Effects of gender stereotypes and stereotype threat on children's performance on a spatial task

Christine K. Shenouda, Judith H. Danovitch

Dans Revue internationale de psychologie sociale 2014/3 (Tome 27), pages 53 à 77
Éditions Presses universitaires de Grenoble

ISSN 0992-986X
ISBN 9782706122880

RIPS / IRSP, 27 (3-4), 53-77 © 2014, Presses universitaires de Grenoble

Effects of gender stereotypes and stereotype threat on children's performance on a spatial task

Effets des stéréotypes de genre et de la menace du stéréotype sur les performances des enfants à une tâche spatiale

*Christine K. Shenouda**
*Judith H. Danovitch***

Abstract

Although boys and girls start out performing similarly on science, technology, engineering, and mathematics (STEM)-related school subjects, the gap between them widens as they mature, leading to the underrepresentation of women in STEM fields at the college level and in the workforce. One explanation for this phenomenon is stereotype threat: that females are intimidated by the persistent negative stereotypes about women in STEM fields. Examining the inception and development of stereotype threat among young children is crucial for understanding and bridging the STEM gender gap. Two experiments investigated the effect of stereotype threat on performance on a spatial task among children ages 4 to 10. After a gender identity activation

Résumé

Bien qu'au départ garçons et filles performent au même niveau dans les disciplines scolaires liées à la science, la technologie, l'ingénierie et les mathématiques (STEM), l'écart entre eux se creuse à mesure qu'ils grandissent, ce qui conduit à la sous-représentation des femmes dans les domaines des STEM aux niveaux universitaire et professionnel. Une explication de ce phénomène est la menace du stéréotype : les femmes sont intimidées par les stéréotypes négatifs persistants sur les femmes dans les domaines des STEM. L'examen de la création et du développement de la menace du stéréotype chez les jeunes enfants est essentiel pour comprendre et combler le fossé entre les sexes dans les STEM. Deux expériences ont étudié l'effet de la menace du

Key-words
Stereotype threat, gender, spatial skills, preschool, STEM

Mots-clés
Menace du stéréotype, genre, habiletés spatiales, préscolaire, STEM (science, technologie, ingénierie et mathématiques)

This work is based on dissertation research submitted by the first author to Michigan State University in partial fulfillment of the requirements for the PhD, which was supported by a Dissertation Completion Fellowship. We would like to thank the staff, parents, and students at the schools in Ann Arbor, East Lansing, Monticello, and Champaign who participated. We also thank Megan Fisher, Margaret Levasseur, Jenna Shier, and other members of the MSU Knowledge in Development Lab for their assistance and Lauren Harris, Cary Roseth, and Isis Settles for comments on previous versions of the manuscript.

* Department of Psychology, University of Illinois at Urbana-Champaign, 603 E. Daniel St., Champaign, IL, 61820. E-mail: cshenoud@illinois.edu
** Department of Psychological and Brain Sciences, University of Louisville, Life Sciences Building, Room 347, Louisville, Kentucky 40292. E-mail: j.danovitch@louisville.edu

manipulation, children replicated a series of designs with LEGO blocks. In Experiment 1 (N = 22), 4-year-old girls whose gender identity was activated were significantly slower than girls whose gender identity was not activated. In Experiment 2 (N = 160), boys and girls in kindergarten through third grade did not perform differently in stereotype-threat conditions or a condition where gender was not activated. However, girls were consistently slower at the block-construction task than boys. In addition, older children's speed was predicted by their stereotypical attitudes towards LEGO blocks, whether their favorite toy was LEGO blocks, and how frequently they played with blocks at home. Results suggest that gender stereotypes affect children's performance on a gendered activity as early as age 4 and may be continuously affecting older girls' performance. Implications for understanding precursors of the STEM gender gap are discussed.

stéréotype sur la performance à une tâche spatiale chez des enfants âgés de 4 à 10 ans. Après une manipulation de l'activation de l'identité de genre, les enfants devaient copier une série de dessins avec des blocs LEGO. Dans l'expérience 1 (N = 22), les filles de 4 ans dont l'identité sexuelle a été activée sont significativement plus lentes que les filles dont l'identité sexuelle n'a pas été activée. Dans l'expérience 2 (N = 160), des garçons et des filles de la maternelle à la troisième année ne se comportent pas différemment dans des conditions de menace du stéréotype ou dans une condition où le genre n'a pas été activé. Toutefois, les filles sont systématiquement plus lentes dans la tâche que les garçons. En outre, la vitesse des enfants plus âgés est prédite par leurs attitudes stéréotypées à l'égard des blocs LEGO, par le fait que les blocs LEGO sont leurs jouets préférés, et par la fréquence à laquelle ils jouent avec des blocs à la maison. Les résultats suggèrent que les stéréotypes de genre affectent la performance des enfants dans une activité de genre dès l'âge de 4 ans, et peuvent continuer à avoir une incidence sur la performance de filles plus âgées. Les implications pour la compréhension des précurseurs de l'écart entre les sexes dans les STEM sont discutées.

Historically, women have been underrepresented in science, technology, engineering, and mathematics (STEM) careers. Some have argued that this is because women are innately less apt than men at the cognitive skills required to excel at STEM fields (e.g., Baron-Cohen, 2004; Benbow & Stanley, 1980). Others have opposed this view, citing the similar performance of males

and females from infancy through maturity on cognitive tasks required to excel at science and math, such as learning about objects and numbers (see Spelke, 2005). Instead, one potential explanation of the STEM gender gap is that gender stereotypes have detrimental effects on girls' beliefs and performance, and these effects result in a STEM gap emerging in childhood and widening through maturity.

Stereotype threat (ST) is the phenomenon of decreased performance by a group of people when their category membership has been activated, provided that this category is stereotyped as lacking skill at the task at hand (Steele & Aronson, 1995; see Aronson & McGlone, 2009, for review). With respect to math performance, among adults, women reminded of their gender or of math-gender stereotypes perform worse on math tests than men and than women in control conditions (e.g., Beilock, Rydell, & McConnell, 2007; Shih, Pittinsky, & Ambady, 1999). This threat is removed when women are told that a difficult test does not yield gender differences (Spencer, Steele, & Quinn, 1999).

The limited research available demonstrating ST on math and math-related tasks among children has shown that ST exists among school-girls (e.g., Ambady, Shih, Kim, & Pittinsky, 2001; Galdi, Cadinu, & Tomasetto, 2014; Huguet & Régner, 2007; Neuville & Croizet, 2007; Tomasetto, Alparone, & Cadinu, 2011; see Régner, Steele, Ambady, Thinus-Blanc, & Huguet, 2014, for a review) and that it emerges in early elementary schooling (Ambady et al., 2001; Galdi et al., 2014; Tomasetto et al., 2011). Ambady et al. (2001) found that activating ethnic identity in a group of Asian American girls, ages 5 through 13, caused them to perform better on a math test, but activating gender identity caused them to perform worse. Neuville and Croizet (2007) have further shown that among French 7- and 8-year-olds, girls whose gender was activated performed significantly better on easy math problems and significantly worse on difficult problems than children in a control condition. How a task is presented to participants can also affect their performance. Eleven- to 13-year-old girls were better able to recreate a geometric shape from memory when it was presented as a test of artistic skill rather than of geometry ability (Huguet & Régner, 2007).

Recent studies have also shown mediating and moderating effects that influence susceptibility to ST in children. For example, Galdi et al. (2014) have shown that ST among 6-year-old girls was mediated by their automatic association of males and math on an implicit association test (IAT), even when they did not show explicit math stereotypes. In addition, mothers' gender stereotypes about math ability moderated 5- to 7-year-old girls' susceptibility to ST on math tests (Tomasetto et al., 2011).

Overall, these studies suggest that stereotypes can have a detrimental effect on girls' performance on math-related tasks, which are typically stereotyped as masculine. Such threat is removed when the focus is diverted to another category with a positive stereotype in that domain (such as ethnicity, Ambady et al., 2001) or the task is labeled in a way that does not activate a stereotype (Huguet & Régner, 2007). Interestingly, children's explicit stereotypes are not a determining factor in children's susceptibility to ST. Rather, it is their implicitly held beliefs that are related to the effects of ST on their performance (Ambady et al., 2001; Galdi et al., 2014; Huguet & Régner, 2009).

One limitation of the research to date is that the youngest children studied have been 5-year-olds. In fact very few studies have looked at ST in younger children. Because research on ST with children has been focused on academic performance, particularly math tests, it may not have been of interest to study younger children, as they are less exposed to math and its stereotypes. However, measuring whether young children are susceptible to ST in domains where they have already acquired gender stereotypes, such as toy and play preferences, can address important questions about the nature and origins of ST. Children learn gender stereotypes by age 2 (e.g., Serbin, Poulin-Dubois, & Eichstedt, 2002), but have children internalized these stereotypes sufficiently that these stereotypes have negative effects on their performance? It is important to examine these potential effects early in development as children may already be potentially missing learning opportunities due to fear of failure as a result of learned stereotypes (Davies, Spencer, & Steele, 2005).

In two experiments, we examine the emergence and development of ST in children ages 4 to 10 using a play-based task. This

study addresses a number of questions. First, when does gender-based ST start to emerge? Stereotype awareness is a prerequisite for ST (Aronson & Good, 2003; Shelvin, Rivadeneyra, & Zimmerman, 2014) and preschoolers are aware of gender stereotypes (Ruble, Martin, & Berenbaum, 2006), yet the question remains whether these stereotypes already affect their performance. By including 4-year-old preschoolers, the current study also examines if children exhibit ST before beginning formal education. We hypothesized that if preschool girls associate the task at hand with males, whether explicitly or implicitly, then even girls as young as age 4 should exhibit ST when their gender is activated.

Second, we examine whether ST is manifested in a domain that is not typically considered in an evaluative manner. Much of ST research, especially with children, focuses on math ability. Math ability is typically stereotyped as "boys are better than girls" which implies differences in intelligence and skill. ST effects follow clearly from this stereotype: girls are not as good as boys at math, therefore girls perform poorly when their identity is activated. However, toy preferences are not typically considered in a performance-related manner: boys are not perceived as being better at playing with blocks, they just prefer blocks. In addition, building with LEGO blocks is an excellent performance task for children because it is a quantifiable activity, and constructing a shape out of LEGO blocks based on a model requires significant working memory capacity, which is depleted by ST (Chan & Rosenthal, 2014; Régner et al., 2010; Schmader, Johns, & Forbes, 2008). Furthermore, LEGO blocks are stereotypically associated with boys. Thus, we hypothesized that girls would display ST even on a block construction task, which could potentially have major implications for their future STEM performance.

Block construction is a spatial task closely related to math (see Mix & Cheng, 2012) and block-construction skills in preschool predict math performance in middle and high school (Wolfgang, Stannard, & Jones, 2001). There is also a strong relationship between block-reproduction success and mental rotation (Brosnan, 1998; Jahoda, 1979) – a skill in which boys outperform girls as early as infancy (Moore & Johnson, 2008; Quinn & Liben,

2008). In fact, spatial reasoning is one of the few cognitive skills where there are consistent gender differences (Hyde, 2005), and women and girls show ST on mental rotation tasks (Moè, 2009; Moè & Pazzaglia, 2006; Neuburger, Jansen, Heil, & Quaiser-Pohl, 2012). Moreover, spatial training can improve children's math performance (Cheng & Mix, 2014) and narrow the gender gap among undergraduates (Feng, Spence, & Pratt, 2007; Miller & Halpern, 2013). Thus, construction with LEGO blocks may be a precursor for STEM-related skills and interest in STEM fields.

Third, this study addresses the relationship between children's performance on a block-construction task and their attitudes and stereotypes about gendered toys and activities. Similar to Galdi et al. (2014), we expected girls who hold stronger gender stereotypes to perform worse on block construction when their gender was activated. Finally, we examine whether there are gender and developmental differences in susceptibility to ST. Although we expected girls to show susceptibility to ST, boys' performance was harder to predict. In previous research, boys' performance on stereotyped tasks has been shown to either increase slightly (Ambady et al., 2001) or not change at all (Neuville & Croizet, 2007) when gender identity is activated. Developmentally, we expected girls to show increased susceptibility to ST as they became more aware of gender stereotypes with age.

Experiment 1

Method

Participants

Twenty-two preschool girls were randomly assigned to the experimental condition (n = 11, M_{age} = 4.56, range: 4.02-5.15) or the control condition (n = 11, M_{age} = 4.66, range: 4.13-4.86). The majority of children were non-Hispanic Caucasian-Americans.

Materials and procedure

The experiment consisted of three tasks in the following order: gender-identity activation, block construction, and assessment of stereotyping. Each child was tested individually.

Gender-identity activation. Children were handed a picture to color in an opaque folder, allowing the experimenter to remain blind to the condition. Similar to Ambady et al. (2001), girls in the experimental condition colored a picture of a girl holding a doll and girls in the control condition colored a picture of trees.

Block-construction task. After three minutes of coloring, children put the picture back in the folder and the experimenter introduced the block task. After completing a practice shape, children were given five minutes to complete as many shapes as they could. There were 15 possible shapes to complete containing two to seven pieces and they became progressively more difficult. All LEGO pieces used in the block task were light or dark blue, and came in a range of sizes. Children were instructed to work as fast as they could but also make sure to be accurate.

Assessment of stereotyping. To gauge whether children implicitly believed that boys were more skilled at block-construction than girls, children completed an implicit stereotype measure (based on Ambady et al., 2001). Children heard a brief story about a child whose gender was not identified who won a block-construction competition and they were asked to repeat the story. The experimenter noted whether children referred to the child in the story using a masculine, feminine, or neutral pronoun.

To measure children's overall explicit attitudes towards toys, children completed the activities subset of the Preschool Occupations, Activities, and Traits scale (POAT-AM, Liben & Bigler, 2002) and an explicit attitude measure about LEGO blocks. Children indicated if boys only, girls only, or both boys and girls should play with each of 14 toys (6 feminine, 6 masculine, 2 neutral, Liben & Bigler, 2002). To measure explicit stereotypes specifically about LEGO blocks, children also used the POAT-AM scale measure to judge who should play with three sets of LEGO blocks: multicolored blocks, blue blocks, and pink blocks.

Results and discussion

Block-construction task

Children's performance on the block-construction task was measured in terms of speed (how many shapes were attempted) and accuracy (how many shapes were completed accurately as a proportion of total shapes attempted). An independent-samples *t*-test revealed a significant effect of condition on speed, $t(20) = 3.77$, $p = .001$, such that children in the experimental condition were significantly slower than children in the control condition (experimental: $M = 8.18$ shapes, $SD = 1.94$; control: $M = 11.00$ shapes, $SD = 1.55$). However, there were no differences in accuracy, $t(20) = 0.34$, $p = .74$ (experimental: $M = .70$, $SD = 0.097$; control: $M = .72$, $SD = 0.12$). Thus, girls whose gender identity was activated were slower on the block-construction task than girls in the control condition. Girls whose gender identity was activated may have been threatened by the instructions to work fast and get the shapes right, causing them to work more slowly in order to check their work and guarantee its accuracy.

Assessment of stereotyping

Children were more likely to refer to the block-competition winner using a masculine pronoun (59%) than a feminine pronoun (14%) or referring to the child neutrally (27%), $\chi^2(2, N = 22) = 7.18$, $p = .028$.

To analyze children's explicit stereotypes, two composite scores of 0 to 6 were generated for the masculine and feminine activities subsets of the POAT-AM. Children received 1 point for indicating a stereotypical answer and 0 for other answers. A repeated-measures ANOVA with POAT Item Category (masculine, feminine) as the within-subjects variable and Condition (experimental, control) as the between-subjects variable did not reveal a main effect of Item Category nor Condition (Fs $\leqslant 1.52$, ps $\geqslant .23$). Thus, children's answers were similar across conditions and on masculine and feminine items. In fact, children's answers were not significantly different from chance on either masculine or feminine items (ts $\leqslant 1.65$, ps $\geqslant .11$). In this case, chance was the likelihood of choosing the stereotypical answer (boys only for

masculine items, girls only for feminine items) versus choosing either the counterstereotypical answer or choosing "both".

When asked who should play with multicolored, blue, and pink blocks, girls were more likely to answer with "both boys and girls" or with "girls only" than they were to answer with "boys only" for multicolored blocks ($\chi^2(1, N = 22) = 8.91, p = .003$) and blue blocks ($\chi^2(1, N = 22) = 6.55, p = .011$). However, girls were more likely than chance to answer with "girls only" for pink blocks ($\chi^2(1, N=22) = 11.64, p = .001$). There were no effects of condition for either of the three block items (Fisher's Exact tests, ps $\geq .21$). Thus, even though boys typically spend more time playing with blocks than girls (Goble, Martin, Hanish, & Fabes, 2012), girls in this study did not explicitly describe playing with LEGO blocks as a masculine activity.

Taken together, these results suggest that preschool girls are susceptible to ST even when engaging in an activity that is not typically considered in an evaluative manner. Girls in the experimental condition were significantly slower than girls in the control condition, suggesting that the gender-identity activation procedure affected their performance. Although girls did not show strong stereotypical beliefs on the POAT-AM or explicit questions about blocks, they implicitly associated superior block competence with males. The discrepancy between children's implicit and explicit answers is consistent with similar findings in the literature (Ambady et al., 2001; Galdi et al., 2014, Huguet & Régner, 2009). However, it may be further explained by slight variations in the implicit and explicit block stereotyping measures. The implicit measure implied competence whereas the explicit measure merely asked about who should play with these blocks. Perhaps children believe that both boys and girls should play with blocks but that boys are better at building with blocks. This issue is addressed in Experiment 2.

It should be noted that the sample in this experiment was fairly small. Although ST effects were detected even among this small sample, the lack of differences in performance accuracy and stereotyping could be the result of the lack of power. In addition, because boys were not included, this study does not examine potential gender differences on the block-construction task among preschoolers. These issues are addressed in Experiment 2.

Experiment 2

Experiment 1 provided initial evidence that 4-year-old girls are susceptible to ST, even if they do not hold explicit stereotypes about masculine and feminine activities in general and about blocks in particular. How does this effect change with age and how would gender-identity activation affect boys' performance on a block-construction task? To answer these questions, Experiment 2 included boys and girls in kindergarten through third grade. We expected older children to be more knowledgeable about stereotypes about blocks than younger children and thus girls to be more susceptible to ST with age. Boys were included to investigate how their performance is affected by gender-identity activation and whether there are gender differences in performance on the block-construction task, regardless of condition.

One limitation of the gender-identity activation procedure in Experiment 1 was that coloring a picture of a girl holding a doll may have activated not only girls' gender identity but also toy stereotypes, thus inflating the ST effect. To address this possibility, Experiment 2 included a second experimental condition where girls colored a picture of a girl only, without a stereotypical toy. The design was similar to Experiment 1, with a number of changes. First, because the main effect of ST in Experiment 1 was on speed rather than accuracy, Experiment 2 used a more precise measure of speed - the time in which children completed eight shapes. Second, to ensure that children sufficiently processed the experimental manipulation, children in Experiment 2 were asked to confirm which picture they colored at the end of the experiment. Third, in addition to the POAT-AM, children completed the personal activities measure (POAT-PM; Liben & Bigler, 2002). This measure was used to assess whether children's personal preferences for gendered toys and activities, not just their gender stereotypes, were related to their performance on the block-construction task. In addition, parents completed a brief questionnaire about their child's toy preferences in order to understand the relationship between toy preferences and performance.

Method

Participants

Participants included children in kindergarten and first grade (n = 76, M_{age} = 6.33, range: 5.15-7.26, 37 girls) and children in second and third grade (n = 82, M_{age} = 8.25, range: 7.18-10.00, 41 girls). Most children were identified by parents as non-Hispanic Caucasian-Americans.

Materials and procedure

Children were randomly assigned to one of two gender activation conditions or a control condition. The experiment consisted of three segments in the following order: gender-identity activation, block-construction, and assessment of stereotyping. Each child was tested individually. Parents of 144 children also completed a brief questionnaire prior to their child's participation.

Gender-identity activation. As in Experiment 1, children colored a picture using crayons with the experimenter blind to the condition. Girls colored either a picture of trees, a standing girl, or a girl holding a doll. Boys colored either the same picture of trees, a standing boy, or a boy holding a truck. The instructions were identical to Experiment 1. When children completed all segments of the experiment, they were also given a sheet with thumbnails of all three coloring pictures (depending on child's gender) and were asked to circle the picture they colored, to guarantee that they had paid attention to the picture and recalled it. Two boys in the younger age group (in the boy-only coloring page condition) failed to circle the correct picture and were eliminated from all analyses.

Block-construction task. Using the same shades and sizes of LEGO blocks as Experiment 1, children first completed a three-piece practice shape. Then, the experimenter instructed the child to complete eight shapes as fast as possible and to make sure to recreate the shapes accurately. The experimenter noted the time taken to complete the shapes and whether children completed each shape accurately. After children completed the LEGO task, children rated the difficulty of the game using a four-point scale.

Assessment of stereotyping. Children completed the implicit stereotype measure from Experiment 1, the activities personal and attitude subsets of the POAT-AM and POAT-PM, (Liben & Bigler, 2002) and an explicit measure about LEGO blocks.

The POAT-PM measure comprises 14 items, most of which overlap with the POAT-AM. Administration followed Liben and Bigler's (2002) procedures. Children used three faces to indicate how much they liked toys and activities ("a lot", "some", or "not at all"). Children completed the personal measure prior to the attitude measure. Children were then asked whether boys, girls, or both would be better at playing with multicolored, blue, and pink LEGO blocks, using the same scale as the POAT-AM. Although the POAT-AM and PM scales are directed at preschoolers, they were used with this older age group, instead of the COAT scales (Liben & Bigler, 2002) as they contain toys and activities relevant to the block-construction task.

Parent questionnaire: Parents completed a brief questionnaire where they indicated their child's favorite toy and whether their child had access to any of the following toys at home and, if so, how often the child played with these toys (never, rarely, sometimes, frequently, all the time): LEGO blocks, dolls, cars/trucks, and a kitchen or tea set. Parents were not aware that their child would be playing with LEGO blocks during the study.

Results and discussion

Block-construction task

A 3 (Condition: neutral, girl or boy with toy, girl or boy only) x 2 (Child's Gender) x 2 (Age Group) ANOVA was conducted to analyze children's speed on the block-construction task. This full-factorial design revealed no main effect of Condition $F(2, 148) = 0.83$, $p = .44$, $\eta_p^2 = .01$. However, there was a main effect of Age Group, $F(1, 148) = 37.59$, $p < .001$, $\eta_p^2 = .20$, such that older children were faster overall than younger children (younger: $M =$ 200s, $SD =$ 75s; older: $M =$ 145s, $SD =$ 45s). There was also a main effect of Child's Gender, $F(1, 148) = 9.66$, $p = .002$, $\eta_p^2 = .06$, such that boys were faster than girls (boys: $M =$ 159s, $SD =$ 61s; girls: $M =$ 185s, $SD =$ 71s). In addition, there was a signifi-

cant Age Group x Gender interaction, $F(1, 148) = 4.51$, $p = .035$, $\eta_p^2 = .03$ (see Figure 1). To follow up on this interaction, planned *t*-tests comparing girls' and boys' speed in each of the age groups revealed a significant difference in speed among the younger children, $t(76) = 2.94$, $p = .004$, but not among older children, $t(80) = 0.94$, $p = .35$ (Bonferroni-adjusted *p*-value to achieve significance for these two *t*-tests is $p = .025$). There was also a significant Condition x Age Group interaction, $F(2, 148) = 5.54$, $p = .002$, $\eta_p^2 = .06$; however, breaking down the interaction did not reveal any significant effects. None of the other interactions were significant, $Fs \leqslant 0.07$, $ps \geqslant .94$. Thus, whereas speed did not differ by condition, older children were faster than younger children and boys were faster than girls, but older girls' performance did not differ from older boys' performance.

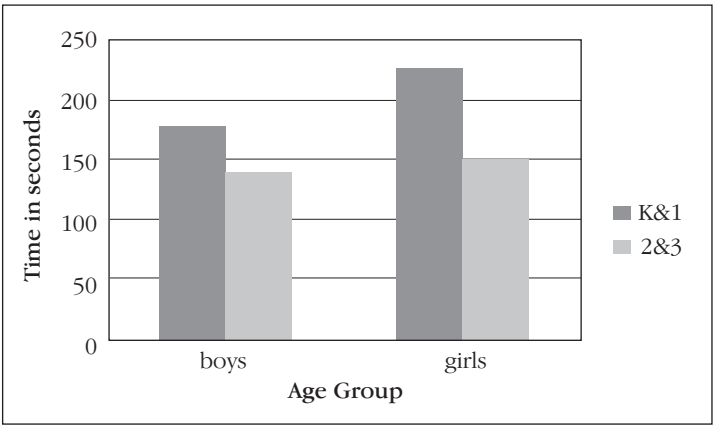

FIGURE 1:
Children's speed on the block-construction task by age (K&1: Kindergarten and 1st grade, 2&3: 2nd and 3rd grade) and gender, collapsed across conditions as performance was not affected by condition.

Children's accuracy on the block-construction task (proportion of correct shapes) was analyzed using the same factorial design. This analysis revealed a main effect of Age Group, $F(1, 148) = 8.24$, $p = .005$, $\eta_p^2 = .05$, such that children in the older age group completed the shapes more accurately than children in the younger age group (younger: $M = .86$, $SD = .20$; older: $M = .93$, $SD = .12$). There were no other significant main effects or interactions, $Fs \leqslant 2.3$, $ps \geqslant .12$. Thus, accuracy on the task was not affected by gender or condition.

Children's ratings of the task difficulty were converted to a 1 to 4 scale, with 1 corresponding to very easy and 4 corresponding to very difficult. Overall, children rated the task as easy ($M = 1.86$, $SD = 0.76$) even though the experimenter introduced the task as a "hard game." A 2 (Gender) × 2 (Age Group) × 3 (Condition) ANOVA revealed a significant main effect of Gender, $F(1, 154) = 9.49$, $p = .002$, $\eta_p^2 = .06$, such that girls rated the game as harder than boys (boys: $M = 1.69$, $SD = 0.70$; girls: $M = 2.05$, $SD = 0.77$). There were no significant main effects of Age Group or Condition (Fs $\leqslant 2.47$, ps $\geqslant .091$). Thus, girls were more likely to indicate that the game was difficult. In fact, there was a significant correlation between children's speed and their difficulty rating, $r(158) = .229$, $p = .001$, such that children who took longer to complete the task rated the task as more difficult.

Assessment of stereotyping

On the implicit stereotype measure, 69% of children used a masculine pronoun to refer to the block competition winner, $\chi^2(1, N = 154) = 99.9$, $p < .001$ (four children refused to answer this question). Older children were more likely than younger children to use a masculine pronoun and boys were more likely to do so than girls, χ^2s $\geqslant 8.3$, ps $\leqslant .004$. In fact, comparing children's responses against chance (chance being using a masculine pronoun versus using a feminine or neutral pronoun) revealed that older and younger boys were more likely than chance to use a masculine pronoun (χ^2s ≥ 8.8, ps $\leqslant .003$), whereas girls' responses were not different from chance (χ^2s $\leqslant 2.95$, ps $\geqslant .086$). Thus, whereas boys implicitly believed that boys were better at building with blocks, girls did not have a consistent implicit stereotype.

To assess children's explicit stereotypes, repeated-measures ANOVAs were used to analyze children's responses to the POAT attitude and personal measures separately. Initial analyses revealed no significant effects of condition on children's responses, and thus only Child's Gender and Age Group were included as between-subjects variables. The Item Category (feminine or masculine) was the within-subjects variable. For the attitude measure, there was a significant effect of Item Category, $F(1, 154) = 10.22$, $p = .002$, $\eta_p^2 = .06$, where children were more

stereotypical on feminine items than masculine items. There was also a significant main effect of Child's Gender, $F(1, 154) = 4.93$, $p = .028$, $\eta_p^2 = .031$, such that boys were more stereotypical than girls, and a significant main effect of Age Group, $F(1, 154) = 4.19$, $p = .042$, $\eta_p^2 = .027$, indicating that older and younger children had different response patterns. A significant Item Category × Age Group interaction further explained this effect, $F(1, 154) = 12.84$, $p < .001$, $\eta_p^2 = .077$, showing that younger children were more stereotypical on masculine items than older children ($t(156) = 3.36$, $p = .001$) and that their responses were equally stereotypical on feminine items ($t(156) = 0.93$, $p = .35$). Taken together, children's opinions about activities and toys (POAT-AM) suggest that they have gender stereotypes about both masculine and feminine activities, that younger children are more stereotypical about masculine items, and that boys are more stereotypical than girls.

For the personal measure, a 2 (Item Category – masculine or feminine) × 2 (Child's Gender) × 2 (Age Group) ANOVA revealed a significant main effect of Item Category, $F(1, 139) = 91.14$, $p < .001$, $\eta_p^2 = .396$, with children preferring masculine items over feminine items. There was also a significant main effect of Age, $F(1, 139) = 14.73$, $p < .001$, $\eta_p^2 = .096$, such that younger children indicated liking items more than older children. This effect would be expected, given that these items primarily target preschoolers. There was also a significant main effect of Child's Gender $F(1, 139) = 4.09$, $p = .045$, $\eta_p^2 = .029$, and a significant Item Category × Child's Gender interaction, $F(1, 139) = 166.48$, $p < .001$, $\eta_p^2 = .55$, which showed that boys preferred masculine items and girls preferred feminine items. There were no other significant interactions, but there was a significant negative correlation between preference for masculine and feminine items, $r = -.324$, $p < .001$, suggesting that children less strongly preferred masculine activities if they more strongly preferred feminine activities, and vice versa.

Overall, children held stereotypical views about which gender would be more competent at building with all three sets of blocks (χ^2s ≥ 27.1, ps $< .001$; see Figure 2). They were more likely to respond that "boys only" were better at building with multicol-

ored and blue blocks and "girls only" were better at building with pink blocks. Whereas boys' and girls' responses for pink blocks were similar ($\chi^2(1, N = 158) = 0.93, p = .30$), boys were more likely to respond "boys only" than girls when evaluating the multicolored blocks and blue blocks (χ^2s ≥ 6.54, ps $\leq .01$). Both boys' and girls' consistent association of girls with pink blocks suggests that the color pink overturns the established stereotype of boys being better at building with LEGO blocks. Furthermore, when analyzed by age group, older children held more stereotypical views about multicolored blocks ($\chi^2(1, N = 158) = .15, p = .042$), but their response patterns were similar for blue and pink blocks (χ^2s ≤ 0.59, ps $\geq .44$).

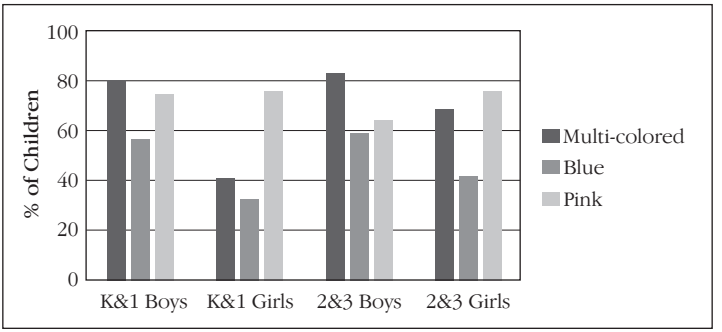

FIGURE 2: Percentage of children who answered "Who is better at building with blocks?" stereotypically for multi-colored ("boys only"), blue ("boys only") and pink blocks ("girls only"). (K&1: Kindergarten and 1st grade, 2&3: 2nd and 3rd grade)

Parent questionnaire

Parents of 28% of boys indicated that LEGO blocks were their child's favorite toy whereas parents of only one girl did so. When asked about the availability and frequency of playing with LEGO blocks, dolls, cars and trucks, and kitchens and tea sets, 51% of boys' parents and 71% of girls' parents reported that all four types of toys were available in their homes (see Table 1). Reported frequency of play was converted to a 5-point scale where 1 corresponded to "never" and 5 corresponded to "all the time" for each type of toy. The frequency of children's play with these toys followed stereotypical patterns, where boys were described as playing with LEGO blocks and cars/trucks more often than girls (ts $\geq$ -3.39, all ps $\leq .001$), and girls as more likely to play with dolls and kitchen/tea sets than boys (ts ≥ 3.59, all ps $\leq .001$).

Parents may have been conforming to gender norms rather than reporting actual frequencies, especially given evidence that parents consistently encourage gender-typed activities for their children (Lytton & Romney, 1991). However, even if this were the case, these results are indicative of what parents in this population expect of their children and how this may affect children's attitudes about STEM.

	Boys (n = 75)		Girls (n = 80)	
	Availability	Frequency	Availability	Frequency
LEGOs	97%	2.37	84%	1.79
Dolls	71%	1.18	100%	2.52
Cars/Trucks	99%	2.21	84%	1.51
Kitchen/Tea Set	56%	1.47	93%	2.11

TABLE 1:
Parent reports of availability of toys and mean scores representing frequency at which children play with these toys, as reported by parents (1=never, 5=all the time).

Block construction, stereotyping, and toy preferences

A multiple regression examined whether children's speed on the LEGO task could be predicted from stereotyping levels on the POAT-AM, implicit and explicit block stereotypes, the child's favorite toy as indicated by parents (LEGO or not LEGO), and how often children played with LEGO blocks. Including only Gender and Age Group in the model predicted 21% of the variance (R^2 = .207, $F(2, 127) = 16.34$, $p < .001$). Adding the remaining five variables listed above resulted in a total prediction of 37% of the variance (R^2 = .369, $F(2, 127) = 10.02$, $p < .001$). In the first model, Gender significantly predicted speed ($\beta = .22$, $p = .007$), as did Age Group ($\beta = -.39$, $p < .001$). In the second model, children's implicit block stereotype and stereotyping level on the POAT-AM did not significantly predict speed ($ps \geqslant .53$). However, explicit block stereotype ($\beta = .29$, $p = .001$), the child's favorite toy ($\beta = -.23$, $p = .007$), and how often children played with LEGO blocks ($\beta = -.19$, $p = .017$) were significant predictors. These results suggest that after controlling for the effects of age and gender, both boys and girls were slower on the block-construction task if they thought boys were better at building with LEGO blocks, if LEGO blocks were not their favorite toy, and if they did not play with LEGO blocks frequently.

General discussion

The goal of this study was to examine the effects of gender stereotypes on children's attitudes and performance during the preschool and early elementary school years. As hypothesized, in Experiment 1, 4-year-old girls completed a toy-based block-construction activity more slowly when their gender identity was activated, even though they did not express explicit gender stereotypes about blocks. Although the sample in Experiment 1 was small and the design does not allow for gender comparisons, calling for caution in our conclusions, this is the first evidence that stereotype threat may be manifested in preschool girls.

Contrary to our hypothesis, in Experiment 2, girls in kindergarten through third grade were not slower or less accurate in stereotype-threat conditions compared to a control condition. They were, however, significantly slower than boys of the same age. Children's performance on the block-construction task in Experiment 2 may reflect a ceiling effect among older children, for whom the task may have been too easy. Thus, a more difficult spatial-construction task may still reveal ST effects in this age group. Despite the presence of a potential ceiling effect, the fact that elementary school children's performance on the block-construction task was predicted by their preferences for and explicit stereotypes about LEGO blocks suggests that there is a relationship between children's beliefs about the task and their performance.

As this study included the youngest age group studied thus far in the ST literature, the finding that preschool girls whose gender identity was activated were vulnerable to ST when recreating shapes with LEGO blocks has important implications for understanding girls' trajectories. If girls are already threatened by their identity at age 4, they may avoid engaging in activities that make them uncomfortable (Frome, Alfeld, Eccles, & Barber, 2006). Because block construction is closely related to math (Mix & Cheng, 2012), this could translate into girls potentially missing opportunities to improve their spatial skills, and hold repercussions for the development of math skills (Cheng & Mix, 2014). It

may even be one of the precursors of women's underrepresentation in STEM fields later in life.

Unlike preschoolers, girls ages 5 to 10 did not show differences in performance under gender-activation conditions on the block-construction task. One explanation for this is that preschoolers are genuinely more susceptible to ST in this task than elementary-school girls. This explanation seems unlikely given that other studies have shown ST across different age groups (e.g., Ambady et al., 2001; Galdi et al., 2014; Huguet & Régner 2007; Neuville & Croizet, 2007). Also, as children learn stereotypes, these stereotypes are more likely to have an effect on their performance (Galdi et al., 2014). Another explanation is that the block-construction task was too easy for elementary-school children, producing a ceiling effect. ST may act differently on easy and difficult tasks and girls sometimes show ST on difficult tasks but not on easy ones (Neuville & Croizet, 2007). This may be a likely explanation given the diminished gender differences among the older age group and children's ratings of the task as "easy."

A third explanation is that girls were under continuous ST as a result of the experimenter describing the task as difficult before the gender-identity activation. This description may have evoked anxiety among girls and activated the stereotype, regardless of condition. This explanation is supported by the consistent gender gap on the block-construction task, where girls were slower overall than boys in both elementary-school age groups. Ganley et al. (2013) have offered a similar interpretation of their results. They did not find ST effects on math tests in a large sample of girls in fourth through twelfth grade, but they did find consistent gender differences on two of their three experiments. This interpretation is also consistent with previous research showing that without identity activation, women perform worse on math tests than men, and their performance can be improved by alleviating the threat, such as indicating that the test does not yield gender differences (Smith & White, 2002; Spencer et al., 1999). One way to test this continuous-threat hypothesis on a block-construction task would be to alleviate girls' stress by activating a positive stereotype, for example by using pink blocks. Both adults and children associate the color pink with females

(Cunningham & Macrae, 2011; Ma & Woolley, 2013) and children in the current study consistently associated pink blocks with girls, both in terms of who should play with them and competence building with them. If girls' performance is hampered by ST, then the use of pink blocks should cause girls to associate the task with their own gender. This could help alleviate the threat and narrow, if not eliminate, the gender gap. Such threat alleviation has been successful in previous research with adults (Spencer et al., 1999) and 11- to 13-year-old girls (Huguet & Régner, 2009).

Unlike previous research showing that children's susceptibility to ST is related to their implicit rather than explicit stereotypes (Ambady et al., 2001; Galdi et al., 2014), there was no relationship between implicit stereotypes about block-construction ability and performance on the block-construction task among any age group in the current study. In contrast, older children's explicit stereotypes about multicolored blocks predicted their speed on the block-construction task, such that both boys and girls were slower if they believed that boys were better at block construction with typical sets of LEGOs. This relationship among girls is consistent with ST research, but it may seem counterintuitive for boys. That said, Cimpian, Mu, and Erickson (2012) found that children's performance on an item-circling task and a mental-rotation task was negatively affected by generic statements about boys' or girls' ability on the task, regardless of whether the statement was positive or negative. By extension, in the current study, boys who believed their group was better at this task may have been anxious about meeting this expectation, thus performing more slowly.

The current findings have a number of implications for educators and parents. By showing that girls as young as age 4 may be susceptible to ST, these findings highlight the danger that stereotypes present for girls in developing the skills necessary for excelling at STEM fields. In addition, by showing that children's preference for and experience with LEGO blocks predicts higher performance on the block task, the present study supports other emerging evidence that encouraging children to engage in play activities related to STEM fields could hold benefits. Recently, there has been a surge in the number of toys directed at girls that

involve activities typically associated with boys. For example, in 2012, LEGO launched a new line of blocks targeting girls that incorporate pink and purple colors and characters such as princesses. This line has become one of LEGO's most popular products (McNally, 2013). Although toy companies may only be seeking to maximize their profit, capturing girls' attention with toys that are related to the development of STEM skills could help narrow the gender gap. Likewise, encouraging parents and educators to promote spatial activities among children at an early age could benefit both girls and boys by improving their spatial skills that have been shown to relate to math and science performance.

References

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.

Aronson, J., & Good, C. (2003). The development and consequences of stereotype vulnerability in adolescents. In F. Pajares & T. Urdan (Eds.), *Adolescence and education*: *Vol. 2. Academic motivation of adolescents* (pp. 299-330). Greenwich, CT: Information Age.

Aronson, J., & McGlone, M. (2009). Stereotype and social identity threat. In T. Nelson (Ed.), *The handbook of prejudice, stereotyping, and discrimination*. New York: Guilford.

Baron-Cohen, S. (2004). *Essential difference: Male and female brains and the truth about autism*. New York: Basic Books.

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General*, *136*, 256-276.

Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact?. *Science*, 210, 1262-1264.

Brosnan, M. J. (1998). Spatial ability in children's play with Lego blocks. *Perceptual and Motor Skills, 87*, 19-28.

Chan, N. H. W., & Rosenthal, H. E. S. (2014). Working memory moderates stereotype threat effects for adolescents in Hong Kong. *International Review of Social Psychology / Revue Internationale de Psychologie Sociale, 27,* 103-118.

Cheng, Y. L. & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development,* 15(1), 2-11.

Cimpian, A., Mu, Y., & Erickson, L. C. (2012). Who is good at this game? Linking an activity to a social category undermines children's achievement. *Psychological Science, 23*(5), 533-541.

Cunningham, S. J., & Macrae, C. N. (2011). The colour of gender stereotyping. *British Journal of Psychology*, *102*, 598-614.

Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of Personality and Social Psychology*, 88, 276-287.

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science, 18*(10), 850-855.

Frome, P. M., Alfeld, C. J., Eccles, J. S., & Barber, B. L. (2006). Why don't they want a male-dominated job? An investigation of young women who changed their occupational aspirations. *Educational Research and Evaluation, 12*(4), 359-372.

Galdi, S., Cadinu, M., & Tomasetto, C. (2014). The roots of stereotype threat: When automatic associations disrupt girls' math performance. *Child Development*, *85*(1), 250-263.

Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., & Perry, M. (2013). An examination of stereotype threat effects on girls' mathematics performance. *Developmental Psychology*, *49*(10), 1886.

Goble, P., Martin, C. L., Hanish, L. D., & Fabes, R. A. (2012). Children's gender-typed activity choices across preschool social contexts. *Sex Roles, 67*, 435-451.

Huguet, P., & Régner, I. (2007). Stereotype threat among schoolgirls in quasi-ordinary classroom circumstances. *Journal of Educational Psychology*, *99*(3), 545-560.

Huguet, P., & Régner, I. (2009). Counter-stereotypic beliefs in math do not protect school girls from stereotype threat. *Journal of Experimental Social Psychology, 45*(4), 1024-1027.

Hyde, J. S. (2005). The gender similarities hypothesis. *American psychologist*, *60*(6), 581-592.

Jahoda, G. (1979). On the nature of difficulties in spatial-perceptual tasks: Ethnic and gender differences. *British Journal of Psychology*, *70*, 351-363.

Liben, L. S., & Bigler, R. S. (2002). The developmental course of gender differentiation: Conceptualizing, measuring, and evaluating constructs and pathways. *Monographs of the Society for Research in Child Development, 67, (2, Serial No. 269).*

Lytton, H., & Romney, D. M. (1991). Parents' differential socialization of boys and girls: A meta-analysis. *Psychological Bulletin*, *109*(2), 267.

Ma, L., & Woolley, J. (2013). Young children's sensitivity to speaker gender when learning from others. *Journal of Cognition and Development, 14*(1), 100-119.

McNally, M. (2013). Successful LEGO strategy delivers continued strong growth. Retrieved April 8, 2014 from http://aboutus.lego.com

Miller, D. I., & Halpern, D. F. (2013). Can spatial training improve long-term outcomes for gifted STEM undergraduates?. *Learning and Individual Differences*, *26*, 141-152.

Mix, K. S., & Cheng, Y. L. (2012). The relation between space and math: Developmental and educational implications. *Advances in Child Development and Behavior*, *42*, 197-243.

Moè, A. (2009). Are males always better than females in mental rotation? Exploring a gender belief explanation. *Learning and Individual Differences*, *19*, 21-27.

Moè, A., & Pazzaglia, F. (2006). Following the instructions! Effects of gender beliefs in mental rotation. *Learning and Individual Differences*, *16*, 369-377.

Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants a sex difference. *Psychological Science, 19*(11), 1063-1066.

Neuburger, S., Jansen, P., Heil, M., & Quaiser-Pohl, C. (2012). A threat in the classroom: Gender stereotype activation and mental-rotation performance in elementary-school children. *Zeitschrift für Psychologie*, *220*(2), 61.

Neuville, E., & Croizet, J.-C. (2007). Can salience of gender identity impair math performance among 7-8 years old girls? The moderating role of task difficulty. *European Journal of Psychology of Education, 22*(3), 307-316.

Quinn, P. C., & Liben, L. S. (2008). A sex difference in mental rotation in young infants. *Psychological Science*, *19*, 1067-1070.

Régner, I., Smeding, A., Gimmig, D., Thinus-Blanc, C., Monteil, J. M., & Huguet, P. (2010). Individual differences in working memory moderate stereotype-threat effects. *Psychological Science, 21*(11), 1646-1648.

Régner, I., Steele, J. R., Ambady, N., Thinus-Blanc, C., & Huguet, P. (2014). Our future scientists: A review of stereotype threat in girls from early elementary school to middle school. *International Review of Social Psychology / Revue Internationale de Psychologie Sociale, 27,* 13-51.

Ruble, D. N.; Martin, C. L., & Berenbaum, S. A. (2006). Gender development. In: Eisenberg, N., (Ed.), *Handbook of Child Development*. New York: Wiley.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*, 336-356.

Serbin, L. A., Poulin-Dubois, D., & Eichstedt, J. A. (2002). Infants' response to gender-inconsistent events. *Infancy*, *3*, 531-542.

Shelvin, K. H., Rivadeneyra, R., & Zimmerman, C. (2014). Stereotype threat in African American children: The role of Black

identity and stereotype awareness. *International Review of Social Psychology / Revue Internationale de Psychologie Sociale, 27,* 175-204.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, *10*, 80-83.

Smith, J. L., & White, P. H. (2002). An examination of implicitly activated, explicitly activated, and nullified stereotypes on mathematical performance: It's not just a woman's issue. *Sex Roles*, 47, 179-191. doi:10.1023/A:1021051223441

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science?: A critical review. *American Psychologist*, *60*(9), 950-958.

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.

Steele, C.M., & Aronson, J. (1995). Stereotype Threat and the intellectual test-performance of African-Americans. *Journal of Personality and Social Psychology*, *69*(5), 797-811.

Tomasetto, C., Alparone, F. R., & Cadinu, M. (2011). Girls' math performance under stereotype threat: the moderating role of mothers' gender stereotypes.*Developmental Psychology*, *47*(4), 943.

Wolfgang, C. H., Stannard, L. L., & Jones, I. (2001). Block play performance among preschoolers as a predictor of later school achievement in mathematics. *Journal of Research in Childhood Education*, 15, 173-180.