Gender stereotypes about intellectual ability emerge early and influence children's interests: extension and replication

by

Kian Golestaneh*, Megha Krishnan*, Pearl Lee*, Ailinh Nguyen*, Robin Stewart*, Jocelyn Xie*

A thesis submitted for Undergraduate Laboratory

in

Psychology and Cognitive Science

of the

University of California, Berkeley

Undergraduate mentor:

Huang H. Ham

Graduate student mentor:

Roya Barhaloo

May 2020

Acknowledgements

We would like to thank these special people for their contributions that made this paper possible:

The faculty advisor -- Professor D'Esposito

Our graduate student mentor -- Roya Barhaloo for her insightful guidance.

Our mentor -- Ham H. Huang for his uplifting support and constructive discussions.

The board members of the Psychology and Cognitive Sciences Division in the Undergraduate Laboratory at Berkeley -- Hareen Seerha, Shreeya Garg, Rajan Parikh, Lexi Zhou, Ashish Ramesh, Arushi Sahay, Josephine Widjaja, and Samadi Karunasundera for keeping us up to date on logistics and giving helpful feedback.

Abstract

Negative gender stereotypes targeting women's intellectual ability are a deterrent to women's pursuit of prestigious careers in fields whose members believe brilliance is required for success. Previous research has documented that children start to develop gender stereotypes about brilliance at the age of six. This can have lifelong effects as people are drawn towards gendered activities, careers, etc. Statistical replication with preexisting data from Bian et al. (2017) supports this claim. To further examine the validity and sufficiency of the measures in the original study, we develop an extension by implementing two new statistical models. Lastly, we propose future work to address the possible role of masculine appearances in the perception of brilliance.

Keywords: Gender stereotype; Intelligence; Masculinity; Child development

Gender stereotypes about intellectual ability emerge early and influence children's interests: extension and replication

Introduction

It is important to examine the effect of gender stereotypes because of its suspected relationship to gender inequality across society. For instance, there is a lack of female PhDs in fields such as physics or philosophy, where members agree natural talent is necessary for success (Storage et al. 2016, Myer et al. 2015, and Leslie et al. 2015). The lack of acceptance and belonging that persists because of stereotypes further discourage women in pursuit of study in these fields (Good, 2012). Stereotypes also present a concern by being reflected subconsciously in actions, as demonstrated by parents unintentionally encouraging scientific thought and engagement more with sons than daughters as well as influencing their estimation of their child's overall intelligence (Tenenbaum et al. 2003, Furnham & Bunclark, 2006). The original study focused on a method of gender evaluation observing how children judge other members of the same sex, and hypothesized that the gender stereotype of brilliance emerge in children as young as 6. Previous studies have also compared gender differences through both the participant's self-evaluation as well as their judgment of other genders (Furnham & Bunclark, 2006). We expected to receive similar results as the original study for our replication. We performed different statistical approaches as an extension to further test whether the outcomes for the experiment were due to chance. We also proposed another extension that supplements the findings of the original study by using clothing associated with a gender; this study would test our hypothesis that children as young as 6 would exhibit gender stereotypes of brilliance in response to these visual cues. We will summarize previous work, discuss our statistical replication and extension, and conclude with proposed future directions. Future work should consider the role of masculine appearance in perceptions of intelligence.

Replication and Extended Analysis

Background

Four separate studies were conducted in order to observe the development of the stereotype that associates males, not females, with brilliance. Study 1 aims to assess the development of this stereotype across 5, 6, and 7-year-olds. Study 2 presents a follow-up of the previous study by examining whether school achievement is involved in the children's perception of brilliance. With the establishment of the developmental trajectory of gender-brilliance stereotypes in Study 1 and 2, Study 3 investigates whether children's gendered beliefs toward brilliance impact their interest in novel games that are labeled for children who are either really smart or really hardworking. Study 4 further examines whether such an impact of gendered differentiation is significant at 6-years-old, the presumed age at which children start to develop own-gender brilliance perceptions.

Methods

Study 1

The study first looks to discern whether children demonstrate endorsement of the stereotype of equating males to brilliance. The researchers who conducted the study believed that girls especially would experience a drastic shift between age groups from associating brilliance to their own gender to being less likely to make that association when they become older. In order to test this hypothesis, children are led through a series of screener questions to demonstrate an understanding of the words “smart” and “nice” before being asked to complete three tasks. The three tasks are given in a randomized order and would reveal the frequency the participants attribute a trait like brilliance or niceness to members of the same gender.

A total of 96 children were participants in this study and there were equal amounts of boys and girls. The children were ages 5, 6, or 7; mostly from middle-class backgrounds; and 75% white. The children were initially assessed for their endorsement of the gender stereotype that brilliance is associated with males. The “niceness” trait was added for comparison because it represents a stereotype that contrasts with the brilliance stereotype (Fiske et al, 2002).

In task (i), the children listened to two stories (given in random order) about a person whose gender was not indicated; one was about an individual who was “really, really smart” and the other was about an individual who was “really, really nice.” The children were then instructed to pick one of 4 pictures, 2 of which were women and 2 were men, to identify who they believed was the protagonist of the story. The pictures were matched for attractiveness and showed the people in professional attire.

For task (ii), 6 pictures were shown one by one, each depicting two people. The children had two practice trials using two photos with same-gender pairs. The following four pictures had one woman and one man. In all trials, the children were told one of the two in the photo was smart or nice and they chose who they believed matched the trait.

In task (iii), the children were asked to place each picture of a high heeled shoe, a hammer, the word “smart”, and the word “nice” next to one of the 4 pictures of individuals they believed the trait or picture corresponded to.

For each trial, a score of 1 was given for a child picking someone of the same gender, and 0 was given otherwise. The overall score was the average of the child's score.

Study 1 Conclusion

The results show a shift in perception of brilliance at 6 years of age. Five-year-old boys and girls exhibit equal own-gender preferences. However, 6- and 7-year-old girls, when compared to boys, are less likely to associate brilliance to their own gender. We successfully replicated the general effect for 5- 6- and 7-year-old children (Wald $X^2 = 0.02$, $p = 0.89$ for 5 year olds as well as Wald $X^2 = 7.56$, $p = 0.006$ for 6 and 7 year olds). The three-way interaction of trait, gender, and age was significant in the replication. The graph (Figure 1) that plots the stereotype score against age was also successfully replicated.

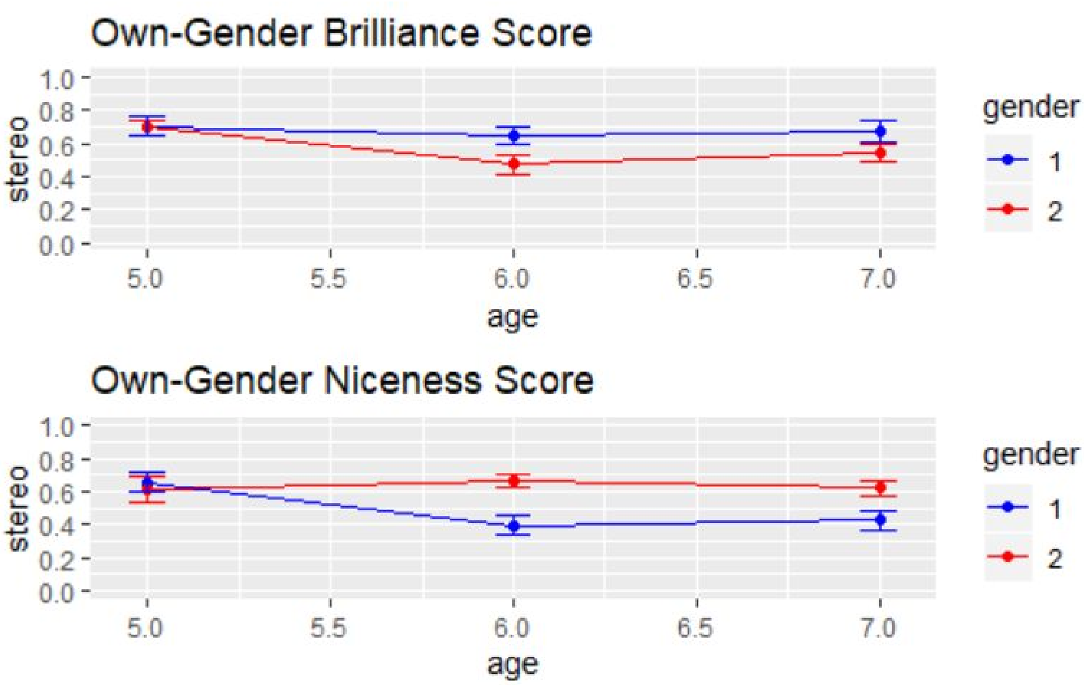

Figure 1. Results of Study 1 boys' (blue) and girls' (red) gender stereotype score for brilliance and niceness by age group (5-, 6-, and 7-years old). Error bars represent ± 1 SE.

Study 1 Extension Analysis

To provide additional support for the results, we attempted to convert the stereotype scores into values that can be inputted into a logistic regression model. The logistic regression model assumes the error of the dependent variable (in our case, stereotype score) to follow a binomial distribution, which provides a better assumption of the data as it is more representative of the options available in the original study; the original study had assumed a normal distribution and made the options appear as continuous rather than categorical, so it was not indicative of the data gathered in the actual study. Using a WaldChi test, we received highly significant results as seen with the p-values received in our logistic regression and Wald X^2 test for 6- and 7- year olds. As we observed a negative correlation in our logistic regression, this trend demonstrates that boys have higher stereotype scores in study 1 (Wald $X^2 = 21.2$, p = 4e-0.6). These results indicate boys as more likely to choose options correlating to their own gender compared to girls. For the 5-year old, we received non-significant results for the stereotype score due to a large p-value (Wald $X^2= 0.074$, $p = 0.785$). The test for 6- and 7-year-old children also have a negative correlation, but in this case, the result is significant and indicates that girls are less likely to choose options correlating to their own gender (Wald X^2= 29.659, p= 5e-0.8). Overall, the results uphold the originally stated conclusions for study 1.

Another methodology-driven extension employed cross-validation (CV), one of the resampling techniques, to test the effectiveness of machine learning models and to evaluate a model with limited data. Root Mean Square Error (RMSE) is the standard deviation of the residuals, which measures the distance between the data points and the regression line. In other words, RMSE, a measure of how spread out these residuals are, assesses the data concentration around the line of best fit. In our extension, we employed RMSE for cross-validation as an assessment of how the test data is concentrated around the line of best fit for the linear model we created with the training data. Furthermore, we created confidence intervals from the means and standard deviations of the SMEs to ensure that the findings were significant.

For the first part of Study 1, we performed 10-fold cross-validation on linear models by creating a subset of data, Study 1a, that looked at subjects aged 5. We first created 10 folds by randomly splitting Study 1a. In addition, we created a for-loop for establishing one fold as the test data and then fit our models to the remaining nine folds. We further proposed a null and an alternative hypothesis for each result found in the study. The alternative hypothesis stated that the stereotype score was influenced by gender. The null hypothesis is that no other variable influenced the stereotype score. Based on each of these models, we made predictions and used them, in addition to the testing data, to generate the RMSEs for both models. Averaging the RMSEs, we further constructed confidence intervals for these means.

The mean of the RMSEs for the alternative hypothesis (adding gender in the model affects interest) is 0.23729650. The mean of the RMSEs for the null hypothesis (adding gender in the model does not affect interest) is 0.23249582. The confidence interval is (-0.01204477, 0.02164614), which contains 0. Because this interval contains 0, this result is not significant as this means that 0 is a reasonable possibility for the true value of the difference.

For the second finding in Study 1, we repeated the same process for Study 1a, except this subset, Study 1b, looked at subjects ages 6 and 7. The mean of the RMSEs for the alternative hypothesis (adding gender in the model affects the stereotype score) is 0.23560490. The mean of the RMSEs for the null hypothesis (adding gender in the model does not affect the stereotype score) is 0.23473267. The confidence interval is (-0.01257856, 0.01432302), which contains 0. Because this interval contains 0, this result is not significant as this means that 0 is a reasonable possibility for the true value of the difference.

Study 2

To explore the reason behind the drop for girls in their own-gender association for brilliance, the experiment turns to look at beliefs about academic achievement as a possible element in evaluating a gender's intellectual ability. Study 2 has a similar design as that of study 1 but with some exceptions: there was a larger sample size of 144 children; each task was separated into 2 blocks, one with pictures of men and women and one with pictures of boys and girls; task (iii) was omitted (the task where children matched images like a hammer or words like "smart" to one of four pictures), and children's perception of school achievement for boys and girls were analyzed. To assess the perception of school achievement, children were shown 4 pictures of unfamiliar children, 2 of which were boys and 2 were girls, and asked who they believed would get the best grades in school. Afterward, the four pictures were all replaced and the children chose the photo they believed showed the person who was first in their class. Finally, the children were not shown any photos but were asked the same two questions about school achievement. They then had to choose between the 2 verbal options of girls or boys. The same measurements were taken as Study 1 where a score of 1 is assigned if the participant chose options of the same gender, and a score of 0 if otherwise.

Study 2 Conclusion

The results are in a similar vein as Study 1 in that 5-year-old boy and girl participants tended to pick their own gender. The results also show that 6- and 7- year-old girls were more likely to select their own gender as having high grades compared to the boy participants, but the researchers determined that there was no significant correlation between performance at school and perceptions of brilliance in relation to girls' drop in own-gender scores. The researchers found no significant difference in the gender-brilliance scores of 5-year-old boys and 5-year-old girls (Wald $X^2 = 0.01$, $P = 0.94$). Through replication (shown in Figure 2), the results were similar (Wald $X^2 = 0.01$, $P = 0.93$). They used another Wald X^2 test to see whether a significant difference emerged starting at age 6 (Wald x^2 = 9.63, P = 0.002), and when the attempt was made to replicate the data the results followed the same trend (Wald $X^2 = 8.492$, $P = 0.004$).

Figure 2. Results of Study 2 boys' (blue) and girls' (red) gender stereotype score for brilliance and niceness by age group (5-, 6-, and 7-years old). Error bars represent ± 1 SE.

Study 2 Extension Analysis

As with Study 1, we used a logistic regression model to ascertain the validity of the original study's results. For Study 2, we further broke down the groups: one for gender versus stereotype while the other looked at the relationship between gender and grades. Looking at gender versus stereotype for children aged 5, our results showed that there was no significant proof (Wald $X^2 = 0.0051$, $P = 0.9428$) that at the age of 5, gender could be a predictor of stereotype scores, aligning with the findings of the original paper. Within the group of children aged from 6 to 7, we found significant results (Wald $X^2 = 9.5461$, $P = 0.002004$) once again falling in line with the results from the original paper. Finally, we observed the effect of introducing a new variable: target. This new variable observed if there was a difference in stereotype scores when each task was separated into two blocks: men and women or boys and girls. Here the results agreed with that of the original paper showing that the target variable does not predict stereotype score (Wald $X^2 = 1.1608$, $P = 0.5597$). The next variables we compared were stereotype scores and grades to see the connection between gender and perception of school achievement. For the first look, we compared the stereotype score versus the grade data as a whole and found significant results (Wald $X^2 = 9.7177$, $P = 0.001825$) showing that there was a correlation between perceptions of school achievement and perceptions of brilliance for both boys and girls. But searching within sub-groups of the data set we can see a different result. The data shows us that this correlation is only significant for the boys while on the contrary not showing significance for the girls' group (boys: Wald $X^2 = 20.096$, $P = 7.365\text{e-}06$) (girls: Wald $X^2 = 1.4171$, $P = 0.2339$). This tells us that a significant correlation between perceptions of school achievement and perceptions of brilliance exists largely in the boys' group rather than the girls.

Study 2 followed the same procedure of performing cross-validation in Study 1, except using a different subset, Study 2a, that looked at subjects aged 5. The alternative hypothesis was that the stereotype score was influenced by gender. The null hypothesis is that no other variable influenced the stereotype score. The mean of the RMSEs for the alternative hypothesis is 0.29562527, while that of the null hypothesis is 0.29583017. The confidence interval is (-0.02324312, 0.02283331), which contains 0, meaning that our result was not significant.

For the second finding in Study 2, we repeated the same process for Study 2a, except this subset, Study 2b, looked at subjects ages 6 and 7. The mean of the RMSEs for the alternative hypothesis is 0.320867532. The mean of the RMSEs for the null hypothesis is 0.321322483. The confidence interval is (-0.006759086, 0.005849184). This result is not significant.

Replication of Supplementary Materials for Studies 1 and 2

To further examine the effect of variables that are not mentioned in the original paper on children's stereotype and interest scores for Studies 1 and 2, we utilized the demographic data provided by the authors' supplementary materials (shown in Figure 3). To do this, we focused on primarily two variables: the average of the education level of the subject's parents, and the total income of the household. The education average was on a scale of 10 to 20, with 10 being less than a high school diploma and 20 being a professional degree. The income variable was the gross income of the household.

For Studies 1 and 2, our linear regression models assessed how the education average of the subject's parents and the total household income might independently affect the subject's stereotype score. For education average, we standardized and ran a summary of the model to evaluate the slope for eduave. The results showed a slope (0.011821) whose p-value (0.8216) was greater than 0.05, signifying the result as nonsignificant. For income, we repeated the same procedure as that of the assessment of education average. The results showed a slight positive slope (0.00859). However, the results were again nonsignificant as shown by the p-value (0.8732).

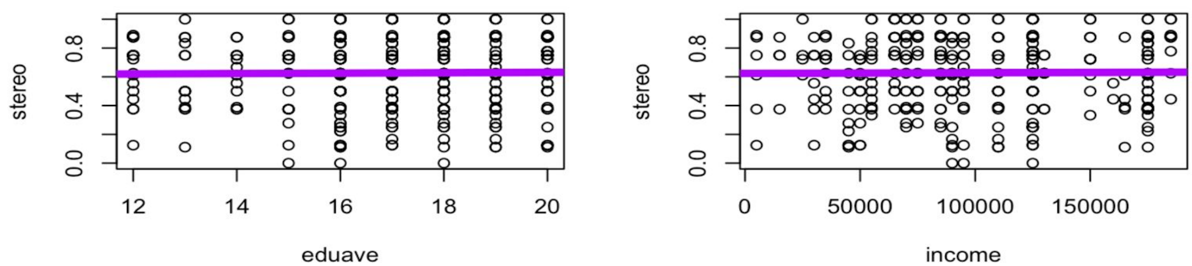

Figure 3. Children's stereotype scores are submitted to linear regression with the average education level of the parents and the total income of the household.

Study 3

In an effort to examine the effect of gendered beliefs about brilliance on children's interest between the Smart Game and Try-Hard Game, Study 3 introduced 64 children aged 6 and 7, of which half were boys and half were girls, to the two games. These games were labeled by the experimenter as being either for "really, really smart" children or for those that "try really, really hard."

Study 3 Conclusion

The results showed that girls showed less interest in the Smart Game and more interest in the Try-Hard Game. Their own-gender score for this task was compared to the own-gender score of Study 1 and Study 2 and it was similar.

The children interests were measured in the two games, and results show that girls were more interested in the game for those who try hard compared to the boys (Wald $\chi^2 = 4.02$, $P = 0.045$), and less interested in the game meant for those who are smart (Wald $\chi^2 = 0.53$, $P = 0.47$). The results of the girls' preferences were checked using a multiple regression model in R, but our output failed to replicate the study's results. Another result showed that girls' own-gender brilliance perceptions were lower than the boys' perceptions ($t = 2.40$, $P = 0.020$). In R, a t-test with a linear regression model was used to check these results, but once again the replication failed. Yielding a larger significance than that of the reported data cast doubt on the replicability of the reported results. A replication of the graphical representation of the results was also carried out (Figure 4).

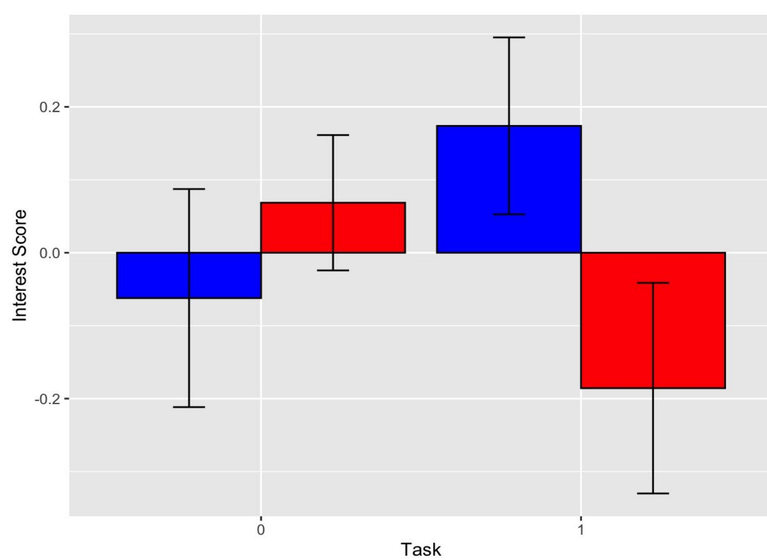

Figure 4. Results of Studies 3: Boys' (blue) and girls' (red) interest score of the games introduced in Study 3. The independent variable is the Task (0 = Smart Game, 1 = Try-Hard Game). Error bars represent ± 1 SE.

Study 3 Extension Analysis

Repeating the computation of cross-validation as that of Study 1, we created a new subset of data, Study 3a, that looked at subjects who were shown the game for smart children. The alternative hypothesis was that the interest score was influenced by gender. The null hypothesis was that no other variable influenced the interest score. The mean of the RMSEs for the alternative hypothesis is 0.80911518 and that of the null hypothesis is 0.83600246. The confidence interval (-0.07750660, 0.02373205) contains zero, which indicates the lack of significance of the result.

For the second finding in Study 3, we repeated the same process for Study 3a, except this subset, Study 3b, looked at subjects who were shown the game for hard-working children. The mean of the RMSEs for the alternative hypothesis is 0.74253496. The mean of the RMSEs for the null hypothesis is 0.71327367. Zero is included in the confidence interval (-0.05456672, 0.11308930), leading us to believe this result is nonsignificant.

Study 4

To assess the prediction of the age of six as the turning point for children's differentiation of brilliance perception, Study 4 zeroed in on a larger sample size of 96 5-and-6-years-old girls and boys. Study 4 followed identical procedures as that of Study 3, except the subjects were only instructed about the game labeled as for children with brilliance. It confirmed a decline in interest among 6-year-old girls in comparison to boys toward the Smart Game.

Study 4 Conclusion

Results (shown in Figure 5) show no significant gender differences in interest among 5-year-olds (Wald $\chi^2 = 0.55$, $P = 0.46$) while ideas about brilliance start to impact children's interest at the age of 6 (Wald $\chi^2 = 3.66$, $P = 0.056$) (Bian et al., 2017). The measurement of an interest score was assessed by averaging instances in which a subject picked the game designed for "really, really smart" children. In performing a random-effect meta-analysis of gender differences in 6- and 7-year-olds' interest in the smart game, the study combined results from Study 3 to estimate the effect size for the gender difference between 6-and-7-year-olds' interest toward the smart game ($d = 0.51$, 95% confidence interval = [0.17, 0.92], $P = 0.008$). (Bian et. al, 2017).

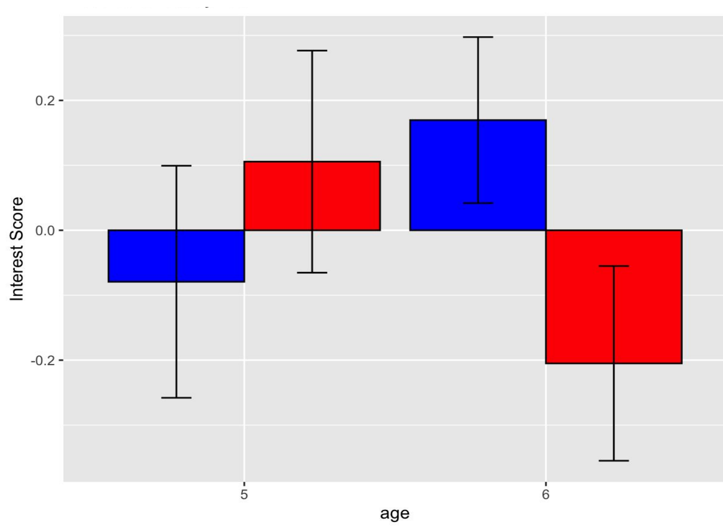

Figure 5. Results of Study 4: Boys' (blue) and girls' (red) interest score of the games introduced in Study 4. The independent variable is the Age (5-and-6-years-old). Error bars represent ± 1 SE.

Study 4 Extension Analysis

In the replication of the Wald Chi-square test for 5-year-olds, results of different values but following the general trend were collected, indicating the lack of significance in gender-interest scores between 5-year-old boys and girls (Wald $\chi^2 = 0.56$, P = 0.46). Different results were again received in the replication of the Wald Chi-square test for 6-year-olds but were able to demonstrate the presence of gender difference in interest toward the brilliance game among 6-year-old boys and girls (Wald $\chi^2 = 3.62$, P = 0.057). Replication of random-effect meta-analysis also yielded different results but successfully revealed a significant (p=0.005) gender difference in interest between 6-and-7-year-old boys and girls (d= 0.64, 95% confidence interval = [0.18, 1.11], P = 0.005). A replication of the graphical representation of the results was also constructed with the aid of ggplot (Fig. 3).

Performing the same cross-validation process as explained in Study 1, we constructed a subset of data, Study 4a, that looked at 5-year-olds who were shown the game for smart children. The alternative hypothesis was that the interest score was influenced by gender. The null hypothesis is that no other variable influenced the interest score. The mean of the RMSEs for the alternative hypothesis is 0.82436975. The mean of the RMSEs for the null hypothesis is 0.81836398. Zero is included in the confidence interval (-0.06115407, 0.07316560), leading us to believe that this result is not significant.

For the second finding in Study 4, we repeated the same process for Study 4a, except this subset, Study 4b, looked at 6-year-olds who were shown the game for smart children. The mean of the RMSEs for the alternative hypothesis is 0.6546080. The mean of the RMSEs for the null hypothesis is 0.6688051. The confidence interval is (-0.1955491, 0.1671548) and includes zero, meaning that the result is nonsignificant.

Replication of Supplementary Materials for Studies 3 and 4

In corroborating the correlation between the demographic data and the children's interest score, our linear regression models looked at how the education average of the subject's parents and the total household income might independently affect the subject's interest score (shown in Figure 6). We followed the same procedures as Studies 1 and 2, and our results showed that for our model looking at education average, the slope was slightly negative (-6.981e-02), however, the p-value was greater than 0.05, and our results were again nonsignificant. Income, on the other hand, the slope (0.04810) rendered a nonsignificant result with the p-value (0.596) being greater than 0.05.

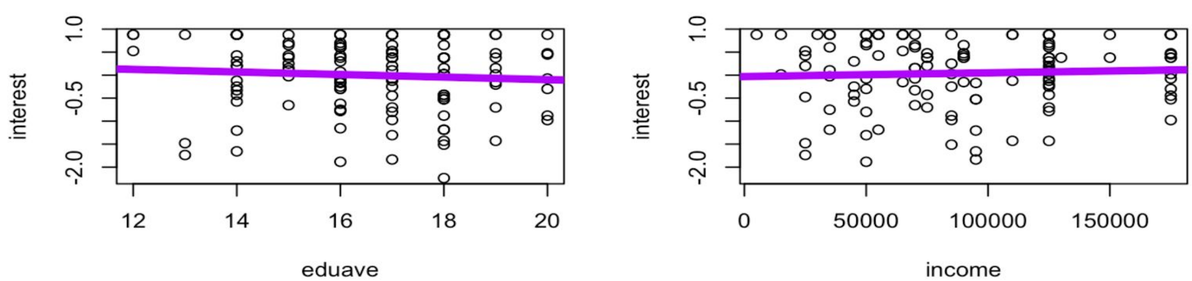

Figure 6. Children's interest scores are submitted to linear regression with the average education level of the parents and the total income of the household. Scatterplots with a regression line are generated and shown. Positive slopes reflect a positive correlation between selected variables.

Extended Study Design Proposal

Background

Longstanding notions of leadership have been closely attached to attributes associated with masculinity (Johnson et al., 2008). Perceptions of gender-role characteristics and transactional leadership are often a primary factor in the underrepresentation of females in fields that demand effective leadership (Johnson et al., 2008, Meyer 2015). Among the pervasiveness of gender-stereotypical attributes, the type of attire has shown a historical influence on perceived professionalism in both academia and applied communities (Furnham et al. 2013). Prior studies have probed into the stereotypical mechanisms and proposed the role congruity theory of prejudice toward female leaders; such is demonstrated when women were found to require both feminine as well as masculine leader behaviors to be evaluated as effective leaders (Johnson et al., 2008). Very limited research, however, examines the genesis of such stereotypes among children at the stage of development (Koenig et al. 2011, Good et al. 2012). However, it was demonstrated in previous studies that older children differentiate between feminine and masculine cues in applying stereotypes in making judgments about people (Biernat, Monica. 1991). To address the question of whether attire formulates a gendered stereotype of a professional leader figure for children at the developmental phase, the follow-up study will select children within the age range of 5 to 7 as the study subjects as opposed to the traditional assessment on adult subjects. The measurements and methodology to test for the children's association with a certain gender will be similar to that of Study 1 with the critical adjustment of including only photos of women portrayed with either masculine or feminine visual cues. This change is done in order to measure the extent to which children attribute the brilliance = males stereotype. Through this modification, we hope to contribute to previous research by exploring whether gender-linked alterations in appearance would influence the perception of brilliance in children.

We predict to observe a similar trend in older girls' own-gender brilliance score as seen in the original study with the expectation that the stereotype of brilliance can also be attributed to visual masculine cues that can transfer its effect to a female wearer. For instance, children would see pictures of a woman wearing a suit or other masculine clothing and be more likely to attribute traits associated with masculinity to them compared to women who wear more feminine clothing such as a skirt or dress.

Methods

Preliminary

12 screener questions are given to ascertain the children's concept of "brilliance" and "niceness" as per the original study. Both traits are represented in equal amounts and presented in both separate blocks and counterbalanced order. Each question will describe the behavior of an unfamiliar child then ask the participant if the trait is associated with that child. For example, "This child can solve math problems really quickly" may be paired with "Is this child smart, not smart, or are you unsure?". Children are gently corrected if they match the trait incorrectly.

Study

The participants would be a group of 96 children that are 5-, 6-, and 7-years-old with an equal amount of boys and girls. The children will be asked to complete a series of three tasks that will be given in random order.

Task (i) is carried out in a similar fashion to the original study. Children would listen to two stories that would also be given in random order: one story is about a smart person while the other is about a nice person. The stories themselves do not indicate the gender of the protagonist. After each story, four photos of women will be laid out in random order. 2 photos will have women with feminine clothing (such as formal dresses or skirts) and 2 photos will show women with masculine clothing (such as ties or suits), and the children would be asked to choose who they believe is the protagonist of each story. A score of 1 is assigned if the child chose options of women in masculine clothing, and a score of 0 if they chose options with women wearing masculine clothing. The photos used in this task, as well as the other two tasks, are of white women, normed for attractiveness ("How attractive does this person look?"), and the clothing is normed for masculinity ("How masculine are the clothing?") and professional appearance ("How formal do the clothes appear?") by a sample of adults through Mechanical Turk.

Task (ii) involves showing 6 photos one at a time. Each picture has two individuals. All trials have pictures that show women, but the first two will wear masculine clothing while the rest wear clothing such as dresses or skirts to indicate femininity. For every photo, the children are told that one of the two individuals were "really, really smart" (for 3 of 6 trials) or "really, really nice" (for the rest of the trials) and asked to choose which of the two they believe have the given trait. Again, a score of 1 would be assigned if the child chose options of women in masculine clothing, and a score of 0 if they chose options with women wearing masculine clothing.

Task (iii) ask children to complete 3 puzzles. Each puzzle will show a 2 row x 4 column sheet. The first row is occupied by 4 pictures, which will be replaced for each puzzle. All four pictures are of women. For one pair of photos with women, both will wear a suit or masculine clothing (camo, rugged clothing). The rest will wear a dress or a skirt to indicate femininity. Each child is given 4 pieces one by one and asked to fill in the spaces below the photos. One piece has the word "smart", one piece has the word "nice", one piece is a picture of a high heel shoe, and one piece is a picture of a hammer. As before, a score of 1 is assigned if the child chose options of women in masculine clothing, and a score of 0 will be given if they chose options with women wearing masculine clothing.

Conclusion

The original study reveals that the stereotype of equating brilliance to males develops at an early age and influences children's interests. To build upon this finding, we proposed an extension that could determine if the same trend in stereotype development can be observed in children when there are only visual masculine cues represented by clothing and not by the presence of men themselves. As with the original study, it would be of further interest to do an additional extension of performing the same studies in different cultural environments and comparing the results (Bian et al. 2017). As with the original study, these extensions have the potential of clarifying the nature of gender stereotypes women encounter that contribute to their avoidance of certain career options.

References

Bian, L., Leslie, S.-J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, *355*(6323), 389–391. doi: 10.1126/science.aah6524

Biernat, Monica. (1991). Gender stereotypes and the relationship between masculinity and femininity: A developmental analysis. *Journal of Personality and Social Psychology,* Vol. 61, Iss. 3: 351-365. doi:10.1037/0022-3514.61.3.351

Furnham, A., Chan, P. S., & Wilson, E. (2013). What to wear? The influence of attire on the perceived professionalism of dentists and lawyers. *Journal of applied social psychology*, *43*(9), 1838–1850. https://doi.org/10.1111/jasp.12136

Furnham, A., & Bunclark, K. (2006). Sex differences in parents' estimations of their own and their children's intelligence. *Intelligence, 34*(1), 1-14. doi:10.1016/j.intell.2005.05.005

Johnson S.K., Murphy S.E., Zewdie S., Reichard R. (2008). The strong, sensitive type: Effects of gender stereotypes and leadership prototypes on the evaluation of male and female leaders. *Organizational Behavior and Human Decision Processes*, Vol. 106, doi: 10.1016/j.obhdp.2007.12.002

Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin*, 137(4), 616–642. Doi: 10.1037/a0023557

Storage D., Horne Z., Cimpian A., Leslie S. (2016). The Frequency of "Brilliant" and "Genius" in Teaching Evaluations Predicts the Representation of Women and African Americans across Fields. *PLoS ONE* 11(3): e0150194. https://doi.org/10.1371/journal.pone.0150194

Tenenbaum, Harriet & Leaper, Campbell. (2003). Parent-Child Conversations about Science: The Socialization of Gender Inequities?. *Developmental psychology.* 39. 34-47. 10.1037/0012-1649.39.1.34.

Leslie SJ., Cimpian A., Meyer M., Freeland E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science* 347(6219):262-5. doi: 10.1126/science.1261375.

Meyer M., Cimpian A., Leslie SJ. (2015). Women are underrepresented in fields where success is believed to require brilliance. *Front Psychol* 6: 235 10.3389/fpsyg.2015.00235

Good C., Rattan A, Dweck CS (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *J Pers and Soc Psychol* 102(4): 700–717.

R References

R: R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

tidyverse: Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

car: John Fox and Sanford Weisberg (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

dplyr: Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2019). dplyr: A Grammar of Data Manipulation. R package version 0.8.3. https://CRAN.R-project.org/package=dplyr

tibble: Kirill Müller and Hadley Wickham (2019). tibble: Simple Data Frames. R package version 2.1.3. https://CRAN.R-project.org/package=tibble

lme4: Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

magrittr: Stefan Milton Bache and Hadley Wickham (2014). magrittr: Forward-Pipe Operator for R. R package version 1.5. https://CRAN.R-project.org/package=magrittr