THE ROOTS OF GENDER GAPS:
INVESTIGATING THE DEVELOPMENT OF GENDER STEREOTYPES ABOUT INTELLIGENCE

BY
LIN BIAN

DISSERTATION
Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology in the Graduate College of the University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:
Professor Renée Baillargeon, Chair
Associate Professor Andrei Cimpian, Director of Research
Professor Dov Cohen
Professor Eva Pomerantz
Professor Sarah-Jane Leslie, Princeton University
Abstract

Women are consistently underrepresented in certain careers—not just in mathematics and science but also in fields such as philosophy and music composition. In my dissertation, I focus on a recently discovered factor that may contribute to this imbalance: Women are culturally stereotyped as being less intellectually gifted than men, and as a result they may face a number of obstacles when pursuing degrees and professions that are portrayed as requiring intellectual giftedness (Leslie, Cimpian, Meyer, & Freeland, 2015). Prior research on this potential mechanism has focused exclusively on adults; thus, little is known about its developmental roots. The earlier children acquire the notion that brilliance is a male quality, the stronger its influence may be on their aspirations. Our present research shows that, by the age of 6, girls are less likely than boys to believe that members of their gender are “really, really smart” – a child-friendly way of referring to brilliance (Studies 1 and 2). Also at 6, girls begin to shy away from novel activities said to be for children who are “really, really smart” (Studies 3 and 4). Moreover, this stereotype undermines both adults’ and children’s evaluation of females’ capabilities for activities portrayed as requiring brilliance (Studies 5 and 6). These studies speak to the early acquisition of cultural ideas about brilliance and gender, and to the immediate impact that these stereotyped notions have on people’s interests and evaluations.
Acknowledgements

Looking back into the years in graduate school, I feel greatly blessed for the love, support and guidance I have received from my mentors, collaborators, parents and friends, without whom this project would not have been nearly successful.

First, I would like to express the deepest appreciation to my advisor, Andrei Cimpian, who has been an extraordinary mentor who gave me persistent guidance during the entire process. Andrei is one of the most brilliant and hardworking people I know. Working with him is inspiring, enjoyable and fruitful, through which I am consistently growing to become a real researcher in the beloved field of developmental psychology. I would also like to express my genuine gratitude to my co-advisor, Renée Baillargeon. I have been very fortunate to work with Renée since the summer of 2012. The joy and enthusiasm she has for baby research was contagious and motivational for me, especially during tough times in the Ph.D. pursuit. Her enormous support for my research projects and long-term career growth has been priceless.

I would like to thank my other committee members and collaborators. Special thanks to Sarah-Jane Leslie, who has always been very warm and supportive. I appreciate of having the chance to collaborate multiple projects with her and get her input on research ideas. I am also grateful to Dov Cohen and Eva Pomerantz, who are tremendously generous with their knowledge and provides insightful feedback on my dissertation. I have greatly benefited from their fresh perspectives on the subject. I would also like to thank Marjorie Rhodes for allowing me the great freedom to explore my research interest, as well as providing many thoughtful ideas and insightful comments during our discussions. I also thank Cindy Fisher, Dan Hyde, Eva Telzer, and Karen Rudolph for their advice, support, and encouragement throughout my time in graduate school.
Special thanks are given to all members in the Cognitive Development Lab (Shelbie Sutherland, Larisa Hussak, Christina Tworek, Daniel Storage, Zachary Horne, Erika Salomon, Melis Muradoglu, Matt Workman, and Rachel Leshin), the Infant Cognition Lab (Jaclyn Aldridge, Danielle Parrish, Melissa Yako, Maayan Stavans, Kyong-sun Jin, Peipei Setoh, Stephanie Sloane, Melody Dawkins, Fransisca Ting and Yi Lin) at the University of Illinois, and the Conceptual Development and Social Cognition Lab (Tara Mandalaywala, Emily Foster-Hanson and Kathryn Yee) at New York University. The groups have been a source of friendships as well as advice and collaboration. It was such a pleasure to spend the past several years working with them.

My time during graduate school was made enjoyable in large part due to the many friends who have been my faithful company all along this voyage. I am especially grateful to Jiangmeng Zhang and Congcong Chen for sharing all the pains and joys with me during these years.

Most of all, I am forever thankful to my parents, Mr. Yibing Bian and Ms. Yuxia Duan. They have been role models for me with their genuine passion and persistent dedication in pursuing their own careers. More importantly, their unconditional love and streaming encouragement has accumulated to be a lasting source of my courage and confidence.

I gratefully acknowledge the funding sources that made this work possible. This project was supported by the Dissertation Completion Fellowship provided by the Graduate College at the University of Illinois, the Robert Larsen Grant for Research in Career Development, and the National Science Foundation. I also would like to give many thanks to the families, children, schools and museums who participated in this research.
# Table of Contents

Chapter 1: Introduction.............................................................................................................1

Chapter 2: The development of the gender stereotype about intelligence.............................17  
  Study 1 ..........................................................................................................................17  
  Study 2 ..........................................................................................................................23

Chapter 3: The effect of the “brilliance ≠ females” stereotype on children’s interest ...............30  
  Study 3 ..........................................................................................................................31  
  Study 4 ..........................................................................................................................34

Chapter 4: The effect of the “brilliance ≠ females” stereotype on adults’ and children’s  
  evaluation of others’ capabilities ..................................................................................38  
  Study 5 ..........................................................................................................................38  
  Study 6 ..........................................................................................................................42

Chapter 5: General Discussion ............................................................................................48

Tables and Figures ...............................................................................................................52

References ............................................................................................................................69
Chapter 1: Introduction

Over the past few decades, there has been enormous interest in exploring the reasons underlying women’s consistent underrepresentation in science, technology, engineering, and math (“STEM” fields; Hill, Corbett, & St Rose, 2010) compared to their greater involvement in non-STEM disciplines, such as social sciences and humanities (e.g., Ceci & Williams, 2007; Halpern et al., 2007). However, a closer inspection of the data on women’s involvement in a variety of disciplines, both STEM and non-STEM, reveals a reality that is more complex than one might expect given the standard focus of the discussion. Although there are indeed fewer women in STEM fields than in non-STEM fields overall, it is also important to note that there is great variability in women’s representation within the two domains. For example, in a recent report by the National Science Foundation (NSF, 2013), about half of the PhD recipients in molecular biology and neuroscience were women, but women were awarded only about 20% of the PhDs in physics and computer science. Similar variation is present in the social sciences and humanities (SocSci/Hum). For example, although women were awarded more than 70% of PhDs in art history and psychology, statistics collected by the U.S. National Center for Educational Statistics showed that women make up only 21% of full-time philosophy faculty (Division APAP, 2011). To explain this more-complex pattern in gender distributions, Leslie, Cimpian, et al. (2015) recently proposed that women might be particularly likely to be underrepresented in fields where success is portrayed as depending on brilliance, which women are perceived as less likely to possess than men (Bennett, 1996, 1997; Kirkcaldy, Noack, Furnham, & Siefen, 2007; Tiedemann, 2000; Upson & Friedman, 2012). Here, I aim to examine this proposal from a developmental perspective, investigating in particular the emergence of the stereotype linking men more than women with intellectual talent. Revealing the development of children’s
gendered ideas about intelligence fills an important gap in the literature. Moreover, these findings could enable researchers and policymakers to devise efficient interventions to reduce the impact of these negative stereotypes on young girls’ career aspirations.

Before describing this research, I will first review two broad perspectives on the current gender imbalances in career participation, particularly in STEM fields. I refer to them as the “biological” and “societal” perspectives.

The Biological and the Societal Perspectives

The biological perspective. The biological perspective makes two critical claims: (1) men and women are inherently different in terms of their cognitive and socioemotional makeup, and (2) these differences influence men’s and women’s aspirations, performance, and career choices, leading to the current gender disparities. Despite consensus on these main claims, researchers propose various hypotheses regarding which specific aspects men and women differ on. Here, I focus on three main hypotheses suggesting that men and women are inherently different with respect to their cognitive abilities (e.g., mathematical and spatial skills; e.g., Halpern et al., 2007; Hyde, Fennema & Lamon, 1990; Kell, Lubinski, Benbow & Steiger, 2013; Wai, Lubinski, & Benbow, 2009), cognitive styles (e.g., Baron-Cohen, 2002; Connellan, Baron-Cohen, Wheelwright, Ba’tki, & Ahluwalia, 2001), and preferred lifestyles (e.g., Ceci & Williams, 2011; Hayes & Bigler, 2013; Lubinski, Benbow, Shea, Eftekhari-Sanjani & Halvorson, 2001).

To illustrate, the “cognitive ability” hypothesis suggests that men are naturally endowed with higher mathematical and spatial abilities than women, especially at the upper end of the distribution (e.g., Halpern et al., 2007; Hyde, Fennema & Lamon, 1990). One piece of evidence supporting this gender difference was found in a meta-analysis involving 3 million subjects,
where Hyde et al. (1990) reported that males were better than females at solving mathematical-reasoning problems in adolescence and adulthood. Moreover, this gender difference was even bigger in high-end samples consisting of talented people, which accords with the argument that men are more variable in a number of traits (e.g., mathematical reasoning abilities) compared to women (e.g., Halpern et al., 2007; Hyde et al., 1990). Since many fields are highly selective, the overrepresentation of males in the right tail of the distribution is regarded as a plausible explanation for the greater number of males than females in these areas.

The “cognitive style” hypothesis argues that men and women naturally differ in at least two dimensions in their cognitive styles, which are “systemizing” and “empathizing” (Baron-Cohen, 2002). In particular, the male brain is better at systemizing, in that men are predisposed to learn about objects and mechanical relationships, have a strong drive to analyze the rules underlying systems they observe in the world, as well as to predict the output of such systems. In contrast, the female brain is better at empathizing, in that women are predisposed to learn about people and their emotions, have a strong drive to understand people’s mental states and thoughts, and are able to respond intuitively to these emotions in an appropriate way. To examine this hypothesis, Connellan et al. (2002) tested whether newborns would display these preferences by showing them an expressive person and a similarly sized inanimate object. Consistent with this hypothesis, male newborns looked longer at the object, whereas female newborns looked longer at the person. Next, to examine how such gender differences relate to people’s tendency to enroll into different disciplines, Billington, Baron-Cohen, and Wheelwright (2007) reported that cognitive style strongly predicts people’s entry into a physical science or humanities subject: Specifically, students with stronger systemizing and weaker empathizing tendencies are more likely to pursue careers in physical science. These findings are used to argue that men and
women may be inherently suited for different types of careers (Baron-Cohen, 2002; Connellan et al., 2002).

The “preferred lifestyle” hypothesis claims that men and women tend to prioritize different aspects of life, in that men often value career over family while women often value family over career. These different work-life priorities direct them to choose different careers, and perhaps that is why women opt out of many fields with heavy workloads, such as STEM (e.g., Ceci & Williams, 2011). As evidence, many more female graduate students than male graduate students expressed the opinion that having a flexible part-time job is important to them (Lubinski et al., 2001). These findings accord with a national survey collected in Britain (Hakim, 2006) and a recent survey conducted in the United States (Hayes & Bigler, 2013), in which women prioritized family over career more than men did. Although researchers sometimes disagree about the exact source of these differential lifestyle preferences, one common view is that the differences exist because women and men are inherently, biologically different (e.g., Ceci, Williams, & Barnett, 2009; Hakim, 2006). Hakim (2006), for example, suggests that the gender differences in work-life orientations are driven by evolutionary forces rather than sociological factors, given the fact that these differences “persist long after the equal opportunities revolution of the 1960s and 1970s gave women equal rights to assess higher education and all positions…” (p. 280).

Although the above-mentioned findings provide some evidence in favor of a biological perspective on women’s underrepresentation in STEM, there are also valid reasons to doubt the two main claims that anchor this perspective. To reiterate, the first claim is that women and men are inherently different on a number of dimensions (cognition, preferences, etc.) that are relevant to their career choices. However, many of these differences are unlikely to be innate, which
opens the door for alternative, sociocultural explanations.¹ For example, Spelke (2005) provides compelling evidence that newborn boys and girls are similarly competent in terms of their reasoning about basic aspects of the world, such as objects and numbers. Additionally, in the comprehensive meta-analysis by Hyde et al. (1990), boys do not outperform girls on math tests until adolescence. Moreover, women’s representation in the STEM domains varies from culture to culture and even from region to region, suggesting that sociocultural factors play a role in the emergence of gender disparities. For example, Pope and Sydnor (2010) investigated the relationship between (1) gender gaps in math performance across various regions of the United States and (2) the extent to which Americans in that region displayed more sexist attitudes (e.g., agreed with the idea that women should take care of the home). As predicted by the sociocultural perspective, they found a positive relation between the two factors: the greater the sexism in a region, the larger the gender gaps between boys’ and girls’ math achievement in that region. Similarly, Guiso, Ferdinando, Sapienza, & Zingales (2008) reported that the more equally women are treated in a country, the smaller the gender gap in math performance, including at the right tail of the distribution. Thus, it is possible that the reported gender differences in mathematical and spatial abilities are a result of differential social input for boys and girls.

The second claim of the biological perspective is that the differences between men and women lead to the current gender gaps in representation. However, recent findings shed doubt on this claim as well. For example, there is no relation between the extent to which a field requires systemizing over empathizing and the male-female PhD ratio in that field (Leslie, Cimpian et al., 2015). Moreover, women are also underrepresented in some disciplines that arguably do not require a high level of mathematical or spatial abilities, such as philosophy and music.

¹ I do not discuss here the literature on the effects of hormones on cognition. This literature is vast, but the results are at best mixed (for reviews, see Ceci et al., 2009; Halpern et al., 2007).
composition (e.g., NSF, 2013). On the basis of results such as these, it seems unlikely that
differences in men’s and women’s natural aptitudes and inclinations can provide a satisfactory
account of the variability in gender gaps across fields.

The societal perspective. According to the societal perspective, people’s beliefs,
performance, and career aspirations are shaped by a range of sociocultural factors, ultimately
leading to gender imbalances in participation (e.g., Bennet, 1996, 1997; Diekman, Brown,
Johnston & Clark, 2010; Guiso et al., 2008; Kirkcaldy, Noack, Furnham, & Siefen, 2007;
Sugimoto, Lariviere, Gingras, & Cronin, 2013; Milkman, Akinola & Chugh, 2012, 2015; Moss-
Racusin et al., 2012; Pope & Sydnor 2010; Sheltzer & Smith, 2014; Tiedemann, 2000; Upson &
Friedman, 2012; Wennerås & Wold, 1997). The social factors concerning women’s participation
can be grouped into two different kinds of stereotypes. The first class of stereotypes concerns the
overall culture of a field, and includes stereotyped beliefs about the characteristics of the field’s
typical members, the typical work environment, and the field’s values (e.g., Cheryan, Master, &
Melzoff, 2015). Women might be less likely to pursue fields whose culture is perceived to be
inconsistent with the image that women are encouraged to adopt by the norms of current society.
For example, people hold the impression that mathematicians and engineers are socially
awkward, that their work is not people-oriented, and that they value “a spark of genius,” which is
incongruent with how many women perceive themselves. Therefore, women may find those
fields less appealing and may be less likely to pursue them as a result. For instance, after
interacting with a stereotypical computer scientist (e.g., a person who wore glasses and a t-shirt
that read “I code therefore I am,” and was a Star Wars fan), women were less interested in
learning more about computer science and less confident that they’d be able to succeed in the
field (Cheryan, Siy, Vichayapai, Drury & Kim, 2011). This reaction was mediated by their
feelings of dissimilarity to the representative of the field and by a lack of sense of belonging. Moreover, this negative effect persisted for up to two weeks after this brief initial exposure (Cheryan, Drury & Vichayapai, 2013).

The second class of stereotypes relevant to gender gaps consists of negative beliefs against women’s abilities (e.g., Ambady, Shih, Kim, & Pittinsky, 2001; Leslie, Cimpian, et al., 2015). These stereotypes are importantly different from the stereotypes about the culture of a field. Whereas the latter contain information about the prototypical member of a field, the former focus on females’ abilities, regardless of field. To illustrate, our culture holds a pervasive, negative stereotype against women’s mathematical abilities (e.g., Ambady et al., 2001; Boucher, Rydell & Murphy, 2015; Kirkcaldy, Noack, Furnham, & Siefen, 2007; Tiedemann, 2000). For example, men typically estimate their own analytical and practical intelligence to be higher than women do (Kirkcaldy et al., 2007), despite the fact that there are no mean-level differences between men and women on these dimensions (Aluja-Fabregat, Colom, Abad, & Juan-Espinosa, 2000; Colom, García, Juan-Espinosa, & Abad, 2002; Saggino et al., 2014). This stereotype also extends into school contexts. Elementary school teachers perceived boys as more capable of logical thinking than girls, even when boys and girls performed equally well at math (Tiedemann, 2000). Additionally, this stereotype influences teachers’ attributions of students’ math performance: They believe it is the extra effort that girls spend on math problems that enables them to achieve a level of math performance comparable to boys’—without this extra effort, they believe that girls would fall behind (e.g., Robinson-Cimpian, Lubienski, Ganley, Copur-Gencturk, 2014). These differential perceptions are likely to influence children’s conceptions of their own mathematics abilities and may eventually steer girls away from participating in STEM fields (e.g., Jacobs, Davis-Kean, Bleeker, Eccles, & Malanchuk, 2005).
Although the existing evidence for the societal perspective helps explain women’s lower involvement in STEM fields, it overlooks the fact that there is great variability in women’s representation within STEM and within the non-STEM domains. This complicated distribution of gender gaps naturally gives rise to the question of whether women’s representation in a particular field might also be influenced by features that cut across the STEM vs. non-STEM divide.

**The Field-specific Ability Belief Model**

In a recent New York Times article titled “Google, Tell Me. Is my Son a Genius?” (Stephens-Davidowitz, 2014), the author tallied anonymous Google searches and found that parents were two and a half times as likely to search “Is my son gifted?” as “Is my daughter gifted?” More generally, parents tended to make more intelligence-related searches about their boys than about their girls. Although it seems like this might be an isolated example, much other research supports the idea that our culture holds a broad stereotype against women: that they are less intellectually gifted than men (e.g., Bennett, 1996, 1997; Kirkcaldy et al., 2007; Tiedemann, 2000; Upson & Friedman, 2012). For example, people tend to underestimate women’s intelligence while overestimating men’s (e.g., Beloff, 1992; Furnham, Reeves, & Budhani, 2002; Rammstedt & Rammsayer, 2000), even though the actual intelligence of men and women is in fact not different (Aluja-Fabregat et al., 2000; Colom et al., 2002; Saggino et al., 2014). Moreover, research focusing on self-estimated intelligence indicates women themselves have lower perceptions of their intelligence than men do. For instance, women have been shown to estimate their IQs lower than men estimate theirs, regardless of their actual intelligence levels (Hamid & Lok, 1995).

Given the fact that women are stereotyped as having less raw intellectual talent than men,
Leslie, Cimpian, and their colleagues (e.g., Leslie, Cimpian et al., 2015; Meyer, Cimpian & Leslie, 2015) recently proposed that the ability beliefs that prevail in each discipline or profession may influence women’s representation. To elaborate, some disciplines are more likely than others to endorse the idea that high-level success is a matter of raw intelligence rather than hard work and dedication. In philosophy, for example, there are widespread messages suggesting that success is largely determined by whether one possesses a spark of genius (e.g., Marshall, 2013). Since women are perceived as being less likely to possess these innate talents, they may be more vulnerable to a field’s belief suggesting brilliance is the key to success. The stereotypes against their intellectual abilities may lead women in brilliance-focused fields to feel a lower sense of belonging, which may affect their motivation as well (e.g., Good, Rattan, & Dweck, 2012). The “brilliance ≠ females” stereotype may also make women the targets of bias from current members of brilliance-focused fields, which may in turn create an inhospitable environment for them. As a result, women might become less likely to pursue these areas in the long term. To summarize, this account consists of two components: 1) the field-specific ability belief component, which emphasizes the variability that exists between fields in their beliefs about the necessity of innate talent for success, and 2) the widespread societal stereotype against women’s intelligence. I will refer to this hypothesis as the Field-specific Ability Belief (“FAB”) hypothesis.

Leslie, Cimpian, and their colleagues (2015) found initial evidence of this hypothesis in a nationwide study of academics. In this study, they recruited 1820 faculty, post-doctoral fellows, and graduate students from different research universities across the United States. Participants were asked to report what they believe is necessary to achieve success in their own field. Consistent with the FAB account, the more a field emphasized brilliance, the lower women’s
representation was at the PhD level in this field. Moreover, ability beliefs predicted women’s representation above and beyond several alternative explanations, such as differences among fields in work-life balance or the extent to which they focus on people vs. objects (e.g., Ceci, Ginther, Kahn, & Williams, 2014; Diekman et al., 2010; Ferriman, Lubinski, & Benbow, 2009). Experimental studies provide converging evidence for this hypothesis (e.g., Bian, Cimpian, Leslie, & Murphy, under review; Emerson & Murphy, 2015; Good, Rattan, & Dweck, 2012; Smith, Lewis, Hawthorne, & Hodges, 2013). For example, in a series of experiments by Bian et al. (under review), college students and Mechanical Turk workers were provided with a range of hypothetical educational and professional opportunities (e.g., major, internship, job) that were portrayed as requiring either brilliance or dedication, depending on the condition. Some experiments also included baseline conditions in which no additional information was provided about these opportunities. Then, participants’ motivation to pursue these opportunities was measured. Women (but not men) reported lower motivation towards the activities said to be for people of high intellectual ability, indicating that the messages emphasizing the key role of innate talents in achieving success undermine women’s motivation and interest.

These results illustrate the power of the FAB model to account for several features of women’s involvement across the spectrum of professions. First, the model provides new insights into women’s underrepresentation in STEM, where beliefs about brilliance were commonly endorsed. Second, the model suggests that field-by-field variability in ability beliefs may explain, at least in part, why women have made more inroads in some STEM fields (e.g., molecular biology) than others (e.g., physics; NSF, 2013). Third, the FAB model suggests an explanation for the variability in female representation within the social sciences and humanities. Although discussions of gender in academia often overlook these fields (but see Haslanger,
2008), the social sciences and humanities in fact exhibit at least as much variability in participation by women as STEM fields do (NSF, 2013). Again, this variability in representation closely tracks the inter-field variability in ability beliefs. In sum, the FAB model provides a potentially important lens through which to view the gender imbalances that are still so common in academia and the workplace.

Despite the robust evidence for the FAB hypothesis, a key question remains unanswered: How might the “brilliance ≠ females” stereotypes interact with a field’s ability messages to discourage women’s participation? Discussing the mechanisms involved will also lay the foundation for the present developmental investigation of the FAB proposal. There are at least two plausible pathways: The stereotypes and ability beliefs might jointly (1) undermine women’s motivation to pursue fields that emphasize brilliance and (2) give rise to bias in evaluating women’s competence. I will present some evidence for each of these mechanisms and then describe how I investigated them in young children in the present studies.

**Potential Mechanisms**

First, the “brilliance ≠ females” stereotype might affect women’s self-concept, leading them to doubt their ability to succeed in professions that are portrayed as requiring such brilliance (e.g., Bennett, 1996, 1997; Bian et al., under review; Gunderson, Ramirez, Levine, & Beilock, 2012). This lowered self-estimation may decrease their sense of belonging and increase their anticipated anxiety in these fields, which in turn is likely to lower their motivation to pursue careers in them (e.g., Bian et al., under review; Emerson & Murphy, 2015; Good, Rattan & Dweck, 2012; Murphy & Dweck, 2010; Smith et al., 2013). For instance, Good et al. (2012) found that when female students perceived their classmates as (1) holding a stereotype against women’s math abilities and also (2) viewing math ability as a fixed trait, they were less likely to
feel they belonged in math. Their lower sense of belonging in turn predicted less interest in pursuing math in the future. This effect extends to business settings as well. Emerson and Murphy (2015) found that, after receiving negative feedback, women were more likely to disengage from a company where a high level of intelligence was valued than from a company where sustained effort was valued. Additional analyses revealed that women’s lack of interest in the intelligence-focused company was a result of their lower sense of belonging, which was negatively affected by their expectations of being stereotyped in this environment.

A second potential mechanism is the following: The two elements of the FAB model (i.e., the “brilliance ≠ females” stereotype and the field-specific ability beliefs) may create a more unwelcoming environment that hinders women’s participation. Specifically, given the stereotypes against women’s intellectual abilities, members of brilliance-focused fields may exhibit bias against them, providing them with fewer opportunities, lower salaries, and fewer accolades (e.g., Sugimoto et al., 2013; Milkman et al., 2012, 2015; Moss-Rausin, et al., 2012; Sheltzer & Smith, 2014; Wennerås & Wold, 1997; but see Williams & Ceci, 2015, for inconsistent results). For instance, before entering academia, women are less likely to get responses from faculty when they contact them to discuss research opportunities (Milkman et al., 2012, 2015). Speaking to the gender bias in opportunities and salary, when faculty members in biological and physical sciences were asked to evaluate the suitability of a male or a female applicant with identical backgrounds for a lab manager position, they rated the male as more suited for the position, were more likely to offer the male mentoring, and provided him a higher starting salary (e.g., Moss-Racusin et al., 2012). The female was seen as less suited for the position, even though she was exactly as qualified as the male applicant. Even after women surmount these challenges and obtain advanced degrees in brilliance-focused fields, they may
still face discrimination with respect to their academic productivity and the value of their research. For instance, Sugimoto et al. (2013) reported that articles are cited fewer times when women are in the most prominent author positions than when men are. Moreover, Wennerås and Wold (1997) analyzed the data from the peer review system of the Swedish Medical Research Council and found that females were less likely to be awarded postdoctoral fellowships, and were perceived as less competent than males who were, in fact, equally productive.

**The Present Research**

Although the FAB hypothesis provides a promising account of the current gender disparities across fields, investigations of this hypothesis that focus exclusively on adult participants are likely to miss a crucial piece of the puzzle: Cultural messages about the presumed cognitive profiles of males and females are likely to be available and influential throughout development. If children absorb and act on these ideas, then—by the time they reach adulthood—many capable girls are likely to have already veered away from fields that are believed to require intellectual talent. Thus, investigating the acquisition and the impact of the “brilliance ≠ females” stereotype on children’s development is likely to enhance our understanding of the root causes of gender gaps across academia and the industry, as well as our ability to intervene effectively at the point where the problem originates.

Therefore, in the present research, I investigate (1) the development of this negative stereotype against women’s intelligence (Chapter 2), (2) how this stereotype affects children’s motivation, especially concerning activities portrayed as requiring high levels of intellectual ability (Chapter 3), and (3) how this stereotype affects adults’ and children’s evaluation of males’ and females’ capabilities (Chapter 4). Before discussing these studies, I review the prior
literature on the development of the gender stereotypes about males’ and females’ cognitive abilities.

Although no previous studies have directly investigated the development of this broad gender stereotype about intelligence, the existing developmental literature suggests that children pick up on gender stereotypes about toy preferences even as early as age 2 or 3 (e.g., Ambady et al., 2001; Cvencek, Meltzoff, & Greenwald, 2011; Liben, Bigler, & Krogh, 2001, 2002; Parsons, Adler & Kaczala, 1982; Signorella, Bigler, & Liben, 1993). When they reach elementary school age, children also begin to acquire stereotypes about dimensions of intellectual competence such as mathematics and reading. For example, a recent study by Cvencek et al. (2011) suggested that girls have learned and internalized the stereotype against their mathematics abilities by second grade (see also Ambady et al., 2001). More strikingly, one study showed that six-year-old girls’ math performance decreased after being incidentally exposed to the gender stereotype about math (Galdi, Cadinu & Tomasetto, 2014). After coloring a picture of a boy successfully solving a math problem and a girl failing to do so, young girls performed worse than boys on a math test. These results indicate an influence of the negative stereotype targeting girls’ math ability, which was activated by the picture (even though presumably the girls themselves were not consciously aware of its activation). Other studies similarly showed that 5- to 7-year-old girls’ math performance decreases when their gender identity is activated (e.g., Neuville & Croizet, 2007; Tomasetto, Alparone, & Cadinu, 2011). An interesting commonality between the studies just described is this sensitive age range, from 5 to 7 years of age, during which the effects of gender stereotypes about the intellectual domain (in particular, mathematics) seem to first surface in children’s development and begin to affect their behaviors.
Importantly, however, no studies we know of have investigated the development of “brilliance ≠ females” stereotype, which is broader and potentially much more pernicious than the stereotype about mathematics ability. Moreover, how this stereotype interacts with brilliance-focused messages, which are the other key component of the FAB framework, to affect children’s reactions (e.g., motivation) and to create obstacles for them (e.g., bias from others) is also unclear. In order to better understand these issues, I conducted six studies that investigate the development and consequences of gender stereotypes about intelligence.

Overview of the Studies

In Chapter 2, I conducted two studies to investigate the development of the stereotypes against women’s intelligence. For this purpose, I created several implicit tasks tapping into children’s beliefs about which gender is smarter. These tasks were administered to children aged 5 to 7 to chart the development of their awareness of the association between “brilliance” and males.

In Chapters 3 and 4, I explored the possible mechanisms through which this stereotype might influence children’s interests. In particular, Chapter 3 investigated the development of children’s motivation toward novel activities (e.g., games) that are portrayed as requiring high levels of intellectual ability. For this purpose, I created two novel games that were portrayed as requiring either “brilliance” or “dedication” (counterbalanced across children). I then assessed children’s motivation towards the two games and their endorsement of the “brilliance ≠ females” stereotypes. The FAB model predicts that girls should begin to show less enthusiasm than boys towards the “brilliance” activities at around the same age when they show evidence of having learned the “brilliance ≠ females” stereotype (see Chapter 2). At a more fine-grained level, we also predict that the precise extent to which girls endorse the “brilliance ≠ females” stereotype
(as measured in our study) should be negatively related to their motivation towards activities portrayed as requiring intellectual ability. More generally, children who hold more negative stereotypes against their own gender’s intelligence should be less motivated to pursue these activities.

Chapter 4 examined the consequences of the “brilliance ≠ females” stereotype for adults’ and young children’s explicit evaluation of others’ competence. In Study 5, adult participants were asked to nominate two acquaintances for a job position that was described as requiring either brilliance or sustained effort. In Study 6, I adapted this procedure for use with 5- to 7-year-olds. In particular, I created two novel games with different descriptions depending on the condition children were randomly assigned to: In one condition, children were told that these games are for children who are “really, really smart,” whereas in the other condition (which served as a baseline) children heard no information about who the game is for. Children were then asked to select teammates with whom to play these games from among a set of unfamiliar children (half boys and half girls). The FAB model predicts that adults and 6- to 7-year-old children (who have shown some evidence of having learned the “brilliance ≠ females” stereotype) would show a bias in favor of males in the conditions where brilliance is emphasized.
Chapter 2: The development of the gender stereotype about intelligence

In this chapter, I describe two studies conducted to investigate the development of the stereotype against women’s intelligence in 5- to 7-year-olds. I chose this age range because (1) before 5, children may have difficulty grasping concepts that are as complex as intellectual capacity (e.g., Nicholls, 1978), and (b) by age 7, children generally acquire most of their culture’s gender stereotypes (e.g., Signorella et al., 1993). In addition, this wide age range allowed me to test when this stereotype originates and chart its development. In Study 1, I examined the extent to which young children associate intelligence with (adult) males. In Study 2, I explored the generality of these results by examining whether this stereotype also influences children’s thinking about young boys’ and girls’ intelligence, not just adult men’s and women’s.

Study 1

Participants

Participants were 32 five-year olds (\(M_{\text{age}} = 5.55\) years, \(SD = 0.30\)), 32 six-year olds (\(M_{\text{age}} = 6.50\) years, \(SD = 0.31\)) and 32 seven-year olds (\(M_{\text{age}} = 7.46\) years, \(SD = 0.33\)). Half of them were boys and half of them were girls. Children were recruited in a small Midwestern city. Twenty-three additional children were tested but excluded from the final sample because of poor performance on a set of screener questions (\(n = 19\), see below), because they refused to finish the study (\(n = 3\)), or because they were more than 2.5 standard deviations from the condition mean (\(n = 1\))^2.

Across Studies 1 to 4, demographic information was available for 75% of the families. The racial/ethnic composition of the sample mirrored that of the community in which this research was conducted: 78% of the children were European American, 7% Asian American, 5%...

---

^2 We used a uniform 2.5 SD outlier exclusion criterion across studies. The results remained the same without excluding the outliers.
African American, 3% Latino or Hispanic, and 7% multi-racial. The median household income was $90,000 (range = $5,000 - $185,000). Eight-two percent of the parents in the sample had at least a bachelor’s degree. Neither race nor socio-economic status (SES) moderated the reported effect\(^3\) (for more details, see Page 32).

**Materials and procedure**

Children were tested individually in a quiet room in our lab or at their school by a single experimenter. The experimenter videotaped the child sessions and recorded the children’s responses on an answer form. The experiment consisted of two parts: (1) a screener phase where I gauged the child’s understanding of the relevant concepts (e.g., intelligence), and (2) three stereotype tasks to assess children’s endorsement of the “brilliance ≠ females” beliefs.

**Screener phase.** The experimenter began the study by telling the child that they would talk about what “smart” and “nice” (a control trait) mean. “Nice” was chosen as a comparison attribute for “smart” because it is as familiar to children but likely not stereotyped in the same way in our society (Eagly & Mladinic, 1994; Fiske, Cuddy, Glick & Xu, 2002). Then, the experimenter said, “I’m going to tell you about some children I know and ask you if you think they’re smart/nice”. Children were provided with twelve screener questions, half of them gauging their understanding of “smart” and half of them gauging their understanding of “nice.” The “smart” and “nice” questions were presented to children as separate blocks whose order was counterbalanced. Each of these questions was about a behavior of an unfamiliar child accompanied by the child’s picture. The experimenter read off descriptions of each pictured child while putting the picture itself behind a little tent so that the subject could not see the gender of the child presented in the picture. We concealed the picture so that children’s answers to these

---

\(^3\) Because the racial/ethnic minority group in our samples was small, we might not have enough power to detect the moderation effect. Thus, the moderation results reported in Studies 1-4 and 6 should be interpreted with caution.
questions would not be influenced by the gender of the child in the pictures, and thus these answers would not interfere with the subsequent main tasks, where we measured children’s gender stereotypes. Next, children were asked to determine if the described child had the target trait (smart vs. nice). For the “smart” screener questions, the experimenter asked the child if they thought the child in the picture was “smart, not smart, or are you not sure?” by using a scale showing a thumbs up, a thumbs down, and a puzzled look. Four descriptions fit the definition of the word “smart” (e.g., this child learns things really fast), and the other two served as fillers (e.g., this child watches really funny cartoons; see Table 1). For the “nice” screener questions, the same scale was used but the question was changed to test if the subject thought the described child was “nice, not nice, or are you not sure?” Again, there were four items that fit the definition of “nice” (e.g., this child likes to help other people), and two fillers (e.g., this child plays on the swings; see Table 1). These answers were used to determine if children understand the meaning of the key terms “smart” and “nice.” Children who did not answer correctly at least two thirds of the questions either for the “smart” or for the “nice” screener questions were excluded from the analysis ($n = 19$).

**Stereotype tasks.** After completing the two screener blocks, the experimenter administered three stereotype tasks that tapped children’s implicit stereotypes about males and females. These tasks were presented in a random order.

**Gender-neutral story task.** This task consisted of two stories, each of which described a “special person” while leaving this person’s gender unspecified (see Table 2 for full text). One story was about a “really, really smart” person at the experimenter’s workplace who “comes up with answers much faster and better than anyone else.” The other one was about a “really, really nice” person at the experimenter’s workplace who “likes to help others with their problems and
is friendly to everyone at the office.” No clues as to the protagonist’s gender were provided in either story. After telling the story, the experimenter laid out four pictures in a line—two of which were adult females, and two of which were adult males. The gender of the depicted individuals alternated across the four pictures. These pictures, as well as those used in the other tasks and in Study 2, were normed on a separate group of adult participants (N = 29) such that there were no significant differences between the males and the females in terms of their attractiveness or other superficial features (e.g., how professionally they were dressed). Additionally, great care was taken to match facial expressions, styles of clothing, etc., across the pictures of males and females. All pictures depicted white individuals, as the strength of gender stereotypes about intelligence may differ depending on the race of the targets (e.g., Ambady et al., 2001; Rowley, Kurtz-Costes, Mistry & Feagans, 2007).

After hearing the experimenter’s story, children were asked to guess which one of the four people might be the person in the story. If children chose a person of the same gender as themselves (e.g., if a girl picked a woman), they were assigned a score of 1 for that trial; otherwise, they were assigned a 0.

**Guessing task.** Children were asked to play a guessing game in which they were shown six pairs of individuals and asked to guess which one had more of a certain trait. In the first two trials, the two individuals were of the same gender (matching the participant’s own gender). These two trials served as practice for the children, and, perhaps more importantly, they also helped camouflage the purpose of our study. For the next four trials, children were shown a man and a woman.

On each of the six trials, children were told that one of the two people was “really, really smart” (three trials) or “really, really nice” (the other three trials), and they were asked to guess
which of the two people in the picture is this target person. The order of the pictures was counterbalanced to avoid confounding trait content with particular pictures. Similar to the first task, children’s responses were scored as a 1 on a trial if they chose the person of the same gender as themselves and 0 otherwise.

**Puzzle task.** Children were asked to complete a series of three “matching puzzles.” Each puzzle consisted of a two by four table (two rows \( \times \) four columns), with the top row of the table filled with pictures of two females and two males, in alternation. The pictures of individuals differed from puzzle to puzzle. However, the four pieces the child was asked to place in the bottom row of the table remained the same: one piece had the word “smart” on it, one piece had the word “nice,” one piece had a picture of a high heel (stereotypically feminine), and the fourth had a picture of a hammer (stereotypically masculine). Children were given the pieces one by one (and told the word on them, for the “smart” and “nice” puzzle pieces) and were then asked to put these pieces in one of the empty slots on the bottom of the puzzle so as to “match” the pictures of the men and women at the top. Children’s answers were scored as a 1 if they matched the piece showing “smart” or “nice” with a person of the same gender as themselves, and 0 otherwise.

At the end of the sessions, children were thanked for their participation and praised for their responses. They also received a small reward for participating in the study.

**Results and Discussion**

Our key prediction was of a three-way interaction between Trait, Gender, and Age: For the trait “smart,” we expected that both boys and girls would favor their own gender at the age of 5, which accords with many prior studies showing that young children believe their gender group is the best on most dimensions (e.g., Ambady et al., 2001; Dunham, Baron, & Banaji, 2011). As
children get older, however, girls may become less likely than boys to attribute “being really, really smart” to their own gender, leading to a significant Gender × Age interaction for this attribute. I did not expect the same pattern to emerge for trait “nice,” hence the prediction of a Gender × Age × Trait interaction.

We submitted children’s stereotype scores to a multilevel mixed-effects linear model with trait (smart vs. nice; level-1 predictor), gender (boys vs. girls; level-2 predictor), and age (5- vs. 6- vs. 7-year-olds; level-2 predictor), plus all possible interaction terms, as categorical fixed effects and a random intercept for participants. The models were computed with the *mixed* command in Stata 14.1.

The analyses revealed a significant interaction between Gender and Trait, Wald $\chi^2 = 15.09, p < .001$. Two follow-up tests indicated that boys were overall more likely than girls to associate “smart” with their own gender, Wald $\chi^2 = 9.51, p = .002$, but boys were less likely than girls to associate “nice” with their own gender, Wald $\chi^2 = 5.81, p = .016$.

Nevertheless, our main goal was to examine the *development* of children’s beliefs on these matters. As predicted, we found a significant three-way interaction between Age, Gender, and Trait, Wald $\chi^2 = 10.56, p = .005$. Four follow-up tests compared boys’ and girls’ stereotype scores about each trait separately for younger (5-year-old) and older (6- and 7-year-old) children, which revealed that at 5, girls and boys were not different in their tendency to regard their own gender as smarter (Wald $\chi^2 = 0.02, p = .893$) or nicer (Wald $\chi^2 = 0.39, p = .531$) than the other gender (see Figure 1A and Table 3). These results are consistent with the overwhelming in-group positivity previously observed in boys and (especially) girls across early and middle childhood (Ambady et al., 2001; Dunham et al., 2011). Despite this strong tendency to view one’s gender in a positive light, 6- and 7-year-old girls were less likely to associate “smart” with their own
gender than boys were, Wald $\chi^2 = 8.10$, $p = .004$ (see Figure 1A and Table 3). With respect to the attribute “nice,” boys at ages 6 and 7 were less likely to associate it with their own gender than girls, Wald $\chi^2 = 17.70$, $p < .001$, (see Figure 1B and Table 3). These developmental patterns were found consistently across the three stereotype tasks.

These results suggest that our culture’s gender stereotypes about intelligence may begin to influence children’s beliefs about which gender is “really, really smart” at a very young age—perhaps as early as age six. In addition, the stereotype associating females with being nice seems to follow a similar developmental trajectory.

**Study 2**

The primary goal of Study 2 was to examine whether this “brilliance ≠ females” stereotype also influences children’s thinking about young boys’ and girls’ intelligence, not just men’s and women’s. A second goal of Experiment 2 was to explore the potential factors that might lead to gender differences in children’s beliefs about who is “really smart.” One plausible influence on children’s ideas about intelligence is their perceptions regarding who gets better grades in school. Thus, I examined whether children’s perceptions of gender differences in school achievement were related to their endorsement of the “brilliance ≠ females” stereotype.

**Participants**

Participants were 48 five-year olds ($M_{age} = 5.50$ years, $SD = 0.33$), 48 six-year olds ($M_{age} = 6.48$ years, $SD = 0.25$) and 48 seven-year olds ($M_{age} = 7.45$ years, $SD = 0.28$). Half of them were boys and half of them were girls. Children were recruited in a small Midwestern city. Thirty-one additional children were tested but excluded from the final sample because they failed the screener questions ($n = 28$), refused to complete the study ($n = 2$) or were more than 2.5 standard deviations from the condition mean ($n = 1$).
Materials and procedure

The procedure of Study 2 was identical to that of Study 1, with three exceptions. First, each task was separated into two blocks: one with pictures of men and women (identical to those in Study 1), and the other with pictures of boys and girls. The order of the two blocks was counterbalanced. The pictures of boys and girls used in this study were normed for attractiveness (“How attractive does this child look?”) and age (“How old do you think this child is?”) in the same sample of 29 Mechanical Turk adults that rated the pictures used in Study 1. Second, because the addition of the child targets increased the length of the sessions, we omitted the puzzle task to avoid taxing children’s attention spans. Third, at the end of the sessions, we assessed children’s perceptions of boys’ and girls’ school achievement. Children first saw 4 pictures of unfamiliar children (2 boys and 2 girls) and were asked, “If you had to make a guess, who do you think gets the best grades in school?” With another set of 4 pictures, they were then asked, “If you had to make a guess, who do you think is first in their class?” Finally, participants were asked the same 2 questions again, except this time they had to choose between 2 verbally-presented options: “boys or girls?” Responses across these 4 items were coded as in the stereotype tasks (same-gender choice = 1; other-gender choice = 0) and averaged.

Results and discussion

We predicted that the results would parallel those in Study 1. We did not have strong a priori predictions regarding whether children’s gender stereotypes about intelligence would depend on the age of the target (child vs. adult).

We first submitted children’s stereotype scores to a multilevel mixed-effects linear model with trait (smart vs. nice; level-1 predictor), target (adults vs. children; level-1 predictor), gender (boys vs. girls; level-2 predictor), and age (younger vs. older children; level-2 predictor), plus all
possible interaction terms, as categorical fixed effects and a random intercept for participants. The models were computed with the *mixed* command in Stata 14.1.

The results revealed a significant interaction between Gender and Trait, $\chi^2 = 45.87$, $p < .001$. Two follow-up tests indicated that, as in Study 1, boys were overall more likely than girls to associate “smart” with their own gender, $\chi^2 = 6.21$, $p = .013$, but they were less likely than girls to associate “nice” with their own gender, $\chi^2 = 31.71$, $p < .001$.

The main goal was, again, to examine the developmental timeline of children’s association between brilliance and males, as well as whether this association is moderated by the age of the target (children vs. adults). As predicted, we found a significant three-way interaction between Age, Gender, and Trait, $\chi^2 = 7.73$, $p = .021$. As illustrated in Figure 1C, follow-up tests indicated that at the age of 5, girls and boys were not different in their tendency to regard their own gender group as smarter than the other gender group, $\chi^2 = 0.01$, $p = .936$. However, 6- and 7-year-old girls were less likely to associate “smart” with their own gender than boys, $\chi^2 = 9.63$, $p = .002$ (see Figure 1C and Table 3). This pattern did not differ significantly by whether children rated adult vs. child targets, $\chi^2 = 1.53$, $p = .46$. With respect to the attribute “nice,” both younger ($\chi^2 = 4.31$, $p = .038$) and older boys ($\chi^2 = 29.35$, $p < .001$) were less likely than girls to associate it with their own gender (see Figure 1D and Table 3). Again, these developmental patterns were found consistently across the two stereotype tasks.

What might explain the drop in girls’ evaluation of their gender’s intellectual abilities? Although many factors are likely involved, in Study 2 we tested whether this drop is associated with differences between younger (5-year-old) and older (6- and 7-year-old) girls in their perceptions of their school achievement—information that is in principle relevant to judging
intelligence. In contrast with the drop in brilliance scores, there was no significant difference between younger and older girls in the likelihood of selecting same-gender children (i.e., other girls) as having top grades, $t(70) = 0.22, p = .826$ (see Figure 2). Older girls were actually more likely to select same-gender children as having top grades than older boys were, $t(94) = 4.41, p < .001$, consistent with the reality that girls get better grades in school than boys at this age (Voyer & Voyer, 2014). Nevertheless, there was no significant correlation between girls’ perceptions of school achievement and their perceptions of brilliance, $r(72) = .11, p = .34$ (for boys: $r(72) = .38, p = .001$). Thus, girls’ ideas about who is brilliant are not rooted in their perceptions of who performs well in school. However, other aspects of children’s experiences in school, such as teachers’ attitudes and biases (e.g., Beilock, Gunderson, Ramirez, & Levine, 2010; Robinson-Cimpian et al., 2014) may still be implicated in the development of this stereotype.

**The distributions of children at each age who associated males with brilliance.** To characterize these distributions, we combined the data from Studies 1 and 2, and categorized children into three groups based on the extent to which they favored their own gender as brilliant across the 6 responses they had to make (2 responses per task × 3 tasks): Children were classified into the “favoring their own gender” category if they selected members of their own gender as brilliant on at least 4/6 trials, into the “favoring the other gender” category if they selected members of the other gender as brilliant on at least 4/6 trials, and as “neutral” otherwise. Examination of distribution of individual children who fell into these three categories at age 5, 6, and 7 reinforces the conclusions of the analyses reported above (see Table 4). For example, 16 out of 40 girls aged 6 were categorized as “favoring the other gender” compared to only 2 out of 40 girls aged 5, $\chi^2(1, N = 80) = 14.67, p = .001$. However, the number of boys in the “favoring the other gender” category was similar (and low) across these two ages (5-year-olds: 8/40; 6-
year-olds: \(7/40\), \(\chi^2(1, N = 80) = 0.09, p = .958\). Similar patterns were found when comparing 5-year-olds’ responses with 7-year-olds’. In summary, these distributional analyses provided converging evidence for the developmental trajectories.

**Moderation analyses with race/ethnicity and socioeconomic status (SES).** Do children from different racial/ethnic backgrounds have different beliefs about which gender is “really, really smart” and which is “really, really nice”? Because each of the racial/ethnic minority groups in our sample (e.g., African Americans, Latinos) was small, we combined them into a single group for purposes of this analysis. To increase the power to detect differences by race/ethnicity, we also pooled the data from Studies 1 and 2. We then submitted these data to a multilevel mixed-effects linear model with race/ethnicity (white children vs. children of color; level-2 predictor), trait (smart vs. nice; level-1 predictor), gender (boys vs. girls; level-2 predictor), and age (5- [younger] vs. 6- and 7-year-olds [older]; level-2 predictor), plus all possible interaction terms, as categorical fixed effects and a random intercept for participants. Contrary to the idea that the development of the stereotypes investigated here varies by racial/ethnic group, we found that race/ethnicity did not significantly moderate the key three-way interaction among Trait, Gender, and Age, Wald \(\chi^2 = 0.15, p = .70\). Inspection of the means revealed broadly similar developmental patterns. For example, own-gender brilliance scores decreased with age for both white girls (\(M_{younger} = .67\) vs. \(M_{older} = .49\)) and girls of color (\(M_{younger} = .83\) vs. \(M_{older} = .60\)).

Do children from high- vs. low-SES backgrounds have different beliefs about which gender is “really, really smart” and which is “really, really nice”? To examine this question, we first created a composite SES measure by (1) standardizing the average education level of the parent(s) (which had been converted to years of education prior to standardizing) and the total
income of the household, and then (2) averaging these two scores (education and income) into a composite SES variable. Next, we performed the same multilevel analysis as above, except that race/ethnicity was replaced by SES. Again, we found that SES did not significantly moderate the key three-way interaction among trait, gender, and age, Wald $\chi^2 = 0.58, p = .45$. Inspection of the means revealed similar developmental trends for high- and low-SES children. For example, both high-SES (+1 SD) girls and low-SES (−1 SD) girls showed age-related drops in their own-gender brilliance scores (high-SES girls: $M_{younger} = .71$ vs. $M_{older} = .49$; low-SES girls: $M_{younger} = .71$ vs. $M_{older} = .53$).

**Conclusion.** In conclusion, the first two studies provide consistent evidence suggesting that, starting at around the age of 6, many children already associate brilliance with males. Consistent with previous research on stereotypes about mathematics, our findings revealed that girls’ tendency to associate intelligence with their own gender begins to decrease when they are first exposed to school, whereas boys perceive their own gender as smarter across this developmental milestone. One possible explanation for this divergence may be that school entry brings with it greater exposure to cultural messages about the greater intellectual abilities of males. For example, parents predict that their sons’ IQ is higher than their daughters’, regardless of which child’s IQ is actually higher (Neto & Furnham, 2011; Kirkcaldy, 2007). These beliefs might be transferred to children via parent-child daily interactions about competence and achievement (Cimpian, Arce, Markman, & Dweck, 2007; Pomerantz & Kempner, 2013), which may become more frequent when children enter school. Similarly, young girls face negative stereotypes against their intelligence in schools as well, since teachers have the same perceptions as parents—that a boy’s IQ and natural intelligence surpasses that of an equally-achieving girl. Moreover, in the media, most “brilliant” professionals are still portrayed by men, whereas
women are more likely to portrayed in roles that highlight their nurturing traits (e.g., nurses, babysitters) (e.g., Paek, Nelson, & Vilela, 2010). Interestingly, our findings revealed that girls’ ideas about intelligence are not related to their perceptions of school achievement. Although young girls perceive their gender as doing better in school, this perception has no bearing on who they perceive to be “really, really smart.” Further research is needed to examine the stark discrepancy between the two types of perceptions.

Given our findings so far, the period before the age of 6 might be a sensitive stage during which interventions might be most effective in buffering children against the potential negative effects of these stereotypes. By addressing this problem at its developmental root, we may have a better chance of putting young men and women on equal playing fields with regard to how they perceive their own intellectual abilities. By doing so, we might also reduce the gender gaps that persist in many fields and careers.
Chapter 3: The effect of the “brilliance ≠ females” stereotype on children’s interest

In this chapter, I investigate how the “brilliance ≠ females” stereotype influences children’s motivation toward novel activities (e.g., games) that are portrayed as requiring high levels of intellectual ability. The FAB model makes two predictions: First, given the results above, brilliance-focused messages should undermine 6- and 7-year-old girls’ (but not 5-year-old girls’) motivation, just as they undermine the motivation of women in our adult studies (Bian et al., under review). Second, children’s endorsement of the “brilliance ≠ females” stereotype should predict their motivation towards activities portrayed as requiring brilliance: Children who hold more negative stereotypes against their own gender’s intelligence should be less motivated to pursue these activities.

These predictions are reinforced by two lines of work that pertain to gender development: First, the literature on the development of gender identity suggests that, from a young age, children are sensitive to the “typical” characteristics of their gender group and conform to these characteristics (for reviews, see Martin & Ruble, 2004; Martin, Ruble, & Szkrybalo, 2002; Rutland, Killen, & Abrams, 2010). Thus, rather than being suspicious of implicit suggestions that girls are “hard workers” and boys are “the smart ones,” young girls may in fact willingly incorporate such stereotyped beliefs into their self-concepts. Second, there is considerable evidence that gender stereotypes shape children’s behaviors, including their activity choices (e.g., Ambady et al., 2001; Cimpian, Mu, & Erickson, 2012; Martin, Eisenbud, & Rose, 1995; McKown & Weinstein, 2003; Neuville & Croizet, 2007; Wigfield & Eccles, 2000). In other words, children may act on the basis of the stereotypes they learn from those around them and from the broader cultural messages to which they are exposed.
Study 3

In this study, 6- to 7-year-old children were introduced to two novel games, one said to be for “children who are really, really smart” and the other for “children who try really, really hard.” Then, I measured children’s motivation toward each game. At the end, children received a modified version of the gender-neutral story task used in Studies 1 and 2 to assess their gender stereotypes about intelligence.

Participants

Participants were 32 six-year olds ($M_{age} = 6.50$ years, $SD = 0.31$) and 32 seven-year olds ($M_{age} = 7.44$ years, $SD = 0.33$). Half of them were boys and half of them were girls. Children were recruited in a small Midwestern city. One additional children were tested but excluded from the final sample because they were more than 2.5 standard deviations from the condition mean.

Materials and procedure

Children were introduced to two novel games (“zarky” and “impok”) in counterbalanced order. For each game, the experimenter showed children a picture of it and briefly described its “rules” (see Table 5). Crucially, one game was said to be for “children who are really, really smart,” and the other was said to be for “children who try really, really hard.” As a comparison/control for “brilliance,” I chose “dedication” because (1) effort is an important ingredient for success, and (2) this dimension is not stereotypically associated with males more than females (e.g., Smith et al., 2013; Tiedemann, 2000), so we can expect it to behave differently. Each game was presented in “smart” format for half of the participants and in “try hard” format for the other half. To ensure that children encoded the crucial “smart” vs. “try hard” information about each game, the experimenter asked them to recall it before proceeding to the main set of questions and corrected them if necessary.
Next, the experimenter asked 4 questions designed to gauge children’s interest in the game (e.g., “Would you want to play the zarky/impok game, or would you not want to play it?”; see Table 6). The order of the questions was randomized. After the first two questions, children were provided with a reminder of the relevant ability information (i.e., that the game is for children who are “really, really smart” vs. who “try really, really hard”). Responses to the four questions were standardized (so that they are on the same scale) and then averaged.

After an abbreviated, simplified set of screener questions (which all children passed), we assessed children’s brilliance stereotypes with the gender-neutral story from Studies 1 and 2. However, before children selected the protagonist of the story from among the 4 pictures provided (as in Studies 1 and 2), we also asked them to repeat the story and then coded the gender of the pronouns they used. This task was adapted from Ambady et al. (2001). The final stereotype score in this study was an average of these two items (the pronouns they used and the pictures they selected; standardized before averaging).

At the end of the sessions, children received a thorough debriefing that was designed to convey that effort and hard work are the key to success (e.g., “If you try really hard and practice a lot, you can be good at *any* game you want”).

**Results and discussion**

The FAB hypothesis makes two major predictions. First, there should be an interaction between Gender and Condition (smart vs. try-hard; within subject) on children’s motivation toward the novel activities. Specifically, the 6- and 7-year-old girls in this study should be less interested in the smart game (but not the try-hard game) than boys. Second, children’s endorsement of the “brilliance ≠ females” stereotype should be related to their motivation toward
the smart game: Children who hold more negative stereotypes against their own gender’s intelligence should be less motivated to pursue the “brilliance-focused” activities.

To test the first prediction, I submitted children’s interest scores (combined across the 4 questions) to a multilevel mixed-effects linear model with game (smart vs. try-hard; level-1 predictor), gender (boys vs. girls; level-2 predictor), and age (6- vs. 7-year-olds; level-2 predictor), plus all possible interaction terms, as categorical fixed effects and a random intercept for participants. As expected, the interaction between gender and game was significant, Wald $\chi^2 = 5.42, p = .020$. Two follow-up tests comparing boys’ and girls’ interest in each game revealed that 6- and 7-year-old girls were less interested than boys in the game for smart children, Wald $\chi^2 = 4.02, p = .045$, but not in the game for hard-working children, Wald $\chi^2 = 0.53, p = .47$ (see Figure 3A and Table 7).

Before testing the second prediction, I conducted an independent-samples $t$ test to examine if we could replicate the previous results. Consistent with our previous findings, 6- and 7-year-old girls’ own-gender brilliance perceptions were significantly lower than boys’, $t(61) = 2.40, p = .020$. Next, to explore whether the gender differences in interest are related to children’s beliefs about brilliance, I submitted the data to a bootstrapped (10,000 replications) product-of-coefficients mediation test by using the PROCESS macro in SPSS 22 (Model 4; Hayes, 2013). Children’s gender (0 = boys, 1 = girls) was the independent variable in this analysis; the own-gender brilliance score was the mediator; and children’s interest in the smart vs. the try-hard game (a difference score) was the dependent variable. The results suggested that these stereotyped beliefs mediated the relationship between children’s gender and their interest in the game for brilliant (vs. persistent) children, indirect effect = $−.11$, 95% confidence interval =
Thus, young children’s emerging notions about who is likely to be brilliant are one of the factors that guide their decisions about which activities to pursue.

These findings provide the first piece of evidence suggesting that the messages emphasizing the importance of intellectual talent to success affect girls’ motivation via their beliefs about such talent. It is possible that, by the time they reach adulthood, many capable girls have already veered away from fields that are generally believed to require raw intellectual talent, which are some of the most prestigious and lucrative careers in our society.

Study 4

In Study 4, I compared 5- and 6-year-old boys’ and girls’ interest in novel games said to be “for children who are really, really smart”. The prediction was that 5-year-old boys’ and girls’ interest in these games would not differ, since their ideas about brilliance are not yet differentiated (Figure 1A and 1C). In contrast, 6-year-old girls’ interest was predicted to be lower than boys’, in line with the results of Study 3.

Participants

Participants were 48 five-year olds ($M_{age} = 5.40$ years, $SD = 0.31$) and 48 six-year olds ($M_{age} = 6.52$ years, $SD = 0.31$). Half of them were boys, and half of them were girls. Children were recruited in a small Midwestern city. Two additional children were tested but excluded from the final sample because they were more than 2.5 standard deviations from the condition mean.

Materials and procedure

The procedure of Study 4 was identical to that of Study 3, except children were only told about the game “for children who are really, really smart.” Half of the children saw the “zarky” game, and half saw the “impok” game.
Results and discussion

I predicted that we would replicate the results of the preceding study with this study’s sample of 6-year olds. However, since children at age 5 show little awareness of the “brilliance ≠ females” stereotype, I predicted that boys and girls of this age would not differ in their interest in the brilliance game. Overall, this would give rise to a Gender × Age interaction. To investigate these predictions, children’s interest scores were submitted to a linear regression with gender (boys vs. girls), age (5- vs. 6-year-olds), and their interaction as categorical predictors. Standard errors were bootstrapped (10,000 replications). As expected, the interaction between Gender and Age was marginally significant, Wald $\chi^2 = 3.12, p = .078$. Follow-ups testing gender differences in children’s interest at each age level revealed no significant gender differences in interest among 5-year-olds, Wald $\chi^2 = 0.55, p = .46$, and a trend in the predicted direction among 6-year-olds, Wald $\chi^2 = 3.66, p = .056$ (see Figure 3B and Table 7). Combining the samples of 6- and 7-year-olds from Studies 3 and 4 with a random-effects meta-analysis (Cumming, 2013), we estimated the magnitude of the difference in boys’ vs. girls’ interest toward the game for brilliant children to be $d = .51$, 95% confidence interval = [.13, .88], $p = .008$.

Overall, Studies 3 and 4 provide coherent evidence suggesting that, beginning at the age of 6, girls start to shy away from novel activities said to be for children who are “really, really smart.” Moreover, their interest towards these activities was predicted by their beliefs about which gender is brilliant. Thus, these results speak to the immediate impact that these stereotyped notions have on children’s interests.

Moderation analyses with race/ethnicity and SES. To examine if children’s race/ethnicity moderates the observed gender differences in interest toward the smart game, we submitted this dependent variable (pooled across Studies 3 and 4) to a linear regression with
race/ethnicity (white children vs. children of color), gender (boys vs. girls), age (5- [younger] vs. 6- and 7-year-olds [older]), and all their two- and three-way interactions as categorical predictors. Similar to the analysis of children’s stereotypes, children’s racial/ethnic backgrounds did not significantly moderate the key Age × Gender interaction in children’s interest toward the smart game, Wald $\chi^2 = 1.48, p = .22$. For example, 6- and 7-year-old girls displayed lower interest in the smart game than 6- and 7-year-old boys regardless of their racial/ethnic group (white children: $M_{boys} = .25$ vs. $M_{girls} = -.22$; children of color: $M_{boys} = .19$ vs. $M_{girls} = -.23$).

Like race/ethnicity, SES did not significantly moderate the Age × Gender interaction in children’s interest toward the smart game, Wald $\chi^2 = 0.01, p = .96$. For example, 6- and 7-year-old girls’ interest in the smart game was lower than 6- and 7-year-old boys’ interest regardless of whether they came from high-SES ($M_{boys} = .08$ vs. $M_{girls} = -.22$) or low-SES ($M_{boys} = .34$ vs. $M_{girls} = -.22$) families.

**Alternative explanations.** We considered two possible alternative explanations for the results of Studies 1–4. One alternative explanation suggests that, because boys are sometimes held back from entering the formal schooling system (Huang, 2015), their understanding of intellectual ability may be delayed relative to girls’ (Eccles, Midgley, & Adler, 1984), which may inflate boys’ confidence about their brilliance (Butler, 2005). To examine this alternative, we combined the schooling status information across Studies 1 to 4, which was available for 34% of the children. Of the children for whom we had this information, 5% were in preschool, 21% in kindergarten, 11% in kindergarten or first grade (some schools in our sample combined these grades), 42% in first grade, and 20% in second grade or higher. Contrary to this explanation, the boys and girls in our sample did not enter school at different ages (e.g., the average chronological age for first-grade boys and girls was 6.87 and 6.72 years, respectively;
\( t(55) = 1.28, p = .204 \). (The children in the combined kindergarten/first grade program were not included in these analyses.) Moreover, own-gender brilliance scores did not differ for boys who had vs. had not already entered first grade (\( M_{before} = .70 \text{ vs. } M_{after} = .67; t = 0.33, p = .743 \)), but these scores differed for girls (\( M_{before} = .71 \text{ vs. } M_{after} = .56; t = 2.16, p = .037 \)).

Another alternative explanation suggests that, because women are subject to stronger modesty norms than men (Rudman, 1998), perhaps 6- and 7-year-old girls’ lower interest in the games for brilliant children (Studies 3 and 4) was due to an increase in concerns about modesty. Contrary to this alternative, children in the age range we tested are notoriously boastful about their abilities (Butler, 2005). Moreover, the difference in boys’ vs. girls’ interest in the brilliance games was specifically mediated by their perceptions about brilliance, identifying these stereotyped perceptions (rather than modesty) as an underlying mechanism. Notably, our measure of the “brilliance ≠ females” stereotype is not susceptible to the modesty explanation: Modesty norms dictate that a woman should not boast about her own smarts (Smith & Huntoon, 2013; Mazei, Hüffmeier, Freund, Stuhl macher, Bilke, & Hertel, 2015), whereas we asked children to judge whether other people were smart. Overall, these alternatives do not provide convincing explanations for our results.
Chapter 4: The effect of the “brilliance ≠ females” stereotype on adults’ and children’s
evaluation of others’ capabilities

In this chapter, I explored how the “brilliance ≠ females” stereotype influences adults’
and children’s evaluation of boys’ and girls’ competence, especially for activities that are
portrayed as requiring high levels of intellectual ability. Specifically, I examined whether
participants doubt whether women will be successful in such activities and are thus less likely to
select or nominate women for them.

The FAB model makes two predictions: First, adults and children should be more likely
to select or nominate males (vs. females) for activities that are said to require brilliance, but not
for other, control activities (e.g., ones are said to require dedication), suggesting that the
messages emphasizing the importance of intellectual talent to success give rise to bias favoring
men over women (Studies 5 and 6). Second, this bias should be moderated by people’s beliefs
about brilliance. The more strongly participants endorse the stereotypes that associate brilliance
with males, the more likely they should be to offer opportunities that seemingly require brilliance
to males (Study 5).

Study 5

Participants

Participants ($N=347; M_{age}=34.58; 140$ men and 207 women) were recruited from
Amazon’s Mechanical Turk service. They were paid $0.50 for participation. Sixty-seven
additional participants were excluded because their IP addresses indicated they resided outside of
the US ($n=12$), because they explicitly reported (during debriefing) that they had not paid
attention ($n=2$), because they didn’t provide responses on the main dependent measure ($n=1$),
or because they failed an attention check ($n=52$; see below).
Materials and procedure

All participants completed the study online, on Qualtrics. At the beginning, participants were asked to imagine a hypothetical scenario that they are “working in the human resources department of a big company” that “is looking to fill a couple of positions in their workforce.” Then they were randomly assigned to one of the two conditions: Brilliance or Dedication.

Next, we presented participants with a description of the ideal candidates that the company is looking for. Participants in the Brilliance condition were told that the ideal candidates should “have a high IQ, superior reasoning skills, and a knack for big, bold ideas”, whereas participants in the Dedication condition were told that the ideal candidates should “be highly motivated, have an outstanding work ethic and a superior commitment to doing their work as well as possible” (see Table 8 for full job descriptions). After reading the description, participants were given a memory test in which they were asked to select from a longer list the required skills mentioned in the description. Their responses served as an attention check. Participants were retained in the analyses if they (1) selected at least one of the traits relevant to their condition, and (2) did not select any of the traits relevant to the other condition (e.g., “intelligent” in the Dedication condition). Fifty-two participants did not meet these criteria and were excluded from the analyses.\footnote{The results remained the same if we included the data from the participants who failed the attention check.}

Participants were then provided a reminder of the requirements to make sure they fully encoded the information at the core of our manipulation.

Participants were then asked to nominate two people for the job they read about. In each nomination round, they were asked to suggest a relative, a friend, or an acquaintance for the job. The gender of this recommended person was the key dependent variable of our study. Responses were scored as a 1 if participants suggested a man, and 0 if they suggested a woman. In order to
mask the purpose of the study, participants were asked to provide other demographic information about the person they nominated, not just their gender (e.g., first name, age, relationship to the participant). Participants were also asked to justify why they selected this person. To maximize the likelihood that they considered the brilliance vs. dedication requirements while making the selections, we asked participants to think about the required traits and reminded them of the description before they nominated the second teammate.

Next, we assessed participants’ endorsement of the stereotype against women’s intellectual abilities using eight items (e.g., “On average, men tend to have higher intellectual capacities than women”; see Table 9). These items were averaged into an overall stereotype score ($\alpha = .90$).

Finally, participants answered a brief demographics questionnaire, and they received a debriefing.

**Results and discussion**

The dependent variable in the study was the gender of the nominated person on each round (male = 1, female = 0). Our overall prediction was that participants would be more likely to recommend males than females for the Brilliance job, but not the Dedication job. In the context of this task, this broad prediction translates into two more-specific possibilities: First, we expect a significant main effect of Condition (i.e., males should be recommended more often for the Brilliance job than for the Dedication job). Second, we might also expect a significant interaction between Condition and Stereotype Endorsement, with the bias to favor males for the Brilliance job being particularly strong for people who are high (vs. low) in endorsement of brilliance stereotypes.

To test these predictions, we conducted a mixed-effects logistic regression using the
xtmelogit command in Stata. In this logistic regression, we modeled the gender of the recommended acquaintance on each round as a function of (1) the gender of the participant (men vs. women; level-2 predictor), (2) condition (Brilliance vs. Dedication; level-2 predictor), (3) nomination round (first vs. second; level-1 predictor), as well as (4) stereotype endorsement (continuous variable; level-2 predictor). In addition, the model included all two-, three-, and four-way interactions among the variables and a random intercept for participants.

Consistent with our first prediction, we found a significant main effect of condition, \( b = -0.50, SE = 0.16, z = -3.07, p = 0.002 \), as participants were more likely to nominate a male for the Brilliance job (\( M = 0.59, SD = 0.03 \)) than for the Dedication job (\( M = 0.48, SD = 0.03 \); see Figure 5). In fact, participants were more likely to recommend a male than a female for the Brilliance job, \( t(168) = 3.60, p < 0.001 \), but not for the Dedication job, \( t(177) = 0.81, p = 0.420 \).

Moreover, the analyses also revealed a significant interaction among Brilliance vs. Dedication condition, stereotype endorsement, and nomination round, \( b = -0.37, SE = 0.19, z = -2.00, p = 0.045 \). Specifically, people were more likely to recommend a male for the Brilliance job than for the Dedication job on the first round, regardless of the extent to which they endorse the stereotype associating brilliance with males (low-stereotype group \( [SD \text{ below the mean}] \): \( z = -3.18, p = 0.001 \); high-stereotype group \( [SD \text{ above the mean}] \): \( z = -2.03, p = 0.042 \)). However, participants’ favoritism towards males in the second round depended on their endorsement of the stereotype. In particular, people in the high-stereotype group (but not in the low-stereotype group) were more likely to nominate males for the Brilliance position than for the Dedication position, \( z = -2.06, p = 0.040 \).

In addition, the analyses revealed several other main effects and interactions. First, there was a significant main effect of participants’ gender, \( b = 1.00, SE = 0.17, z = 5.93, p < 0.001 \),
indicating that male participants were more likely to recommend male acquaintances than female participants overall, which accords with people’s general tendency to prefer their own gender (e.g., Dunham et al., 2011). Second, the main effect of stereotype endorsement was marginally significant, $b = 0.09, SE = .05, z = 1.85, p = .064$, suggesting that the more people agree with the stereotype against women’s intellectual abilities, the more likely they were to nominate men as candidates in both conditions. This finding is consistent with the literature on gender discrimination (e.g., Moss-Racusin et al., 2012). Third, we found a significant effect of nomination round, $b = -.34, SE = .16, z = -2.11, p = .035$, reflecting people’s tendency to recommend more males on the first round than on the second round. Lastly, the analyses uncovered a marginally significant two-way interaction between condition and nomination round, $b = .62, SE = .33, z = 1.91, p = .057$. Specifically, participants tended to choose more males for the Brilliance job than for the Dedication job on the first selection, $z = -3.53, p < .001$, but not on the second selection, $z = -0.68, p = .496$.

In summary, participants were more likely to nominate men than women for “brilliance required” positions, but not “dedication required” positions, suggesting that messages emphasizing the importance of brilliance to success give rise to bias favoring men over women. Moreover, this bias is moderated by the stereotypes against women’s intellectual abilities. Specifically, the bias in favor of men is stronger among people who are more likely to endorse these stereotypes. Overall, these results are consistent with our hypothesis that men are being favored over women in contexts where success is thought to depend largely on being brilliant.

**Study 6**

The primary goal of Study 6 was to investigate whether the biases favoring males in “brilliance required” contexts are already present in childhood. If even young children hold these
biases, then—from early on—girls may start facing stronger barriers than boys in their pursuit of intellectually-demanding educational opportunities, which might eventually lead to the current gender imbalance. Thus, investigating the development of these biases is likely to enhance our understanding of the roots of gender disparities, as well as our ability to intervene effectively at the point where the problem originates.

A secondary goal here was to test an alternative explanation for the results of Study 5. In principle, one might argue that the Brilliance vs. Dedication differences found in Study 5 are just due to a preference for women in dedication-oriented contexts (rather than a preference for men in brilliance-oriented contexts). This alternative is not entirely consistent with the results of Study 5, since participants were actually gender-neutral in their recommendations in the Dedication condition (48% men, 52% women) rather than favoring females. In contrast, they did favor males in the Brilliance condition (59% men, 41% women). Nevertheless, the present experiment provides a more direct test of this alternative explanation by comparing a Brilliance condition with a Baseline condition, in which neither brilliance nor dedication was mentioned. Our prediction is that children would provide fewer opportunities to girls in the Brilliance condition, but not in the Baseline condition.

Moreover, we asked children to evaluate unfamiliar targets. In contrast, participants in Study 5 nominated familiar people for certain positions. This change was implemented so that participants could not rely on their prior knowledge of the individuals they are selecting. To the extent that children in Study 6 choose fewer girls for the Brilliance than the Baseline game, they would do so because of their abstract beliefs about boys vs. girls (i.e., their stereotypes), not because of their specific knowledge about certain individuals.

**Participants**
The participants were 64 five-year-old children ($M_{age} = 5.44$ years, $SD = 0.28$; 32 girls and 32 boys), 64 six-year-old children ($M_{age} = 6.52$ years, $SD = 0.33$; 32 girls and 32 boys), and 64 seven-year-old children ($M_{age} = 7.46$ years, $SD = 0.29$; 32 girls and 32 boys). Children were recruited in a small Midwestern city in the United States. They were tested either in a university lab ($N = 57$) or in a quiet room at their school ($N = 135$). Two additional children were tested but excluded from the sample because they refused to complete the study.

Demographic information was available for 71% of the families. The racial/ethnic composition of the sample mirrored that of the community in which this research was conducted: 69% of the children were European American, 9% Asian American, 5% African American, 3% Latino or Hispanic, 1% other (parents did not specify), and 12% multi-racial. The median household income was $90,000 (range = $4,500 - $185,000). Seventy-seven percent of the parents in the sample had at least a bachelor’s degree. Neither race nor socio-economic status (SES) moderated the reported effect (for more details, see page 52).

**Materials and procedure**

Children were told about two novel games (“Zorb” and “Tever”). The experimenter introduced each game by showing children a picture and providing them with a brief description (see Table 10). After hearing about each game, the children who had been randomly assigned to the Brilliance condition were provided with the following information: “This game is not for everyone. It’s only for children who are really, really smart. Only smart children can be good at this game.” Children in the Baseline condition were not provided with this information. To ensure that children encoded the manipulation, the experimenter then asked children in the Brilliance condition to recall who would be good at this game and corrected them if needed.
Next, children were told that “this is a team game” and that they need to choose three teammates to play the game with. They were then shown six pictures of unfamiliar children (three boys and three girls), and they were asked to pick three children as their teammates. The pairing between the pictures of children and particular games (Zorb vs. Tever) was counterbalanced across subjects so that no perceptual cues about these children consistently co-occurred with the descriptions or other features of the games. After the first two selections, children in the Brilliance condition were provided with a reminder of the relevant ability messages to make sure they could still consider this information in making their selections. On each selection, children received a score of 1 if they chose a boy, and a score of 0 if they chose a girl.

After completing all three selections, children received a thorough debriefing (e.g., “If you try really hard and practice a lot, you can be good at any game you want”). At the end of the sessions, children were thanked for their participation and praised for their responses.

Results and discussion

As in Study 5, we submitted our data to a mixed-effects logistic regression using the `xtmelogit` command in Stata. Specifically, we modeled the gender of the chosen teammate on each round as a function of (1) the gender of the participant (boy vs. girl; level-2 predictor), (2) condition (Brilliance vs. Baseline; level-2 predictor), (3) selection round (first vs. second vs. third; level-1 predictor), and (4) the age of the participant (level-2 predictor). In addition, the model included all two-, three-, and four-way interactions among the variables and a random intercept for participants.

Unsurprisingly, we found a significant effect of gender, $b = 2.58$, $SE = 0.25$, $z = 10.25$, $p < .001$, suggesting that boys were more likely than girls to choose boys as teammates. In fact,
both boys’ and girls’ tendency to choose their own gender was significantly higher than chance (boys: \(t(95) = 7.94, p < .001\); girls: \(t(95) = 11.25, p < .001\)). These results are consistent with the general tendency to favor one’s own gender group (e.g., Dunham et al., 2011).

Despite this strong favoritism towards one’s own gender, there was a significant interaction between gender and condition, \(b = -0.98, SE = 0.46, z = -2.11, p = .035\). In particular, boys’ tendency to pick boys as teammates did not vary significantly by condition, but girls tended to pick more boys for the Brilliance than for the Baseline game (see Figure 6). These findings suggest that the messages emphasizing brilliance promote young children’s (especially girls’) biases favoring males.

The analyses also uncovered a significant interaction between condition and selection round, \(b = .44, SE = .19, z = 2.28, p = .022\). Follow-up tests revealed that children’s tendency to pick boys did not differ in the two conditions on the first \(z = -0.72, p = .474\) or the second selections \(z = 0.62, p = .533\), but they picked more boys for the Brilliance than for the Baseline game on the third selection \(z = 2.29, p = .022\).

We found the interaction between gender and condition was not moderated by age, \(b = -0.47, SE = 0.53, z = -0.88, p = .379\). This result contrasts with those of Studies 1 and 2, in which 5-year-olds behaved differently from older children. What might explain this apparent discrepancy? First, this teammate-selection task might be more implicit than the stereotype tasks used in Studies 1 and 2, because (1) children were not asked to directly indicate which person is smart and (2) they were provided with more options (three boys and three girls). The implicitness of the task might diminish the tendency to select one’s own group and allow the bias favoring males to surface even in young children’s selections. In addition, choosing the right teammate in this study has direct consequences for whether children’s team will win, which might serve as
extra impetus for children to rely on their (stereotyped) beliefs about who is brilliant rather than ingroup favoritism to make their selections.

**Moderation analyses with race/ethnicity and SES.** To examine if children’s race/ethnicity moderates the observed gender by condition interaction on children’s evaluation of others’ competence, we submitted our data to a linear regression with race/ethnicity (white children vs. children of color), gender (boys vs. girls), age (continuous variable), condition (smart vs. baseline), and all their two-, three-, and four-way interactions as predictors. The results revealed that children’s racial/ethnic backgrounds did not significantly moderate the gender × condition interaction in children’s tendency to select teammates for the smart game, Wald $\chi^2 = 0.14, p = .710$. For example, girls tended to pick more boys for the smart game than for the baseline game, regardless of their racial/ethnic group (white girls: $M_{smart} = .26$ vs. $M_{baseline} = .16$; girls of color: $M_{smart} = .43$ vs. $M_{baseline} = .28$). Similarly, SES did not significantly moderate the age × gender interaction in children’s interest toward the smart game, Wald $\chi^2 = 0.69, p = .407$. For example, girls tended to pick more boys for the smart game than for the baseline game, regardless of whether they came from high-SES ($M_{smart} = .35$ vs. $M_{baseline} = .18$) or low-SES ($M_{smart} = .27$ vs. $M_{baseline} = .18$) families.

In summary, messages that emphasize the importance of innate talent to success gave rise to a bias favoring males in children’s behavior, as they tended to provide more opportunities to boys than to girls in these contexts. These results are consistent with our hypothesis that women may be the targets of bias in fields where success is thought to depend largely on being brilliant. Over time, then, girls may face stronger barriers than boys in their pursuit of prestigious, intellectually-demanding careers.
Chapter 5: General Discussion

The present work marks a significant advance in our scientific understanding of the developmental course of gender stereotypes about intelligence and their pernicious consequences for young children’s activity choices and evaluations. Our findings so far seem to suggest that these stereotypes might be passed on to children as early as the first year of school. In addition, these stereotypes begin to influence children’s attitudes and behaviors as soon as they are acquired. Specifically, they seem to undermine young girls’ interest in activities said to be for “smart kids” and bias children’s evaluations of others’ competence for such activities.

Studies 1 and 2 are the first to examine the development of the stereotype that associates males more than females with intellectual brilliance. Although prior work has explored many aspects of gender development as well as of stereotyping in early childhood, this is the first investigation of the harmful “brilliance ≠ females” stereotype that, we hypothesize, discourages women from pursuing many prestigious fields and occupations. Further, Studies 3 and 4 are the first to examine the developmental course of boys’ and girls’ attitudes toward, and engagement with, activities that are said to require high levels of intellectual ability. Although prior work has examined children’s attitudes toward activities requiring particular domain-specific skills (e.g., mathematics; e.g., Wigfield & Eccles, 2000), no research so far has documented, as we do, a divergence in boys’ and girls’ attitudes toward the broader class of activities thought to require high levels of raw intellectual ability. Moreover, Studies 5 and 6 are the first to examine the effect of brilliance-focused messages on adults’ and children’ implicit bias in evaluating females’ competence. In addition, Study 6 provides a novel examination of the developmental course of children’s bias in evaluating boys’ and girls’ suitability for activities that place a strong emphasis on intelligence. Although prior work has examined many aspects that could influence children’s
evaluation and affiliation choices (e.g., ingroup favoritism), this research presents the first examination of the developmental roots of the bias against females in careers portrayed as requiring intellectual talent. Since these careers often hold the highest status and are most valued in our society, this research illuminate some of the developmental roots of the current gender-based status inequities.

**Practical Implications.** Given the potential effects of these stereotypes, it is important for researchers, parents, and teachers to be aware of their developmental timeline so that we as a society can consider how we might best minimize their impact. To speculate, it is possible that neutralizing either of the components of the FAB model in young children would bring about gains in girls’ participation. First, conveying to children that men and women do not differ in their intellectual potential would make it less likely for brilliance-oriented messages to differentially threaten girls. Such an intervention would in fact be consistent with the current literature on this topic, which suggests gender parity in intellectual ability (e.g., Flynn, 2012; Hyde & Mertz, 2009; Spelke, 2005; for a different view, see Baron-Cohen, 2002; Deary, Thorpe, Wilson, Starr, & Whalley, 2003; Hedges & Nowell, 1995). A second, complementary intervention strategy would be to alter the ability belief component—or, at least, the messages sent to young children about how to achieve success. In particular, highlighting the importance of sustained effort (vs. raw talent) might reduce the impact of the “brilliance ≠ females” stereotype on girls’ motivation and the bias against their competence. In an effort-focused environment, women and girls who are targets of stereotypes against their intelligence might nevertheless believe they are capable of succeeding and might thus maintain strong motivation for the careers they are pursuing. In addition, an environment that emphasizes effort and growth would be less likely to signal to women that they will be judged on the basis of existing
stereotypes about their abilities (e.g., Emerson & Murphy, 2015). In the absence of this situational threat, it may be easier for women to join and persist in a field. More generally, the idea that abilities can be cultivated (rather than being fixed) is likely to have beneficial long-term effects for both women’s and men’s ability to sustain involvement with a field despite the unavoidable challenges that arise along the way (e.g., Dweck, 1999, 2006). In this way, we may have a chance of alleviating the negative consequences of the “brilliance ≠ females” stereotype and putting men and women on an equal playing field with regard to their access to various careers.

**Future directions.** These studies inspire further questions, which we plan to investigate in future work. For example, what are the sources of the “brilliance ≠ females” stereotype? Why does the stereotype seem to emerge (given our pilot data) between the ages of 5 and 6? Is exposure to school environments perhaps implicated in the acquisition of this stereotype? Where do children get the idea that males are more intellectually gifted than females, especially since girls usually get better grades in school than boys do? In our present work, we found children’s perceptions of which gender performs better at school do not predict their beliefs about which gender is brilliant, but many other aspects of children’s experiences in school, such as teachers’ attitudes and biases (e.g., Beilock, Gunderson, Ramirez, & Levine, 2010; Robinson-Cimpian, Lubienski, Ganley, & Copur-Gencturk, 2014) may still be implicated in the development of this stereotype.

We can also ask whether race might moderate the content of these stereotyped beliefs. In the present work, we assessed children’s stereotypes about white children and adults. Would children’s stereotypes differ for targets that belong to other race/ethnicity groups (and perhaps
depending on children’s own race/ethnicity group as well)? It will be also important to test whether these findings extend beyond a middle-class, majority-white US cultural context.

**Conclusions.** Overall, the six studies provide a comprehensive investigation of the developmental aspect of the Field-specific Ability Beliefs hypothesis. These findings, coupled with prior research (e.g., Bian et al., under review; Leslie, Cimpian, et al., 2015) suggest that this theoretical model has the potential to explain the complex, field-by-field pattern of women’s representation across the entire academic spectrum. They also suggest that consequential stereotyped beliefs take root early, and thus that early intervention is crucial if we are to equalize the opportunities available to young women and men.
Table 1

The screener questions used in Studies 1–2

“Smart” Block
- The child can always answer even the hardest questions from the teacher.
- The child learns things really fast.
- The child can solve really hard puzzles.
- The child figures things out really quickly.
- The child watches really funny cartoons. (R)
- The child practices sports all the time. (R)

*Question:* Is the child smart, not smart, or are you not sure?

*Scoring:* Mean of 6 items (1 = “smart”; 0 = “not smart” or “not sure”)

“Nice” Block
- The child likes to help other people.
- The child always shares their toys with other children.
- The child tries to make other children feel better when they are sad.
- The child likes to give hugs to family and friends.
- The child likes to listen to music. (R)
- The child plays on the swings. (R)

*Question:* Is the child nice, not nice, or are you not sure?

*Scoring:* Mean of 6 items (1 = “nice”; 0 = “not nice” or “not sure”)

*Note.* The items marked with an (R) were reverse-scored. Item order was randomized within each scale. Children who scored less than 0.67 on either block were excluded from analyses.
Table 2
The Gender-Neutral Stories Used to Assess Children’s Stereotypes in Studies 1 and 2

<table>
<thead>
<tr>
<th>Trait</th>
<th>Story about an Adult (Studies 1 and 2)</th>
<th>Story about a Child (Study 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smart</td>
<td>There are lots of people at the place where I work. But there is one person who is really special. This person is really, really smart. This person figures out how to do things quickly and comes up with answers much faster and better than anyone else. This person is really, really smart.</td>
<td>When I was your age, there were lots of children at the kindergarten where I went. But there was one child who was really special. This child was really, really smart. This child learned things very quickly and could answer even the hardest questions from the teacher. This child was really, really smart.</td>
</tr>
<tr>
<td>Nice</td>
<td>There are lots of people at the place where I work. But there is one person who is really special. This person is really, really nice. This person likes to help others with their problems and is friendly to everyone at the office. This person is really, really nice.</td>
<td>When I was your age, there were lots of children at the kindergarten where I went. But there was one child who was really special. This child was really, really nice. This child shared their toys with everyone else, and really cared about the other kids. This child was really, really nice.</td>
</tr>
</tbody>
</table>
Table 3
Boys’ and Girls’ Stereotype Scores in Studies 1 and 2 (Standard Deviations in Parentheses)

<table>
<thead>
<tr>
<th>Age</th>
<th>Gender</th>
<th>Study 1</th>
<th>Study 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Smart</td>
<td>Nice</td>
</tr>
<tr>
<td>5-year-olds</td>
<td>Boys</td>
<td>0.71 (0.22)</td>
<td>0.66 (0.22)</td>
</tr>
<tr>
<td></td>
<td>Girls</td>
<td>0.69 (0.19)</td>
<td>0.61 (0.31)</td>
</tr>
<tr>
<td>6-year-olds</td>
<td>Boys</td>
<td>0.65 (0.20)</td>
<td>0.40 (0.25)</td>
</tr>
<tr>
<td></td>
<td>Girls</td>
<td>0.48 (0.24)</td>
<td>0.67 (0.15)</td>
</tr>
<tr>
<td>7-year-olds</td>
<td>Boys</td>
<td>0.68 (0.26)</td>
<td>0.43 (0.24)</td>
</tr>
<tr>
<td></td>
<td>Girls</td>
<td>0.54 (0.21)</td>
<td>0.62 (0.18)</td>
</tr>
</tbody>
</table>
Table 4
The number of children in each category in Studies 1 and 2

<table>
<thead>
<tr>
<th></th>
<th>Boys</th>
<th>Girls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5-yr-olds</td>
<td>6-yr-olds</td>
</tr>
<tr>
<td>Favoring own gender</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>Neutral</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Favoring other gender</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>
Table 5
The Games Used to Assess Children’s Interest in Studies 3 and 4

Zarky

I want to tell you about this game that I ask children to play sometimes. It’s called Zarky, and it’s a lot of fun. In this game, what you have to do is to bring the red pieces from this side to this side, one piece at a time, without going in a straight line and without getting them stuck in between the blue pieces. Oh, and here is something else about the Zarky game, and this is important so make sure you’re paying attention. This game is not for everyone. It’s only for children who are really, really smart [who try really, really hard]. Only smart [hardworking] children can be good at this game.

Impok

I want to tell you about this game that I ask children to play sometimes. It’s called Impok, and it’s a lot of fun. In this game, what you have to do is to figure out how to get the big pyramids next to each other in the black squares and get the small pyramids next to each other in the white squares in only ten moves and without crossing the grey squares. Oh, and here is something else about the Impok game, and this is important, so make sure you’re paying attention. This game is not for everyone. It’s only for children who are really, really smart [who try really, really hard]. Only smart [hardworking] children can be good at this game.

Note. Note. In Study 3, each of the games was presented in the “smart” format to half of the children and in the “try-hard” format to the other half.
Table 6
The Four Questions Used to Assess Children’s Interest in Studies 3 and 4

1) Imagine I had the Zarky/Impok game right here, in front of you. Would you want to play the Zarky/Impok game, or would you not want to play it?
   
   [if “yes”]
   Would you sort of want to play it (= 3), or really want to play it (= 4)?
   
   [if “no”]
   Would you sort of not want to play it (= 2), or really not want to play it (= 1)?

2) Do you like the Zarky/Impok game, or do you not like it?
   
   [if “yes”]
   Do you sort of like it (= 4), like it (= 5), or really like it (= 6)?
   
   [if “no”]
   Do you sort of not like it (= 3), not like it (= 2), or really not like it (= 1)?

3) Imagine you are playing the Zarky/Impok game. Would playing Zarky/Impok make you feel happy or sad?
   
   [if “happy”]
   Would it make you feel sort of happy (= 4), happy (= 5), or really happy (= 6)?
   
   [if “sad”]
   Would it make you feel sort of sad (= 3), sad (= 2), or really sad (= 1)?

4) If you had a chance to do something tomorrow, would you play the Zarky/Impok game (= 1) or would you do something else (= 0)?

Note. The numerical scoring of each option is indicated in parentheses. Question order was randomized across children. Because the questions used different scales, responses to each were standardized before averaging.
<table>
<thead>
<tr>
<th>Age</th>
<th>Gender</th>
<th>Study 3</th>
<th>Study 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Smart</td>
<td>Try-hard</td>
</tr>
<tr>
<td>5-year-olds</td>
<td>Boys</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Girls</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>6-year-olds</td>
<td>Boys</td>
<td>0.20 (0.71)</td>
<td>–0.09 (0.81)</td>
</tr>
<tr>
<td></td>
<td>Girls</td>
<td>–0.17 (0.77)</td>
<td>0.10 (0.49)</td>
</tr>
<tr>
<td>7-year-olds</td>
<td>Boys</td>
<td>0.15 (0.69)</td>
<td>–0.03 (0.90)</td>
</tr>
<tr>
<td></td>
<td>Girls</td>
<td>–0.21 (0.88)</td>
<td>0.04 (0.58)</td>
</tr>
</tbody>
</table>
Table 8
*Job Descriptions Used in Study 5*

<table>
<thead>
<tr>
<th>Description condition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Brilliance condition</strong></td>
</tr>
<tr>
<td>We are looking for applicants to fill two new positions in our workforce. These positions will involve a lot of responsibility and the opportunity to join one of our newest, and most exciting, departments.</td>
</tr>
<tr>
<td>Because of the work we do, we are looking for candidates who have a high IQ, superior reasoning skills, and a knack for big, bold ideas. That is, we’d like to hire someone whose intellectual abilities stand out from those of their peers. Our work environment values and emphasizes employees' natural intelligence, expecting everyone to push their inborn smarts to the limit. Therefore, we are especially interested in candidates who demonstrate an inherent aptitude for this position.</td>
</tr>
</tbody>
</table>

| **Dedication condition**    |
| We are looking for applicants to fill two new positions in our workforce. These positions will involve a lot of responsibility and the opportunity to join one of our newest, and most exciting, departments. |
| Because of the work we do, we are looking for highly motivated candidates with an outstanding work ethic and a superior commitment to doing their work as well as possible. That is, we’d like to hire someone who has demonstrated significant and sustained dedication in their past positions. Our work environment values and emphasizes employees’ strivings and their consistent effort to achieve goals, expecting everyone to continuously improve their work performance. Therefore, we are especially interested in candidates who demonstrate continual passion for the job. |
Table 9

*Items in the Stereotype Endorsement Measure Used in Study 5.*

- One is more likely to find a male with a genius-level IQ than a female with a genius-level IQ.
- Extreme intellectual brilliance is more common in men than in women.
- On average, men tend to have higher intellectual capacities than women.
- Even though it’s not true of everyone, males are generally born with greater raw intelligence than females.
- The reason why there are few female philosophers is that women tend to think more practically.
- Men and women have complementary cognitive skills: Men are better at understanding objects and mechanical systems, whereas women are better at understanding people and their emotions.
- Even though it may not be politically correct to say it, males and females might be naturally suited for different kinds of intellectual activities.
- Males’ and females’ biology has an effect on their cognitive abilities (even though the differences might be small).

**Scoring:** Mean of 8 Likert-type scale items (1 = disagree strongly; 9 = agree strongly)
Table 10
The Games Used to Assess Children’s Bias in Study 6

Zorb

I want to tell you about this game that I ask children to play sometimes. It’s called Zorb, and it’s a lot of fun. In this game, what you have to do is to figure out how to get 2 small coins and 2 big coins in each green tube. But you also need to make sure to have all the big ones at the bottom and all the small ones at the top without using any of the pink tubes more than once. [Oh, and here is something else about the Zorb game, and this is important, so make sure you’re paying attention. This game is not for everyone. It’s only for children who are really really smart. Only smart children can be good at this game.]

Tever

I want to tell you about this game that I ask children to play sometimes. It’s called Tever, and it’s a lot of fun. In this game, what you have to do is to figure out how to put as many pencils and erasers as you can into these holes in just one minute. But you also need to have the pencils in a straight line and the erasers in a triangle shape without having the straight lines cross the triangles. [Oh, and here is something else about the Tever game, and this is important, so make sure you’re paying attention. This game is not for everyone. It’s only for children who are really really smart. Only smart children can be good at this game.]

Note. Both games were presented in the “smart” format to half of the children.
### Table 11
The Proportions of Children's Choices of Boys as Teammates in Study 6 (Standard Deviations in Parentheses)

<table>
<thead>
<tr>
<th>Age</th>
<th>Gender</th>
<th>Study 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Smart</td>
</tr>
<tr>
<td>5-year-olds</td>
<td>Boys</td>
<td>0.61 (0.31)</td>
</tr>
<tr>
<td></td>
<td>Girls</td>
<td>0.21 (0.21)</td>
</tr>
<tr>
<td>6-year-olds</td>
<td>Boys</td>
<td>0.75 (0.27)</td>
</tr>
<tr>
<td></td>
<td>Girls</td>
<td>0.30 (0.20)</td>
</tr>
<tr>
<td>7-year-olds</td>
<td>Boys</td>
<td>0.68 (0.22)</td>
</tr>
<tr>
<td></td>
<td>Girls</td>
<td>0.40 (0.23)</td>
</tr>
</tbody>
</table>
Fig. 1. Boys’ (blue) and girls’ (red) stereotype scores in Study 1 (A and B) and Study 2 (C and D), by age group (5- vs. 6- vs. 7-year-olds). The error bars represent ± 1 SE.
Fig. 2. Boys’ (blue) and girls’ (red) average proportions of selecting children of the same gender as having top grades in Study 2, by age group (5- vs. 6- vs. 7-year-olds). The error bars represent ± 1 SE.
Fig. 3. Boys’ (blue) and girls’ (red) interest (average of standardized responses to 4 questions) in novel games in Study 3 (A) and Study 4 (B). The main independent variable (task in Study 3, age in Study 4) is bolded. The error bars represent ± 1 SE.
**Fig. 4.** The difference between boys and girls in their interest toward the smart vs. the try-hard game was mediated by their own-gender brilliance scores. Standardized coefficients are depicted. *p < .05, + p < .10
Fig. 5. Average proportions of nominating males as job candidates for the Brilliance (Orange) and Dedication (Green) job in Study 5. The error bars represent ± 1 SE.
Fig. 6. Average proportions of boys suggested as teammates in Study 6, by Brilliance (Orange) vs. Baseline (Green) condition and the gender of participants. The error bars represent ± 1 SE.


Bian, L., Cimpian, A., Leslie, S-J., & Murphy, M. (under review). “You need to be a genius to do that!” Messages about Intellectual Talent Discourage Women from Pursuing Educational and Professional Opportunities.


differences in IQ at age 11: The Scottish mental survey 1932. *Intelligence, 31*(6), 533-542.


Nicholls, J. G. (1978). The development of the concepts of effort and ability, perception of academic attainment, and the understanding that difficult tasks require more ability. *Child
development, 800-814.


