Arch Sex Behav (2015) 44:1233–1242
DOI 10.1007/s10508-014-0400-5

ORIGINAL PAPER

Effects of Gender Color-Coding on Toddlers' Gender-Typical Toy Play

Wang I. Wong · Melissa Hines

Received: 22 August 2013 / Revised: 18 August 2014 / Accepted: 18 August 2014 / Published online: 30 September 2014

Abstract Gender color-coding of children's toys may make certain toys more appealing or less appealing to a given gender. We observed toddlers playing with two gender-typical toys (a train, a doll), once in gender-typical colors and once in gender-atypical colors. Assessments occurred twice, at 20–40 months of age and at 26–47 months of age. A Sex × Time × Toy × Color ANOVA showed expected interactions between Sex and Toy and Sex and Color. Boys played more with the train than girls did and girls played more with the doll and with pink toys than boys did. The Sex × Toy × Color interaction was not significant, but, at both time points, boys and girls combined played more with the gender-atypical toy when its color was typical for their sex than when it was not. This effect appeared to be caused largely by boys' preference for, or avoidance of, the doll and by the use of pink. Also, at both time points, gender differences in toy preferences were larger in the gender-typical than in the gender-atypical color condition. At Time 2, these gender differences were present only in the gender-typical color condition. Overall, the results suggest that, once acquired, gender-typical color preferences begin to influence toy preferences, especially those for gender-atypical toys and particularly in boys. They thus could enlarge differences between boys' and girls' toy preferences. Because boys' and girls' toys elicit different activities, removing the gender color-coding of toys could encourage more equal learning opportunities.

W. I. Wong · M. Hines
Department of Psychology, University of Cambridge, Cambridge, UK

W. I. Wong (✉)
Department of Psychology, University of Hong Kong, Pokfulam, Hong Kong
e-mail: iwwong@hku.hk

Keywords Gender stereotyping · Sex-typing · Gender differences · Toy preferences · Color preferences

Introduction

Preferences for gender-typical toys, such as toy vehicles and dolls, show large gender differences (Hines, 2010) which appear by age 2 years (Alexander, Wilcox, & Woods, 2009; Servin, Bohlin, & Berlin, 1999; Snow, Jacklin, & Maccoby, 1983; Zosuls et al., 2009) and increase across childhood (Golombok et al., 2012). These gender differences in children's toy preferences are among the largest of all behavioral gender differences (Hines, 2010; Hyde, 2005).

Preferences for the colors pink and blue also show gender differences, with girls generally liking pink more than boys do and boys generally liking blue more than girls do (Chiu et al., 2006; Picariello, Greenberg, & Pillemer, 1990). Gender differences in color preferences emerge later in life than gender differences in toy preferences (Jadva, Golombok, & Hines, 2010). The gender difference in preference for pink begins to appear at about 2 years of age and, like the gender differences in toy preferences, grows larger across early childhood (LoBue & DeLoache, 2011). The colors pink and blue also co-vary with the gender-typicality of children's toys; boys' toys are often colored blue and girls' toys are often colored pink (Cunningham & MaCrae, 2011).

The value of using colors to distinguish boys' and girls' toys has been debated by researchers (e.g., Paoletti, 2012; Ruble, Lurye, & Zosuls, 2010) and in society more generally (Jenkins, 2011; Orenstein, 2011; Paul, 2011). A common concern is that associating colors with gender may create arbitrary gender stereotypes (e.g., that blue is for boys and pink is for girls) that intensify gender differences. Another concern is that gender-coding by color may have developmental

implications because children spend much of their time learning through toy play and different toys offer different learning opportunities.

For instance, play with boy-typical toys elicits spatial activities and may therefore enhance spatial skills. In contrast, play with girl-typical toys elicits social play and may therefore enhance verbal and social skills. Large and consistent gender differences between boys' and girls' toy preferences have thus been hypothesized to contribute to gender differences in later spatial cognitive and social development outcomes (Block, 1983; Caldera, Huston, & O'Brien, 1989; Sprafkin, Serbin, Denier, & Connor, 1983). Because playing with boy-typical toys is thought to enhance spatial skills important for success in science and mathematics, some researchers and parents advocate encouraging girls to play with boy-typical toys in attempts to narrow gender gaps (e.g., Eliot, 2009; Fine, 2010; Serbin, Connor, & Iler, 1979). Although not discussed as extensively, encouraging boys to play with girl-typical toys could enhance their social and verbal skills. Thus, the gender color-coding of toys has been thought to magnify gender differences in cognitive and social developmental outcomes (e.g., Orenstein, 2011; Paul, 2011). As yet, however, there is no empirical evidence that gender color-coding of gender-typical toys alters boys' or girls' interest in them or enlarges the differences between boys' and girls' toy interests.

One prior study attempted to examine this question by studying 12- to 24-month-olds (Jadva et al., 2010). Jadva et al. assessed looking time at pictures of a car versus a doll, in gender-typical colors and in gender-atypical colors. Boys were found to look at the car more than girls did and girls were found to look at the doll more than boys did, regardless of whether the toy was pink or blue. This finding could be interpreted to show that the use of gender color-coding does not affect children's gender-typical toy preferences. However, the infants in Jadva et al. did not show gender differences in preferences for pink or blue, perhaps because they were too young. This interpretation is consistent with later evidence suggesting no significant gender differences in color preferences for children younger than 2 years of age (LoBue & DeLoache, 2011). The children in the study by Jadva et al. might therefore have been too young to show any effect of gender-typical colors.

The current study examined whether gender-typical colors influence gender-typical toy preferences in toddlers, at an age when gender differences in both toy preferences and color preferences are present. Preferences were assessed by observing play with real toys instead of looking time at pictures. Participants were studied on two occasions, separated by 6–8 months, to see if the magnitude of any effect of gender-typical colors on gender-typical play increased with age. The main hypotheses under investigation were (1) preferences for toys would be increased by the use of a color consistent with the child's sex and reduced by the use of a color inconsistent with the child's sex and (2) the use of gendered colors would increase the size of differences in toy preferences between boys and girls.

Method

Participants

A total of 56 boys and 70 girls, aged 20–40 months, from a university town in the United Kingdom took part at time point 1 (T1). Mean ages in months were: boys, 28.52 ($SD = 5.79$); girls, 29.22 ($SD = 5.51$). Of these children, 40 boys and 59 girls were followed up after 6–8 months (T2), when they were aged 26–47 months. Mean ages in months were: boys, 35.40 ($SD = 5.42$); girls, 36.10 ($SD = 5.78$). Each child took part with one parent (M age in years = 35.89, $SD = 5.75$). Approximately 80 % of the parents had completed tertiary education.[1] The boys did not differ from the girls in age, birth order, parental education or parental age.

Procedure

Toy Preferences

Stimuli included a toy vehicle and a doll because vehicles and dolls are strongly gender-typed toys (Zosuls et al., 2009). Specifically, a train served as the boy-typical toy and a soft doll served as the girl-typical toy. The toys were provided in two color conditions: gender-typical colors (i.e., pink doll, blue train) and gender-atypical colors (i.e., blue doll, pink train). In both conditions, children played with the toys on the floor for 4 min. The order of the two color conditions was counterbalanced and the left–right position of the toys was randomized for each presentation. Participants were instructed to play as they normally would. For T1 and T2, 7 tests and 14 tests were conducted at the participants' homes, respectively. All other tests were conducted in the laboratory, where the researcher watched recordings from two video recorders fixed on the playroom ceiling from an adjacent room that was also equipped with a one-way mirror. When testing was conducted at home, the researcher stayed in the same room with the participants and videotaped the play using a portable video recorder, but participants were told not to interact with the researcher when playing with the toys. If the child approached the researcher, the researcher told him/her to return to play. Play with the toys was similar when observed at home and in the laboratory and was similar whether data from home visits

[1] The education levels were assessed using UNESCO classification, which defines tertiary education as education beyond post-secondary non-tertiary education. Tertiary education includes 1st stage (not leading directly to an advanced research qualification) and 2nd stage (leading to an advanced research qualification).

Table 1 Parental responses and examples for observed toy play

Code	Example
Positive	
Comment favorably	e.g., "It's a lovely doll"
Play along	e.g., Push train back and forth to child
Show physical affection	e.g., Laugh
Initiate play	Offer toy e.g., "Would you like to play with the doll?"
Give approval	e.g., "Okay, you can play with that"
Facilitate play	Suggest how to play or ask questions and give comments that sustain play e.g., "Let's push the train to the wall"
Negative	
Comment negatively	e.g., "That's not fun"
Interfere play	e.g., "Don't do that"
Suggest alternative play	Offer alternative toy when child is engaged with another toy e.g., "Let's play with the train instead"
Ignore play	Ignore child's request to engage in play or get attention
Refuse to play	Explicit refusal to play e.g., "I am not going to play with that"
Neutral	
Watch attentively	Pay close attention to child's play and not giving any other responses
Does not interfere play	Not watching but respond to child when requested
Give instruction	e.g., "Bring the train back onto the mat"
Neutral comments	e.g., "It has four wheels"
Help	Help at child's request

were included, so data collected in the laboratory and from home visits were combined.[2]

As in previous observational studies on gender-typical toy play (e.g., Berenbaum & Hines, 1992; Pasterski et al., 2005), play during the first 3 min of each condition was coded as "play with the train," "play with the doll," or "no play." The remaining minute was coded if parts of the first 3 min were unscorable (e.g., the child went out of sight). Play was coded if the child was engaged meaningfully or was in deliberate contact with a toy. Touching a toy to put it away was not counted as play nor was accidental physical contact. Physical contact did not always have to be constant for play to be coded, because some forms of play involved intermittent physical contact. When a child played with two toys simultaneously, play with each toy was coded separately. For coding purposes, each condition was divided into 36, 5-s intervals. To adjust for individual differences in total play time, the score for each toy was the number of intervals playing with that toy divided by the total number of intervals playing with either toy. A second rater coded 40 play sessions. Inter-rater reliabilities were: blue train ($r = .95$), pink doll ($r = .96$), blue doll ($r = .95$), and pink train ($r = .72$).

[2] More details about the analyses are available from the corresponding author upon request.

Table 2 Sex × Time × Toy × Color analysis of variance

Source	*Df*	*F*	*P*
Sex	1	4.12	.044
Time	1	12.84	<.001
Toy	1	332.21	<.001
Color	1	1.31	ns
Sex × Time	1	.95	ns
Sex × Toy	1	26.27	<.001
Sex × Color	1	4.08	.023
Time × Toy	1	9.76	.002
Toy × Color	1	6.72	.011
Time × Color	1	.02	ns
Sex × Time × Toy	1	3.95	.049
Sex × Time × Color	1	1.16	ns
Sex × Toy × Color	1	.09	ns
Time × Toy × Color	1	.06	ns
Sex × Time × Toy × Color	1	.08	ns
Error	124		

Parental Responses

Children were observed playing with a parent because it is difficult to separate young children from their parents. To examine the possibility that parental responses confounded children's play, parental positive, negative, and neutral responses to play with each toy were recorded using codes adapted from previous studies (e.g., Fagot, 1983; Pasterski et al., 2005) (see Table 1). To adjust for individual differences in the total amounts of play and responses to each toy, each response was converted to a proportion (i.e., number of a type of response to a certain toy/total number of responses to that toy). A second rater coded 40 play sessions. Inter-rater reliabilities for the blue train were: positive ($r = .86$), negative ($r = .79$), and neutral ($r = .92$). Inter-rater reliabilities for the pink doll were: positive ($r = .71$), negative ($r = .98$), and neutral ($r = .84$). Inter-rater reliabilities for the blue doll were: positive ($r = .71$), negative ($r = .85$), and neutral ($r = .70$). Inter-rater reliabilities for the pink train were: positive ($r = .73$), negative ($r = .84$), and neutral ($r = .72$).

Data Analysis

We first used a Sex (boy/girl) × Time (T1/T2) × Toy (train/doll) × Color (pink/blue) repeated measures ANOVA to evaluate for

Table 3 Estimated statistics for pair-wise contrasts of significant main effects and interactions in the ANOVA

Sex					
Sex (I)	Sex (J)		Mean difference (I–J)	SE	*p*
Male	Female		−.024*	.012	.044
Toy					
Toy (I)	Toy (J)		Mean difference (I–J)	SE	*p*
Train	Doll		.467*	.026	<.001
Time					
Time (I)	Time (J)		Mean difference (I–J)	SE	*p*
1	2		−.046*	.013	<.001
Sex × Toy					
Toy	Sex (I)	Sex (J)	Mean difference (I–J)	SE	*p*
Train	Male	Female	.107*	.024	<.001
Doll	Male	Female	−.156*	.032	<.001
Sex × Toy					
Sex	Toy (I)	Toy (J)	Mean difference (I–J)	SE	*p*
Male	Train	Doll	.599*	.038	<.001
Female	Train	Doll	.336*	.034	<.001
Sex × Color					
Color	Sex (I)	Sex (J)	Mean difference (I–J)	SE	*p*
Blue	Male	Female	.005	.020	ns
Pink	Male	Female	−.053*	.017	.002
Sex × Color					
Sex	Color (I)	Color (J)	Mean difference (I–J)	SE	*p*
Male	Blue	Pink	.045*	.021	.036
Female	Blue	Pink	−.012	.019	ns
Time × Toy					
Toy	Time (I)	Time (J)	Mean difference (I–J)	SE	*p*
Train	1	2	.009	.015	ns
Doll	1	2	−.100*	.027	<.001
Toy × Color					
Toy	Color (I)	Color (J)	Mean difference (I–J)	SE	*p*
Train	Blue	Pink	−.004	.015	ns
Doll	Blue	Pink	.037*	.018	.038
Sex × Time × Toy					
Sex	Toy	Time (I)	Time (J)	Mean difference (I–J)	SE	*p*
Male	Train	1	2	.031*	.015	.039
	Doll	1	2	−.147*	.038	<.001
Female	Train	1	2	−.013	.023	ns
	Doll	1	2	−.053	.037	ns

* The mean difference is significant at the .05 level. An interaction can be viewed from different perspectives and the choice of perspective can be based on theoretical relevance (Vander Stoep & Johnson, 2009). Therefore, for Sex × Toy and Sex × Color interactions, all possible simple main effects are presented because of their theoretical relevance. For the rest, simple main effects are presented if an effect is significant at some levels and not significant at others, or when an effect is in opposite directions at different levels

gender differences in toy and color preferences and to evaluate the main hypothesis that there would be Sex × Toy × Color interactions. The ANOVA was supplemented with planned paired-sample *t* tests to examine how preferences for each type of toy differed between color conditions, first within each sex and then in boys and girls combined. We also tested whether gender differences were similar across color conditions using planned independent sample *t* tests. These additional analyses allowed us to test our main hypothesis, that gender-typed colors would influence toy choices more directly. In total, 27 children did not take part at T2. At T1, three boys and two girls did not complete both conditions; at T2, these numbers were three boys and one girl. Values were missing completely at random and unrelated to the child's age, sex or test location (Little's MCAR: $\chi^2 = 58.04, df = 57$). Missing values were imputed with maximum likelihood using Expectation Maximization (2,500 iterations, convergence criterion = .0001).

Results

Sex × Time × Toy × Color ANOVA

Table 2 shows the *F*-statistics for all main effects, two-way interactions, three-way interactions, and the four-way interaction. Table 3 shows estimated statistics for the pair-wise contrasts of the significant main effects and simple main effects that followed up on the interactions. There were significant main effects of Sex (girls > boys), $p = .044$, Time (T2 > T1), $p < .001$, and Toy (train > doll), $p < .001$, and significant two-way interactions between Sex and Toy, $p < .001$, Sex and Color, $p = .023$, Time and Toy, $p = .002$, and Toy and Color, $p = .011$. There also was a significant Sex × Time × Toy interaction, $p = .049$.

The simple main effects for the two-way interactions showed that boys preferred the train more than girls did and

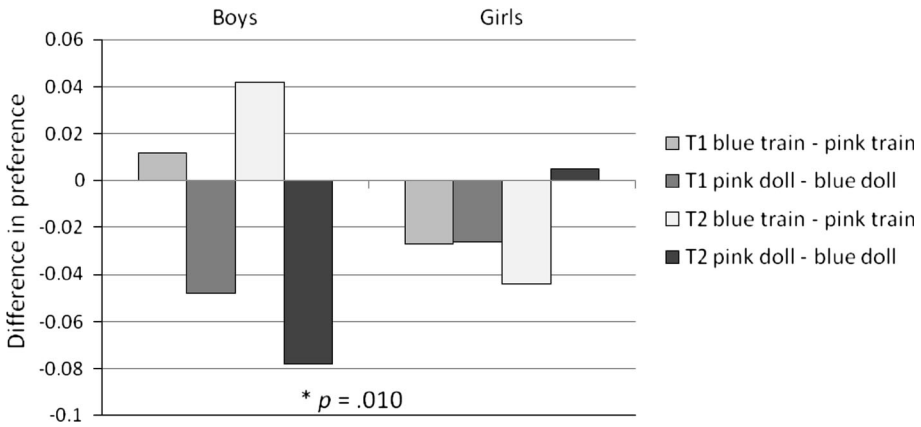

Fig. 1 Differences in preferences between color conditions (preference for one toy minus preference for another) within each sex

girls preferred the doll more than boys did. Although both boys and girls preferred the train over the doll, this difference was larger in the boys. Girls also preferred pink toys more than boys did although boys and girls did not differ in their preference for blue toys. In addition, boys preferred blue toys to pink toys, but girls preferred pink toys and blue toys equally. Preference for the train remained the same over time, but preference for the doll increased from T1 to T2. For the train, color had no effect; for the doll, it was more preferred when it was blue than when it was pink.

Finally, simple main effects analysis suggested that the three-way (Sex × Time × Toy) interaction occurred because boys' preference for the train decreased and their preference for the doll increased over time while girls' toy preferences did not change over time; the two-way (Sex × Toy) interaction was significant at both time points, reflecting the typical gender differences in preferences for these toys, with boys showing greater interest in the train than girls and girls showing greater interest in the doll than boys, all $ps < .05$. The hypothesized three-way (Sex × Toy × Color) interaction was not significant, nor was the four-way (Sex × Time × Toy × Color) interaction.

Differences in Preferences for Each Type of Toy between Conditions

Within Sex

We supplemented the ANOVA with planned paired-sample *t* tests to test our main hypothesis that gender color-coding would affect toy preferences. At T2, boys preferred the doll significantly less when it was pink than when it was blue, $t(55) = 2.40$, $p = .010$. Other comparisons were generally in the predicted direction, but did not differ significantly between the two conditions. An exception to this general pattern was girls' T1 preference for the doll, where the mean for the blue doll was nonsignificantly higher than the mean for the pink doll. Figure 1 shows the differences in preferences between color conditions within each sex.

Boys and Girls Combined

The pattern of means in the within-sex analysis suggested that more effects of color might be statistically significant if the sample size were larger. Therefore, we combined boys and girls to examine the effect of gender-typed colors on play with gender-typical and gender-atypical toys.[3] For this analysis, we created four preference variables based on the gender typicality of the toy and its color in relation to the child's sex: typical toy and typical color (T–T, i.e., boys' preference for the blue train and girls' preference for the pink doll), typical toy and atypical color (T–A, i.e., boys' preference for the pink train and girls' preference for the blue doll), atypical toy and typical color (A–T, i.e., boys' preference for the blue doll and girls' preference for the pink train), and atypical toy and atypical color (A–A, i.e., boys' preference for the pink doll and girls' preference for the blue train). The hypothesis was that for both the gender-typical and the gender-atypical toy, a color typical for the child's sex would increase preference.

Figure 2 shows the differences in preferences between color conditions in boys and girls combined. Preference for the gender-atypical toy was greater when it had a color typical for the child's sex than when it did not. This was true for T1, $t(125) = 1.84$, $p = .034$, $d = .16$, as well as T2, $t(125) = 2.75$, $p = .004$, $d = .25$.[4] For gender-typical toys, however, changing the color did not have a significant effect on children's preferences.

Gender Differences in Different Conditions

We also explored whether gender color-coding would increase gender differences in toy preferences using planned independent sample *t* tests. Table 4 shows the descriptive statistics for toy preferences in the two color conditions at each time point and the effect sizes for the gender differences. At T1, there were significant gender differences in preferences for the blue train,

[3] This analysis came at the suggestion of a reviewer.

[4] Cohen's *d* was calculated with the formula: $t/\sqrt{N}$.

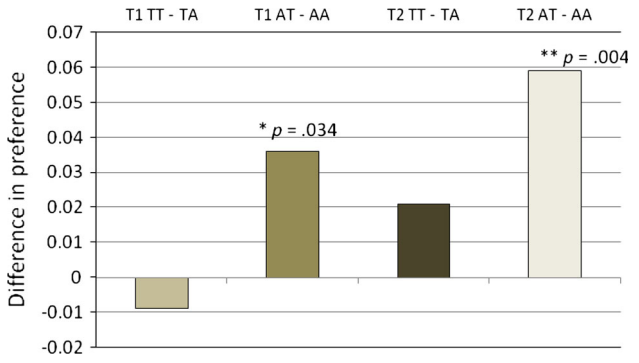

Fig. 2 Differences in preferences between color conditions (preference for one type of toy minus preference for another) in boys and girls combined. *TT* gender-typical toy with a color typical for the child's sex, *TA* gender-typical toy with a color atypical for the child's sex, *AT* gender-atypical toy with a color typical for the child's sex, *AA* gender-atypical toy with a color atypical for the child's sex

Table 4 Descriptive statistics for boys' and girls' toy preferences at each time point

	Boys ($n = 56$)		Girls ($n = 70$)		
	M	*SD*	*M*	*SD*	*d*
T1					
Gender-typical color condition					
Blue train	.91	.11	.77	.21	.81***
Pink doll	.20	.21	.41	.27	−.86***
Gender-atypical color condition					
Pink train	.90	.17	.79	.21	.57***
Blue doll	.24	.26	.44	.28	−.74***
T2					
Gender-typical color condition					
Blue train	.90	.13	.77	.23	.68***
Pink doll	.33	.22	.48	.31	−.55***
Gender-atypical color condition					
Pink train	.86	.18	.81	.20	.26
Blue doll	.41	.29	.47	.28	−.21

*** $p < .001$. Play can total more than 1.0, because children sometimes used two toys at the same time. Cohen's *d* was calculated with the formula: (Mean 1 – Mean 2)/pooled SD

Table 5 Parental responses

	Boys ($n = 56$)		Girls ($n = 70$)	
	M	*SD*	*M*	*SD*
T1				
Blue train				
Positive	.50	.20	.50	.20
Negative	.04	.06	.06	.08
Neutral	.46	.20	.44	.19
Pink doll				
Positive	.60	.24	.65	.22
Negative	.03	.07	.02	.04
Neutral	.37	.23	.34	.21
Pink train				
Positive	.46	.20	.50	.20
Negative	.06	.08	.06	.08
Neutral	.48	.19	.44	.19
Blue doll				
Positive	.65	.23	.64	.24
Negative	.02	.06	.02	.05
Neutral	.33	.22	.34	.23
T2				
Blue train				
Positive	.46	.22	.51	.20
Negative	.02	.03	.03	.04
Neutral	.52	.22	.47	.21
Pink doll				
Positive	.58	.25	.61	.22
Negative	.03	.09	.02	.05
Neutral	.40	.24	.38	.21
Pink train				
Positive	.46	.21	.50	.21
Negative	.03	.04	.02	.03
Neutral	.52	.21	.47	.21
Blue doll				
Positive	.59	.30	.65	.29
Negative	.04	.10	.02	.05
Neutral	.38	.30	.33	.28

$t(108) = 5.13$, $p < .001$, the pink doll, $t(124) = 5.03$, $p < .001$, the pink train, $t(124) = 3.91$, $p < .001$, unequal variances, and the blue doll, $t(124) = 3.97$, $p < .001$. At T2, there were significant gender differences only in preferences for the blue train, $t(110) = 3.91$, $p < .001$, and the pink doll, $t(122) = 3.15$, $p = .001$, unequal variances. In terms of effect sizes, by Cohen's (1988) standard, the expected gender differences in preferences for the train and the doll were large at T1 ($d = .81$ and $-.86$ in order) when the toys were in gender-typical colors, but only moderate to large, ($d = .57$, and $-.74$ in the same order) when they were in gender-atypical colors. Similarly, at T2, effect sizes for the gender differences for the same toys in the same order were moderate to large when the toys were in gender-typical colors ($d = .68$ and $-.55$), but small when they were in gender-atypical colors ($d = .26$ and $-.21$).

Parental Responses

Analyses of parental responses (see Table 5) suggested that parents did not confound their child's play. First, none of the parental responses differed significantly depending on the child's sex, all $ps > .05$, suggesting that parents did not encourage girls to play with dolls more than they did boys nor did they encourage boys to play with trains more than they did

girls. Parental responses also did not correlate with child behavior in a way that would suggest that parental encouragement and discouragement shaped gender-typed play—out of 32 correlations (between boys' and girls' preference for each of the four toys and positive and negative parental responses at T1 and T2), there was only one negative correlation between negative parental responses and child preference (for the blue train in girls at T1) and only one positive correlation between positive parental responses and child preference (for the pink doll in girls at T2). This lack of parental influence on children's play may be due to the short duration of the play and because the parents gave largely positive or neutral responses and very few negative responses, as in other studies (e.g., Fagot, 1978).

Discussion

This study provided some evidence that gender color-coding influences young children's gender-typical toy play. Although a four-way ANOVA did not show a significant Sex × Toy × Color interaction as had been predicted (or a Sex × Time × Toy × Color interaction), specific findings from planned comparisons suggested an effect of color. First, at both of two time points, when data for boys and girls were combined, gender-atypical toys in colors typical for the child's sex were preferred to gender-atypical toys in colors atypical for the child's sex. Second, at T2, boys' and girls' toy preferences differed significantly only when toys were in gender-typical colors and not when they were in gender-atypical colors. Third, effect sizes for gender differences in toy preferences were larger when the toys were in gender-typical colors than when they were in gender-atypical colors. Our discussion of these findings will focus first on how the findings for gender differences in toy and color preferences resemble those seen in previous studies. We will then expand the discussion of our findings regarding the effects of gender-typical colors on children's preferences for different types of toys. Finally, we will discuss the theoretical and practical implications of our findings, as well as their limitations.

Gender-Typical Toy and Color Preferences

We found gender differences in toy preferences similar to those reported in prior studies. The ANOVA revealed the expected significant Sex × Toy interaction reflecting boys' greater preference than girls for the train and girls' greater preference than boys for the doll. In addition to looking at these between-sex preferences, researchers have sometimes looked at within-sex preferences for gender-typical toys and we did so too. Here, we found that both boys and girls preferred the train to the doll. Other researchers have also often found inconsistent within-sex preferences, as opposed to between-sex preferences, for gender-typical toys, probably because toys' attractiveness, independent of their gendered characteristics, can influence within-sex preferences (Hines & Alexander, 2008). For instance, both Jadva et al. (2010) and Serbin et al. (2001) found that 12-month-old boys preferred female to male toys. For slightly older infants, Serbin et al. (2001) found that 23-month-old girls preferred male to female toys while Jadva et al. (2010) found that 24-month-old girls showed equal preferences for a male toy and a female toy. Similarly, for older children, Berenbaum and Hines (1992) found that 3- to 8-year-old girls liked male and female toys equally and Idle, Wood, and Desmarais (1993) found that 3- to 5-year-old girls preferred male to female toys.

Boys and girls also differed in their preferences for toys of different colors, as indicated by a Sex × Color interaction. This interaction occurred because girls showed a greater preference than boys for the pink toys. There was no significant difference between boys and girls in their preference for the blue toys. Older children have typically been found to show gender differences in preferences for both pink and blue (Chiu et al., 2006; Picariello et al., 1990). Our finding was consistent with previous research suggesting that pink is more gender-typed than blue (Cunningham & MaCrae, 2011; Leinbach, Hort, & Fagot, 1997). Because the preference for the blue toys showed no significant gender difference, the effects of color on toy preferences that we observed were likely the result of the toys being pink or not pink.

Effect of Color on Toy Preferences

We found that, for boys and girls combined, color reversal affected preference for the gender-atypical toy at both time points. This effect appeared to largely reflect the effect of color on boys' preference for/avoidance of the doll; at T2, boys avoided the doll significantly more if it was pink than if it was blue. Color reversal also reduced the between-gender differences in toy preferences at both time points and the gender differences were nonsignificant in the atypical color condition at T2. These findings contrast with the non-significant Sex × Toy × Color and Sex × Time × Toy × Color interactions in the ANOVA. However, it is generally more difficult to detect higher-level than lower-level interaction effects in an ANOVA. The planned paired *t* tests that tested change in preference for each toy and the independent *t* tests that tested the gender differences in individual color conditions examined the color effect more directly, so we will discuss these findings.

First, color affected preference for a gender-atypical toy but not preference for a gender-typical toy. Perhaps preferences for gender-typical toys are more due to inborn preferences for certain affordances such as opportunity for movement and not so much to peripheral cues such as gender labeling by color. The finding that non-human primates who have never been exposed to human toys or gender color-coding also show gender differences

in preferences for gender-typical toys but no consistent avoidance of gender-atypical toys (Alexander & Hines, 2002; Hassett, Siebert, & Wallen, 2008) is compatible with this explanation. Perhaps young children have no strong avoidance of gender-atypical toys per se and their avoidance of gender-atypical toys is influenced by cues such as color. Future studies can test this possibility.

Boys were more affected by color than were girls. Perhaps this is because boys receive stronger social pressure than girls in general in regard to gender-typical behavior (Feinman, 1981; Lytton & Romney, 1991; Maccoby & Jacklin, 1974). Also, girl-typical toys are more likely to be blue than boy-typical toys are to be pink (Cunningham & MaCrae, 2011). These factors may have led to our findings that boys showed more gender-typical preferences than girls for both toys and colors. For example, boys preferred the blue toys over the pink toys while girls preferred these toys similarly. A color switch between pink and blue might thus have a stronger effect on boys' than on girls' toy preferences.

We also tested whether color reversal reduced the between-sex differences in toy preferences. The gender differences in the gender-atypical color condition were smaller than those in the gender-typical color condition (moderate to large vs. large at T1, and small vs. moderate to large at T2). At T2, the gender differences in preferences for both toys became non-significant when the colors were gender atypical.

These effects of color also help explain some unexpected ANOVA interactions. First, a Sex × Time × Toy interaction showed that boys' toy preferences became less gender-typed, while girls' toy preferences were stable, over time. This finding contrasts with the expectation that children's gender-typical toy preferences would grow over time, although, as mentioned, past findings on age-related changes in toy preferences were not always consistent. In this study, boys' toy preferences could become less gender-typical if they were affected by color and the effect of color increased with age. In line with this explanation, the effect of color on the gender-atypical toy was larger at T2 than at T1 and more evident in boys. Second, a Toy × Color interaction showed that children preferred the blue doll to the pink doll. This finding could be driven by the finding that boys' preference for the doll increased when it was blue.

Theoretical Implications

Two mechanisms that could contribute to the effects of color on children's toy preferences are gender labeling and affective associations between toys and colors. In regard to gender labeling, children show greater interest in objects that have been labeled verbally for children of their sex than in objects that have been labeled verbally for the other sex (Masters et al., 1979; Ruble et al., 2007). The frequent coloring of toys in gender-typical colors could allow the colors to function as visual gender labels that signal to children which toy is appropriate for them. Children begin to understand gender labels at around 2 years of age (Ruble et al., 2007), consistent with our finding that gender color-coding affected toy preferences of children aged 20–47 months. Jadva et al. (2010) did not find an effect of color perhaps because children in Jadva et al. were younger than 2 years of age and so may not yet respond to gender labels. In regard to affective associations, color preferences in adults can be explained largely by their liking of entities associated with the colors (Palmer & Schloss, 2010). For instance, students, especially those who like their university, prefer colors associated with their university to those associated with other universities (Schloss, Poggesi, & Palmer, 2011). If similar processes occur in children, boys and girls may also acquire preferences for gender-typical colors by playing with toys in these colors.

An effect of color on gender-typical toy play does not mean that gender-typical toy play requires gender color-coding. As mentioned, gender-typical toy preferences appear to be initiated by preferences for certain toy affordances (Alexander & Hines, 2002; Hassett et al., 2008). Moreover, gender differences in toy preferences emerge before those in color preferences (Jadva et al., 2010). Therefore, in contrast to claims that gender-typical color preferences are inborn (Hurlbert & Ling, 2007) and underlie gender-typical toy preferences (Alexander, 2003), they probably arise in part from playing with gender-typical toys. As children grow older, preferences for gender-typical colors may also further increase preferences for gender-typical toys.

Limitations

Greater power may be needed to better test the effect of gender-typical colors on children's gender-typical toy preferences. For instance, the results of a four-way ANOVA did not support the main hypothesis of interest by showing a significant Sex × Toy × Color interaction. Additional analyses did, however, provide some support for an effect of color on children's toy preferences, perhaps because more power is required to detect high level ANOVA interactions as the ANOVA interaction controls for all main effects, lower-level interactions, and the grand mean (Rosnow & Rosenthal, 1988). Given our interest in planned contrasts of means as opposed to unplanned contrasts and residuals, the planned supplementary analysis may be more appropriate for the evaluation of the color effect (Rosnow & Rosenthal, 1988, 1989).

Also, the observed effect of gender-typical colors was not large and it was smaller than the effect of sex on color and toy preferences. For instance, effect sizes for the effect of color on gender-atypical toy preferences were $d = .16$ and .25 at T1 and T2, respectively, suggesting that color has a small influence on toy preferences at the ages studied. The boys and girls in this study may have been too young to show stronger

effects of color because they were only beginning to differ in their color preferences and to prefer gender-typical colors over other colors (Jadva et al., 2010; LoBue & DeLoache, 2011). Color preferences become more gender-typical across early childhood until at least age 4 or 5 years (LoBue & DeLoache, 2011). At the same time, children's cognitive understanding of gender continues to develop. By about age 2 years, most children have gender identity, meaning that they know that they are girls or boys. By about age 4–5 years, they have gender stability, meaning that they know this will remain the same over time. Finally, by about age 7 years, they have gender consistency, meaning that they understand that gender remains stable across situations (Ruble et al., 2007). In addition, findings on children's preferences for gender-typical clothing in general (which includes gender-typical colored clothing) suggest that gender-typical preferences become more rigid as children come to understand their gender identity and its stability, although they then relax following the acquisition of understanding of gender consistency (Halim et al., 2014). Most of our children were between 2 and 3 years of age and thus would not understand that their gender was stable (Ruble et al., 2007). Effects of color might be stronger when children enter the gender stability phase and their gender-typical color preferences strengthen. Therefore, it would be of interest to test older samples of children and to examine how their understanding of gender identity, stability, and consistency relates to the impact of gender-typical colors on their toy preferences.

We used two specific toys whose colors were manipulated experimentally and observed children's toy play with a parent present. Future studies could test the generalizability of our findings by using different toys or procedures. In addition, hypothesis testing can only reject or fail to reject the null hypothesis; it cannot accept the null hypothesis. Thus, the non-significant gender differences in toy preferences in the T2 atypical color condition should not be over-interpreted as an indication that color reversal can eliminate gender differences in toy preferences. For example, larger samples may find reduced yet significant gender differences in preferences for gender-typed toys in gender-typical colors.

Bonferroni adjustments of probabilities are sometimes used in studies involving multiple comparisons to avoid increasing Type I error rates. We did not use an adjustment of this type for two primary reasons. First, such adjustments are not suitable for planned comparisons (Rice, 1989) and most of the comparisons we made were planned. Second, the use of such adjustments to reduce Type I error can produce other problems (Cabin & Mitchell, 2000; Kusuoka & Hoffman, 2002; Miller, 1981; Perneger, 1998; Rice, 1989). For instance, the decision on whether to apply the adjustments and the number of tests to be adjusted are highly arbitrary, the research question addressed by these adjustments (that all null hypotheses are true simultaneously) is often not of interest to researchers, a given comparison can be interpreted differently depending on how many other tests are performed, and Type II error rates increase as a result of reducing Type I error rates and both types of errors are important.

Although we did not use Bonferroni adjustments, if used, the finding that color affected preference for the gender-atypical toy would remain significant at T2 even when considering all four comparisons at T1 and T2, the finding that boys' preference for the doll was affected by color at T2 would remain even when considering all four comparisons in boys, and the conclusion that color reversal eliminated the gender differences in toy preferences at T2 would remain even when considering all eight comparisons at T1 and T2. Therefore, our results are unlikely to be the product of Type I error.

Conclusion

This study provided initial evidence that gender color-coding influences the toy preferences of young children. Although young children's preferences for gender-typical toys were unaffected by color, they were more likely to play with gender-atypical toys that had a color typical for their sex than ones that did not. The effects we saw were small, but we tested children at an age when gender-typical color preferences are still growing (Halim et al., 2014; LoBue & DeLoache, 2011); effects in older children might well be larger. Because play with different toys provides different learning opportunities, gender color-coding may magnify gender differences in cognitive and social abilities and removing gender color-coding could encourage more gender-similar play patterns and abilities. Our findings encourage further studies of the impact of using gender-typical colors on children's toy preferences. In particular, it would be useful to study the impact of gender color-coding on older children who have stronger gender-typical color preferences and to examine relationships of this impact to the acquisition of progressive stages of gender identity understanding.

Acknowledgments This research was supported by the Gates Cambridge Trust. Data were presented at the 2012 annual meeting of the International Academy of Sex Research in Lisbon, Portugal. We would like to thank the parents, children, children's centers and nurseries who contributed data to the research, and Mihaela Constantinescu who helped with data coding.

References

Alexander, G. M. (2003). An evolutionary perspective of sex-typed toy preferences: Pink, blue, and the brain. *Archives of Sexual Behavior, 32*, 7–14.

Alexander, G. M., & Hines, M. (2002). Sex differences in response to children's toys in nonhuman primates (*Cercopithecus aethiops sabaeus*). *Evolution and Human Behavior, 23*, 467–479.

Alexander, G. M., Wilcox, T., & Woods, R. (2009). Sex differences in infants' visual interest in toys. *Archives of Sexual Behavior, 38*, 427–433.

Berenbaum, S. A., & Hines, M. (1992). Early androgens are related to childhood sex- stereotyped toy preferences. *Psychological Science, 3*, 203–206.

Block, J. H. (1983). Differential premises arising from differential socialization of the sexes. *Child Development, 54*, 1335–1354.
Cabin, R. J., & Mitchell, R. J. (2000). To Bonferroni or not to Bonferroni: When and how are the questions. *Bulletin of the Ecological Society of America, 81*, 246–248.
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.
Chiu, S. W., Gervan, S., Fairbrother, C., Johnson, L. L., Owen-Anderson, A. F. H., Bradley, S. J., et al. (2006). Sex-dimorphic color preference in children with gender identity disorder: A comparison to clinical and community controls. *Sex Roles, 55*, 385–395.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
Cunningham, S. J., & MaCrae, C. N. (2011). The colour of gender stereotyping. *British Journal of Psychology, 102*, 598–614.
Eliot, L. (2009). *Pink brain, blue brain: How small differences grow into troublesome gaps—and what we can do about it*. Oxford: Oneworld.
Fagot, B. I. (1978). The influence of sex of child on parental reactions to toddler children. *Child Development, 49*, 459–465.
Fagot, B. I. (1983). *Interactive behavior code*. Unpublished instrument.
Feinman, S. (1981). A status theory of the evaluation of sex-role and age-role behavior. *Sex Roles, 10*, 445–456.
Fine, C. (2010). *Delusions of gender*. New York, NY: W. W. Norton & Company.
Golombok, S., Rust, J., Zervoulis, K., Croudace, T., Golding, J., & Hines, M. (2012). Continuity in sex-typed behavior from preschool to adolescence: A longitudinal population study of boys and girls aged 3–13 years. *Archives of Sexual Behavior, 41*, 591–597.
Halim, M. L., Ruble, D. N., Tamis-LeMonda, C. S., Zosuls, K. M., Lyrye, 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.
Hassett, J. M., Siebert, E. R., & Wallen, K. (2008). Sex differences in rhesus monkey toy preferences parallel those of children. *Hormones and Behavior, 54*, 359–364.
Hines, M. (2010). Sex-related variation in human behavior and the brain. *Trends in Cognitive Sciences, 14*, 448–456.
Hurlbert, A. C., & Ling, Y. (2007). Biological components of sex differences in color preference. *Current Biology, 17*, 623–625.
Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*, 581–592.
Idle, T., Wood, E., & Desmarais, S. (1993). Gender role socialization in toy play situations: Mothers and fathers with their sons and daughters. *Sex Roles, 28*, 679–691.
Jadva, V., Golombok, S., & Hines, M. (2010). Infants' preferences for toys, colors and shapes: Sex differences and similarities. *Archives of Sexual Behavior, 39*, 1261–1273.
Jenkins, J. (Producer). (2011). *Fighting the power of pink [Radio broadcast]*. Retrieved from http://www.bbc.co.uk/programmes/b010y39d.
Kusuoka, H., & Hoffman, J. I. E. (2002). Advice on statistical analysis for circulation research. *Circulation Research, 91*, 662–671.
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.
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.
Lytton, H., & Romney, D. M. (1991). Parents' differential socialization of boys and girls: A meta-analysis. *Psychological Bulletin, 109*, 267–296.
Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.
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.
Miller, R. G. (1981). *Simultaneous statistical inference*. New York, NY: McGraw-Hill.
Orenstein, P. (2011). *Cinderella ate my daughter*. New York, NY: HarperCollins Publishers.
Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of color preferences. *Proceedings of the National Academy of Sciences, 107*, 8877–8882.
Paoletti, J. B. (2012). *Pink and blue: Telling the boys from the girls in America*. Bloomington: Indiana University Press.
Pasterski, V., Geffner, M. E., Brain, C., Hindmarsh, P., Brook, C., & Hines, M. (2005). Prenatal hormones and postnatal socialization by parents as determinants of male-typical toy play in girls with congenital adrenal hyperplasia. *Child Development, 76*, 264–278.
Paul, A. M. (2011). Is pink necessary? *The New York Times*. Retrieved from http://www.nytimes.com/2011/01/23/books/review/Paul-t.html.
Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *British Medical Journal, 316*, 1236–1238.
Picariello, M. L., Greenberg, D., & Pillemer, D. B. (1990). Children's sex-related stereotyping of colors. *Child Development, 61*, 1453–1460.
Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution, 43*, 223–235.
Rosnow, R. L., & Rosenthal, R. (1988). Focused tests of significance and effect size estimation in counseling psychology. *Journal of Counseling Psychology, 35*, 203–208.
Rosnow, R. L., & Rosenthal, R. (1989). Definition and interpretation of interaction effects. *Psychological Bulletin, 105*, 143–146.
Ruble, D. N., Lurye, L. E., & Zosuls, K. M. (2010). Pink frilly dresses and early gender identity. *Princeton Report on Knowledge*, *2*(2). Retrieved from http://www.princeton.edu/prok/issues/2-2/pink_frilly.xml.
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.
Schloss, K. B., Poggesi, R. M., & Palmer, S. E. (2011). Effects of university affiliation and "school spirit" on color preferences: Berkeley vs. Stanford. *Psychonomic Bulletin and Review, 18*, 498–504.
Serbin, L. A., Connor, J. M., & Iler, I. (1979). Sex-stereotyped and nonstereotyped introductions of new toys in the preschool classroom: An observational study of teacher behavior and its effects. *Psychology of Women Quarterly, 4*, 261–265.
Serbin, L. A., Poulin-Dubois, D., Colburne, K. A., Sen, M. G., & Eichstedt, J. A. (2001). Gender stereotyping in infancy: Visual preferences for and knowledge of gender-stereotyped toys in the second year. *International Journal of Behavioral Development*, *25*, 7–15.
Servin, A., Bohlin, G., & Berlin, L. (1999). Sex differences in 1-, 3-, and 5-year-olds' toy- choice in a structured play-session. *Scandinavian Journal of Psychology, 40*, 43–48.
Snow, M. E., Jacklin, C. N., & Maccoby, E. E. (1983). Sex-of-child differences in father-child interaction at one year of age. *Child Development, 54*, 227–232.
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.
Vander Stoep, S. W., & Johnson, D. D. (2009). *Research methods for everyday life: Blending qualitative and quantitative approaches*. New York: Wiley.
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*, 688–701.