink and blue: The development of color preferences as a distinct gender-typed behavior in toddlers. *Archives of Sexual Behavior, 44*(5), 1243–1254. <https://doi.org/10.1007/s10508-015-0489-1>.
- 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*, 807–815. <https://doi.org/10.1007/s10508-008-9403-4>.
- Zosuls, K. M., Ruble, D. N., Tamis-LeMonda, C. S., Shrout, P. E., Bornstein, M. H., & Greulich, F. K. (2009). The acquisition of gender labels in infancy: Implications for gender-typed play. *Developmental Psychology, 45*, 668–701. <https://doi.org/10.1037/a0014053>.
- Zuckerman, C. (2017, January). Pink and blue: Coloring inside the lines of gender. *National Geographic*. Retrieved from http://www.nationalgeographic.com/magazine/2017/01/pink-blue-project-color-gender/?utm_source=Facebook&utm_medium=Social&utm_content=link_fb20170105ngm-gendercolor&utm_campaign=Content&sf49585588=1.