THE DEVELOPMENT OF INTERGROUP BIAS

by

Antonya Marie Gonzalez

A.B., Washington University in St. Louis, 2013
M.A., University of British Columbia, 2015

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in
THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Psychology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

April 2019
© Antonya Gonzalez, 2019
The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

The Development of Intergroup Bias

submitted by Antonya Gonzalez in partial fulfillment of the requirements for

the degree of Doctor of Philosophy in Psychology

Examining Committee:

Dr. Andrew Baron
Supervisor

Dr. Toni Schmader
Supervisory Committee Member

Dr. Darko Odic
Supervisory Committee Member

Dr. Kristin Laurin
University Examiner

Dr. Laurie Ford
University Examiner

Additional Supervisory Committee Members:

Supervisory Committee Member

Supervisory Committee Member
Abstract

In adults, intergroup biases, such as racial attitudes and gender stereotypes, have been clearly linked to biased behavior. However, attempts to change intergroup bias in adulthood have been relatively unsuccessful, leading researchers to consider whether bias change might be more effective earlier in development. Indeed, children as young as age three show evidence of intergroup bias, and by age six, children have often internalized the biased attitudes and stereotypes of their culture. The following dissertation further examines the development of intergroup bias in order to understand how best to target bias change in childhood. First, in a series of three studies, I test an alternative method of measuring children’s implicit gender stereotypes called the Preschool Auditory Stroop in order to disentangle the specific associations that underlie implicit bias. The first two studies validate the use of this method and indicate that children as young as age three have implicit gender associations about the attributes and toys associated with boys and girls. The third study indicates that this method may be less likely to detect bias than category-based measures like the Implicit Association Test. Next, across four samples, I examine the effect of preschool children’s beliefs about math and gender on their math-related performance. Results conducted on the combined dataset indicate that while only a small number of girls have stereotypes associating math with boys, these girls perform significantly worse on a test of Approximate Number System accuracy when it is framed as a math test rather than a game or an eyesight test. Finally, the last set of studies examine the efficacy of a counter-stereotypical exemplar exposure intervention to reduce bias across development. As compared to adults, 5 to 12-year-old children appear to require less explicit instruction to change their bias. Taken together, this work provides novel insights into the nuanced development of intergroup bias and the malleability of bias in childhood.
Lay Summary

The early emergence of racial and gender bias has motivated researchers to find methods of reducing intergroup bias in childhood. In the following dissertation, I investigate a number of outstanding issues in the development of intergroup bias to further understand the best ways to change harmful attitudes and stereotypes. This dissertation shows that children have automatic stereotypes about gender as early as age three, and that the gender stereotypes children endorse at this age can significantly impact their behavior. Thus, even as early as preschool, intergroup bias is present and has harmful effects on children’s behavior. I also find that exposing children to counter-stereotypical exemplars can significantly reduce their automatic and unconscious racial bias, and that this method is easier to implement in children than adults. As such, this work suggests that to prevent bias from negatively affecting behavior throughout the lifespan, future interventions should target children.
Preface

I am the primary author of the work presented in this dissertation, which was conducted in collaboration with colleagues. I was responsible for study design, data collection, data analysis and drafting manuscripts. Additional contributions for each chapter are described below.

Chapter 1: Introduction

I am the primary author of this chapter, with contributions from my supervisor, A. S. Baron.

Chapter 2: Decoupling Implicit Associations (Studies 1-3)

A version of this chapter is being prepared for submission: Gonzalez, A. M., Block, K., Oh, H. J. J., Bizzotto, R. & Baron, A.S. Measuring implicit gender stereotypes in early childhood using the Preschool Auditory Stroop.

I am the primary author of this work: I designed the studies, supervised data collection, conducted data analyses, and drafted the manuscript. K. Block and A. S. Baron contributed to study design and data interpretation. H. J. Oh and R. Bizzotto collected data and helped with manuscript preparation. All authors edited the manuscript.

Chapter 3: Intergroup Bias and Behavior (Studies 4-7)

A version of this chapter is being prepared for submission: Gonzalez, A. M., Odic, D., Schmader, T., & Baron, A.S. Gender stereotypes impair preschool girls’ intuitive number sense.

I am the primary author of this work: I designed the studies, supervised data collection, conducted data analyses, and prepared the manuscript. All authors edited the manuscript and contributed to study design and data interpretation.
Chapter 4: Malleability of Intergroup Bias (Studies 8-9)

A version of this chapter is being prepared for submission: Gonzalez, A.M., Steele, J.R., Chan, E., Lim, S., & Baron, A.S. *Developmental differences in implicit racial bias change.*

I am the primary author of this work: I designed the studies, supervised data collection, conducted data analyses, and drafted the manuscript. J. R. Steele and A. S. Baron contributed to study design and data interpretation. E. Chan and S. Lim collected data and helped with manuscript preparation. All authors edited the manuscript.

Chapter 5: General Discussion

I am the primary author of this chapter, with contributions from my supervisor, A. S. Baron.

The research presented in this dissertation was approved by the UBC Behavioural Ethics Board under certificate H10-00147.
# Table of Contents

Abstract........................................................................................................................................... iii

Lay Summary ...................................................................................................................................... iv

Preface............................................................................................................................................... v

Table of Contents ............................................................................................................................. vii

List of Tables ...................................................................................................................................... xi

List of Figures ..................................................................................................................................... xii

Acknowledgements ............................................................................................................................ xiii

Chapter 1: Introduction ....................................................................................................................... 1
  1.1 Overview...................................................................................................................................... 1
  1.2 Implicit and Explicit Intergroup Bias ......................................................................................... 2
  1.3 Development of Intergroup Bias............................................................................................... 5
    1.3.1 Development of Gender Stereotypes .................................................................................. 6
    1.3.2 Development of Racial Attitudes ....................................................................................... 9
  1.4 The Current Research ............................................................................................................... 11

Chapter 2: Disentangling Implicit Associations .................................................................................... 16
  2.1 Synopsis...................................................................................................................................... 16
  2.2 Introduction................................................................................................................................. 16
  2.3 Study 1 ....................................................................................................................................... 22
    2.3.1 Method ............................................................................................................................... 22
      2.3.1.1 Participants .................................................................................................................. 22
      2.3.1.2 Procedure ..................................................................................................................... 23
    2.3.2 Results ............................................................................................................................... 25
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3.2.1</td>
<td>Reaction Time Scoring</td>
<td>25</td>
</tr>
<tr>
<td>2.3.2.2</td>
<td>D-Scoring</td>
<td>27</td>
</tr>
<tr>
<td>2.3.3</td>
<td>Discussion</td>
<td>30</td>
</tr>
<tr>
<td>2.4</td>
<td>Study 2</td>
<td>30</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Method</td>
<td>32</td>
</tr>
<tr>
<td>2.4.1.1</td>
<td>Participants</td>
<td>32</td>
</tr>
<tr>
<td>2.4.1.2</td>
<td>Procedure</td>
<td>33</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Results</td>
<td>33</td>
</tr>
<tr>
<td>2.4.3</td>
<td>Discussion</td>
<td>36</td>
</tr>
<tr>
<td>2.5</td>
<td>Study 3</td>
<td>36</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Method</td>
<td>37</td>
</tr>
<tr>
<td>2.5.1.1</td>
<td>Participants</td>
<td>37</td>
</tr>
<tr>
<td>2.5.1.2</td>
<td>Procedure</td>
<td>38</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Results</td>
<td>39</td>
</tr>
<tr>
<td>2.5.3</td>
<td>Discussion</td>
<td>40</td>
</tr>
<tr>
<td>2.6</td>
<td>General Discussion</td>
<td>41</td>
</tr>
</tbody>
</table>

Chapter 3: Intergroup Bias and Behavior ................................................................. 54

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Synopsis</td>
<td>54</td>
</tr>
<tr>
<td>3.2</td>
<td>Introduction</td>
<td>54</td>
</tr>
<tr>
<td>3.3</td>
<td>Method</td>
<td>57</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Participants</td>
<td>57</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Procedure</td>
<td>57</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Measures</td>
<td>58</td>
</tr>
<tr>
<td>3.4</td>
<td>Combined Sample Results</td>
<td>60</td>
</tr>
</tbody>
</table>
3.4.1 Math-Gender Beliefs ................................................................. 61
3.4.2 ANS Task Performance ............................................................ 61
3.5 Individual Study Results ............................................................... 64
  3.5.1 Study 4 Results ................................................................. 64
  3.5.2 Study 5 Results ................................................................. 65
  3.5.3 Study 6 Results ................................................................. 67
  3.5.4 Study 7 Results ................................................................. 70
3.6 Discussion .................................................................................. 71

Chapter 4: Malleability of Intergroup Bias ......................................... 82
  4.1 Synopsis .................................................................................... 82
  4.2 Introduction ............................................................................... 82
  4.3 Study 8 .................................................................................... 88
    4.3.1 Method ............................................................................... 88
      4.3.1.1 Participants ................................................................. 88
      4.3.1.2 Procedure ................................................................... 90
      4.3.1.3 Measures ................................................................... 91
    4.3.2 Results ............................................................................... 93
      4.3.2.1 Time 1 Analyses .......................................................... 93
      4.3.2.2 Change from Time 1 to Time 2 ..................................... 95
    4.3.3 Discussion ......................................................................... 97
  4.4 Study 9 .................................................................................... 97
    4.4.1 Method .............................................................................. 99
      4.4.1.1 Participants ................................................................. 99
      4.4.1.2 Procedure .................................................................. 100
List of Tables

Table 2.1 List of words used in Study 1. ........................................................................................................ 45
Table 2.2 List of words used in Study 2. ........................................................................................................ 46
Table 2.3 List of words used in Study 3. ........................................................................................................ 47
Table 2.4 PAS and IAT D-Scores by gender (Study 3). ................................................................................. 48
Table 3.1 Differences between Study 4-7....................................................................................................... 74
Table 3.2 Means and standard deviations for beliefs and performance......................................................... 75
Table 3.3 Table of coefficients by experiment and analysis type................................................................. 76
List of Figures

Figure 2.1 Average reaction times for older children (ages 6-7) using Most criteria................. 49
Figure 2.2 Average reaction times for younger children (ages 3-4) using Most criteria.......... 50
Figure 2.3 Mean PAS D-scores by type (Study 1)................................................................. 51
Figure 2.4 Mean PAS D-scores by type (Study 2)................................................................. 52
Figure 2.5 Mean PAS and IAT D-scores (Study 3)................................................................. 53
Figure 3.1 Example of two trials from the Approximate Number System (ANS) Task......... 79
Figure 3.2 Girls’ ANS task performance by condition......................................................... 80
Figure 3.3 Boys’ ANS task performance by condition......................................................... 81
Figure 4.1 IAT D-Scores for Children and Adults at Time 1 (Study 8)................................. 109
Figure 4.2 IAT D-Scores for Children and Adults at Time 2 (Study 8)................................. 110
Figure 4.3 IAT D-Scores for Adults (Study 9)................................................................. 111
Acknowledgements

As any graduate student can attest, it is impossible to complete a dissertation without the guidance and support of others. Firstly, I would like to thank my advisor, Dr. Andrew Baron, for his advising throughout my graduate career. Without your mentorship, I would not have been able to achieve my goals as a scholar. Thank you for encouraging me to elevate my thinking, work harder, and have confidence in my own abilities. Though I still have much to learn, your guidance has helped me to feel accomplished in what I have achieved thus far and enthusiastic to continue my scholarship journey in future years. I would also like to thank the members of my dissertation committee: Dr. Toni Schmader and Dr. Darko Odic. I truly appreciate the support and encouragement you have provided over the past several years, and for your mentorship, which has made me a better scholar. Additional thanks to Dr. Toni Schmader for welcoming me into her lab and for being a research role model over the years. I would also like to thank the collaborators on this work. First, I would like to thank Dr. Jennifer Steele, who has been a supportive mentor from afar and has helped guide my research goals and interests. Additionally, I would like to thank my good friend and collaborator Katharina Block, whose accomplishments have motivated me throughout my graduate career. I would also like to thank my research assistants, particularly Julie Oh, Evelyn Chan, Riley Bizzotto, and Sarah Lim, who went over and above the call of duty to help collect the data presented in this dissertation. Finally, I would like to thank my family: my mother, my father, and my brother, who support me from across the country, and my partner Tristan, for his daily encouragement. Thank you for always believing in me and supporting me through the ups and downs. I am grateful for your presence in my life.

The research presented in this dissertation was funded by a grant from the Social Sciences and Humanities Research Council to A. S. Baron and a Vanier Canada Graduate Scholarship to A. M. Gonzalez.
Chapter 1: Introduction

1.1 Overview

Human existence is rife with instances of intergroup prejudice and discrimination that often stem from beliefs about social categories such as race, gender, religion, and nationality. From a social cognitive development perspective, these beliefs stem from two broad sources: cognitive processes and cultural input. In order to make sense of the world around us, we are predisposed to form associations between social groups and attributes (e.g. Bigler & Liben, 2007; Liberman, Woodward, & Kinzler, 2017). Even in infancy, we show preferences for unknown individuals based on race and gender, two categories that are particularly privileged in our social evaluations (e.g. Bar-Haim, Ziv, & Lamy, 2006; Kinzler, Shutts, & Correll, 2010; Quinn, Yar, Kuhn, Slater, & Pascalis, 2002). By age six, we have internalized social group associations from our culture and community, and express attitudes and stereotypes based on race and gender (e.g. Baron & Banaji, 2006; Bian, Leslie, & Cimpian, 2017; Cvencek, Meltzoff, & Greenwald, 2011; Raabe & Beelmann, 2011).

The early emergence of intergroup bias has led to recent interest in understanding its developmental trajectory, and the potential for bias change in childhood. This interest is practically motivated by a desire to prevent intergroup bias from affecting behavior across the lifespan, and theoretically motivated by recent work suggesting that adulthood may not be an optimal period in development to change bias (e.g. Baron, 2015; Gonzalez, Steele, & Baron, 2017). The purpose of this dissertation is to further examine the development of intergroup bias starting from its emergence in childhood, with the ultimate goal of developing interventions to reduce harmful biases. Specifically, I will present three papers that explore essential contributions to our understanding of intergroup bias change: the content of intergroup associations, the effect of these associations on behavior, and the malleability of intergroup bias.
In introducing this dissertation, I will provide theoretical background for the conceptualization of implicit and explicit forms of bias and a review of our current understanding of the development of intergroup bias. The introduction will conclude by motivating the research conducted in this dissertation and providing an overview of subsequent chapters.

1.2 Implicit and Explicit Intergroup Bias

Despite an increased desire for social equality, our society is plagued with systemic bias and inequality that affects the opportunities and daily experiences of members of marginalized groups. For example, a recent poll (2017) conducted in the United States found that 43% of women said they had experienced discrimination based on gender. These numbers were even more striking for racial minorities; 52% of Latinx and 71% of Black Americans surveyed said they had experienced discrimination (Pew Research, 2016). As this continued discrimination and prejudice is linked to ongoing economic and health disparities by race and gender, as well as interpersonal violence toward members of marginalized groups, researchers have sought to investigate the psychological underpinnings of intergroup bias.

Though the presence of bias is most apparent when it is expressed verbally, more subtle forms of bias also shape behavior without our conscious knowledge. These automatic, less conscious forms of bias can be conceptualized as implicit bias, which is a more difficult form of bias to control than its explicit counterpart. Many researchers conceptualize intergroup attitudes and stereotypes using this dual process model (Greenwald & Banaji, 1995; Greenwald & Banaji, 2017), and studies examining intergroup bias through this lens indicate that implicit and explicit bias are distinct constructs. Numerous studies have validated measures of implicit and explicit bias and shown that they tap into distinct associations and differentially predict biased behavior (e.g. Greenwald et al., 2009). For example, higher levels of explicit bias have been shown to
negatively affect interracial interactions through verbal expressions, while implicit bias affects non-verbal behavior signals (e.g. Dovidio, Kawakami, & Gaertner, 2002; McConnell & Leibold, 2001). In addition to predicting distinct behaviors, implicit and explicit bias have a low positive correlation, suggesting that while these constructs are related, they can diverge (Hofmann, Gawronski, Gschwender, Le, & Schmitt, 2005; Nosek, Banaji, & Greenwald, 2002). Accordingly, it is possible for implicit and explicit biases to contradict one another, particularly when biases are viewed as socially inappropriate, as in the case of racism, sexism, and other forms of social group bias (Raabe & Beelmann, 2011). As such, it is critical for researchers to examine optimal methods to change both implicit and explicit forms of bias.

Both forms of intergroup bias have been directly linked to prejudiced and discriminatory behavior toward members of marginalized groups. Intergroup bias has been shown to affect the quality of interactions, and influence behaviors such as hiring, voting, and medical treatment of outgroup members (Green et al., 2007; Greenwald, Poehlman, Uhlmann, & Banaji, 2009). For example, research has shown that subtle stereotypes about women’s competency predict bias in hiring selections; people who stereotype women as less competent are less likely to hire a woman for a job position (Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012). In general, bias is more likely to shape behavior toward others when it can either be “justified” or fails to be “suppressed” (Crandall & Eshleman, 2003). For example, in contexts where biased behavior is strongly discouraged based on cultural norms, the relationship between bias and behavior may be weaker than in contexts where these norms are less prevalent.

Concerningly, intergroup bias can also limit opportunities by constraining the behavior of members of marginalized groups. This relationship has been shown robustly through a phenomenon called stereotype threat, which occurs when reminders of negative in-group stereotypes lead to underperformance on stereotype-relevant tasks (Steele, 1997). A number of
studies have shown that stereotypes about race and academic ability can impair Blacks’ performance on tests (e.g. Blascovich, Spencer, Quinn, & Steele, 2001; Steele & Aronson, 1995). Similar effects have been found for women; after reminders of stereotypes associating math more with men, women tend to underperform on math assessments (e.g. Nguyen & Ryan, 2008; Spencer, Steele, & Quinn, 1999; Schmader & Johns, 2003; Walton & Spencer, 2009). Recent work has also shown that these stereotypes not only affect performance in stereotyped domains, but also affect individual interest (e.g. Cheryan, Master, & Meltzoff, 2015; Shapiro & Williams, 2012). As such, members of stereotyped groups may be less likely to enter certain domains, constraining their opportunities and decisions.

These detrimental consequences have motivated recent interventions to change intergroup bias. Specifically, since implicit bias can influence behavior without conscious knowledge, researchers have focused on trying to change this type of bias. However, it appears that implicit bias change in adulthood is often short-lived and does not always correspond with behavioral change (Forscher et al., 2016; Forscher, Mitamura, Dix, Cox, & Devine, 2017; Lai et al., 2016). For example, in a recent comparison of nine interventions to reduce implicit racial bias, researchers found that all nine interventions successfully reduced bias immediately after implementation. However, after a delay of several hours, bias magnitude returned to pre-intervention levels (Lai et al., 2016). Another meta-analysis of almost 500 studies showed that while interventions can successfully change implicit and explicit bias in adults short-term, bias change did not lead to behavioral change (Forscher et al., 2016).

A possible explanation for these findings is that implicit associations may be too rigid to change; perhaps when these associations are acquired, they are permanently stored and inevitably activated in certain contexts. However, an alternative explanation is that implicit associations may be too rigid to change in adulthood, but less rigid at earlier points in
development. Adults have had a lifespan of reinforcement from cultural messages of bias, while in contrast, children have had much less exposure. Thus, it may be worthwhile to consider changing intergroup bias earlier in development, before implicit associations have been continually reinforced over many years. Similar to the build-up of dental plaque on teeth, this reinforcement of bias across the lifespan may make these biases and behaviors more difficult to “clean” away (Baron, 2015).

1.3 Development of Intergroup Bias

Given the evidence that adulthood may not be the optimal time in development to attempt to change intergroup bias, researchers have sought to understand the nature of bias across development. As early as preschool, children demonstrate explicit bias reflective of cultural stereotypes (e.g. Brown & Johnson, 1971, Doyle & Aboud, 1995; Martin & Ruble, 2004). Generally, explicit bias remains stable until middle childhood, and then often decreases across development (Raabe & Beelman, 2011; but see also Leitner, Hehman, & Snowden, 2018). This decrease in the expression of bias may be a result of impression management; in many cultures, as children get older, they learn that expressing bias is not acceptable. In conjunction with increasing executive function, which allows children to better inhibit prepotent responses (Carlson, Moses, & Claxton, 2004), these cultural messages may lead older children and adults to suppress their biases.

In contrast to explicit intergroup bias, which is affected by motivation and ability to suppress bias, implicit intergroup bias is less susceptible to these influences across development. Implicit racial attitudes and gender stereotypes are often conceptualized as developmentally invariant, with very little change in magnitude after biases are acquired (Baron, 2015). Indeed, adults have robust levels of implicit racial and gender bias (e.g. Nosek et al., 2009; Lai et al.,
2016), and the magnitude of these biases is comparable to children as young as age six (Baron & Banaji, 2006; Cvencek et al., 2011).

However, recent work suggests that cultural influences across development may lead to differences in the magnitude of children’s implicit bias (Gibson, Rochat, Tone, & Baron, 2017; Steele, George, Williams, & Tay, 2018; Williams & Steele, 2017). For example, racial socialization has been shown to influence the implicit ingroup bias of young adults (Gibson et al., 2017). Specifically, Black young adults who attend Historically Black Colleges, where Black identity is celebrated and positively reinforced, have stronger levels of implicit pro-Black preference than those who do not attend these colleges. Thus, while implicit bias is easily acquired and difficult to change, it is highly dependent on cultural information, which might result in variability in the developmental trajectory of implicit bias across cultural contexts (see Baron, 2015). This is further supported by evidence of regional and international variability in adult levels of implicit bias (Nosek et al., 2009; Payne, Vuletich, & Lundberg, 2017).

In summary, while explicit and implicit bias are both shaped by cultural context, as compared to explicit bias, there is little evidence that implicit bias decreases across development without intervention (see Chapter 4 for additional literature). As such, while both forms of bias require attention, the persistence of implicit bias after acquisition implicates a need to target these underlying associations. The following sections will detail the development of explicit and implicit gender and racial bias, and then present several outstanding issues in our understanding of the development of intergroup bias.

1.3.1 Development of Gender Stereotypes

In recent years, there has been increased attention to women’s underrepresentation in science and STEM-related fields, exemplified by social media movements like...
#thisiswhatascientistlookslike and government initiatives like 30 by 30. These forms of action seek to tackle a prominent social group bias within our culture; a gender bias that associates math and science more with men and boys than women and girls. The origins of these gender stereotypes are present by age two, when children are acutely sensitive to gender as a social category and begin to associate different attributes with boys and girls (Martin & Ruble, 2004). In preschool, children recognize their gender identity and use that identity to actively seek out activities and toys that are associated with their gender in-group. By age six, children have implicit and explicit beliefs about the abilities associated with boys/men and girls/women; research on explicit gender stereotypes has shown that children at this age believe boys are more likely than girls to be brilliant (Bian et al., 2017) and that boys are better at math and science than girls (Cvencek et al., 2011; Cvencek, Meltzoff, & Kapur, 2014; del Río & Strasser, 2013; Lummis & Stevenson, 1990; Master, Cheryan, Moscatelli, & Meltzoff, 2017).

Importantly, the developmental trajectory of implicit and explicit gender stereotypes is somewhat unclear due to inconsistent findings across development. For example, due to a competing in-group preference that leads children to believe their gender is better, children often explicitly state that their own gender is better at math and science, which may serve to mask awareness of gender stereotypes (see Régner, Steele, Ambady, Thinus-Blanc, & Huguet, 2014). As a result, while some studies show children’s explicit endorsement of gender stereotypes, in others, children believe their own group is academically superior. In contrast to explicit gender stereotypes, implicit stereotypes emerge relatively early and remain stable across development. Implicit gender stereotypes about academics appear to be present by ages six to ten in several developed countries, with both boys and girls associating boys more with math and girls more with reading (Cvencek et al., 2014; Passolunghi, Ferreira, & Tomasetto, 2014; Steffens, Jelenec, & Noack, 2010).
Across development, implicit gender stereotypes sometimes correspond with explicit gender stereotypes (Nosek et al., 2002). In a large-scale online study, researchers found that adults both implicitly and explicitly associated math more with men and liberal arts more with women. Interestingly, these stereotypes conflict with broader explicit beliefs, as many individuals profess a desire for gender equality (Pew Research, 2010). One reason that explicit stereotypes may sometimes persist in the domain of gender, is that gender differences may be viewed as more acceptable to discuss explicitly. For example, children in middle childhood believe gender is a function of meaningful group differences (Rogers & Meltzoff, 2017). As a result of the broader public acceptance of gender differences, it is possible that gender stereotypes may be particularly difficult to change, even in childhood. Further discussion of this possibility can be found in Chapter 5.

In childhood, stereotypes about gender have been linked to experiences of stereotype threat, similar to those found in adults. One of the first studies investigating the effects of stereotype cues on children’s performance tested Asian-American girls aged 5-7 and 11-13. This study examined whether contextually activating girls’ gender identity would remind them of gender stereotypes, and consequently, impair math performance. As compared to girls in a control condition (who colored neutral images), girls who colored images designed to prime female gender identity performed worse on a standardized math test, indicating that they had internalized gender stereotypes and were negatively affected when these stereotypes were activated (Ambady, Shih, Kim, & Pittinsky, 2001). This effect has been conceptually replicated in other studies with girls as young as age five, demonstrating that stereotype-based performance effects can emerge relatively early in development (e.g. Galdi, Cadinu, & Tomasetto, 2014; Muzzatti & Agnoli, 2007; Tomasetto, Alparone, & Cadinu, 2011). However, it is important to note that a number of studies have failed to find these effects, highlighting a need to further
explore the necessary conditions for this type of threat to occur (see Flore & Wicherts, 2015; Ganley et al., 2013). This issue is further discussed below and empirically tested in Chapter 3.

In summary, our current understanding of the development of gender stereotypes leaves open a number of questions. Firstly, as there is substantial variability in stereotype acquisition across different samples, it remains unclear what factors influence children’s internalization of gender stereotypes. Secondly, the trajectory of stereotypes about gender and academics remains confounded; the vast majority of studies examining these biases have measured a “boy=math” stereotype at the same time as measuring a “girl=reading” stereotype. As such, it is unclear whether children are acquiring one stereotype or the other, or both. While the majority of literature has framed the development of gender stereotypes as an acquisition of a “boy=math” stereotype, it is equally possible that young children might first acquire a “girl=reading” stereotype. Further discussion of this issue can be found below and in Chapter 2.

1.3.2 Development of Racial Attitudes

Racial bias is undoubtedly a prominent issue facing North American society; ongoing acts of prejudice against people of color underscore the need to address harmful racial attitudes and stereotypes. Concerningly, our tendencies of social categorization predispose us to show racial bias as early as infancy (Liberman et al., 2017). At three months, infants demonstrate a preference for individuals from familiar racial groups, indicative of their nascent ability to use race to tell people apart (Bar-Haim et al., 2006). Indeed, by age three, both implicitly and explicitly, children show a robust preference for individuals of their own race (Dunham, Chen, & Banaji, 2013; Raabe & Beelmann, 2011; Xiao et al., 2015). This in-group preference appears to be an evolutionary adaptation to identify one’s own group and associate it with positivity;
children’s implicit in-group bias emerges very quickly, and in relation to seemingly arbitrary group membership (Dunham, Baron, & Carey, 2011; Baron & Dunham, 2015).

An alternative perspective on the emergence of in-group racial bias, the Perceptual-Social Linkage Hypothesis, posits that this in-group preference is a result of experience. Since infants have more exposure to own-race faces across development, and these faces are often paired with positive experiences, infants form an in-group racial bias that continues into childhood (Lee, Quinn, & Pascalis, 2017). It is plausible that both perspectives are true, and that in-group racial preference in childhood may stem from both a cognitive bias to associate one’s own group with positivity (Dunham et al. 2011), as well as repeated exposure to in-group members paired with positivity (Lee et al., 2017).

Importantly, continued expression of in-group bias across development is dependent on cultural information about the status of one’s group. In contrast to children who are members of high-status racial groups, children who are members of lower-status racial groups tend to show no in-group preference when their in-group is pitted against a group of higher status (Baron & Banaji, 2009; Dunham, Baron, & Banaji, 2007; Gibson, Robbins, & Rochat, 2015; Newheiser & Olson, 2012). Furthermore, in places where racial status disparities are particularly pronounced, children from low-status racial groups actually prefer the high-status racial group within their culture (Dunham, Newheiser, Hoosain, Merril, & Olson, 2014; Shutts et al., 2011).

Like in-group preference, this preference for high-status groups appears to stem from an evolutionary bias, as infants begin to demonstrate preference for high-status individuals as early as six months (Pun, Birch, & Baron, 2016). Specifically, infants assume that high-status individuals will prevail in social confrontations and are likely to have more resources (Enright, Gweon, & Sommerville, 2017; Mascaro & Csibra, 2012; Thomsen, Frankenhuis, Ingold-Smith, & Carey, 2011). This translates into social preference later in development, when preschool
children positively evaluate high-status individuals who have more resources (Shutts, Kinzler, Katz, Tredoux, & Spelke, 2016; Horwitz, Shutts, & Olson, 2014).

Thus, as development progresses, implicit racial attitudes are influenced by two distinct preferences: a preference for one’s in-group and a preference for higher-status racial groups within one’s culture (Baron, 2015; Dunham, Baron, & Banaji, 2008; Dunham, Chen, & Banaji, 2013; Newheiser, Dunham, Merrill, Hoosain, & Olson, 2014; Raabe & Beelmann, 2011). Explicitly, these preferences decrease after middle childhood: between ages eight and ten, explicit preference for high-status in-groups decreases (Raabe & Beelmann, 2011). In contrast, the implicit form of this bias persists across development if cultural messages about racial status remain stable (e.g. Gibson et al., 2017; Steele et al., 2018). As such, individuals who are not explicitly prejudiced in later childhood and adulthood may still have implicit racial bias.

As racial bias is pervasive and relatively immune to societal norms, researchers have recently explored ways to successfully reduce implicit racial bias. This research has focused on adults, and a number of interventions, such as exposure to counter-stereotypical exemplars, have successfully reduced bias immediately after intervention implementation (see Lai et al., 2016). Recently, researchers have begun to refocus intervention efforts on children, and successful bias change has been observed with children ages 3-5 as well as age 9-12 (Gonzalez, Steele, & Baron, 2017; Qian et al., 2016). However, it remains unclear whether implicit racial bias can be changed in middle childhood (i.e. ages 6-8), or whether methods to change bias are more effective in childhood or adulthood. These questions are further detailed below and explored in Chapter 4.

1.4 The Current Research

Coupled with the detrimental consequences of intergroup bias in adulthood, evidence of the early emergence of racial and gender bias has led researchers to focus on developing
interventions to reduce bias in childhood. The developmental trajectory of racial attitudes and gender stereotypes suggest that these biases are in place by age six and persist across development. Furthermore, work with adults suggests that intergroup bias is relatively difficult to change later in development, potentially due to repeated reinforcement of bias. Taken together, this work makes a case for the development of interventions to reduce intergroup bias in childhood, as this might be the earliest and most optimal time to intervene in order to decrease the influence of bias on behavior across the lifespan. However, a number of outstanding issues limit our understanding of the development of intergroup bias, and further investigation of these issues is essential to the design of targeted and effective bias interventions.

Firstly, as a result of confounding implicit measures, our understanding of the developmental trajectory of distinct implicit associations remains relatively unclear. Explicit measures can easily disentangle distinct associations by asking about only one association at a time (e.g. when testing an association of boys=math, one could ask “Do you think boys or girls are better at math?”). In comparison, implicit associations are more difficult to assess, and the most popular form of measuring these associations is through use of the Implicit Association Test (IAT; see Chapter 2). The IAT measures bias by pitting two associations against each other; for example, when measuring children’s stereotypes about boys and math, this stereotype is often measured in conjunction with a stereotype associating girls with reading. In order to appropriately design interventions to decrease implicit intergroup bias, we must first identify which association drives the bias effects observed in childhood. Chapter 2 of this dissertation will expand upon this issue further and provide a potential experimental method of disentangling and charting the trajectory of children’s implicit gender stereotypes.

Specifically, Chapter 2 consists of three studies seeking to disentangle distinct implicit associations. The studies detailed in this chapter adapt a method called the Auditory Stroop for
use with preschool children, and then extend this method to measure children’s stereotypes about math, reading, and gender (Most, Sorber, & Cunningham, 2007). By disentangling the distinct associations that are often conflated by measures like the Implicit Association Test, this method will allow researchers to examine which sources contribute to the development of specific implicit gender associations, as well as which of these associations predict biased behavior. Studies 1 and 2 test the validity of this method using well-known gender stereotypes, while Study 3 independently measures children’s gender associations with math and their gender associations with reading.

In addition to the lack of clarity concerning the content of implicit associations, very few studies have actually examined the relationship between intergroup bias and behavior in childhood. As a result, the conditions under which intergroup bias affects children’s behavior are relatively understudied. Past research suggests that children do indeed show biased behavior toward others (e.g. Aboud, 1993; Fishbein & Imai, 1993), and their own choices and behavior can be constrained when stereotypes are activated (Ambady, Shih, Kim, & Pittinsky, 2001; Galdi, Cadinu, & Tomasetto, 2014, Muzzatti & Agnoli, 2007; Tomasetto et al., 2011). However, no studies thus far have looked at the moderating role of children’s endorsement of stereotypes in order to directly link the magnitude of children’s intergroup bias to behavioral outcomes. Chapter 3 will examine this issue in greater depth using the case study of gender stereotypes, and their influence on young girls’ math-related performance.

In Chapter 3, in a series of four studies, I investigate the relationship between explicit stereotypes and children’s performance on a math-related performance measure. Studies 4-7 test whether children’s explicit stereotypes about math and gender predict their performance on a measure of Approximate Number System (ANS) accuracy, which is an innate cognitive capacity underlying formal mathematic reasoning (Chen & Li, 2014; Feigenson, Libertus, & Halberda,
2013; Halberda, Mazzocco, & Feigenson, 2008; Libertus, Odic, & Halberda, 2012; Starr, Libertus, & Brannon, 2013). These studies also examine the moderating role of contextual activation by measuring whether the relationship between bias and behavior is stronger under conditions of threat.

The final outstanding issue examined in this dissertation concerns the malleability of implicit bias across development. Past work with adults suggests implicit bias might be more malleable in childhood as compared to adulthood, but this remains an empirical question (Baron, 2015). To create developmentally appropriate interventions to reduce implicit bias, researchers must examine the efficacy of these interventions across development and understand the conditions that elicit bias change at different stages. Chapter 4 of the dissertation contains an in-depth discussion of the malleability of intergroup bias across development and an examination of differences between children and adults in the efficacy of implicit racial bias change.

In Chapter 4, I present two studies exploring the malleability of implicit intergroup bias across development. These studies seek to conceptually replicate earlier work showing that exposure to counter-stereotypical exemplars can change implicit racial bias in children, as well as examining whether increasing the racial salience of these exemplars can improve the efficacy of this manipulation with younger children (Gonzalez, Steele, & Baron, 2017). Further, to test whether brief interventions can induce bias change beyond the limits of priming effects (e.g. Roskos-Ewoldsen, Roskos-Ewoldsen, & Carpentier, 2009), these studies assess the magnitude of bias both immediately and an hour after exemplar exposure. Thus, Studies 8 and 9 seek to answer the following questions: (1) Can implicit intergroup bias be reduced across development using a brief intervention of exposure to counter-stereotypical exemplars? (2) Does this reduction in implicit intergroup bias last beyond immediate post-intervention testing in children and adults?
In summary, the research presented in this dissertation is motivated by a desire to change intergroup bias at the optimal point in development using effective and efficient interventions. The different issues tackled in each chapter of this dissertation make up important pieces of a larger puzzle to understand how best to change intergroup bias. Each of these theoretical contributions will form the foundations of a greater understanding of the development of intergroup bias, which in turn, will allow future research to identify age-appropriate interventions that target specific intergroup biases and interrupt their detrimental effects on behavior. In the fifth and final chapter of this dissertation, I discuss the limitations and implications of this work, as well as future directions for lines of inquiry.
Chapter 2: Disentangling Implicit Associations

2.1 Synopsis

The majority of research examining the development of intergroup bias has used the Implicit Association Test (IAT), which measures two associations simultaneously. As such, it is often unclear which association drives implicit bias at any given point in development. Charting the trajectory of distinct implicit associations is essential to our understanding of the development of intergroup bias and our ability to directly target harmful associations for bias change. The following three studies test the validity of the Preschool Auditory Stroop (PAS), a potential method of measuring distinct non-evaluative implicit gender associations. The first two studies demonstrate that 3 to 4-year-old and 6 to 7-year-old children show implicit gender stereotypes using this measure and are faster to respond when stereotypically feminine words are paired with female voices and stereotypically masculine words are paired with male voices. These results suggest that this methodology can be used to disentangle the gender stereotypes of children as early as preschool. The third study extends this method to separately examine children’s stereotypes about math and reading. While children showed evidence of stereotypes on the IAT, they did not show stereotypes on the PAS, suggesting that the categorical vs. exemplar-based nature of these two methods may lead to differences in the detection of stereotypes.

2.2 Introduction

Across cultures, gender is one of the earliest social categories that children represent, and one that children often privilege above other social categories (Kinzler et al., 2010; Martin & Ruble, 2004; 2010). As early as six months old, infants have rudimentary representations of gender, and match male and female voices to the faces of men and women (Fagan & Singer,
1979; Miller & Eimas, 1983). As development progresses and children enter preschool, they associate different toys and activities with boys and girls and prefer toys associated with their own gender (Kuhn, Nash, & Brucken, 1978; Leinbach, Hort, & Fagot, 1997; Weinraub et al., 1984). By the time children enter elementary school, they have acquired gender stereotypes about the roles and abilities of boys and girls; at age six, children believe that boys are more likely to be “really, really smart” and associate math and science domains more with boys than girls (Bian et al., 2017; Cvencek et al., 2011; Master et al., 2017).

Endorsement of these gender stereotypes has been shown to shape children’s goals and behaviors. Young girls and boys have values and career interests that are stereotypically associated with their own gender (Block, Gonzalez, Schmader, & Baron, 2018; Croft, Schmader, Block, & Baron, 2014; Weisgram, Bigler, & Liben, 2010). Furthermore, they pursue and avoid activities in accordance with stereotypes. For example, girls who associate boys with brilliance are more likely to avoid activities intended for smart people (Bian et al., 2017), and girls who endorse stereotypes about STEM express less interest in STEM-related activities such as robotics and computer programming (Master et al., 2017). In addition to constraining children’s choices, these stereotypes can impair children’s achievement in counter-stereotypical domains; when stereotypes about math are activated in test settings, young girls do worse on subsequent math assessments, demonstrating stereotype-congruent performance (Ambady et al., 2001; Neuville & Croizet, 2007; Tomasetto, Alparone, & Cadinu, 2011). The early emergence of gender stereotypes and their detrimental consequences highlight the need to understand when these biases develop and influence children’s academic and career choices.

Recently, researchers have focused on examining implicit stereotypes, as these associations are less susceptible to social desirability bias (see Greenwald & Banaji, 2017), and predict behavior distinctly from explicit stereotypes (Greenwald, Poehlman, Uhlmann, & Banaji,
As early as age six, children’s implicit biases are representative of cultural stereotypes about gender, particularly in the domain of math, where children associate math more with boys than girls (Cvencek et al., 2011; 2014; Passolunghi, Ferreira, & Tomasetto, 2014; Steffens, Jelenec, & Noack, 2010). As such, it is worthwhile to investigate children’s implicit stereotypes in conjunction with their explicit stereotypes in order to understand the distinct developmental trajectories of these biases and their influence on behavior.

The majority of studies examining the development of implicit gender stereotypes have used the Implicit Association Test (IAT) to measure children’s implicit associations. During the IAT, participants must categorize images or words related to a target group and a comparison group (e.g. boy/girl) as well as two contrasting concepts (e.g. math/reading; Greenwald, McGhee, & Schwartz, 1998). On the first set of test trials, participants decide if each stimulus belongs to a category on the left (e.g. boy or math) or on the right (e.g. girl or reading). On the second set of test trials, participants categorize stimuli again, but with the category-concept pairing reversed (e.g. girl or math on left, boy or reading on right). Thus, difference scores generated from the IAT give an overall score of how much one pairing of children’s category-concept associations (e.g. boy=math/girl=reading) compares to the opposite pairing (e.g. girl=math/boy=reading).

While this method is undoubtedly informative, it fails to decouple the influence of distinct associations. For example, while the IAT might tell us that a White participant associates White people with positivity and Black people with negativity more strongly than they associate White people with negativity and Black people with positivity, it is unclear whether this is driven by ingroup positivity or outgroup negativity (see Brewer, 1999; Hewstone, Rubin, & Willis, 2002). Research on children’s implicit attitudes suggests that decoupling these associations tested by the IAT is crucial to understanding the nature of bias and bias change (Williams &
Distinct implicit associations may have very different developmental trajectories; recent work shows that children display ingroup favoritism before they begin to display outgroup derogation, and this ingroup positivity emerges distinctly from outgroup negativity in infancy (Buttelmann & Bohm, 2014; Pun et al., 2017). Knowing the independent developmental trajectory of implicit associations can help us to understand the role of different sources of bias and understand how distinct associations contribute to biased behavior.

Methods such as the Affective Misattribution Procedure (AMP) and the Implicit Racial Bias Test (IRBT) present alternatives to the IAT that can help to decouple implicit attitudes. In the child-friendly AMP, participants are presented with inkblots, and must decide whether the inkblot is “nice” or “not so nice” (Williams & Steele, 2017). Before presentation of each inkblot, a picture of an individual from one of the target racial groups is shown. The premise of this procedure is that children will be primed by the stimuli they see and judge the inkblot in an affectively congruent manner. Using this method, researchers have found that while younger White majority children (5-8 years old) demonstrate ingroup positivity, this preference was not present in older children (9-12 years old) (see also Degner & Wentura, 2010). Further, neither older nor younger children had significant levels of outgroup negativity. However, when levels of implicit racial bias were tested using the IAT, participants had significant levels of pro-White bias. These results suggest that in childhood, implicit racial bias may be driven by positivity toward ingroup and high-status racial groups rather than outgroup and low-status negativity.

The other method that has been used to test children’s implicit racial bias, and that has the potential to disentangle children’s ingroup positivity and outgroup negativity, is the Implicit Racial Bias Test (IRBT) (Qian et al., 2016; 2017). During this task, children are presented with a face from either their own racial group or a racial outgroup in the center of an iPad screen. Beneath each face, there is a smiling face and a frowning face that children are instructed to
press on certain trials. For the congruent trials, children are instructed to touch a smiling face when they see a same race face and a frowning face when they see an other race face. During the incongruent trials children are told to do the opposite. Unlike the Implicit Association Test (IAT) this method only requires children to learn and attend to one association at a time, making the task simpler for young children. The scoring of this method is based on the D-score, or difference score, calculated for the Implicit Association Test. This calculation uses the difference between congruent and incongruent trials divided by their combined standard deviation to produce a standardized difference score between trials, similar to Cohen’s d. Use of a D-score allows for comparisons between participants who have different overall reaction times, which is particularly useful when comparing children across development (Greenwald, Nosek, & Banaji, 2003).

While these methods can be used to disentangle implicit evaluative associations, thus far, they have not been adapted to examine non-evaluative associations and their distinct components. When measuring children’s stereotypes about gender, such as the cultural stereotype associating math more with boys than girls, the vast majority of implicit studies thus far have used the IAT, which pits a stereotype associating math with boys against a stereotype associating reading with girls (e.g. Cvencek et al., 2011). Thus, while children appear to have a stereotype associating math with boys, it is unclear how much of their bias is driven by a stereotype associating reading with girls. As the vast majority of efforts to change children’s academic gender stereotypes have focused on increasing girls’ engagement with math, it is important to consider that these gender divisions in academic interest and achievement may partially stem from boys failing to engage with reading and related subjects (Andre, Whigham, Hendrickson, & Chambers, 1999; Martinot, Bagès, & Désert, 2012).
One option for disentangling biases measured by the IAT is the use of the Quadruple Process Model (Quad Model), which uses modeling to examine four distinct processes involved in an IAT response: the automatic activation of an association, the ability to determine a correct response, the success at overcoming automatically activated associations, and the influence of a general response bias (Conrey, Sherman, Gawronski, Hugenberg & Groom, 2005). Using multinomial modeling, the probability that each of these processes is activated in a response can be estimated, and the observed data can be compared against this estimate. By fitting the quad model to the data and teasing apart the distinct associations that are activated automatically, researchers can estimate the strength of participants’ distinct implicit biases. However, this method depends only on error rates, and does not consider response latencies. As such, this modeling method may not be optimal for use with children, as it does not account for the biases of children who take longer to respond, or even children whose error rates are more reflective of developmental differences than bias.

A method that presents a potential solution to disentangling gender stereotypes is the Auditory Stroop, which operates on the principle of cognitive interference when attending to one feature over another (Most et al., 2007). During this task, participants hear words in a male or female voice and must identify the gender of the voice. The results of the original study indicated that both adults and children ages 7-8 were slower to identify the gender of the voice when the content of the word conflicted with gender stereotypes. As children only categorize based on two dimensions during this task (male/female), it is possible to compare reaction times to the different types of words children hear, allowing for quantification of distinct biases. Moreover, young children might find this method easier than tasks like the IAT, where they must remember four different categories. As this method does not require categorization of words, it also lends
itself to testing implicit associations for more complex categories that children may find difficult to lexicalize.

The first study of the present paper aims to adapt this technique in order to measure preschool-aged children – the youngest age when children have been shown to report implicit intergroup bias. Our first goal was to conceptually replicate the results of the original Auditory Stroop with children ages 6-7 years old. Our second goal was to test the effectiveness of the Preschool Auditory Stroop (PAS) with younger children (ages 3-4 years old) and examine potential developmental differences. To test the face validity of this measure, we chose to measure children’s reaction times in response to words associated with common gender stereotypes. If young children respond more quickly when the gender stereotypicality of the word matches the gender of the voice, this success with overt gender stereotypes would suggest that this method could be used to investigate more subtle gender associations. Accordingly, we hypothesized that both early elementary school and preschool children would be faster to categorize the voice gender when the word content was stereotypically congruent (e.g. “pink” spoken in a “girl” voice), as opposed to stereotypically incongruent (e.g. “football” spoken in a “girl” voice).

2.3 Study 1

2.3.1 Method

2.3.1.1 Participants

Our sample included 228 participants: 114 4-year-olds (59 females, \( M=4.07 \) years, \( SD=0.55 \)) and 114 7-year-olds (58 females, \( M=6.92 \) years, \( SD=0.55 \)). Our goal was to recruit 60 children per gender and age group, and we stopped testing when we believed we had met that goal. Participants were recruited from a community-based science center and tested onsite in a
A room dedicated to behavioral science research. A legal guardian provided consent for all participants. Separately, 72 children were excluded (62 3-4-year-olds and 10 6-7-year-olds) for failing to complete the task \( N = 52 \), randomly pressing keys throughout the study \( N = 6 \), computer errors \( N = 4 \), experimenter error \( N = 4 \) or having an error rate below chance levels \( \leq 50\%; N = 6 \). This exclusion rate (approximately 24%) is consistent with other studies conducted with developmental populations in museum settings (see Gonzalez, Dunlop, & Baron, 2017).

All participants included in the study spoke English at least 30% of the time in their daily lives, as reported by parents. We established this cut-off because much of our local population is multilingual. 62.3% \( (N=142) \) of our sample identified as Caucasian, 16.2% \( (N=37) \) identified as Mixed or multiple ethnicities, 14.5% \( (N=33) \) identified as East Asian or Pacific Islander. Out of the remaining 6.9% \( (N=16) \), six participants identified as South Asian, three identified as First Nations or Aboriginal, three identified as Middle Eastern, three identified outside of the options provided, and one participant identified as Latino. Overall, the population of visitors to Science World has an median income of $75,000 CAD per year. Approximately 85% percent of parents who visit this location have a high school education or higher, and 57% have received a university education or higher.

2.3.1.2 Procedure

Preschool Auditory Stroop (PAS). We adapted the Auditory Stroop used by Most and colleagues (2007) to make it more child-friendly and easier to use with preschool aged children. Specifically, we adapted the word list from Most et al. (2007) by selecting four words from each category that should be most easily recognizable by preschool children: 4 stereotypically masculine (“baseball”, “football”, “rough”, “tough”), 4 stereotypically feminine (“lipstick”,...
“makeup”, “pretty”, “pink”) and 8 neutral words (4 practice words: “apple”, “door”, “draw”, “paper” and 4 test words: “pencil”, “spoon”, “table”, “window”; see Table 2.1). Each word was recorded in an adult male and female voice matched in similar age and accent (native English speakers). Pilot testing indicated that children were able to discern the male and female voices accurately.

To further aid in young children’s ability to complete the task, two JellyBean® response buttons each affixed with a sticker of a cartoon image of a girl or a boy were placed in front of a computer screen. Before the task, children were informed that they would hear words and that each word would be spoken by either a “girl” or a “boy”. Children were asked to identify the gender of the voice by pressing the “girl” button if they heard a female voice and the “boy” button if they heard a male voice. They were also instructed to respond as quickly as possible. The computer screen remained blank during each trial. If a participant incorrectly identified the gender of the voice, a red X appeared on the screen to indicate that they had made an error. To continue, they had to press the correct button. Between trials, a fixation cross was shown in the center of the screen for 1500ms.

Children first completed 10 practice trials in which children heard one of four neutral words spoken by either a male or a female voice (each word was presented 2-3 times in random order). On each trial, children identified the gender of the voice by pressing the corresponding button. Feedback on the accuracy of each response was provided by the experimenter. If the child initially struggled with identification of the voices, the experimenter was allowed to guide them through the practice trials.

Before the test trials, the experimenter reminded the child of the instructions, and told them that they only needed to identify the gender of the voice on each trial. The manner of presentation for the test trials was the same as the practice trials except error feedback was not
provided by the computer following each response, and the experimenter did not engage with the child during this portion of the study. For each test trial, children heard each of twelve different words (four gender neutral, four feminine-stereotypical and four masculine-stereotypical) presented twice in a male and twice in a female voice, resulting in 48 test trials (16 trials presented in gender counter-stereotypical voice (e.g., male voice saying “pink”, female voice saying “football”); 16 trials presented in a gender stereotypical voice (e.g., female voice saying “pink”, male voice saying “football”); and 8 trials of gender-neutral words (e.g., male and female voice saying “apple”).

2.3.2 Results

We first conducted our analyses using the criteria established by Most et al. (2007) to test whether or not we conceptually replicated their results. This scoring procedure reports average reaction times by trial type.

Second, we used a scoring procedure based on the one used by Qian et al. (2016) to score the Implicit Racial Bias Test (IRBT), as this method follows more recent recommendations for analyzing reaction time data in implicit tasks. This scoring procedure reports a D-score which has been used to score the well-established Implicit Association Test (Greenwald et al., 2003). See Equations 1a-1c below for exact calculations used.

2.3.2.1 Reaction Time Scoring

As described by Most et al. (2007) in their original test of the Auditory Stroop, trials were classified into three distinct types: congruent, neutral, and incongruent. Congruent trials were defined as trials where the gender of the voice and the stereotypicality of the word were congruent (e.g. girl voice saying pink), neutral trials were trials where word content was neutral (e.g. girl voice saying table), and incongruent trials were trials where word content was
incongruent (e.g. girl voice saying football). Response times less than 400 milliseconds were excluded from the dataset. Furthermore, response times greater than three standard deviations from each participants’ means were excluded separately for congruent, neutral, and incongruent trials. Additionally, trials where participants made errors in voice classification were excluded. From the remaining trials, we computed mean reaction time scores for each trial type (congruent, incongruent, neutral) for each participant.

**Older children (age 6-7).** To examine potential differences by trial type, we conducted a mixed factorial ANOVA with Trial Type entered as a within-subjects variable, and Child Gender entered as a between-subjects variable. Child Gender did not affect children’s reaction times on the task, $F(1,112) = 0.84, p = 0.360, \eta^2 = .007$, nor did it interact with Trial Type to predict reaction times, $F(2,224) = 0.29, p = .752, \eta^2 = .003$. However, Trial Type did have a significant effect on children’s reaction times, $F(2,224) = 35.57, p < .001, \eta^2 = .24$ (see Figure 2.1).

Simple effects analysis revealed that children were significantly faster to respond to trials when word type was stereotypically congruent with voice gender, as compared to trials when word type was stereotypically incongruent with voice gender ($p < .001$) or when word type was neutral ($p = .001$). Furthermore, children were significantly faster to respond to trials when word type was neutral, as compared to incongruent trials ($p < .001$). These results replicate the pattern of results found by Most et al. (2007) and indicate that this adjusted method captures children’s implicit gender stereotypes at this age.

**Younger children (age 3-4).** Once again, for younger children, we conducted a mixed factorial ANOVA with Trial Type entered as a within-subjects variable, and Child Gender entered as a between-subjects variable. Child Gender did not affect children’s reaction times on the task, $F(1,112) = 0.06, p = .801, \eta^2 = .001$, nor did it interact with Trial Type to predict
reaction times, $F(2,224) = 0.33, p = .717, \eta^2 = .003$. However, Trial Type did have a significant effect on children’s reaction times, $F(2,224) = 13.52, p < .001, \eta^2 = .11$ (see Figure 2.2).

Simple effects analysis revealed that children were significantly faster to respond to trials when word type was stereotypically congruent with voice gender, as compared to trials when word type was stereotypically incongruent with voice gender ($p < .001$) or when word type was neutral ($p = .001$). The difference in response times between neutral and incongruent trial types was marginally significant ($p = .11$). However, the overall difference between congruent and incongruent word trials suggests that children as young as age three have implicitly internalized the stereotypes tested here, and are faster to respond when the stereotypicality of the word matches the voice gender.

2.3.2.2 D-Scoring

**Analytic Approach.** In addition to the analyses reported above, we conducted analyses using D-scores (or difference scores, which reflect more recent recommendations for analyzing implicit reaction time data in children (Baron & Banaji, 2006; Qian et al., 2016). A D-score captures the magnitude of an implicit association by comparing trial types when different concepts are paired together (e.g. girl=pink vs. boy=pink). In addition, using a D-score allows us to compare the magnitude of bias across experiments. D-scores were calculated by computing the difference between average response latencies between different voice and word types, divided by the standard deviation of response latencies across all trials. Unlike the scoring procedure used by Most et al. (2007), this method is less conservative; it does not exclude based on error and excludes mean reaction times greater than 10,000, rather than by excluding trials more than three standard deviations from the mean for each trial type. As such, we are able to include more trials for useable participants, thereby increasing our power.
In addition to adjusting the previous trial exclusion criteria, we also adopted the participant exclusion criteria used by Qian et al. (2016), which is based off of recent D-score calculation recommendations for excluding participants who did not complete the task accurately (Nosek, Bar-Anan, Sriram, Axt, & Greenwald, 2014). Participants were excluded if they had an error rate above 35%, responded to more than 10% of trials in 300 ms or less, or if their mean reaction time was greater than three standard deviations away from the sample mean. Fifteen younger children ages 3-4 were excluded based on these criteria, leaving a final sample of 99 (51 females, 48 males). No older children ages 6-7 were excluded.

In order to differentiate between biases regarding the stereotypically feminine words and the stereotypically masculine words, we calculated a D-score for each word type. This method is a departure from the D-score calculations used by Qian et al. (2016), who calculate one D-score for all trial types. However, we chose to calculate an overall D-score and disentangle these scores to explore whether gender bias was driven by masculine or feminine type words.

Our overall D-score was computed using the reaction time difference between congruent and incongruent trials. D-score for each word type (e.g. feminine or masculine) was computed using the reaction time difference between trials with a female voice and trials with a male voice. We subtracted the consistent trials (e.g. girl voice saying “pink”) from the inconsistent trials (e.g. boy voice saying “pink”) so that a positive D-score would indicate that children were responding faster to gender-consistent trials, and a negative D-score would indicate that children were responding faster to gender-inconsistent trials.

Older children (age 6-7). We first examined potential differences by Child Gender. There were no gender differences in implicit bias magnitude for our overall D-score, \( t(112) = \)

\[ D_{overall} = \frac{RT_{incongruent} - RT_{congruent}}{(SD_{incongruent} + SD_{congruent})} \]
\[ D_{feminine} = \frac{RT_{male=feminine} - RT_{female=feminine}}{(SD_{male=feminine} + SD_{female=feminine})} \]
\[ D_{masculine} = \frac{RT_{female=masculine} - RT_{male=masculine}}{(SD_{female=masculine} + SD_{male=masculine})} \]
.38, \( p = .703 \), feminine words, \( t(112) = -.51, p = .615 \), or masculine words, \( t(112) = .27, p = .787 \), so we collapsed across gender for all subsequent analyses (see Figure 2.3).

As detailed above, the overall D-score examined whether children were faster to respond on congruent or incongruent trials. We found that this score was positive, suggesting that children were significantly faster to respond when word type was stereotypically congruent with voice gender, \( D = .22, t(113) = 7.43, p < .001 \).

To disentangle independent effects of feminine and masculine stereotypical words, we also looked at the separate D-scores. For feminine words, implicit bias was significantly above chance levels, \( D = .29, t(113) = 9.45, p < .001 \), indicating that older children were faster to respond when these words were paired with a female voice. For masculine words, implicit bias scores were also above chance, \( D = .21, t(113) = 7.43, p < .001 \), indicating that children were faster when these words were paired with a male voice. There was a marginally significant difference in the magnitude of these two scores, with children’s feminine stereotypes being stronger than their masculine stereotypes, \( t(113) = 1.85, p = .067 \).

Younger children (age 3-4). For younger children, there were also no significant differences in reaction times by Child Gender for our overall D-score, \( t(97) = .64, p = .527 \), feminine words, \( t(97) = 1.73, p = .087 \), or masculine words, \( t(97) = .836, p = .405 \), so we collapsed across gender for all subsequent analyses (see Figure 3.3).

Again, we found that the overall D-score was positive, indicating that children were significantly faster to respond when word type was stereotypically congruent with voice gender, \( D = .12, t(98) = 6.12, p < .001 \). For feminine words, implicit bias was significantly different from chance, \( D = .10, t(98) = 3.05, p = .003 \), indicating that preschool children were faster to respond when these words were paired with a female voice (as compared to a male voice). For masculine words, bias was again different from chance, \( D = .18, t(98) = 6.67, p < .001 \), indicating that
children were faster when these words were paired with a male voice. There was a marginally significant difference in the magnitude of these two scores, with children’s masculine stereotypes being stronger than their feminine stereotypes, $t(98) = 1.86, p = .066$.

**Developmental differences.** We compared D-scores between younger and older children to see if the magnitude of implicit bias increased across development. For feminine words, older children had significantly more bias than younger children, and associated them more strongly with the female voice, $t(211) = -4.15, p < .001$. For masculine words, there were no significant developmental differences, $t(211) = -0.65, p = .518$.

### 2.3.3 Discussion

In summary, our results replicated those of Most and colleagues (2007) and extended them to a younger age group. Children ages 3-4 and 6-7 were generally faster to respond when stereotypically feminine words were paired with female voices, and stereotypically masculine words were paired with male voices. Accordingly, they were also slower to respond when word type was stereotypically incongruent with voice type. There were also age differences in the magnitude of distinct stereotypes, with older children having significantly stronger bias for female-stereotypical words as compared to younger children. These results suggest that this method can be used to examine and disentangle the implicit gender stereotypes of children as young as age 3.

### 2.4 Study 2

The results of Study 1 indicate that children as young as age three respond faster when the voices they identify are stereotypically congruent with word content (e.g. faster when a girl voice says “pink” and slower when a boy voice says “pink”). Study 2 replicates these results
using computer-generated voices. Additional pilot studies using human voices indicated that children are highly sensitive to the pitch of male and female voices, and this can result in a response bias toward one voice type. For example, deeper male voices can lead children to respond more quickly to male voices, skewing the data toward slower responses for female voices and faster responses for male voices and potentially obscuring the detection of stereotypes. In addition to using computer-generated voices, we also test the use of alternative gender-neutral words (see Table 2.2).

Study 1 also indicated that older children had significantly stronger implicit feminine stereotypes than younger children, but this was not the case for masculine stereotypes. One possible reason this could be the case is that the words “lipstick” and “makeup” may not be as familiar to 4-year-olds. Indeed, these words are used quite infrequently compared to the other words from Study 1 (e.g. “pretty” and “pink”; Google Books, 2018). As such, we replaced these two words with two stereotypically feminine words from the Communicative Development Inventory (CDI). We specifically selected words that are produced by more than half of children by 30 months and are therefore likely to be known by three-year-olds (“doll” and “dress”; WordBank, 2018). We also changed the word “baseball,” as it is not as well-known to Canadian children, and the word “rough”, as it is acoustically similar to the word “tough”. These two words were also replaced with stereotypically masculine words from the CDI known by age three (“trick” and “blue”, WordBank, 2018).

In order to make this method more replication-friendly, we also use novel voice technology to control for voice pitch and speaking rate. Recent developments have led to text-to-speech generators that strongly resemble a human voice (Google Cloud, 2018). Consequently, this study replicates the results of Study 1 using voices generated by Google Text-to-Speech API. Words in the male voice were recorded using Wavenet Voice B, at a pitch of -0.50, and
words in the female voice were recorded using Wavenet Voice C at a pitch of +5.00. These voices were pilot tested using gender-neutral words with a sample of 60 children (30 ages 3-4 and 30 ages 6-7). Results of pilot testing indicated that children as young as age three respond to these voices at comparable rates ($p = .93$) and categorize them at an overall rate of 89% accuracy.

In addition to making the voices more consistent and replicable, we test a “blocking” procedure to facilitate children’s categorization of words. One major difference between the PAS and the IAT is that the PAS does not require that children categorize words. However, blocking these words together might lead to increased detection of stereotypes, particularly if researchers are interested in stereotypes related to an overarching category, rather than individual stimuli. To examine this possibility, half of participants take the PAS in the same manner as Study 1, while the other half have masculine and feminine words separated into blocks.

2.4.1 Method

2.4.1.1 Participants

Our sample included 142 participants: 68 4-year-olds (45 females, $M=4.22$ years, $SD=0.44$) and 72 7-year-olds (39 females, $M=6.76$ years, $SD=0.60$). We preregistered a recruitment goal of 70 children per age group, and we stopped testing when we believed we had met that goal. Participants were recruited from the same community-based science center using the same procedures as Study 1 (see Study 1 for typical sample information). A legal guardian provided consent for all participants. Separately, 54 children were excluded (40 3-4-year-olds and 14 6-7-year-olds) for failing to complete the task ($N = 39$), randomly pressing keys throughout the study ($N = 4$), speaking less than 30% English in daily life ($N = 2$), a language barrier identified by experimenters ($N = 11$) or having an error rate below chance levels ($\leq 50%;$
This exclusion rate (approximately 27%) is consistent with other studies conducted with developmental populations in museum settings (see Gonzalez, Dunlop, & Baron, 2017). Like Study 1, all participants included in the study spoke English at least 30% of the time in their daily lives. Out of the full sample, 43.6% (N=61) identified as Caucasian, 24.3% (N=34) identified as East Asian or Pacific Islander, 20.7% (N=29) identified as more than one race and/or ethnicity. Out of the remaining 11.5% (N=16), five participants identified as South Asian, five identified as Middle Eastern, and six identified as Latino.

2.4.1.2 Procedure

Preschool Auditory Stroop (PAS). The general procedure for the PAS was identical to Study 1 other than differences in the words used (see Table 2.2) and use of the blocking procedure for half of participants. For the participants in the blocking condition, during the test trials, the three different word types were divided into three separate blocks by type. The masculine-stereotypical and feminine-stereotypical blocks were counterbalanced with each other and always appeared with one before and one after the neutral block. Before each block of test trials, children were told “Now you’re going to hear some different words. Remember, if you hear the girl voice, press this button, and if you hear the boy voice, press this button.”

The other half of participants completed the procedure in an identical manner to Study 1.

2.4.2 Results

As preregistered on the Open Science Framework (https://osf.io/5jrkc/), we scored results using the D-score procedure detailed in Study 1 in order to look at children’s stereotypes about male and female stereotypical words separately. Again, participants were excluded if they had an error rate above 35%, responded to more than 10% of trials in 300 ms or less, or if their mean
reaction time was greater than three standard deviations away from the sample mean. Three younger children ages 3-4 were excluded based on these criteria, and two older children ages 6-7, leaving a final sample of 135 (80 females, 55 males).

**All children.** As preregistered, we first conducted a mixed-factorial ANOVA to see whether the magnitude of participant stereotypes differed based on word type (feminine-stereotypical, masculine-stereotypical, neutral), age (3 to 4-year-olds vs. 6 to 7-year-olds), and condition (blocking vs. non-blocking). Results indicated that there were no significant interactions between any of the three variables ($ps > .37$). There was no effect of age category, $F(1,130) = .045, p = .831, \eta^2 < .001$, or condition, $F(1,130) = .117, p = .733, \eta^2 = .001$, but there was a significant effect of word type, $F(2,260) = 15.02, p < .001, \eta^2 = .104$. Post-hoc simple effects analyses indicated that children responded significantly faster to feminine words ($M = .19$) as compared to masculine words ($M = .11; p = .020$) and neutral words ($M = -.02; p < .001$), and faster to masculine words as compared to neutral words ($p = .002$).

The overall $D$-score was positive, suggesting that children were significantly faster to respond when word type was stereotypically congruent with voice gender, $D = .14, t(134) = 8.85, p < .001$. To disentangle independent effects of feminine and masculine words, we also looked at the separate $D$-scores. For feminine words, implicit bias was significantly above chance levels, $D = .19, t(134) = 7.93, p < .001$, indicating that children were faster to respond when these words were paired with a female voice. For masculine words, implicit bias scores were also above chance, $D = .11, t(134) = 4.46, p < .001$, indicating that children were faster when these words were paired with a male voice.

We also compared the $D$-score for neutral words against chance to ensure that children were not faster to respond to either voice type. This $D$-score was not significantly different from
zero, $D = -.02$, $t(134) = 0.69$, $p = .490$, indicating that children did respond to the voices at comparable rates.

**Older children (age 6-7).** As preregistered, we also compared the D-scores for each age group against chance separately. For 7-year-olds, we found that the overall D-score was positive, suggesting that children were significantly faster to respond when word type was stereotypically congruent with voice gender, $D = .14$, $t(69) = 6.59$, $p < .001$. For feminine words, implicit bias was significantly above chance levels, $D = .21$, $t(69) = 6.18$, $p < .001$, indicating that older children were faster to respond when these words were paired with a female voice. For masculine words, implicit bias scores were also above chance, $D = .10$, $t(69) = 3.02$, $p = .004$, indicating that children were faster when these words were paired with a male voice.

**Younger children (age 3-4).** For 4-year-olds, we found that the overall D-score was positive, indicating that children were significantly faster to respond when word type was stereotypically congruent with voice gender, $D = .14$, $t(64) = 5.87$, $p < .001$. For feminine words, implicit bias was significantly different from chance, $D = .17$, $t(64) = 4.98$, $p < .001$, indicating that preschool children were faster to respond when these words were paired with a female voice (as compared to a male voice). For masculine words, bias was again different from chance, $D = .12$, $t(64) = 3.26$, $p = .002$, indicating that children were faster when these words were paired with a male voice.

**Developmental differences.** For exploratory purposes, we compared D-scores between younger and older children to see if the magnitude of implicit bias increased across development. There were no significant differences between younger and older children for feminine words, $t(133) = 0.89$, $p = .374$, or masculine words, there were no significant developmental differences, $t(133) = 0.59$, $p = .558$. 
2.4.3 Discussion

In summary, we found that children ages 3-4 and 6-7 had significant implicit associations between feminine words and female voices, as well as masculine words and male voices. As such, these results replicate the results of Study 1 using voices from the Google API. There were also no differences between the blocking and non-blocking conditions, suggesting that either method could be used to measure children’s implicit gender stereotypes. Furthermore, as children did not associate neutral words significantly with either voice gender, the voices used in this study can be used in future work. In contrast to Study 1, there were no age differences in this study in the magnitude of children’s implicit feminine stereotypes, potentially as a result of using more child-friendly feminine words. Based on these results, Study 2 further confirms the utility of this method with children as young as age 3.

2.5 Study 3

The results of Study 2 are comparable to those of Study 1 and suggest that children have implicit associations between feminine words and voices and masculine words and voices. Thus, it appears that children have implicit stereotypes about gender as early as age three.

Study 3 seeks to extend this work by examining stereotypes that are linked to children’s academic performance (Ambady et al., 2001; Neuville & Croizet, 2007; Passolunghi et al., 2014; Tomasetto et al., 2011). Past work has shown that as early as age six, children endorse a stereotype associating math more with boys than girls (Block, Gonzalez, Choi, Wong, & Baron, 2018; Cvencek et al., 2011; 2014; Passolunghi et al., 2014; Tomasetto et al., 2011). However, this research has used the Implicit Association Test, which tests the association between boys and math at the same time as it tests the association between girls and reading. Adults and children appear to have an explicit stereotype that girls are better at reading than boys (Andre et
and as such, it is plausible that children might also have an implicit stereotype associating girls with reading that partially drives their results on the IAT. Specifically, it is possible that children’s association between girls and reading is stronger than their association between boys and math, but due to the nature of the IAT, this particular pattern would be obscured. The current study will add to the existing literature by examining the developmental trajectory of these two distinct stereotypes in elementary school children.

Based off the results of Study 2, we used the blocking procedure, which produced comparable results to the non-blocking procedure, and allows us to group together math and reading words to facilitate children’s categorization of the different word types. In contrast to Study 2, they will be told explicitly that they are categorizing “math” and “reading” words before each block. However, in comparison to the IAT, this method still requires less categorization on each individual trial. As such, we compare children’s independent math and reading associations with their IAT scores to examine how gender stereotypes about math vs. gender stereotypes about reading may differentially predict bias on the IAT.

2.5.1 Method
2.5.1.1 Participants

Our sample included 48 participants (26 females, \(M=7.52\) years, \(SD=0.84\)). Power analyses using G*Power indicated that this sample size would give us greater than 90% power to detect a within-subjects effect. Participants were recruited from the same community-based science center using the same procedures as Study 1 and 2 (see Study 1 for typical sample information). A legal guardian provided consent for all participants. Separately, one child was excluded for failing to complete the task. Like Study 1 and 2, all participants included in the study spoke English at least 30% of the time in their daily lives. Out of the full sample,
47.9% \( (N=23) \) identified as Caucasian and 22.9% \( (N=11) \) identified as East Asian or Pacific Islander. Out of the remaining 29.2% \( (N=14) \), six participants identified as South Asian, three identified as Latino, and five identified as more than one ethnicity listed.

2.5.1.2 Procedure

**Preschool Auditory Stroop (PAS).** The general procedure for the PAS was identical to the blocking procedure used in Study 2 other than differences in the words used (see Table 3.3). The presentation of the PAS was counterbalanced with the Implicit Association Test.

**Child Implicit Association Test (IAT).** Implicit gender stereotypes were also measured using a child-friendly Implicit Association Test. This test measures the strength of an association between a target category and an attribute. In this IAT, we measured associations between gender (Boy/Girl) and academic subject (Math/Reading). Stimuli for the gender categories were cartoon images of boys and girls, which varied in skin tone, eye, and hair color to represent an ethnically diverse sample. The stimuli for the attribute categories consisted of the same words presented in the PAS (see Table 2.3). These words were presented using the female voice recorded from the Google API. A red “x” appeared whenever a stimulus was categorized incorrectly and disappeared once participants made the correct response. Participants were presented with two JellyBean buttons in front of the monitor that color matched with the side of the screen they were placed in front of (green on the left, red on the right). Participants were told that any time they see an image or hear a word, they should determine which category it belongs in, and press the correct button.

Participants began by categorizing images by gender in 12 practice trials. Next, participants categorized the math and reading words, and completed 20 of these practice trials. After practice trials, children completed a test block with 30 trials where they were presented...
with either an image or a word. To classify these stimuli, children used the same buttons for school subjects and gender (e.g. boy+reading and girl+math). After the critical test block, participants completed another practice block where the boy and girl images were categorized on opposite sides for 20 trials. Finally, participants completed another test block, where the pairing of the attributes and target categories was switched (e.g. boy+reading, girl+math for the first test block, girl+reading, boy+math for the second). Sides were counterbalanced across conditions.

2.5.2 Results

For the PAS, once again scored results using the D-score procedure in order to look at children’s stereotypes about masculine and feminine stereotypical words separately. Again, participants were excluded if they had an error rate above 35%, responded to more than 10% of trials in 300 ms or less, or if their mean reaction time was greater than three standard deviations away from the sample mean. One child was excluded based on this criteria.

For the IAT, we calculated D-scores following the procedures outlined by Greenwald, and colleagues (2003), which have been used extensively in developmental research (e.g. Baron & Banaji, 2006; Dunham, Baron, & Banaji, 2006, 2007; Gonzalez, Steele, & Baron, 2017). We also excluded children if they had an error rate above 35% or responded to more than 10% of trials in 300 ms or less (Nosek et al., 2014).

Preschool Auditory Stroop. We first examined the overall PAS D-score, which measures children’s overall level of bias for congruent (girl=reading, boy=math) and incongruent (girl=math, boy=reading) associations. We found that children did not seem to have implicit associations in either direction, \( D = -.01, t(46) = -1.70, p = .64 \) (see Figure 3.5).

We also looked separately at PAS D-scores for math and reading words. For math words, implicit bias was not significantly different from chance in either direction, \( D = .006, t(46) = \)
0.14, \( p = .89 \), indicating that children did not significantly associate math words with either gender. For reading words, implicit bias was also not significantly different from chance in either direction, \( D = -.06, t(46) = 1.70, p = .10 \), indicating that children also did not associate reading words with either gender. There was no significant differences in the magnitude of these two scores, \( t(46) = -1.08, p = .29 \) (see Figure 3.5).

**Implicit Association Test.** The IAT D-score was positive, and significantly different from chance, \( D = .15, t(46) = 2.10, p = .04 \), suggesting that children associated math more with boys and reading more with girls.

**Exploratory Analyses.** Though our current sample is underpowered to detect these differences, we performed a number of exploratory analyses. We first performed a bivariate correlation to look at the relationship between the PAS and IAT overall D-scores. Results indicated that there was not a significant correlation between these two measures, \( r(45) = .12, p = .42 \). Additionally, we looked at gender differences between the PAS math, reading, and overall D-scores, as well as the IAT D-score. There were no significant differences between boys and girls on any of these measures (\( ps > .18 \); see Table 2.4).

### 2.5.3 Discussion

We found that when using the Preschool Auditory Stroop, children ages 6-8 did not have significant associations between gender and math or reading words. Interestingly, this was not the case when using the Implicit Association Test; when using the IAT, children had a significant implicit association between boy=math and girl=reading. This difference between the two methods could be a result of the additional categorization required on each trial of the IAT; on the PAS, categorization of the math and reading words only occurs at the beginning of the task,
while on the IAT, every other trial requires children to decide whether the word presented is a “math” or “reading” word.

### 2.6 General Discussion

The results of our first two studies indicate that the Preschool Auditory Stroop procedure successfully captures children’s implicit associations between gender-stereotypical words and gender. Similar to validation studies of other implicit measures, we tested the PAS by examining whether children had implicitly internalized ubiquitous stereotypes (e.g. Baron & Banaji, 2006; Greenwald et al., 1998; Payne, Cheng, Govorun, & Stewart, 2005). Children showed evidence of implicit gender stereotypes, as both 3 to 4 and 6 to 7-year-old children who were asked to identify the gender of a spoken voice were faster to respond when word content was stereotypically congruent with voice gender (e.g. girl voice saying “pink”). Thus, word content served to either hinder or facilitate children’s response times. These results conceptually replicate the work of Most et al. (2007), and demonstrate that this modified methodology, which employs buttons and child-friendly labelling rather than voice responses, can be used with children as young as three.

In an additional departure from the methods used in the original Auditory Stroop paper, we performed additional analyses using more recent recommendations for analyzing implicit reaction time data. We kept our analyses consistent with the scoring procedure employed by the IRBT, which draws from traditional IAT D-score analyses, but like the PAS, measures implicit biases in preschoolers using a single categorical dimension (Qian et al., 2016). However, in addition to traditional D-score analyses, we computed additional D-scores disentangling the two different word types. Using D-scores, in comparison to the original reaction time analyses conducted by Most and colleagues, allowed us to make comparisons across development
between children with different average reaction times. Additionally, calculating D-scores for different word types allowed us to detect potential differences in the magnitude of stereotypes. Consistent with our aforementioned results, overall D-score analyses indicated that both preschool and early elementary school children were significantly faster to categorize words on stereotypically congruent trials as compared to stereotypically incongruent trials. Results were comparable when looking at stereotypically masculine and feminine words separately; children were faster to respond to both types of words when they were spoken in the congruent voice gender.

In Study 1, we found that older children showed stronger implicit feminine stereotypes than younger children, while in Study 2, implicit feminine and masculine stereotypes were similar across the age range. This difference between studies is most likely due to our adjustment in word choices. The words “lipstick” and “makeup” may be strongly associated with girls for 6 to 7-year-olds, but 3 to 4-year-olds may not be as familiar with these terms. As such, there do not seem to be any critical differences in implicit stereotype magnitude from ages 3-7. This is the first study to show that children as young as ages 3-4 have implicit stereotypes about gender. Based on the lack of developmental change across this age range, we speculate that the magnitude of these particular gender stereotypes may remain relatively stable across this period of development.

In Study 3, we sought to extend the use of this method and test the magnitude of children’s implicit stereotypes about gender, math, and reading. Results showed that as measured by the PAS, children ages 6-8 did not have significant stereotypes associating either gender with math or reading. In contrast, when associations were measured by the IAT, these same children had a significant stereotype associating boy=math and girl=reading. A plausible explanation for the difference between these two methods is that one measure is exemplar-based while the other
is category-based (Williams & Steele, 2017). While the IAT forces children to categorize individual stimuli into a broader category on every single trial, the PAS only forces children to decide if the individual stimulus is associated with boys or girls. As such, categorization on every trial of the IAT may activate stereotypes more strongly than in the PAS.

This interpretation is consistent with studies on implicit racial bias that have found weaker effects when children are not forced to categorize individuals into racial groups (Williams & Steele, 2017). Previous work has shown that when children complete exemplar-based measures, like the AMP, instead of category-based measures, like the IAT, they do not show evidence of racial outgroup negativity, and in some cases, also lack racial ingroup positivity. Thus, exemplar-based measures like the PAS offer the advantage of disentangling individual implicit associations, but may detect stereotypes at a different rate than category-based measures. While the more ingrained stereotypes tested in Study 1 and 2 may be more easily activated in young children, it is possible that gender stereotypes about math and reading require more contextual activation for children ages 6-8. Future work should look to test the PAS with an older developmental population, such as adults, who may have implicit associations about gender, math, and reading that are more easily activated.

In conclusion, the PAS offers a potential method of disentangling children’s implicit gender associations that can be used with children as young as age 3. While this method may detect stereotypes at a different rate than category-based measures like the IAT, comparisons between the two methods provides insight into the conditions required to activate stereotypes in young children. Not all gender stereotypes may be easily activated in childhood and observing the developmental trajectory of these stereotypes with multiple methods may be the best way to chart stereotype sensitivity. This method may also offer an advantage over the IAT by measuring stereotypes that do not lend themselves as easily to categorization. We recommend use of the
Google voices that were used here, in order to ensure that children respond to voice gender at comparable rates. We also recommend the use of D-scoring, as this method follows more recent implicit scoring recommendations and allows researchers to compare the magnitude of bias across developmental samples. By disentangling distinct associations and facilitating use with young children, we hope that the PAS can offer novel and informative inquiry into the development of implicit cognition.
Table 2.1 List of words used in Study 1.

<table>
<thead>
<tr>
<th>Female-Stereotypical</th>
<th>Male-Stereotypical</th>
<th>Neutral</th>
<th>Practice (Neutral)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipstick</td>
<td>Baseball</td>
<td>Pencil</td>
<td>Apple</td>
</tr>
<tr>
<td>Makeup</td>
<td>Football</td>
<td>Spoon</td>
<td>Door</td>
</tr>
<tr>
<td>Pretty</td>
<td>Rough</td>
<td>Table</td>
<td>Draw</td>
</tr>
<tr>
<td>Pink</td>
<td>Tough</td>
<td>Window</td>
<td>Paper</td>
</tr>
</tbody>
</table>
Table 2.2 List of words used in Study 2.

<table>
<thead>
<tr>
<th>Female-Stereotypical</th>
<th>Male-Stereotypical</th>
<th>Neutral</th>
<th>Practice (Neutral)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dress</td>
<td>Truck</td>
<td>They</td>
<td>This</td>
</tr>
<tr>
<td>Doll</td>
<td>Football</td>
<td>Them</td>
<td>That</td>
</tr>
<tr>
<td>Pretty</td>
<td>Tough</td>
<td>Their</td>
<td></td>
</tr>
<tr>
<td>Pink</td>
<td>Blue</td>
<td>These</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.3 List of words used in Study 3.

<table>
<thead>
<tr>
<th>Math</th>
<th>Reading</th>
<th>Neutral</th>
<th>Practice (Neutral)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adding</td>
<td>Books</td>
<td>They</td>
<td>This</td>
</tr>
<tr>
<td>Counting</td>
<td>Letters</td>
<td>Them</td>
<td>That</td>
</tr>
<tr>
<td>Math</td>
<td>Reading</td>
<td>Their</td>
<td></td>
</tr>
<tr>
<td>Numbers</td>
<td>Story</td>
<td>These</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.4 PAS and IAT D-Scores by gender (Study 3).

<table>
<thead>
<tr>
<th>D-score</th>
<th>Boys (SD)</th>
<th>Girls (SD)</th>
<th>t</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAS Math</td>
<td>.07 (.30)</td>
<td>-.04 (.32)</td>
<td>1.28</td>
<td>.208</td>
</tr>
<tr>
<td>PAS Reading</td>
<td>-.06 (.27)</td>
<td>-.06 (.22)</td>
<td>0.05</td>
<td>.962</td>
</tr>
<tr>
<td>PAS Overall</td>
<td>.03 (.15)</td>
<td>-.04 (.18)</td>
<td>1.38</td>
<td>.175</td>
</tr>
<tr>
<td>IAT Overall</td>
<td>.19 (.45)</td>
<td>.11 (.50)</td>
<td>0.58</td>
<td>.568</td>
</tr>
</tbody>
</table>
Figure 2.1 Average reaction times for older children (ages 6-7) using Most criteria (Study 1).
Figure 2.2 Average reaction times for younger children (ages 3-4) using Most criteria (Study 1).
Figure 2.3 Mean PAS D-scores by type (Study 1).
Figure 2.4 Mean PAS D-scores by type (Study 2).
Figure 2.5 Mean PAS and IAT D-scores (Study 3).
Chapter 3: Intergroup Bias and Behavior

3.1 Synopsis

Few studies have examined the relationship between intergroup bias and behavior in childhood, limiting our understanding of when in development bias begins to impact behavior. Of the studies that have examined the effects of bias on behavior, the majority have tested the effects of a stereotype associating math with boys/men more than girls/women. This study extends previous work by examining the relationship between gender stereotypes and preschool girls’ math-related performance in contexts where stereotypes have been activated. Girls’ math-related performance was tested using a measure of intuitive number sense, which is a universal skill that predicts later math ability. Across a combined dataset of four samples, girls who associated math more with boys performed worse on a number sense task when it was framed as a math test. These results provide evidence that stereotypes can impair girls’ intuitive number sense, and that threat contexts may elicit these effects. In addition to contributing to our theoretical understanding of the relationship between intergroup bias and behavior, these results are practically important, as they indicate that gender stereotypes may affect young girls’ acquisition of formal mathematics concepts and developing interest in math-related fields.

3.2 Introduction

Women continue to be highly underrepresented in mathematics, engineering, and related fields; a pattern that can be partially attributed to the presence of cultural stereotypes associating math more with men than women (National Science Foundation, 2016; Nosek et al., 2009; Organisation for Economic Co-operation and Development, 2015). A large body of correlational and experimental work has linked these stereotypes to a gender gap in math performance and has shown that even subtle reminders of gender stereotypes can cause women to underperform on
tests of their math ability (Miller, Eagly, & Linn, 2015; Nguyen & Ryan, 2008; Spencer, Steele, & Quinn, 1999; Walton & Spencer, 2009). In addition to affecting adult women, gender stereotypes can emerge in elementary school and impair school-aged girls’ math performance when those stereotypes are contextually activated (Ambady et al., 2001; Cvencek et al., 2011; Cvencek et al., 2014; Galdi et al., 2014; Tomasetto et al., 2011). As children’s early experiences with math are likely to inform their later interest and engagement, it is important to identify whether stereotypes begin to impair girls’ math ability even before they enter formal education.

The existing body of research on stereotype-based performance impairments has typically examined effects on formal math tests. Children acquire such skills through a combination of individual interest and educational experience. In contrast, before children enter formal schooling, they have a more basic, universal, and intuitive number sense often termed the Approximate Number System (ANS). The ANS provides us with our gut, intuitive sense of number, and appears to be foundational for later formal, symbolic math abilities. Children and adults who have a very precise number sense perform substantially better on various formal and informal math assessments, even when controlling for working memory, intelligence, and other related variables (Chen & Li, 2014; Feigenson, Dehaene, & Spelke, 2013; Halberda et al., 2008; Libertus et al., 2012; Starr et al., 2013). Despite its universality, the ANS is not encapsulated; adults and children (ages 5-7) who have their ANS temporarily boosted through training or feedback perform better on a subsequent math test, and when ANS acuity is reduced through these methods, they perform worse (DeWind & Brannon, 2012; Hyde, Khanum, & Spelke, 2014; Wang, Odic, Halberda, & Feigenson, 2016).

The current research examines another potential modulation of the ANS: whether cultural stereotypes about women and mathematics can alter young children’s ANS acuity, even prior to extensive exposure to formal mathematics education. Critically, despite a lack of overall sex
differences in ANS capabilities (Spelke, 2005), contextual activation of gender stereotypes might impair the ANS accuracy of girls who have internalized this bias. As this system helps with the acquisition of formal mathematics skills, any stereotype-based impairments of the ANS that start in early childhood, before formal math education, would only compound in degree over time, potentially impairing girls’ acquisition of formal mathematics concepts and their developing interest in math-related fields. Thus, an understanding of how stereotypes affect girls’ more basic numerical cognition is crucial to ensure that girls and boys do not begin their formal math education on unequal footing.

In addition to examining the effect of contextual stereotype activation on children’s ANS performance, we also look at individual variability in children’s gender stereotypes as a predictor of performance. While a number of studies have shown that young girls can be affected by stereotypes in certain contexts, other researchers have recently found mixed results for these stereotype threat effects (Flore & Wicherts, 2015; Ganley et al., 2013). However, these studies do not measure the presence of stereotypes in children, which might be a necessary prerequisite for context to affect children’s math-related performance. As the development of stereotypes about math and gender can vary in early development (e.g. Passolunghi et al., 2014; Steffens et al., 2010), with children internalizing these stereotypes at different ages, it is essential for researchers to measure their presence. As evidenced by work with adults, girls who have not yet internalized stereotypes about math and gender would most likely be unaffected by manipulations intended to impair their math performance through stereotype activation (Schmader, Johns, & Barquissau, 2004).

In the following chapter, the combined results from four samples are presented. Overall, these studies examine the hypothesis that 3-6 year-old preschool girls who have already internalized gender stereotypes about math ability would exhibit impaired ANS accuracy when
the task is described as a measure of math and counting, rather than as a game (Study 4 & 5) or an eye test (Study 6 & 7).

3.3 Method

3.3.1 Participants

We tested a total of 762 children ages 3-6 across four samples (see Table 3.1). Though our main hypotheses focus on girls, in Study 4-6 we also collected data from boys as comparison to test the specificity of effects. Participants were recruited from a community-based science center. Children were excluded for pressing the buttons randomly or in a fixed pattern, scoring below chance levels on the ANS task (< 50% of trials correct), parent or sibling interference, language barriers, and any computer or experimenter error. Our a priori goal was to run 60 useable children per gender and age group (3-4 and 5-6 years) in each study, and we stopped running participants after we believed we had met this goal. Participants were recruited by research assistants who approached potential families in a local science center and after reviewing the study description, sought parental consent and child assent to participate. Children were tested onsite in an area dedicated for behavioral science research.

3.3.2 Procedure

Participants were tested individually in a soundproof room dedicated to behavioral science research. The experiment was presented on a computer using Inquisit™ version 4, and an experimenter read all directions aloud to children. We randomly assigned children to condition by alternating which condition they were in but balanced this assignment across age and gender.

In all studies, children were presented with instructions before the ANS task based on condition. One condition was intended as a control condition, and the other was intended to
prime gender stereotypes about math. Study 4 and 5 had two conditions: the Game Control condition and the Math Test condition. In the Game Control condition, children were given the following instructions: “Now we’re going to play a game. Your job is to try your best”. In the experimental (Math Test) condition, children were given the following instructions: “Now we’re going to test your math ability. This test tells us whether boys or girls are naturally better at math and counting.” In Study 6 and 7, the Math Test condition was identical to Experiment 1. However, to control for priming of gender and possible effects of simply calling the task a ‘test,’ we modified the wording of our control condition. Specifically, in our control condition for this experiment, children were told: “Now we’re going to test your eyesight ability. This test tells us whether boys or girls are naturally better at seeing things quickly.”

Afterwards, all children were presented with the ANS task and instructed to complete the task in the same manner. In Study 4, 5, and 7, after the ANS task, children were presented with explicit questions, which were randomized. In Study 6, the order of presentation for the ANS task and the explicit stereotypes were counterbalanced, with half of participants completing the ANS task first, and the other half of participants answering the explicit questions first. Upon completion of the study, all children were given a sticker for participation, and parents were debriefed on the aims of the research.

3.3.3 Measures

**Approximate Number System (ANS) Task.** We measured each child’s ANS accuracy using the standardized Panamath test (Halberda et al., 2008). Participants were introduced to Big Bird and Grover – two characters drawn on the screen, each of whom had an empty box that was color matched to their character (yellow and blue respectively). Participants were told to decide which character had more dots in their box on each trial. For participants ages four and above,
children pressed a corresponding yellow and blue JellyBean™ button based on which character they thought had more dots. Participants who were three years old were simply asked to point to the character they thought had more dots, and the experimenter would answer for them using the keyboard.

For each trial, two arrays of colored dots (yellow and blue) appeared in their respective boxes for 1500 milliseconds (see Figure 1). To control for the difficulty of the task, children were presented with different numerical ratios based on norms for their age. In Study 4, we used the preprogrammed ratios in the Panamath software. In Study 5-7, ratios were more accurately customized for age norms. Half of trials had cumulative surface area that was congruent with the number of dots, and on the other half of trials, this was incongruent. Children wore headphones during the task and received either positive or negative verbal feedback from the program based on performance on each trial. All children included in our final sample completed 80 trials.

After completing the task, children in the Game Control condition were told: “Great job! We’ve found that boys and girls both really like playing that game.” In the Test conditions, children were told: “Great job! We’ve found that boys and girls do equally well on that test.”

**Explicit Math-Gender Stereotypes.** In order to measure math-gender stereotypes, children were presented with an image of a cartoon boy and girl on the computer screen (Ambady et al., 2001). Children were asked two types of questions regarding their math-gender beliefs, a) which child they thought was better at math and counting and b) which child they thought liked math and counting more. Experimenters would verbally ask “Which person do you think is better at math counting? Do you think this person (on the left) is better at math and counting, this person (on the right) is better at math and counting, or are they the same?” Children were able to respond with one of the three options. Experimenters asked each type of question twice, and the order of questions was counterbalanced. Furthermore, the ethnicity and
skin tone of the cartoon children varied in each trial. For purposes of interpretation, math-gender beliefs were coded in relation to participants’ own gender (0 = opposite gender – math association, 2 = own gender – math association). These questions could have appeared in any order, but questions about ability (‘Which child is better at…’) and interest (‘Which child likes…’) were always blocked together.

**Control Stereotype Measure.** In Study 5, in addition to measuring math-gender stereotypes, we included a control measure to ensure that children were not simply selecting one gender regardless of question content. Children were presented again with images of a boy and a girl on the same screen and asked two questions in the same style as the explicit stereotype measures. First, they were asked which of the two children was better at “daxing” and then which of the two children liked “daxing” more. Some children received questions about explicit math-gender stereotypes first, and others received questions about daxing first.

### 3.4 Combined Sample Results

The results across the four studies followed a similar pattern, thus, for increased power analyses are first presented on the combined dataset (N = 762). This mega-analytic approach is generally preferable to meta-analysis (i.e., estimating the true effect size from sample-level effects), when the raw data are available (Costafreda, 2009; DeRubeis, Gelfand, Tang, & Simons, 1999; Sung et al., 2014). It also in line with a growing preference for fewer well-powered studies (Ioannidis, 2005), and recommendations to pool multiple small samples to boost power when testing higher order interactions and to provide more stable estimates of effect sizes (Schimmack, 2012). Results for each individual experiment are presented below in section 3.4. While not all effects are identical across the four studies, this variation is to be expected within
multi-study data (see Simmons, Nelson, & Simonsohn, 2011; Spellman, Gilbert, & Corker, 2017).

3.4.1 Math-Gender Beliefs

Our first set of analyses examined the magnitude of explicit stereotypes about math and gender in our combined sample (see Table 3.2). Overall, children did not explicitly endorse a stereotype associating males more with math, underscoring the importance of examining the moderating role of these beliefs. We found no gender difference in the magnitude of math-gender beliefs, as boys and girls had comparable average associations between their own gender and math, \( t(459.25) = -0.091, p = .927, d < 0.001 \) (t-test uses corrected values due to unequal variance, \( p = .001 \)). Furthermore, both boys and girls on average explicitly associated their own gender with math, boys: \( t(263) = 41.20, p < .001 \); girls: \( t(497) = 68.09, p < .001 \). There was no difference in the magnitude of beliefs across conditions, as mean levels were comparable across the combined Control and Math Test conditions, \( t(760) = .37, p = .714, d < 0.001 \). Lastly, we found that math-gender beliefs were not correlated with age for girls, \( r = -0.02, p = .662 \) and marginally correlated with age for boys, \( r = .11, p = .064 \), indicating that explicit beliefs about math and gender were not changing significantly across this age range.

3.4.2 ANS Task Performance

Our second set of analyses concerned overall ANS performance and potential age and gender differences on this measure. ANS performance was quantified as children’s overall accuracy across the 80 trials of the task. Across both studies children performed well: on average they correctly answered 80.61% of trials (boys: 77.44%; girls: 82.29%). Consistent with other work on children’s ANS, task accuracy increased with age, \( r = .31, p < .001 \). We also found an
overall gender difference, with girls performing better on the task than boys, \( t(507.80) = -5.59, p < .001, d = .43 \) (t-test uses corrected values due to unequal variance, \( p = .028 \)).

Our third and key set of analyses tested the hypothesis that in the Math Test condition, as a result of activating children’s math-gender stereotypes, a stronger association between one’s own gender and math would predict better ANS performance. We expected no such relation in the combined Control condition. Further, we tested child gender as a potential moderator. To test this hypothesis, we performed a regression analysis with math-gender beliefs (standardized), child gender (dummy coded; 1 = male), and condition (dummy coded; 1 = Math Test) entered as predictors of ANS task performance and controlled for sample. We found that experiment was a significant predictor of ANS performance such that children performed better in Study 5 than Study 4, 6, and 7 (\( \beta = .30, \text{CI}_{95} [.07, .54], p = .010 \)), and better in Study 7 than Study 4, 5, and 6 (\( \beta = .43, \text{CI}_{95} [.19, .68], p < .001 \); see Table S1). However, experiment did not interact with any other variables to predict ANS task performance (\( ps > .13 \)). As a result, all subsequent analyses presented in the manuscript control for experiment.

Analyses on the combined dataset revealed a significant three-way interaction between children’s math-gender beliefs, child gender, and condition predicting performance on the ANS task, \( \beta_{\text{int}} = -.32, \text{CI}_{95} [-.60, -.04], p = .024 \). For girls, there was a significant interaction between math-gender beliefs and condition, \( \beta = .30, \text{CI}_{95} [.12, .48], p = .001 \). Most notably, simple slopes analyses supported the core hypothesis: girls who associated boys with math (-1SD from the mean = 0.79) performed worse in the Math Test condition than the Control condition, \( \beta = -.42, \text{CI}_{95} [-.67, -.17], p = .001 \). This simple effect of condition was non-significant for girls who strongly associated girls with math, +1SD from the mean = 1.63, \( \beta = .18, \text{CI}_{95} [-.07, .43], p = .151 \).
Analyzed differently, in the Math Test condition, girls’ beliefs about gender and math, $M = 1.20, SD = 0.37$, predicted their ANS performance; having a weaker association between girls and math predicted poorer ANS task performance, $\beta = .18, CI_{95} [.05, .31], p = .006$. In the Control condition, girls’ math-gender beliefs were marginally associated with math performance, $\beta = -.12, CI_{95} [-.25, .01], p = .070$.

In contrast, we found no significant interaction between condition and math-gender beliefs predicting performance on the ANS task for boys, who performed similarly regardless of condition or beliefs, $M = 1.23, SD = 0.47; \beta = -.02, CI_{95} [-.23, .19], p = .845$. In other words, while we found an association between gender stereotypes and girls’ universal number sense, we did not find this relation with boys. Further, our manipulation of stereotype salience only affected girls’ ANS performance if they had acquired the stereotype associating males more with math, pointing to a potential mechanism underlying this effect. Follow-up analyses including age as a possible moderator in the model yielded no significant main effects or interactions by age.

In a final set of follow-up simple slopes analyses, we examined the conditions under which a gender difference in performance was observed (see Figures 3.2 and 3.3). These analyses revealed that in two conditions, girls performed significantly better than boys on the task; in the Math condition, when they associated math with their own gender, and the Control condition, when they associated math with the opposite gender: $\beta_{\text{Math Test condition + associate math with own gender}} = -.47, CI_{95} [-.77, -.18], p = .002; \beta_{\text{Control condition + associate math with opposite gender}} = -.43, CI_{95} [-.72, -.14], p = .004$. This association was marginally significant for girls in the control condition who associated their own gender with math: $\beta_{\text{Control condition + associate math with own gender}} = -.25, CI_{95} [-.55, .04], p = .090)$. Girls in the Math Test condition who associated math with the opposite gender
did not perform significantly better than boys (-1SD from the mean), $\beta = -.01, \text{CI}_{95} [-.30, .29], p = .955$.

### 3.5 Individual Study Results

We next present analyses separately for the four different experiments. Results of key analyses are summarized in Table 3.2 and 3.3 along with the mega-analysis of the combined dataset summarized above.

#### 3.5.1 Study 4 Results

**Math-Gender Beliefs.** Our first study acted as a pilot study to examine potential effects of task framing on children’s ANS performance by gender. We found no gender differences in the magnitude of math-gender beliefs (see Table 3.2), as boys and girls had comparable associations between their own gender and math, $t(94) = -.74, p = .46, d = 0.15$. Boys significantly explicitly associated their own gender with math, boys: $t(41) = 2.27, p = .028$, and girls had a marginally significant explicit association between their own gender and math $t(53) = 1.64, p = .11$. Furthermore, there were no differences in the magnitude of beliefs across conditions, as mean levels of math-gender beliefs were comparable across the Game, $M = 1.16$, $SD = .53$, and Math Test, $M = 1.13$, $SD = .51$, conditions, $t(94) = .31, p = .76, d = 0.06$.

**ANS Task Performance.** In a similar manner as our key regression analyses, we entered math-gender beliefs, child gender, and condition as predictors of ANS task performance (see Table 3.3). The three-way interaction between children’s math-gender beliefs, child gender, and condition was non-significant, most likely as a result of our study being very underpowered to detect the interaction, $\beta_{\text{int}} = -.59, \text{CI}_{95} [-1.30, .25], p = .184$. When conducting exploratory analyses to examine the performance of girls in the Math Test condition, we found that girls’
math-gender beliefs significantly predicted their ANS performance, $\beta_{\text{int}} = .40$, CI$_{95} [.03, .67]$, $p = .031$. Based on this finding, we chose to pursue additional follow-up studies with more power.

3.5.2 Study 5 Results

**Math-Gender Beliefs.** We compared children’s explicit math-gender beliefs by gender and condition (see Table 3.2). There was no gender difference in the magnitude of math-gender beliefs, as boys and girls had comparable associations between their own gender and math, $t(227) = -.46$, $p = .65$, $d = -0.06$. Both boys and girls on average explicitly associated their own gender with math, boys: $t(98) = 4.09$, $p < .001$; girls: $t(105) = 4.39$, $p < .001$. Furthermore, there were no differences in the magnitude of beliefs across conditions, as mean levels of math-gender beliefs were comparable across the Game, $M = 1.16$, $SD = .40$, and Math Test, $M = 1.19$, $SD = .40$, conditions, $t(227) = .62$, $p = .54$, $d = 0.08$.

**ANS Task Performance.** We entered math-gender beliefs, child gender, and condition as predictors of ANS task performance (see Table 3.3). The three-way interaction between children’s math-gender beliefs, child gender, and condition was non-significant, most likely as a result of our study being underpowered, $\beta_{\text{int}} = -.37$, CI$_{95} [-.93, .19]$, $p = .193$. However, for girls, there was a significant interaction between math-gender beliefs and condition, $\beta = .46$, CI$_{95} [.04, .88]$, $p = .031$. Girls who associated boys with math performed worse in the Math Test condition than the Game condition, $\beta = -.60$, CI$_{95} [-1.16, -.04]$, $p = .037$. This effect was non-significant for girls who strongly associated girls with math, $\beta = .33$, CI$_{95} [-.25, .90]$, $p = .263$. Thus, it appears
that girls who endorsed math-gender stereotypes associating boys more with math had worse performance when these stereotypes were activated.

Analyzed differently, in the Math Test condition, girls’ beliefs about gender and math predicted their ANS performance; girls with a weaker association between girls and math performed worse on the ANS task than girls with a stronger association, $\beta = .37$, CI$_{95}$ [.06, .69], $p = .020$. These results suggest that when math-gender stereotypes were activated, they predicted girls’ ANS task performance. This was not the case in the Game condition, where girls’ math-gender beliefs were not significantly associated with ANS performance, $\beta = -.09$, CI$_{95}$ [-.37, .19], $p = .530$. We found no significant interaction between condition and math-gender beliefs predicting performance on the ANS task for boys, who performed similarly regardless of condition or beliefs, $\beta = .09$, CI$_{95}$ [-.28, .46], $p = .621$.

In addition to these analyses, we once again looked at lower-order interactions with gender as a moderator. In the Math Test condition, there was a significant interaction between children’s math-gender beliefs and their gender predicting ANS performance, $\beta = -.44$, CI$_{95}$ [-.83, -.04], $p = .032$. There was no interaction between math-gender beliefs and child gender in the Control condition, $\beta = -.06$, CI$_{95}$ [-.46, .33], $p = .750$. Simple slopes analyses indicated that gender did not significantly predict ANS task performance regardless of condition or beliefs ($ps > .093$), suggesting that these interactions were not driven by gender differences in performance.

**Control Measure.** In order to ensure that children were not answering the explicit questions about math based on an overall gender preference, we included a control measure to ensure that children were not simply selecting their own gender regardless of question content. To test whether or not children had an overall gender response bias, we examined whether there
was a correlation between children’s beliefs about math and gender, and their beliefs about a novel word (“daxing”) and gender.

The score for children’s beliefs about daxing and gender was calculated by coding the beliefs in a similar manner to the math-gender beliefs, according to participants’ own gender (0 = opposite gender – daxing association, 2 = own gender – daxing association). The questions about daxing interest and ability were then averaged to create an overall daxing-gender association score ($M = 1.1$, $SD = .56$). We observed no correlation between children’s math-gender association and their daxing-gender association, $r(203) = .02$, $p = .836$, suggesting that children were not simply answering the questions about math and gender based on an overall gender preference.

### 3.5.3 Study 6 Results

**Math-Gender Beliefs.** For Study 6, we again compared children’s explicit math-gender beliefs by gender and condition (see Table 3.2). Boys, $M = 1.26$, $SD = .48$, and girls, $M = 1.24$, $SD = .36$, had comparable associations between their own gender and math, $t(246) = .33$, $p = .74$, $d = 0.05$. Both boys and girls on average explicitly associated their own gender with math, boys: $t(110) = 5.73$, $p < .001$, girls: $t(117) = 7.19$, $p < .001$). Similarly, math-gender beliefs were not significantly different between the Eyesight, $M = 1.26$, $SD = .46$, and Math Test, $M = 1.24$, $SD = .39$, conditions, $t(246) = .41$, $p = .68$, $d = 0.05$.

**ANS Task Performance.** Analyses were conducted in the same way as previous studies (see Table 3.3). The three-way interaction between children’s math-gender beliefs, child gender, and condition was non-significant, $\beta_{int} = -.44$, CI$_{95}$ [-.98, .10], $p = .108$. For girls, there was a marginally significant interaction between math-gender beliefs and condition, $\beta = .38$, CI$_{95}$ [-.05, .81], $p = .086$. However, as our effect size was similar to Experiment 1, we explored the simple
slopes of this interaction to see if results were consistent. Girls who associated boys with math performed worse in the Math Test condition than the Eyesight condition, $\beta = -.68$, CI$_{95}$ [-1.23, -.13], $p = .015$. This effect was non-significant for girls who strongly associated girls with math, $\beta = .08$, CI$_{95}$ [-.50, .66], $p = .827$.

In the Math Test condition, girls’ beliefs about gender and math did not significantly predict their ANS performance, though effect sizes were consistent with previous studies, $\beta = .26$, CI$_{95}$ [-.07, .59], $p = .116$. In the Eyesight condition, girls’ math-gender beliefs were also not significantly associated with ANS performance, $\beta = -.12$, CI$_{95}$ [-.40, .16], $p = .405$. We found no significant interaction between condition and math-gender beliefs predicting performance on the ANS task for boys, who performed similarly regardless of condition or beliefs, $\beta = -.06$, CI$_{95}$ [-.39, .26], $p = .712$.

Once again, we looked at lower-order interactions with gender as a moderator in order to examine potential gender differences. In both the Math Test condition, $\beta = -.33$, CI$_{95}$ [-.74, .07], $p = .109$, and the Control condition, $\beta = .11$, CI$_{95}$ [-.25, .47], $p = .544$, the interaction between children’s math-gender beliefs and their gender was non-significant.

**Order Analyses.** One of our concerns after Study 4 and 5 was that children might respond to our measure of explicit stereotypes based on their own performance (e.g. if the child did well, they might say that their gender does better based on that self-assessment). To address this issue in Study 6, we counterbalanced the order in which the explicit questions were presented. To ensure that results were robust controlling for potential order effects, we also ran the regression analyses controlling for order. Order was not found to be a significant covariate, $\beta = -.12$, CI$_{95}$ [-.38, .14], $p = .355$, and overall results were nearly identical.
As an additional exploratory analysis, we added order as a moderator to look at a potential three-way interaction between gender, condition and order. Consistent with the aforementioned Experiment 2 analyses, there was a main effect of gender, $F(1,221) = 9.04, p = .003, \eta^2 = .04$, such that girls performed better on the ANS task than boys. There were no main effects of condition, $F(1,221) = 1.36, p = .245, \eta^2 = .006$ or order, $F(1,221) = 1.17, p = .281, \eta^2 = .005$ predicting ANS task performance. Additionally, there were no two-way interactions between gender, condition, and order ($ps > .203$). However, there was a three-way interaction between gender, condition and order predicting ANS performance, $F(1,221) = 7.51, p = .007, \eta^2 = .033$. In order to analyze this interaction, we performed analyses for each order separately. We did not conduct regression analyses of variation in stereotype belief for the samples split by order due to inadequate power for these higher order interactions.

**Explicit Questions Before ANS Task.** Our first analysis examined if girls’ ANS accuracy was negatively impacted by stereotypes about boys being better at math when asked about these stereotypes before the task. Consistent with the overall findings of Experiment 2, we found a marginally significant main effect of gender, $F(1,111) = 3.74, p = .056, \eta^2 = .03$, such that girls performed better than boys. There was no main effect of condition, $F(1,111) = .42, p = .521, \eta^2 = .004$ but there was a gender by condition interaction predicting ANS task performance, $F(1,111) = 9.50, p = .003, \eta^2 = .08$. Simple effects analyses indicated that girls performed significantly worse in the Math Test condition, $M = 77.18, SD = 12.86$, as compared to the Eyesight condition, $M = 84.82, SD = 7.18; p = .007$. Thus, asking girls about their math-gender stereotypes before the ANS task may have served to strengthen the manipulation, resulting in an effect of stereotypes on performance regardless of explicit stereotype endorsement.
**Explicit Questions After ANS Task.** Next, we examined whether girls’ ANS acuity was negatively impacted when questions about gender stereotypes came after the task (as they had been placed in Experiment 1). We again found a main effect of gender, $F(1,110) = 5.30$, $p = .023$, $\eta^2 = .05$, such that girls performed better than boys. There was no main effect of condition, $F(1,110) = .96$, $p = .328$, $\eta^2 = .009$, and no gender by condition interaction, $F(1,110) = .93$, $p = .337$, $\eta^2 = .008$. These results indicate when girls were asked about their beliefs after the ANS task, there was no general effect of priming stereotypes on girls’ ANS performance.

### 3.5.4 Study 7 Results

**Math-Gender Beliefs.** Girls on average explicitly associated their own gender with math (see Table 3.2), $t(231) = 8.23$, $p < .001$. There were no differences in the magnitude of beliefs across conditions, as mean levels of math-gender beliefs were comparable across the Eyesight, $M = 1.22$, $SD = .38$, and Math Test, $M = 1.20$, $SD = .40$ conditions, $t(230) = -4.7$, $p = .64$, $d = 0.05$.

**ANS Task Performance.** In a similar manner as our key regression analyses, we entered math-gender beliefs, child gender, and condition as predictors of ANS task performance (see Table 3.3). The interaction between math-gender beliefs and condition was not significant, $\beta = -.12$, CI$_{95}$ [-.08, .38], $p = .53$. As the main effect of interest was significant in our previous studies, but not in this study, $\beta = -.11$, CI$_{95}$ [-.17, .16], $p = .41$, we decided to perform a mega-analysis on the combined dataset to examine this small, but potentially impactful effect of the Math Test condition on the ANS performance of girls who associated math with boys.
3.6 Discussion

The present findings suggest that contextual reminders of learned gender stereotypes about math can affect preschool girls’ performance on the most basic and universal assessments of number intuitions. Though only a small number of girls from the combined dataset showed evidence of a math-gender stereotype (n = 60), for these girls, framing the task as a test of math ability significantly impacted their ANS performance. When stereotypes about math and gender were activated before the ANS task, girls who associated math more strongly with boys performed worse than when the task was framed as a game or a test of eyesight. Thus, girls’ beliefs about math appear to impact their intuitive number sense specifically when these stereotypes are activated in a testing context.

These results are especially striking since boys and girls normally show comparable ANS acuity (Spelke, 2005), suggesting that their intuitive sense of number can be modulated by awareness of cultural stereotypes from at least preschool onward. Thus, these effects are an example of how context and cultural stereotypes can may impair girls’ math performance at an early age through impairment of a core cognitive system. While these results indicate that preschool girls’ number sense can be impacted by stereotypes, boys in our study were unaffected. In general, young girls may be more sensitive to gender stereotypes, as they appear to internalize their parents’ gender biases more than young boys (Croft et al., 2014). Moreover, other work shows that boys are slower to internalize stereotypes about math and gender (Steffens et al., 2010). In line with this evidence, we speculate that boys in our study may have been less sensitive to gender stereotype activation.

Interestingly, these results suggest that there may be an overall gender difference in ANS performance, with girls outperforming boys when stereotypes are not activated. These results are consistent with evidence that school-aged girls often do outperform boys in math, albeit by a
smaller margin than language arts (Voyer & Voyer, 2014). However, other work suggests that there are no gender differences in children’s math performance, and in particular, no gender differences in ANS acuity (Lindberg, Hyde, Petersen, & Linn, 2010; Spelke, 2005). Future work should seek to replicate and explore the causes of this gender difference, as well as examine whether or not young girls’ comparable performance in math relative to boys might actually be underperformance in respect to their potential (Good, Aronson, & Harder, 2008).

As previous research has found mixed results of the effect of stereotypes on children’s math test performance (Flore & Wicherts, 2015), we suspect that this is in part due to variability in the stereotype knowledge and beliefs that children have (Picho & Schmader, 2017; Schmader et al., 2004). While our sample did not, on average, show traditional endorsement of stereotypes about gender and math, this finding is supported by work showing that young children often display in-group favoritism in their explicit responses (Régner et al., 2014). Furthermore, within our data, there is individual variability that is clearly important in predicting girls’ susceptibility to stereotype effects. Future studies should ensure measurement of children’s stereotypes as key moderators of the effect of contextual cues on math performance.

At a surface level, the pattern of results in this study appear comparable to stereotype threat effects that have been found with older girls and adult women (e.g. Ambady et al. 2001; Galdi et al., 2014; Nguyen & Ryan 2008; Tomasetto et al., 2011; Walton & Spencer, 2009) However, the mechanisms behind these effects are most likely different for young girls. In women, stereotype threat effects are proposed to stem from anxiety about confirming stereotypes about one’s own group; this anxiety leads to impaired working memory performance, which results in underperformance on tasks associated with the activated stereotype (Schmader, Johns, & Forbes, 2008). In contrast, for young girls, it seems more likely that those who have stereotypes about math and gender may simply disengage from the task at hand when these
stereotypes are activated. Future research should examine the mechanism behind the stereotype-based performance effects observed in these studies.

Though only a handful of girls were impacted by our stereotype framing, these particular girls may be at risk for reduced performance in mathematics domains when they enter a formal schooling environment. In conjunction with past work, these results suggest that even though both genders start off on a level playing field in terms of foundational math abilities, internalization of math-gender stereotypes may tip the scales quite early in development by decreasing young girls’ ANS accuracy – just as this ability begins to aid them in learning formal mathematical concepts. If contextual activation of stereotypes can impair the basic numerical abilities of preschool girls, these effects might compound across development to prevent girls from achievement in mathematics. Thus, interventions to increase girls’ engagement in math and math-related fields should consider starting very early in development, before gender stereotypes can create a cycle of impaired performance and reduced interest in math.
Table 3.1 Differences between Study 4-7.

<table>
<thead>
<tr>
<th>Study</th>
<th>Control Condition</th>
<th>Adjusted Ratios</th>
<th>Order of Measures</th>
<th>Control Measure</th>
<th>Boys</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 (n = 96)</td>
<td>Game</td>
<td>No</td>
<td>Explicit Second</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>5 (n = 205)</td>
<td>Game</td>
<td>Yes</td>
<td>Explicit Second</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>6 (n = 229)</td>
<td>Eyesight</td>
<td>Yes</td>
<td>Counterbalanced</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>7 (n = 232)</td>
<td>Eyesight</td>
<td>Yes</td>
<td>Explicit Second</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
Table 3.2 Means and Standard Deviations for Math-Gender Beliefs and ANS Task Performance. Note: Math-Gender Beliefs range from 0 to 2, with higher numbers indicating a stronger association between one’s own gender and math. ANS performance is a percentage of correct trials.

<table>
<thead>
<tr>
<th></th>
<th>Study 4</th>
<th>Study 5</th>
<th>Study 6</th>
<th>Study 7</th>
<th>Combined Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math-Gender Beliefs</td>
<td>1.15 (.52)</td>
<td>1.17 (.40)</td>
<td>1.25 (.43)</td>
<td>N/A</td>
<td>1.20 (.42)</td>
</tr>
<tr>
<td>Girls</td>
<td>1.19 (.54)</td>
<td>1.16 (.37)</td>
<td>1.24 (.36)</td>
<td>1.21 (.39)</td>
<td>1.20 (.39)</td>
</tr>
<tr>
<td>Boys</td>
<td>1.11 (.50)</td>
<td>1.18 (.44)</td>
<td>1.27 (.49)</td>
<td>N/A</td>
<td>1.20 (.47)</td>
</tr>
<tr>
<td>Math</td>
<td>1.13 (.51)</td>
<td>1.19 (.40)</td>
<td>1.24 (.40)</td>
<td>1.20 (.40)</td>
<td>1.20 (.41)</td>
</tr>
<tr>
<td>Control</td>
<td>1.16 (.53)</td>
<td>1.15 (.40)</td>
<td>1.27 (.46)</td>
<td>1.22 (.38)</td>
<td>1.21 (.43)</td>
</tr>
<tr>
<td>Girls/Math</td>
<td>1.12 (.57)</td>
<td>1.19 (.35)</td>
<td>1.19 (.32)</td>
<td>1.20 (.40)</td>
<td>1.18 (.39)</td>
</tr>
<tr>
<td>Girls/Control</td>
<td>1.31 (.50)</td>
<td>1.13 (.39)</td>
<td>1.30 (.39)</td>
<td>1.22 (.38)</td>
<td>1.22 (.40)</td>
</tr>
<tr>
<td>Boys/Math</td>
<td>1.15 (.45)</td>
<td>1.19 (.45)</td>
<td>1.29 (.46)</td>
<td>N/A</td>
<td>1.22 (.45)</td>
</tr>
<tr>
<td>Boys/Control</td>
<td>1.08 (.54)</td>
<td>1.17 (.42)</td>
<td>1.24 (.52)</td>
<td>N/A</td>
<td>1.18 (.49)</td>
</tr>
<tr>
<td>ANS Accuracy</td>
<td>77.10 (12.34)</td>
<td>81.12 (11.02)</td>
<td>78.09 (12.21)</td>
<td>N/A</td>
<td>80.61 (11.41)</td>
</tr>
<tr>
<td>Girls</td>
<td>79.05 (14.07)</td>
<td>81.80 (10.73)</td>
<td>80.34 (12.12)</td>
<td>84.09 (9.39)</td>
<td>82.29 (10.93)</td>
</tr>
<tr>
<td>Boys</td>
<td>75.58 (10.68)</td>
<td>80.39 (11.34)</td>
<td>75.71 (11.91)</td>
<td>N/A</td>
<td>77.44 (11.63)</td>
</tr>
<tr>
<td>Math</td>
<td>76.92 (12.23)</td>
<td>80.84 (11.43)</td>
<td>77.10 (12.51)</td>
<td>83.92 (9.22)</td>
<td>81.15 (11.18)</td>
</tr>
<tr>
<td>Control</td>
<td>77.29 (12.58)</td>
<td>81.41 (10.62)</td>
<td>79.15 (11.86)</td>
<td>84.25 (9.58)</td>
<td>80.08 (11.62)</td>
</tr>
<tr>
<td>Girls/Math</td>
<td>79.16 (13.33)</td>
<td>81.09 (10.74)</td>
<td>78.43 (13.12)</td>
<td>83.92 (9.22)</td>
<td>81.49 (11.23)</td>
</tr>
<tr>
<td>Girls/Control</td>
<td>78.87 (15.66)</td>
<td>82.50 (10.76)</td>
<td>82.46 (10.63)</td>
<td>84.25 (9.58)</td>
<td>83.11 (10.58)</td>
</tr>
<tr>
<td>Boys/Math</td>
<td>74.49 (10.66)</td>
<td>80.59 (12.17)</td>
<td>75.63 (11.73)</td>
<td>N/A</td>
<td>77.40 (11.93)</td>
</tr>
<tr>
<td>Boys/Control</td>
<td>76.46 (10.80)</td>
<td>80.15 (10.43)</td>
<td>75.78 (12.19)</td>
<td>N/A</td>
<td>77.47 (11.38)</td>
</tr>
</tbody>
</table>
Table 3.3 Table of coefficients by experiment and analysis type. Condition is coded as 0 = Control, 1 = Test; Gender is coded as 0 = F, 1 = M.

<table>
<thead>
<tr>
<th>Beliefs x Gender x Condition predicting ANS performance</th>
<th>Beta</th>
<th>Study 4</th>
<th>Study 5</th>
<th>Study 6</th>
<th>Study 7</th>
<th>Mega-Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction</td>
<td>β = -.59 (.44)</td>
<td>t = 1.34</td>
<td>p = .184</td>
<td>β = -.44 (.27)</td>
<td>t = 1.61</td>
<td>p = .108</td>
</tr>
<tr>
<td>Beliefs x Condition predicting ANS performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls</td>
<td>β = .57 (.32)</td>
<td>t = 1.75</td>
<td>p = .193</td>
<td>β = .38 (.22)</td>
<td>t = 1.74</td>
<td>p = .083</td>
</tr>
<tr>
<td>Boys</td>
<td>β = -.02 (.30)</td>
<td>t = 0.08</td>
<td>p = .940</td>
<td>β = -.06 (.17)</td>
<td>t = 0.37</td>
<td>p = .712</td>
</tr>
<tr>
<td>Beliefs x Gender predicting ANS performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>β = .25 (.32)</td>
<td>t = 0.77</td>
<td>p = .441</td>
<td>β = .11 (.18)</td>
<td>t = 0.61</td>
<td>p = .543</td>
</tr>
<tr>
<td>Math</td>
<td>β = -.34 (.30)</td>
<td>t = 1.13</td>
<td>p = .260</td>
<td>β = -.33 (.21)</td>
<td>t = 1.61</td>
<td>p = .109</td>
</tr>
<tr>
<td>Condition x Gender predicting ANS performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Own-Gender-Math Association</td>
<td>β = .43 (.64)</td>
<td>t = 0.67</td>
<td>p = .506</td>
<td>β = .74 (.38)</td>
<td>t = 1.96</td>
<td>p = .051</td>
</tr>
<tr>
<td>High Own-Gender Math Association</td>
<td>β = -.75 (.58)</td>
<td>t = 1.29</td>
<td>p = .201</td>
<td>β = -.15 (.38)</td>
<td>t = 0.38</td>
<td>p = .701</td>
</tr>
</tbody>
</table>
### Condition predicting ANS performance

<table>
<thead>
<tr>
<th>Condition</th>
<th>Beta</th>
<th>Study 4</th>
<th>Study 5</th>
<th>Study 6</th>
<th>Study 7</th>
<th>Mega-Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Girls/Low</strong></td>
<td>$\beta = -0.56 (.51)$</td>
<td>$t = 1.14$</td>
<td>$p = 0.258$</td>
<td>$\beta = -0.60 (.29)$</td>
<td>$t = 2.11$</td>
<td>$p = 0.037^*$</td>
</tr>
<tr>
<td><strong>Girls/High</strong></td>
<td>$\beta = 0.56 (.41)$</td>
<td>$t = 1.36$</td>
<td>$p = 0.178$</td>
<td>$\beta = 0.33 (.29)$</td>
<td>$t = 1.12$</td>
<td>$p = 0.263$</td>
</tr>
<tr>
<td><strong>Boys/Low</strong></td>
<td>$\beta = -0.15 (.40)$</td>
<td>$t = 0.37$</td>
<td>$p = 0.711$</td>
<td>$\beta = -0.05 (.28)$</td>
<td>$t = 0.18$</td>
<td>$p = 0.856$</td>
</tr>
<tr>
<td><strong>Boys/High</strong></td>
<td>$\beta = -0.19 (.41)$</td>
<td>$t = 0.46$</td>
<td>$p = 0.644$</td>
<td>$\beta = 0.14 (.27)$</td>
<td>$t = 0.50$</td>
<td>$p = 0.619$</td>
</tr>
</tbody>
</table>

### Beliefs predicting ANS performance

<table>
<thead>
<tr>
<th>Condition</th>
<th>Beta</th>
<th>Study 4</th>
<th>Study 5</th>
<th>Study 6</th>
<th>Study 7</th>
<th>Mega-Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Math/Girls</strong></td>
<td>$\beta = 0.40 (.18)$</td>
<td>$t = 2.19$</td>
<td>$p = 0.031^*$</td>
<td>$\beta = 0.37 (.16)$</td>
<td>$t = 2.36$</td>
<td>$p = 0.020^*$</td>
</tr>
<tr>
<td><strong>Math/Boys</strong></td>
<td>$\beta = 0.06 (.24)$</td>
<td>$t = 0.24$</td>
<td>$p = 0.810$</td>
<td>$\beta = -0.06 (.12)$</td>
<td>$t = 0.49$</td>
<td>$p = 0.622$</td>
</tr>
<tr>
<td><strong>Control/Girls</strong></td>
<td>$\beta = -0.17 (.27)$</td>
<td>$t = 0.63$</td>
<td>$p = 0.532$</td>
<td>$\beta = -0.09 (.14)$</td>
<td>$t = 0.63$</td>
<td>$p = 0.530$</td>
</tr>
<tr>
<td><strong>Control/Boys</strong></td>
<td>$\beta = 0.08 (.18)$</td>
<td>$t = 0.45$</td>
<td>$p = 0.650$</td>
<td>$\beta = -0.15 (.14)$</td>
<td>$t = 1.08$</td>
<td>$p = 0.280$</td>
</tr>
</tbody>
</table>
### Gender predicting ANS performance

<table>
<thead>
<tr>
<th>Beta</th>
<th>Study 4</th>
<th>Study 5</th>
<th>Study 6</th>
<th>Study 7</th>
<th>Mega-Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Study 4</td>
<td>Study 5</td>
<td>Study 6</td>
<td>Study 7</td>
<td></td>
</tr>
<tr>
<td><strong>Math/Low</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta = -0.06$ (.41)</td>
<td>$\beta = 0.41$ (.29)</td>
<td>$\beta = 0.07$ (.27)</td>
<td>N/A</td>
<td>$\beta = -0.01$ (.15)</td>
</tr>
<tr>
<td></td>
<td>$t = 0.15$</td>
<td>$t = 1.44$</td>
<td>$t = 0.26$</td>
<td></td>
<td>$t = 0.06$</td>
</tr>
<tr>
<td></td>
<td>$p = 0.883$</td>
<td>$p = 0.152$</td>
<td>$p = 0.793$</td>
<td></td>
<td>$p = 0.955$</td>
</tr>
<tr>
<td><strong>Math/High</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta = -0.74$ (.42)</td>
<td>$\beta = -0.46$ (.27)</td>
<td>$\beta = -0.59$ (.29)</td>
<td>N/A</td>
<td>$\beta = -0.47$ (.15)</td>
</tr>
<tr>
<td></td>
<td>$t = 1.78$</td>
<td>$t = 1.69$</td>
<td>$t = 2.08$</td>
<td></td>
<td>$t = 3.16$</td>
</tr>
<tr>
<td></td>
<td>$p = 0.078$</td>
<td>$p = 0.093$</td>
<td>$p = 0.039^*$</td>
<td></td>
<td>$p = 0.002**$</td>
</tr>
<tr>
<td><strong>Control/Low</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta = -0.49$ (.49)</td>
<td>$\beta = 0.14$ (.28)</td>
<td>$\beta = -0.12$ (.14)</td>
<td>N/A</td>
<td>$\beta = -0.43$ (.15)</td>
</tr>
<tr>
<td></td>
<td>$t = 0.99$</td>
<td>$t = 0.50$</td>
<td>$t = 0.83$</td>
<td></td>
<td>$t = 2.89$</td>
</tr>
<tr>
<td></td>
<td>$p = 0.326$</td>
<td>$p = 0.615$</td>
<td>$p = 0.013^*$</td>
<td></td>
<td>$p = 0.004**$</td>
</tr>
<tr>
<td><strong>Control/High</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta = 0.01$ (.41)</td>
<td>$\beta = -0.27$ (.29)</td>
<td>$\beta = 0.01$ (.11)</td>
<td>N/A</td>
<td>$\beta = 0.25$ (.15)</td>
</tr>
<tr>
<td></td>
<td>$t = 0.02$</td>
<td>$t = 0.92$</td>
<td>$t = 0.09$</td>
<td></td>
<td>$t = 1.70$</td>
</tr>
<tr>
<td></td>
<td>$p = 0.982$</td>
<td>$p = 0.360$</td>
<td>$p = 0.076$</td>
<td></td>
<td>$p = 0.090$</td>
</tr>
</tbody>
</table>
Figure 3.1 Example of two trials from the Approximate Number System (ANS) Task.

Easier Ratio (5.0)  Harder Ratio (1.4)
Figure 3.2 Girls’ ANS task performance by condition.
Figure 3.3 Boys’ ANS task performance by condition.
Chapter 4: Malleability of Intergroup Bias

4.1 Synopsis

Though past research suggests that implicit intergroup bias may be more malleable in childhood than adulthood, no studies thus far have conducted a direct comparison of bias change across development. The first study in this chapter investigates whether exposure to counter-stereotypical exemplars is more effective at reducing implicit pro-White/anti-Black racial bias in children ages 5-12 as compared to adults. We found that while this method successfully reduced children’s bias both immediately after the intervention and after an hour-long delay, it did not decrease the implicit racial bias of adults. The second study examined whether explicit evaluative instructions might increase the efficacy of this intervention with adults and found that this was the case: racial bias was significantly reduced at both time points for adults who were exposed to counter-stereotypical exemplars and received these instructions. These results suggest that while implicit intergroup bias may be similarly malleable in adulthood and childhood, changing implicit bias in children may be easier, as they require less explicit instruction.

4.2 Introduction

Over the past decade, a growing number of studies have provided evidence for the early emergence of implicit racial bias (e.g. Baron, 2015; Baron & Banaji, 2006; Baron & Banaji, 2009; Dunham et al., 2006, 2007; Newheiser & Olson, 2012; Dunham et al., 2014; Rutland, Cameron, Milne, & McGeorge, 2005; Steele et al., 2018; Williams & Steele, 2017). Children as young as three years old show implicit preferences based on race, and these biases appear to be driven by in-group favoritism as well as a preference for high-status racial groups (Baron, 2015; Dunham et al., 2006, 2007; 2013 Newheiser et al., 2014; Qian et al., 2016; Setoh et al., 2017). Importantly, when cultural messages about high and low status racial groups are pervasive, these
biases remain developmentally stable, with children showing comparable levels of bias to adults (Baron & Banaji, 2006; Dunham et al., 2013; see also Baron, 2015).

For example, in North America, where systemic bias and cultural messages from media sources (e.g. portrayals in the news, movies, television) perpetuate negative stereotypes of Blacks, majority group White children demonstrate an implicit racial preference for Whites over Blacks across development (Baron & Banaji, 2006; Dunham et al., 2013; Gonzalez, Steele, & Baron, 2017; Hall, Hall, & Perry, 2016; Tukachinsky, Mastro, & Yarchi, 2017; Weisbuch, Pauker, & Ambady, 2009). This bias persists into adulthood and is linked to negative behavior toward Blacks, such as unfriendliness in interracial interactions, biased voting, and disparate healthcare decisions (Dovidio et al., 2005; Green et al., 2007; Greenwald et al., 2009; Payne et al., 2008). As a result, recent work has examined whether and how these implicit racial biases can be changed.

Unfortunately, it appears to be relatively difficult to change implicit racial bias in adults (Lai et al., 2016). In a comparison of nine interventions designed to reduce bias, each intervention successfully reduced biases in the short-term, but none successfully changed bias over the course of 24 hours. It has been hypothesized that due to cultural reinforcement of racial biases across the lifespan, it may be easier to change these attitudes in childhood as compared to later in development (Devine, 1989; Greenwald & Banaji, 1995; Rudman, 2004). Consistent with this possibility, several studies with children have successfully reduced implicit racial bias immediately following an intervention, and importantly, these effects appear to have lasted days, weeks, and even years (Neto, Pinto, & Mullet, 2015; Vezzali, Capozza, Giovannini, & Stathi, 2011).

Specifically, two intensive interventions lasting several weeks successfully reduced 10 to 12-year-old children’s implicit racial bias. In one such intervention, Portuguese children were
exposed to twenty music classes aimed at decreasing anti-dark-skin prejudice (Neto et al., 2015) by exposing children to music and musicians from Cape Verde. Children in this music program showed significantly less implicit anti-dark skin prejudice up to two years later, as compared to a control group of children who did not receive this intervention. The other intervention involved multiple sessions of imagined out-group contact over the course of three weeks to decrease anti-immigrant prejudice in Italian children. Children exposed to this intervention had lower levels of implicit bias one-week following the conclusion of the intervention as compared to those in a control condition (Vezzali et al., 2011). Taken together, these studies suggest that intensive interventions involving exposure to positive associations and intergroup contact may be able to reduce children’s implicit bias for longer periods of time.

While these intensive interventions are promising, they require extensive resources to design and implement. As such, researchers have recently examined whether brief interventions, which are more efficient and scalable to administer, can also reduce implicit racial biases. For example, researchers have been able to reduce implicit racial bias in children as young as age 3 years of age using perceptual individuation training (Qian et al., 2017; Xiao et al., 2015). This method operates on the premise of the perceptual-social linkage hypothesis, which posits that difficulties in individuating other-race faces facilitates the generalization of racial bias across outgroup members (Lee, Quinn, & Heyman, 2017). Thus, teaching children to individuate other-race faces should interrupt this process and decrease their negativity toward racial outgroups. Children in these studies were trained to individuate different Black faces (Xiao et al., 2015). Following the individuation training, children had significantly lower levels of anti-Black racial bias relative to before training. Furthermore, when this intervention was administered more than once, levels of bias in children were significantly lower seventy days after the final training.
session (Qian et al., 2017). This work suggests that brief interventions can reduce racial bias in children, and with repeated use, may induce longer lasting bias change.

In the current set of studies, we examined potential developmental differences in the efficacy of an alternative brief intervention: exposure to positive outgroup exemplars. This type of intervention is practical to implement and has been shown to temporarily reduce implicit racial bias in adults across several studies (e.g. Columb & Plant, 2011; Dasgupta & Greenwald, 2001; Lai et al., 2014; 2016; Marini, Rubichi, & Sartori, 2011). Furthermore, as evidence with children suggests that implicit biases can be internalized immediately after exposure to short stories, it is plausible that this method could also be used to reduce existing biases (Gonzalez, Dunlop, & Baron, 2017). In the first study examining this method of bias reduction in children, 5 to 12-year-olds were exposed to short stories about Black adults who contributed positively within their community (Gonzalez, Steele, & Baron, 2017). Compared to children in control conditions, which exposed children to positive White exemplars or positively valenced descriptions of flowers, this brief exposure to positive Black exemplars reduced pro-White/anti-Black racial bias in children nine and up. For children ages eight and younger, this exposure did not successfully reduce bias.

It is possible that this intervention failed to be effective for this younger age group because they did not spontaneously categorize the positive exemplars by race (Pauker, Williams, & Steele, 2016; Shutts, Banaji, & Spelke, 2010). Instead of encoding the people in each story as positive Black exemplars, children may have focused on other social categories like gender or age (Kinzler et al., 2010; Shutts, 2015; Williams & Steele, 2017). As such, the first goal of the current research was to examine whether different counter-stereotypical exemplar exposure, that makes race more salient, would reduce implicit racial bias in both younger (ages 5-8) and older (ages 9-12) children.
The second goal of the current research was to directly compare the efficacy of an implicit racial bias intervention between children and adults. The current body of literature suggests that childhood may indeed be an optimal point in development to reduce implicit racial bias, as studies with children have been effective in bias change over time (Neto et al., 2015; Qian et al., 2017; Vezzali et al., 2011). However, this assessment is based on the results of a handful of indirect studies, as no studies thus far have directly examined whether specific interventions to reduce implicit racial bias are more effective with children as compared to adults. A direct comparison between adults and children would provide much needed insight into the malleability of implicit racial bias and developmental differences in the conditions required to elicit change.

We based the current method of counter-stereotypical exemplar exposure on earlier work with children, with adjustments to vignette content to make race more salient (Gonzalez, Steele, & Baron, 2017). In order to increase race salience, we included both Black and White exemplars in our experimental condition. In contrast to previous work, which only exposed children to positive Black adult exemplars, in the current research both children and adults were exposed to a pair (a boy and a girl) of positive Black child exemplars who engaged in pro-social actions and a pair (a boy and a girl) of negative White child exemplars who engaged in anti-social actions. In the control conditions, both sets of exemplars were White.

Based on the results of previous research, where exposure to positive Black exemplars decreased bias in children age nine and up, we hypothesized that at minimum, older children who were exposed to positive Black exemplars would show lower levels of implicit pro-White/anti-Black racial bias in comparison to those exposed to positive White exemplars. Although we did not have concrete predictions about developmental differences across childhood, given that other interventions have been effective with young children (Qian et al.,
2017; Xiao et al., 2015), it seemed plausible that with these modifications, our intervention would also be effective with younger children.

Our predictions for the effects of this intervention with adults were also mixed. Past work would suggest that exposure to positive Black exemplars can reduce racial bias in adults (e.g. Dasgupta & Greenwald, 2001; Lai et al., 2017); however, these studies have often taken more explicit approaches to bias change, using either well-known exemplars, or supplementing stories with evaluative instructions to help adults internalize counter-stereotypical associations. To our knowledge, no interventions thus far have tested the effectiveness of counter-stereotypical exemplar exposure among adults using a more simplistic, child-friendly story. As a result, we examined potential developmental differences in the efficacy of this intervention between younger children (ages 5-8), older children (ages 9-12) and adults.

In addition to measuring the immediate post-intervention effects of counter-stereotypical exemplar exposure on implicit intergroup bias, we also assessed the potential of this method to produce a reduction in implicit race bias that lasts up to an hour after the brief exposure session. Thus, the implicit racial bias of children and adults was tested twice: first, immediately after exposure to exemplar vignettes and second, one hour after reading the vignettes. The first testing point falls within typical limits of priming effects (usually a maximum of 15-20 minutes; Roskos-Ewoldsen et al., 2009), while the second testing point extends into a time period when priming effects have usually decayed significantly (Higgins, Bargh, & Lombardi, 1985; Srull & Wyer, 1979). Thus, the current study assessed whether exposure to counter-stereotypical exemplars temporarily shifts implicit intergroup bias in a manner similar to priming effects, or whether this change is more long-lasting. If implicit bias reduction failed to last beyond an hour, this would suggest that this manipulation may not be powerful enough on its own to induce long-term bias change.
4.3 Study 8

4.3.1 Method

4.3.1.1 Participants

Child Participants. A total of 376 White and Asian participants who met our preregistered eligibility criteria were recruited from a community science center from August 2016 to November 2017 and received a sticker for participation in the study. Participants were recruited by research assistants walking around the science center or entered the lab space on their own and asked to participate. Parents gave consent for their child’s study participation.

We preregistered a goal of recruiting 120 useable child participants ages 5-8 and 120 useable child participants ages 9-12 (https://osf.io/rjpmz/). We stopped running participants after we believed we had met our goal of 120 participants per age group who completed both Time 1 and Time 2 measures. Separately, 59 additional participants were excluded for significant lack of understanding or random button pressing on the IAT (n = 14), experimenter errors in study protocol (n = 10), reported developmental delays (n = 10), technology errors (n = 5), parent interference with the study (n = 1), and failing to complete the study at Time 1 (n = 16) and for answering less than four out of six comprehension questions correctly (n = 3). In accordance with our preregistration and previous research, we also excluded participants with more than 25% of their response latencies under 300 ms and participants with errors on 25% or more of trials (Baron & Banaji, 2006; Gonzalez, Steele, & Baron, 2017). These exclusions are listed for each sample.

After post data collection exclusions, our final total of participants who completed Time 1 and Time 2 measures is below our goal of 120. Our lab space is set up as an exhibit, and as such, we ran a number of participants who completed the first portion of the study but were
unable to complete Time 2 measures. To increase power, we include these participants in our analysis that only includes Time 1 participants.

**All Time 1 participants.** After the exclusions detailed above, we had a total of 376 useable child participants who completed all of our Time 1 measures. Of these participants, 40 additional exclusions were made based on IAT response latencies and errors at Time 1 only, resulting in a final sample of 336 participants. A post-hoc power analysis using G*Power indicates that this sample size would give us greater than 95% power to detect a medium sized between-subjects interaction (Erdfelder, Faul, & Buchner, 1996). Of these 336 participants, 62.2% identified as Caucasian (n = 209), 28% identified as East Asian (n = 94), and 9.8% identified as mixed race (Caucasian and East Asian, n = 33). The mean age of the 162 younger participants (ages 5-8) was 6.9 years (SD = 1.14), and the mean age of the 174 older participants (ages 9-12) was 10.9 years (SD = 1.05).

**Time 1 and Time 2 subset.** After excluding participants who were unable to complete our Time 2 measures (n = 126), we had a total of 210 useable child participants. Ten additional exclusions were made based on IAT response latencies at and Time 2 (see Data Preparation in Results), resulting in a final sample of 200 useable participants with both Time 1 and Time 2 data. A post-hoc power analysis using G*Power indicates that this sample size would give us greater than 85% power to detect a medium sized between-subjects interaction. Of these 200 participants, 60.5% identified as Caucasian (n = 121), 31.5% identified as East Asian (n = 63), and 8% identified as mixed race (Caucasian and East Asian, n = 16). The mean age of the 97 younger participants (ages 5-8) was 6.9 years (SD = 1.12), and the mean age of the 103 older participants (ages 9-12) was 10.8 years (SD = 1.04).

**Adult Participants.** A total of 119 White and Asian participants who met our preregistered eligibility criteria were recruited from our university human subject pool from
August 2016 to November 2017 and received either course credit or $10 as compensation for
study participation. We preregistered a goal of recruiting 120 useable adult participants as a third
age group to look at the efficacy of this intervention across development. We stopped running
participants after we believed we had met our goal of 120 participants.

Subsequently, 19 participants were excluded for technology and/or experimenter errors in
protocol (n = 6) and failing basic attention checks (n = 13). Four additional participants were
excluded based on IAT response latencies and errors, resulting in a final sample of 96
participants. All participants completed measures at Time 1 and Time 2. Of these participants,
67.7% identified as East Asian (n = 65), 30.2% identified as Caucasian (n = 29), and 2.1%
identified as mixed race (Caucasian and East Asian, n = 2). The mean age of participants was
21.8 years old (SD = 3.88), and all participants were in the process of completing, or had a
university degree.

4.3.1.2 Procedure

Child participants were tested in our lab space at Science World, and each participant was
tested individually. An experimenter read all instructions, stories and explicit questions out loud
to each child. Participants were randomly assigned to either our experimental condition, where
they were read stories about Black prosocial and White antisocial characters, or our control
condition, where they were read about White prosocial and White antisocial characters. After the
stories, children completed the Child Implicit Association Test (Child IAT). For the next hour,
children explored Science World with their parents as they normally would. After an hour, we
buzzed them using a pager system to get them to return to our lab. We then administered another
Child IAT, as well as the same explicit questions.
Adult participants were tested in our lab space at the University of British Columbia, and each participant was tested individually. The order of tasks was identical to child participants. Experimenters read directions for each individual portion of the task, but adult participants read the stories themselves, and completed the tasks on their own. For the hour between Time 1 and Time 2, participants were able to leave the testing room and do whatever they would normally do in that time period.

4.3.1.3 Measures

Vignettes. All participants were presented with three different stories using the same characters (see Appendix A for text). Two characters were a pair of children who engaged in antisocial behavior. These children were White in both conditions. The other two characters were a pair of children who engaged in prosocial behavior. Depending on condition, these positive child exemplars were either White (control condition) or Black (experimental condition). The stories were presented in the same order. Stories also included images of the characters. Drawings of the prosocial characters were matched in posture and affect.

Story Comprehension Questions. After each of the stories, all participants were asked two questions to ensure their understanding of which characters partook in certain actions (see Appendix A).

Child Implicit Association Test. Implicit racial bias was measured using a child-friendly Implicit Association Test. This test measures the strength of an association between a target category and an attribute. In this IAT, we measured associations between race (Black/White) and affect (good/bad). The category reminders for the two racial groups were represented with eight images of racially prototypical Black and White children matched in age and attractiveness (four children of each race; stimuli from Baron & Banaji, 2006, Gonzalez, Steele, & Baron, 2017). A
smiling and a frowning face served as category reminders of “good” and “bad”. Stimuli for the racial groups were the same eight images as the category reminders, presented one at a time. The stimuli for “good” and “bad” were presented acoustically, using words from other Child IAT studies (good words: happy, fun, good, nice, bad words: yucky, sad, mad, mean; Baron & Banaji, 2006; Gonzalez, Steele, & Baron, 2017). These voices were recorded by a female voice speaking each word in an affectively congruent manner. Participants were presented with two JellyBean buttons in front of the monitor that color matched with the side of the screen they were placed in front of (yellow on the left, blue on the right). Participants were told that any time they see an image or hear a word, they should determine which category it belongs in, and press the correct button.

Participants began by categorizing images as either Black or White. They completed 12 of these practice trials. Next, participants categorized words they heard into “good” or “bad”. Participants completed 20 of these practice trials. Following the two practice trials, children completed a test block with 30 trials where they were presented with either an image or a word. To classify these stimuli, children used the same buttons for attributes (good/bad) and target categories (Black/White). For example, during this critical test block, children might have paired Black+good on one button and White+bad on the other. After the critical test block, participants completed another practice block where the sides of the target categories were reversed. In this block, participants only classified images, and completed 20 trials. Finally, participants completed another test block, where the pairing of the attributes and target categories was switched (e.g. Black+good, White+bad for the first test block, Black+bad, White+good for the second). The side for target categories and attributes was counterbalanced across conditions. This measure is designed to measure the strength of the association between paired stimuli by recording children’s reaction times during the two distinct test blocks. Specifically, this test
measures the relative positivity and negativity that participants associate with Black and White racial groups.

4.3.2 Results

For each participant, an IAT D-score was calculated, which represents the magnitude of a participant’s implicit preference for White vs. Black racial groups (see Baron & Banaji, 2006; Greenwald et al., 2003). Our data were coded such that a positive score indicated implicit preference for White over Black, and a negative score indicated an implicit preference for Black over White.

4.3.2.1 Time 1 Analyses

All participants. When all three age groups were entered into a 2 (Condition: Experimental or Control) x 3 (Age Group: Younger, Older, Adult) ANOVA, we found a significant effect of Condition ($F(1,429) = 4.43, p = .036, \eta^2 = .01$) such that participants who were exposed to counter-stereotypical exemplars in the experimental condition ($D = .09, SD = .49$) showed lower levels of bias than those in the control condition ($D = .21, SD = .51$). There was also a marginal effect of Age Group ($F(2,429) = 2.23, p = .11, \eta^2 = .01$), and no Condition x Age Group interaction ($F(2,429) = .64, p = .53, \eta^2 = .003$). However, as these results could be due to differences in intervention efficacy for adults and children, we conducted additional ANOVAs with children and adults only to see if effects of condition would hold in these separate age groups. The ANOVA with child participants only also serves as a conceptual replication of Gonzalez et al. (2017).

Child participants. When including child participants only, we found a significant effect of Condition ($F(1,332) = 6.52, p = .011, \eta^2 = .02$), no significant difference between Younger
and Older Children \( F(1,332) = 0.38, p = .54, \eta^2 = .001 \), and no Condition by Age Group interaction \( F(1,332) = 0.50, p = .48, \eta^2 = .001 \). These results indicate that children in our control condition, who were exposed to stereotypical exemplars \( (D = .19, SD = .50) \), had significantly higher levels of bias than children in our experimental condition, who were exposed to counter-stereotypical exemplars \( (D = .05, SD = .48) \). Thus, our results conceptually replicate the finding from Gonzalez et al., (2017), that exposing children to counter-stereotypical exemplars decreased children’s implicit racial bias. However, unlike the aforementioned paper, there was no difference between older and younger children, suggesting that the intervention used in the current study was effective for both age groups.

Testing mean levels of bias against chance \( (\mu = 0) \) indicated that immediately after the intervention (Time 1), children who were exposed to stereotypical exemplars \( (D = .19, SD = .50) \) had a significant level of pro-White/anti-Black implicit racial bias, \( t(163) = 4.96, p < .001 \). In contrast, children who were exposed to counter-stereotypical exemplars \( (D = .05, SD = .48) \) did not appear to prefer either racial group, \( t(171) = 1.47, p = .14 \).

**Adult participants.** An independent samples t-test indicated that there was not a significant difference in mean levels of implicit pro-White/anti-Black racial bias between adults in the stereotypical exemplar condition \( (D = .26, SD = .56) \) and adults in the counter-stereotypical exemplar condition \( (D = .22, SD = .49) \), \( t(97) = .36, p = .72 \), Cohen’s \( d = 0.08 \). These results suggest that for our sample of adults, exposure to counter-stereotypical exemplars was not effective in reducing implicit racial bias.

Testing mean levels of bias against chance \( (\mu = 0) \) indicated that immediately after the intervention (Time 1), adults who were exposed to stereotypical exemplars \( (D = .26, SD = .56) \) had a significant level of pro-White/anti-Black implicit racial bias \( (t(48) = 3.28, p = .002) \), as did
those who were exposed to counter-stereotypical exemplars ($D = .22, SD = .49$), $t(49) = 3.15, p = .003$.

### 4.3.2.2 Change from Time 1 to Time 2

**All participants.** To look at differences in IAT score directly after our manipulation (Time 1) as compared to one hour later (Time 2), we first conducted a 2 (Time: Time 1 or Time 2) x 2 (Condition: Experimental or Control) x 3 (Age Group: Younger, Older, Adult) mixed-factorial ANOVA. When all three age groups were entered into the ANOVA, no interactions between any of the variables were significant ($ps > .22$). We found no significant main effect of Time ($F(1,290) = 1.49, p = .22, \eta^2 = .005$), but a marginal effect of Condition ($F(1,290) = 3.70, p = .056, \eta^2 = .013$), and a significant effect of Age Group ($F(2,290) = 3.32, p = .038, \eta^2 = .022$). The marginal effect of Condition indicated that participants in the experimental condition ($D = .13, SD = .55$) showed significantly lower levels of bias than those in the control condition ($D = .22, SD = .55$). Simple effects analyses indicate that overall, adults showed significantly higher levels of bias ($D = .26, SD = .55$) than younger children ($D = .12, SD = .56; p = .014$), and a marginally higher level of bias than older children ($D = .16, SD = .55, p = .079$). There was no significant difference between younger and older children ($p = .43$).

**Child participants.** Again, to examine potential age differences for children, we ran the ANOVA with children only. No interactions between any of the variables were significant ($ps > .15$). We also found that there was no significant main effect of Time ($F(1,196) = 1.53, p = .22, \eta^2 = .008$) or Age Group ($F(1,196) = 0.75, p = .39, \eta^2 = .004$). However, there was a significant effect of Condition ($F(1,196) = 6.55, p = .011, \eta^2 = .032$), such that children in who were exposed to counter-stereotypical exemplars had significantly lower levels of implicit racial
bias ($D = .07, SD = .57$) than children who were exposed to stereotypical exemplars ($D = .20, SD = .56$). These results suggest that regardless of child age and post-intervention delay (i.e. Time 1 or Time 2), exposure to counter-stereotypical exemplars significantly reduced children’s implicit racial bias.

To further examine potential effects of Time delay on bias, we tested mean levels of bias against chance ($\mu = 0$) for results at Time 2, an hour after the initial intervention. Children who were exposed to stereotypical exemplars had significant levels of bias one hour after the stories ($t(98) = 4.95, p < .001$). Children who were exposed to counter-stereotypical exemplars also had significant levels of bias one hour later ($t(100) = 2.08, p = .04$). The inconsistency of these results compared with those above suggest that bias may have begun to return to pre-intervention levels after an hour.

**Adult Participants.** When an ANOVA was conducted with adults only, there were no effects of Time ($F(1,94) = 0.12, p = .73, \eta^2 = .001$), Condition ($F(1,94) = 0.001, p = .98, \eta^2 < .001$), or a Time by Condition interaction ($F(1,94) = 0.29, p = .59, \eta^2 = .003$), suggesting that adult levels of implicit racial bias were not different between Time 1 and Time 2, and bias was not significantly decreased by exposure to counter-stereotypical exemplars.

We also tested mean levels of bias against chance ($\mu = 0$) for results at Time 2. Adults who were exposed to stereotypical ($t(46) = 3.09, p = .003$) and counter-stereotypical exemplars ($t(48) = 3.66, p = .001$) had significant levels of implicit pro-White/anti-Black racial bias, further suggesting that this intervention was not effective for adults.
4.3.3 Discussion

In summary, we found that children who were exposed to positive Black exemplars showed significantly less pro-White/anti-Black bias than those in the control condition who were exposed to positive White exemplars. Furthermore, children who were exposed to positive Black exemplars no longer showed significant levels of pro-White/anti-Black bias and had no implicit preference for either racial group immediately following the intervention. In contrast to previous work, there were no age differences between younger and older children. However, we did find that this intervention was not as effective for adults, as there was no significant difference in levels of bias between adults who were exposed to positive Black versus White exemplars, and adults in both conditions had significant pro-White/anti-Black bias. Thus, it appears that while this intervention successfully reduced implicit racial bias in children, it failed to reduce bias for adults. For all participants, levels of bias did not change significantly over the course of an hour.

4.4 Study 9

Consistent with previous research, exposure to counter-stereotypical exemplars reduced implicit bias in children (Gonzalez, Steele, & Baron, 2017). Furthermore, the success of this intervention extended to children below the age of nine, indicating that counter-stereotypical exemplar exposure can reduce bias across the age range of 5-12. These results add to a growing body of work suggesting that implicit racial bias can be changed in childhood (e.g. Qian et al., 2017; Xiao et al., 2015). Results an hour after exposure were mixed; tests against chance suggest that pro-White/anti-Black bias may return after an hour, but bias levels were not significantly different between Time 1 and Time 2. It may be the case that like perceptual individuation, brief exposure to counter-stereotypical exemplars may need to occur more than once to induce longer-lasting effects. Regardless, these results suggest that even after an hour, levels of implicit pro-
White/anti-Black racial bias are lower after counter-stereotypical exemplar exposure as compared to when children have been exposed to stereotypical exemplars.

Importantly, in contrast to other research, this intervention did not reduce bias in adults. As other studies have successfully reduced adults’ implicit racial bias in the short-term using more adult-oriented exposure (see Lai et al., 2016), it is possible that the child friendly nature of our stories may have limited their success. Specifically, adult participants may be more likely to experience reactance in response to the overtness of our stories, resulting in a lack of bias change (Brehm, 1966). The increased racial salience within this intervention may be useful for children, but may make the purpose of the stories too obvious for adults. Another possibility is that our use of child exemplars may not be effective in changing adults’ biases. Adults may subtype child exemplars as being less representative of the broader racial group and consequently, fail to generalize the positivity from these stories (Richards & Hewstone, 2001; Williams & Steele, 2017).

Alternatively, this manipulation may actually not have been explicit enough for adults. Several studies that have successfully changed implicit racial bias in adults have used well-known exemplars (Columb & Plant, 2011; Dasgupta & Greenwald, 2001; Joy-Gaba & Nosek, 2010), a strategy that utilizes existing implicit associations to shift bias. Others have paired unknown exemplar exposure with strategies such as self-involvement in story scenarios, or evaluative instructions (Lai et al., 2014; 2016; Marini et al., 2012). It is possible that these types of additions, all of which may aid participants in internalizing the presented associations, are essential to induce implicit racial bias change in adults.

If adults do require additional information to internalize implicit associations from a counter-stereotypical exemplar story, this would suggest that this type of intervention might be more effective with children. In order to test this hypothesis, we examined whether the addition
of explicit evaluative instructions would lead to the successful reduction of implicit pro-
White/anti-Black racial bias in adults, even when a child-friendly story was used. This work
follows two recent studies testing several interventions to reduce implicit racial bias in adults,
where researchers were able to successfully reduce bias through exposure to counter-
stereotypical exemplars, but specifically included evaluative instructions (e.g. following
exposure to the counter-stereotypical exemplars, participants were told to “think ‘Good’ when
you see the faces of your Black teammates, and ‘Bad’ when you see the white faces from the
cheating team.”; Lai et al., 2016).

All participants in this study were exposed to the counter-stereotypical story from Study
8 where the pro-social children were Black and the anti-social children were White. In one
condition, participants read the story on its own, in the same manner as the experimental
condition in Study 8. In the other condition, after participants read the story, they received
additional instructions to internalize the presented association (Black=good, White=bad). Once
again, to examine bias after a delay, we also tested participant levels of bias immediately after
the intervention (Time 1) and an hour later (Time 2). Thus, our study followed a 2 (Time 1 or
Time 2) x 2 (Condition: Evaluative Instructions or No Instructions) design.

4.4.1 Method

4.4.1.1 Participants

A total of 125 White and Asian participants who met our preregistered eligibility criteria
were recruited from our university human subject pool from March 2018 to September 2018 and
received either course credit or $10 as compensation for study participation
(https://osf.io/rjpmz/). We preregistered a goal of recruiting 120 useable adult participants and
stopped running participants after we believed we had met this goal.
Subsequent participant exclusions were made based on our preregistered criteria. Five participants were excluded due to experimenter errors in study protocol, six participants were excluded for getting less than five out of six story comprehension questions correct, and twelve participants were excluded for failing basic attention checks. Six additional exclusions were made based on IAT response latencies and errors at Time 1 and Time 2, resulting in a final sample of 96 useable participants. A post-hoc power analysis using G*Power indicates that this sample size would give us over 95% power to detect a medium sized interaction (with a correlation of .16 between repeated measures). Of these 96 participants, 79.2% identified as East Asian (n = 76) and 20.8% identified as Caucasian (n = 20). The mean age of participants was 21.8 years old (SD = 3.15), and all participants were in the process of completing, or had a university degree.

4.4.1.2 Procedure

The order of tasks for adult participants was identical to Study 8, with the addition of instructions for participants in the Evaluative Instructions condition. In both conditions, participants were exposed to counter-stereotypical exemplars and read about Black pro-social and White anti-social characters. Participants were randomly assigned to either our Evaluative Instructions condition, where they were given additional instructions to internalize the presented association (Black=good, White=bad) or our No Instruction condition where they are given no additional instructions.

4.4.1.3 Measures

**Evaluative Instructions.** These instructions were provided after comprehension questions, but before the Child IAT. Participants in the Evaluative Instructions condition were
told: “In a moment you are going to complete a task designed to firmly establish in people’s minds, even in difficult and misleading situations, that White = Bad, Black = Good. To make this new task easier, remember the story you just read, and how the White characters participated in negative and antisocial actions, as well as how the Black characters participated in positive and prosocial actions. On the remainder of tasks, think “good” when you see an image of a Black individual and think “bad” when you see an image of a White individual.”

Participants in the No Instruction condition did not receive these instructions and instead proceeded directly to the IAT after reading the vignettes and completing the comprehension questions At Time 2, regardless of condition, no additional instructions were given: participants proceeded directly to the Child IAT.

4.4.2 Results

As done in Study 1, we calculated an IAT D-score for each participant. A 2 (Time 1 or Time 2) x 2 (Condition: Evaluative Instructions or No Instructions) ANOVA revealed no significant main effect of Time ($F(1,94) = .401, p = .53, \eta^2 = .004$), or Time by Condition interaction ($F(1,94) = 2.30, p = .13, \eta^2 = .024$). There was a significant effect of Condition ($F(1,94) = 6.70, p = .011, \eta^2 = .067$) such that participants who were only exposed to counter-stereotypical exemplars ($D = .25, SD = .54$) had significantly more implicit pro-White/anti-Black racial bias than participants who were exposed to counter-stereotypical exemplars and received evaluative instructions ($D = -.07, SD = .52$). These results suggest that compared to the counter-stereotypical exemplar intervention alone, the addition of evaluative instructions significantly reduced adult levels of implicit pro-White/anti-Black racial bias.
Testing mean levels of bias against chance ($\mu = 0$) indicated that immediately after the intervention (Time 1), participants who received additional evaluative instructions ($D = -.07, SD = .52$) showed no implicit preference for either racial group $t(48) = -1.02, p = .31$, while participants who did not receive additional instructions ($D = .25, SD = .54$) showed a significant pro-White/anti-Black implicit racial bias, $t(46) = 3.15, p = .003$, further supporting the conclusion that this intervention was effective with the use of evaluative instructions, but not without. Levels of bias showed a similar pattern one hour after the intervention (Time 2), with participants who received additional evaluative instructions ($D = .08, SD = .51$), $t(48) = 1.07, p = .29$, showing no bias, and those who did not receive additional instructions showing significant bias ($D = .18, SD = .58$), $t(46) = 2.18, p = .04$.

4.4.3 Discussion

We found that as compared to counter-stereotypical exemplar exposure alone, addition of evaluative instructions successfully decreased adults’ implicit pro-White/anti-Black implicit racial bias. Accordingly, participants who were exposed to the stories paired with evaluative instructions did not show implicit preference for either racial group, while those who were exposed to the stories alone had significant pro-White/anti-Black implicit bias. Bias did not change significantly over the course of an hour. Furthermore, additional analyses indicated that levels of bias were comparable between children who were exposed to counter-stereotypical exemplars, and adults who were exposed to counter-stereotypical exemplars and received evaluative instructions.
4.5 **General Discussion**

Across two studies, children and adults were exposed to positive child exemplars in an attempt to change implicit pro-White/anti-Black racial bias. In our first study, immediately after reading the stories, children ages 5-12 who were exposed to positive Black exemplars and negative White exemplars showed significantly less implicit bias than those who were exposed to positive and negative White exemplars. Furthermore, children in this condition no longer showed an implicit preference for White over Black. After an hour delay, levels of bias remained lower for children exposed to positive Black exemplars, though an implicit pro-White preference appears to have returned after this time delay. Importantly, there were no significant differences between younger and older children in intervention effectiveness.

Though this finding contrasts with previous work suggesting that counter-stereotypical exemplar exposure is more effective with older children (ages 9-12) and does not reduce bias in younger children (ages 5-8), it is consistent with other recent work indicating that teaching preschool children (ages 4-6) to perceptually individuate other-race faces can decrease implicit racial bias (Xiao et al., 2015; Qian et al. 2017). It is possible that exposure to counter-stereotypical exemplars could constitute a type of psychological individuation (see Brewer, 1988; Fiske & Neuberg, 1990), potentially leading participants to make novel evaluations of new racial group members rather than relying on stereotypes. Another possibility is that these two methods may operate differently; perceptual individuation interrupts generalization of bias to group members, but exposure to counter-stereotypical exemplars may more directly target the underlying implicit association.

Furthermore, while perceptual individuation can significantly reduce bias in preschoolers for over two months, it is unclear whether or not this method of bias change would be equally effective over time with children above age 6. It is possible that perceptual individuation is
particularly effective with preschool children because they have a weaker understanding of cultural attitudes and stereotypes about racial groups; recent evidence suggests that children’s implicit racial bias may be primarily driven by in-group preference at this age, rather than an understanding of group status within one’s culture (Qian et al., 2016). Perceptual individuation may be effective at disrupting in-group bias, but it remains unclear whether this method successfully counteracts later forms of bias driven by cultural status. Counter-stereotypical exemplar exposure may be more effective with older children by counteracting cultural messages more directly, rather than interrupting the generalization of bias to a racial group.

In contrast to our findings with children, exposure to counter-stereotypical exemplars alone did not significantly reduce adults’ implicit pro-White/anti-Black bias. Adults who were exposed to positive and negative White exemplars and those who were exposed to positive Black and negative White exemplars both had a significant implicit pro-White/anti-Black preference. This bias was present in both conditions immediately after reading the stories, as well as after an hour delay. This finding contrasts with previous research indicating that adults’ implicit racial bias can be changed through counter-stereotypical exemplar exposure; however, none of these prior studies have exposed adults to unknown, child-friendly exemplars without some kind of explicit evaluative instruction to internalize the presented association. As such, the current research suggests that this type of exposure is not enough to change adults’ implicit racial bias.

In our second study, we were able to successfully reduce adults’ implicit racial bias by adding evaluative instructions to our counter-stereotypical exemplar manipulation. In a further replication of Study 1, adults who were only exposed to positive Black and negative White exemplars without any sort of instruction had significant levels of pro-White/anti-Black bias both immediately after story exposure and after an hour-long delay. In comparison, adults who read these stories and then received additional instructions (to internalize a Black=good, White=bad
association) had reduced levels of implicit bias such that they showed no preference for either racial group. This reduction in bias also lasted after an hour, beyond the usual bounds of priming effects (Higgins et al., 1985; Roskos-Ewoldsen et al., 2009; Srull & Wyer, 1979). It is important to note that making the purpose of the IAT apparent to participants through use of these instructions should not be enough to change bias, as it is relatively difficult for participants to control their responses on the IAT without explicit strategies (Banse, Seise, & Zerbes, 2001; Kim, 2003; Steffens, 2004). As such, controllability of response is unlikely to be the driving mechanism behind this bias reduction.

Our findings suggest that some kind of explicit or conscious change may need to be present for counter-stereotypical exemplar exposure to effectively reduce adults’ implicit racial bias. For example, explicit instruction has been particularly effective in studies that seek to break the negative “habit” of prejudice (Devine, 1989; Devine, Forscher, Austin, & Cox, 2012; Monteith, 1993). In these studies, implicit racial bias is significantly decreased after participants are made explicitly aware of their implicit racial bias and are presented with concrete strategies to reduce that bias (e.g. thinking of counter-stereotypical exemplars or taking the perspective of outgroup members). Other studies that make adults more explicitly aware of bias (Rudman, Ashmore, & Gary, 2001), or train them to correct bias (Kawakami, Dovidio, & van Kamp, 2007; Stewart, Latu, Kawakami, & Myers, 2009), have also been successful in bias reduction. Furthermore, studies that have decreased adults’ bias using counter-stereotypical exemplar exposure have either employed these evaluative instructions (Lai et al., 2014; 2016) or used well-known exemplars, which may shift associations more easily (Dasgupta & Greenwald, 2001).

Limitations. The current study results suggest that this intervention was effective with both younger and older children. However, participant means suggest that this intervention may have been more effective for older children, partially due to lower levels of bias in younger
children overall. This lower level of implicit pro-White/anti-Black bias in younger children is consistent with other studies with children in this population, where Blacks comprise only 1.2% of the population (Statistics Canada, 2016). As such, future studies should seek to conduct this intervention with a younger sample that has higher levels of implicit racial bias. If studies with other populations can successfully reduce bias in younger children using this type of intervention, it would further suggest that this method can be employed to reduce implicit racial bias in children ages 5-8.

It also remains unclear which component of our exemplar exposure increased its efficacy with younger children. For example, while our goal was to increase the racial salience of exemplars, we also targeted multiple associations (Black=good, White=bad), and employed child exemplars instead of adults. Any or all of these changes may have contributed to the difference in results between this study and that conducted by Gonzalez and colleagues (2017). As such, future work should seek to narrow down these possibilities to identify the specific mechanisms that allow younger children to encode the associations presented in counter-stereotypical exemplar stories.

Another limitation of the current study was the lack of two additional conditions that would shed important light on the malleability of implicit racial bias across development: an evaluative instruction + counter-stereotypical exemplar condition for children, and an evaluative instruction only condition for adults. It remains unclear whether or not adding evaluative instructions to this intervention might increase its efficacy with children and result in longer-lasting bias change comparable to that of adults. Furthermore, it is unclear whether the evaluative instructions enhanced the efficacy of the counter-stereotypical exemplar exposure, or if the evaluative instructions may have changed bias on their own. Studies have shown that the use of evaluative statements can be effective in changing implicit bias, which would suggest that this
addition could have impacted adults’ bias without the use of counter-stereotypical exemplars (Kurdi & Banaji, 2017). If this is the case, this would still support our hypothesis that adults might need more explicit instruction to effectively change bias. Future studies should examine whether additional modifications to our counter-stereotypical exemplar exposure might improve the efficacy of this intervention with adults, as well as disentangling the effects of exemplar exposure versus evaluative instructions.

**Implications.** Though we ultimately found that both children and adults’ implicit racial bias could be shifted, adults appear to require a stronger “dosage” of exemplar exposure than children. These results suggest that children’s implicit racial bias may be more malleable than that of adults, due to relatively less exposure to cultural stereotypes (Baron, 2015). As a result, exposure to counter-stereotypical exemplars in mainstream media, such as television, movies, or books, may be particularly effective for children, who do not require any additional instruction to internalize positive outgroup associations. In contrast, for adults, counter-stereotypical exemplar exposure may be most effective when paired with diversity training or other forms of explicit instruction (see Bezrukova, Spell, Perry, & Jehn, 2016; Devine et al., 2012).

Coupled with previous research, our results suggest that counter-stereotypical exemplar exposure may be particularly effective for older children, who can encode race more easily, but may also have more malleable implicit associations due to increased cognitive flexibility and relatively less exposure to stereotypes (Baron 2015; Gonzalez, Dunlop, & Baron, 2017; Gonzalez, Steele, & Baron, 2017). As such, it may be worthwhile for interventions using counter-stereotypical exemplar exposure to target children ages 9-12, as they appear to require less input than younger children or adults. To understand whether this is a function of our specific manipulation, or the malleability of older children’s bias, future work should examine the efficacy of other interventions with this particular developmental population.
In conclusion, these results suggest that counter-stereotypical exemplar exposure may be easier to implement in children as compared to adults and may be optimal for use with children ages 9-12. Future work should seek to further examine bias change beyond the course of an hour, as well as the potential of repeated brief interventions to more effectively decrease bias. In order to counteract continuous cultural messages of bias and prejudice and solidify counter-stereotypical associations, it may be essential for bias reduction interventions to occur multiple times. A critical next step in this line of research is the implementation of more longitudinal assessments of children’s bias change as well as inclusion of behavioral measures. In conjunction with the current research, these future studies will shed light on the use of counter-stereotypical exemplar exposure to change children’s bias impactfully.
Figure 4.1 IAT D-Scores for Children and Adults at Time 1 (Study 8). Higher scores represent more pro-White/anti-Black racial bias.
Figure 4.2 IAT D-Scores for Children and Adults at Time 2 (Study 8). Higher scores represent more pro-White/anti-Black racial bias.
Figure 4.3 IAT D-Scores for Adults (Study 9). Higher scores represent more pro-White/anti-Black racial bias.

All participants read positive Black exemplar stories from Study 1.
Chapter 5: Conclusion

5.1 Summary of Key Findings

In this dissertation, I presented three sets of studies examining outstanding issues in the development of intergroup bias. Past research has shown that explicit and implicit intergroup bias is present by age six, and in the case of implicit bias, these biases remain relatively stable across development (e.g. Baron & Banaji, 2006; Dunham et al., 2013). Furthermore, relative difficulty in changing adults’ intergroup bias suggests that it may be worthwhile to try and change bias in childhood, before attitudes and stereotypes have been reinforced across the lifespan (e.g. Forscher et al., 2016; Lai et al., 2016). While this past work has laid foundations for the development of interventions to reduce intergroup bias in childhood, there are still a number of missing pieces that are essential for effective and targeted interventions.

Due to limitations in current methods for measuring implicit bias, the developmental trajectory of distinct implicit associations is still unclear. Measures like the Implicit Association Test (IAT) test multiple associations simultaneously, and as such, it remains unclear what component of intergroup bias drives these associations. Chapter 2 modified and tested an alternative method of measuring implicit gender stereotypes called the Preschool Auditory Stroop (PAS). This method operates on a principle of cognitive interference; children hear words spoken in a male or female voice and must categorize the voice by gender. If the words are stereotypically associated with one gender over the other, children should be faster to categorize when the word content matches the voice gender (e.g. girl voice saying “pink”).

Importantly, this method can be used to disentangle distinct gender stereotypes, as only one association is being tested at a time (e.g. girl=pink), as compared to the IAT, which tests two associations at once (e.g. girl=pink vs boy=blue). Children’s gender stereotypes were quantified using a modification of traditional IAT D-scoring, which calculates a difference score between
stereotypically congruent and incongruent trials (e.g. reaction time differences between girl=pink vs. boy=pink). Results of Study 1 and Study 2 indicated that 4-year-old and 7-year-old children were faster to categorize the gender of a spoken voice when the word content was stereotypically congruent. Specifically, children had significant distinct implicit stereotypes associating stereotypically feminine words with girls and stereotypically masculine words with boys.

Study 3 extended this work to math-gender stereotypes, which have been measured across development using the IAT. As such, there is a lack of clarity regarding the true nature of these stereotypes, and whether they are primarily driven by a boy=math or a girl=reading association. Interestingly, we found key differences between stereotypes measured by the IAT and the PAS, potentially due to the exemplar-based nature of the PAS, which does not require children to attend to the overarching category (i.e. math or reading) on each trial. While children ages 6-8 showed significant stereotypes associating boy=math and girl=reading on the IAT, they did not show either of these implicit associations on the PAS. These results suggest that children’s gender stereotypes about math and reading may not be as easily activated at this point in development. Research suggests that children acquire these stereotypes around ages six to ten (Cvencek et al., 2011; 2014; Passolunghi et al., 2014; Steffens et al., 2010). Thus, at this point in development, these particular gender stereotypes may not be internalized as strongly, and therefore may require additional contextual activation. Specifically, while children most likely have these stereotypes, these associations may not be salient enough to be activated by exemplar-based measures like the PAS.

Another issue limiting the implementation of optimal interventions to reduce bias is a dearth of research on the relationship between intergroup bias and behavior in childhood. While a number of studies with adults suggest a clear link between explicit and implicit bias and behavior (e.g. Dovidio et al., 2012), few studies have examined whether the magnitude of
intergroup bias is predictive of children’s biased behavior. Chapter 3 presented a combined dataset of four studies examining the relationship between young girls’ explicit stereotypes about math and gender, and their performance on a math-related measure of Approximate Number System (ANS) accuracy. The ANS is our intuitive and universal sense of number that is present from birth, and this number sense underlies our formal mathematics abilities acquired in childhood. As such, this ability actually precedes formal math education, and impairment on this measure would suggest that young girls’ math abilities are hindered even before they acquire basic mathematics skills in school.

In Study 4-7, children ages 3-6 were presented with an ANS task, and the instructions given before the task varied by condition. Half of participants received instructions designed as a control condition and were either told that the task they were about to complete was a game, or a test of eyesight ability. The other half of participants were told that the study was a test of math ability and that researchers were interested in whether boys or girls were better at math and counting. This math test condition was designed to activate children’s math-gender stereotypes. Across the combined dataset presented in Study 4-7, there was a significant three-way interaction between condition (i.e. control vs. math test instructions), children’s explicit beliefs about math and gender, and child gender. Decomposing this interaction indicated that for girls who had stereotypes associating math more with boys than girls, there was a significant effect of condition, such that girls whose math-gender stereotypes were activated performed significantly worse than girls who thought the task was a game or a test of eyesight. Analyzed differently, for girls in the math test condition, stereotypes about math and gender predicted their ANS performance, such that girls who associated math more with boys than girls performed worse than girls who associated math with girls.
These results suggest that gender stereotypes can influence girls’ math-related performance as early as age three. Thus, it appears that intergroup bias has the potential to constrain children’s behavior at a very early age and may begin to affect behavior as soon as stereotypes are acquired. This effect of intergroup bias on behavior at such an early age adds further credence to the argument for changing children’s intergroup bias before the effects of this bias can compound over the lifespan. For example, in the case of girls’ math-gender stereotypes, early effects of bias on girls’ ANS capacities might hinder their acquisition of formal mathematics concepts, setting them back in their math achievements relatively early in development. As such, these girls may be less likely to pursue math-related careers in later development, contributing to women’s overall underrepresentation in STEM. In other cases, where intergroup bias might affect children’s actions toward members of low-status or stereotyped groups, repeated instances of prejudiced behavior might normalize this behavior for perpetrators and impair the achievements and mental health of members of marginalized groups.

The final outstanding issue examined in this dissertation is the malleability of intergroup bias across development. Past work with adults and children suggests that implicit intergroup bias may be more malleable in children as compared to adults (e.g. Gonzalez et al., 2017; Lai et al., 2016; Neto et al., 2015; Qian et al., 2017; Vezzali et al., 2011). Specifically, while implicit racial bias change does not appear to last beyond 24 hours in adults, bias change in children has lasted up to months and years later. However, thus far, there have been no direct comparisons testing an intervention to change intergroup bias in both children and adults, and as such, it is an empirical question whether these two populations require comparable conditions for bias change to occur. Chapter 4 conducted a direct comparison between children and adults in the efficacy of counter-stereotypical exemplar exposure to reduce implicit racial bias. This method has been used in previous work with children and adults to successfully reduce bias in immediate post-
intervention testing (e.g. Columb & Plant, 2011; Dasgupta & Greenwald, 2001; Gonzalez, Steele, & Baron, 2017; Lai et al., 2014; 2017).

In Study 8, to reduce children and adults’ implicit pro-White/anti-Black racial bias, participants were presented with positive and negative child exemplars. In the experimental condition, participants were presented with a pair of Black children who engaged in prosocial behavior, and a pair of White children who engaged in antisocial behavior. In the control condition, one set of White children engaged in prosocial behavior, and the other pair engaged in antisocial behavior. Implicit racial bias was then measured both immediately after children read the stories and after an hour-long delay. Results indicated that this manipulation was successful for children, but not for adults. Children who were exposed to positive Black exemplars had lower levels of implicit pro-White/anti-Black racial bias both immediately and after an hour-long delay, as compared to children who were exposed to positive White exemplars. There was no difference in intervention efficacy between younger (ages 5-8) and older (ages 9-12) children, suggesting that this intervention was effective across the age range of 5-12. In contrast, adult levels of bias were comparable regardless of which exemplars they were exposed to, indicating that this intervention was not successful with that population.

Study 9 further examined the conditions required to induce bias change in adults. As the majority of past studies with adults have used either well-known exemplars or included some kind of explicit instructions to successfully reduce implicit racial bias, we hypothesized that adults might require additional instructions to change their bias. Adults were once again exposed to positive Black exemplars and negative White exemplars, with the control condition remaining identical to the counter-stereotypical exemplar condition in Study 8. In the experimental condition, they were given additional instructions to internalize the presented associations of “Black=good, White=bad”. We found that when adults were given evaluative instructions after
the story, their levels of bias were significantly decreased as compared to the condition where they read the story without additional instructions.

Taken together, the results of Study 8 and 9 suggest that adults may need additional explicit guidance to change their existing associations as compared to children. As such, bias may be similarly malleable in both adults and children, but similar to unscrewing a tightly sealed jar, bias change may require more “force” in adults. The relative ease in changing children’s implicit intergroup bias suggests that exposure to counter-stereotypical exemplars may be a particularly effective method of bias reduction for this population. In contrast, it seems likely that alternative interventions that involve more explicit direction for changing bias, such as habit-breaking interventions that train adults to confront and address their biases (Devine et al., 2012), may be more effective methods of bias change later in development.

5.2 Implications and Future Directions

The research presented in this dissertation substantially adds to our understanding of the development of intergroup bias. Chapter 2 and 3 of this dissertation present evidence that stereotypes are acquired relatively early in development, and while there is individual variability in stereotype acquisition and activation that may mask their presence, these stereotypes can significantly affect the behavior of children who do have them. The findings presented in these two chapters further confirm the importance of understanding the contexts that activate children’s implicit association, as these appear to be critical to understanding when and how bias relates to behavior in childhood. Chapter 4 presents important novel evidence regarding the malleability of bias and developmental differences in the conditions required to elicit bias change. Though bias can be changed across development, methods that involve less explicit instruction to confront one’s own bias may be more effective with children than adults. As such,
this research suggests that diversity exposure in children’s everyday lives and in mainstream media may be particularly effective methods to reduce bias in childhood.

As a whole, the work presented in this dissertation further suggests that intergroup bias change in childhood is a worthwhile endeavor. Furthermore, the results presented here are some of the first to investigate these issues in childhood, and as such, lay the groundwork for further investigation of intergroup bias and bias change across development. Below, I detail the theoretical and practical implications of this work and outline future research directions.

5.2.1 Expression of Bias Across Development

*Individual Variability.* The results of Chapter 2 and 3 suggest that there is important individual variability in children’s implicit and explicit gender stereotypes. In Chapter 2, there was evidence of population-level implicit gender stereotypes associating particular toys and attributes with different genders as early as age three. In contrast, only the IAT, which may detect stereotypes more easily than the PAS due to its category-based nature, detected math-gender stereotypes in children ages 6-8. As such, this stereotype was clearly present in some children who did not show bias on the PAS. This result is consistent with the lack of overall gender stereotypes found in the studies from Chapter 3; while not all girls expressed stereotypes associating math more with boys than girls, the girls who did express these stereotypes were susceptible to behavioral effects on their math-related performance. Thus, there may be more individual variability in children’s math-gender stereotypes than previously acknowledged in the literature.

Critically, these results suggest that population-level bias means may not be the appropriate level of analysis for examining the presence and consequences of intergroup bias across development. Even when not all children show evidence of intergroup bias, these
associations may substantially impact the behavior and cognition of the handful of children who do (see Greenwald, Banaji, & Nosek, 2015). Additionally, it is difficult to draw conclusions about the true state of bias across development when drawing from unique geographical locations that may have more or less bias depending on the political and social climate (e.g. Leitner et al., 2018; Payne et al., 2017). In the future, researchers may wish to look at the development of intergroup bias on more of an individual differences level, in order to identify children who have significant levels of racial or gender bias and intervene with those particular children to prevent the perpetuation of biased behavior.

Identifying children who have intergroup bias is also critical to understanding the cultural sources of bias that lead some children to endorse bias at a young age while others do not. For example, future research could examine parents as a source of bias, as previous research has shown that they may they pass stereotypes on to their children at a relatively young age (e.g. Croft et al., 2014; Sinclair, Dunn, & Lowery, 2005). It is possible that parents with higher levels of explicit bias, who talk about issues of gender or race with their children in biased ways, may have children who show intergroup bias earlier than children whose parents are more egalitarian. This may be particularly true in the case of gender, as it is generally more acceptable to discuss gender differences than racial differences (Rogers & Meltzoff, 2017). In the case of racial bias, parents who adopt a colorblind approach, and refuse to confront the negative racial attitudes children are exposed to, may similarly have more biased children at a young age.

Future studies should consider taking a longitudinal approach to this question and looking at individual trajectories of bias from early childhood through adulthood. Researchers could even consider following children from infancy and examining the predictors of when early preferences and social categorization progress into intergroup bias (Bar-Haim et al., 2006; Mahajan & Wynn, 2012; Pun et al., 2017). Identifying the sources of bias that map on to high levels of intergroup
bias endorsement would be useful in the design of interventions to directly confront those biases. For example, if children’s intergroup bias is strongly linked to parents’ levels of bias, interventions could focus on reducing parent bias in conjunction with their children’s. If bias is linked to certain types of media exposure, such as television shows or books that perpetuate negative attitudes and stereotypes, perhaps counter-stereotypical exemplar exposure might be a more effective method of combating these sources.

*Contextual Activation.* In addition to highlighting the importance of individual variability in intergroup bias, the results of Chapter 2 and 3 emphasize the importance of understanding its contextual activation. In Chapter 2, children did not show implicit math-gender stereotypes on an exemplar-based measure of bias but did seem to have significant bias on a category-based measure, suggesting that this bias was in fact present, but only activated when children were forced to think about a categorical association. Furthermore, in Chapter 3, the behavior of girls with stereotypes associating math more with boys was only affected under stereotype threat conditions. Thus, the work in these two chapters suggests that more attention must be drawn to the contextual activation of bias, which might affect identification of a) the presence of bias across development and b) the effects of bias on behavior in childhood.

Previous work with children and adults suggests that context can have important effects on the expression of bias. Specifically, the magnitude of racial attitudes in adults has been shown to vary based on whether or not participants must categorize by race (Olson & Fazio, 2003). As discussed in Chapter 2, studies examining children’s implicit racial bias also find stronger effects when children are forced to categorize outgroup members (Williams & Steele, 2017). Thus, there may be variability in the magnitude of intergroup bias across development based on the methods used by researchers in each study. As such, it is possible that category-based measure of intergroup bias may overestimate the magnitude of bias and/or exemplar-based measures may
underestimate it. By looking at the relationship between these different measures of bias and behavior, researchers may be able to better determine which measures are most useful for bias quantification.

In addition to the effects of context on expression of bias, work with adults and children suggests that contextual activation can determine when bias affects behavior. As discussed in Chapter 3, work with adults and children has shown that gender bias only detrimentally affects women and girls when they are reminded of negative stereotypes (e.g. Ambady et al., 2001; Schmader, 2010). These reminders have varied from more explicit statements, to subtle gender cues. As very few studies have examined the conditions under which gender bias affects children’s behavior, and none have looked at conditions under which racial bias affects behavior, there is room for further exploration of these contexts, which would help researchers to develop targeted interventions to reduce biased behavior.

The aforementioned findings and the work presented in this dissertation are in accordance with the Justification-Suppression Model of bias in adults, which states that certain contexts justify expression of bias and lead to more biased behavior, while other contexts lead to bias suppression (Crandall & Eshleman, 2003). In adults, factors such as the egalitarian norms and values of a society are posited to lead to bias suppression, while factors like situational ambiguity and motivation to uphold social hierarchies might allow adults to justify their expression of bias. These mechanisms may be too cognitively sophisticated to apply to children, but nonetheless, there are undoubtedly social contexts that are more or less likely to lead to bias expression in children. For example, children may not automatically encode an individual’s race during an interaction (e.g. Pauker et al., 2016; Shutts et al., 2011), but when this type of categorization is primed, they may be more likely to engage in biased behavior (Williams & Steele, 2017). The
impact of bias suppression and justification on children remains an understudied avenue of research, and future studies should seek to further explore this issue.

*Interactions between Individual Variability and Context.* Additionally, it is important to note that individual variability and contextual activation appear to work in tandem to influence bias expression and behavior (see Picho & Schmader, 2017; Chapter 3 of this dissertation). Focusing on only one of these components without the other could result in masking of important bias effects. Specifically, none of the studies thus far that have failed to find stereotype threat effects in young girls have examined individual variability in children’s awareness of stereotypes (see Ganley et al., 2013). As such, young girls who do not have these stereotypes would most likely fail to show biased behavior under conditions of threat. Alternatively, studies that have failed to find a relationship between bias and behavior may need to further examine the moderating role of context in this relationship (e.g. Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013). Generally, it seems advisable for researchers to look at both individual variability and contextual activation of bias as distinct, but potentially interacting predictors. This is particularly important in the case of studying the development of intergroup bias; failing to examine a potential interaction between these two variables might result in the underestimation of bias effects in early development, when less children may be aware of stereotypes and these stereotypes may not be as easily activated as those of adults.

5.2.2 *Bias Change Across Development*

*Associative vs. Propositional Change.* The Associative-Propositional Evaluation (APE) model is a theoretical framework for understanding the distinction between implicit and explicit representation (Gawronski & Bodenhausen, 2006). The model distinguishes between associative and propositional processes. Associative processes are defined as automatic reactions that occur
when learned associations are triggered by a stimulus. For example, if an individual has acquired an association between spiders and negativity, encountering a spider will induce a negative affective reaction. These learned associations do not need to be personally endorsed by an individual to become activated, as contextual activation of an association occurs without cognitive reflection or control. Regardless of logical reasoning that a spider is unlikely to harm you, the stimulus encounter will still trigger a “spider=bad” association. Stimuli can have more than one association, and different contexts can activate alternative associations for the same stimulus. These associative processes are proposed to underlie our implicit attitudes, which are automatically activated, even against an individual’s personal beliefs or desires.

In contrast, propositional processes are based off of inferences and require validation of beliefs. Unlike associative processes, individuals must reflect and reason that a proposition is true. These propositional processes evaluate the validity of affective reactions, and after reflection, these associations become propositions. For example, experiencing a negative affective reaction after encountering a spider might lead to the propositional judgment that “I dislike spiders”. Propositional reasoning is thought to underlie explicit attitudes, as these attitudes are only expressed after cognitive reflection. When an association is activated (e.g. spider=bad), but propositional reasoning leads the individual to reject the validity of that association (“Spiders are good”), explicit and implicit cognitions diverge.

The APE model is particularly useful when considering implicit and explicit bias change. According to the model, implicit attitude change stems from either a) changing the underlying association or b) shifting the activation to an alternative association. Underlying change in associative structure is often done through classical conditioning processes, when individuals are exposed to an alternate stimulus pairing repetitively. In contrast, when associations are shifted, there is no need to expose individuals to an alternate stimulus pairing. Instead, particular contexts
can activate different associations that already exist. Explicit attitude change can stem from either implicit attitude change or a change in propositional beliefs, which helps to explain why implicit and explicit attitude change is often asymmetrical.

The mechanisms behind the implicit attitude change observed in Chapter 4 remain unclear; based on the assumptions of the APE model, underlying attitude change is proposed to be slow and gradual, while shifting associations to an alternative association may occur more rapidly. It seems plausible that the mechanism driving the observed bias change is a shift from a “White=Good, Black=Bad” representation to a “White=Bad, Black=Good” representation rather than a direct change to the structure of a “White=Good, Black=Bad” association (see Lai et al., 2016). However, it is interesting to note that children’s preferences were not reversed. After counter-stereotypical exemplar exposure, children did not show a significant implicit preference for either racial group. This could be indicative of change to the underlying association, or it is possible that the shift is to a “White=Good, Black=Good” or “White=Bad, Black=Bad” association that builds upon the prior association. Future work using methods like the AMP, or other exemplar-based measures, may be able to shed light on which components of the association are shifting.

It is also important to consider the nature of these associations across development, and how they might impact bias change. Theorists have proposed that rather than being driven by associations (e.g. “Spiders = bad”), implicit bias may instead be coded as “structured beliefs”, and may have a propositional structure (e.g. “Spiders are bad”; Mandelbaum, 2016). While this could be the case for young children, this may not be how biases originate, as in infancy, biases are arguably implicit, and most likely associative in structure, as they are probably basic associations between co-occurring attributes (see Lee, Quinn, & Pascalis, 2017). Future work should seek to examine the structure of implicit bias across development, and whether implicit
biases might begin as associative processes and progress to propositional structure later in development.

*Strategies for reducing biased behavior.* The research presented in Chapter 4 focuses on changing the underlying implicit association as a strategy for ameliorating the harmful effects of bias on behavior. However, our manipulation does not directly examine links between bias and behavior; future studies should examine whether this reduction in bias corresponds with a reduction in biased behavior. Furthermore, future work should longitudinally examine the impact of repeated interventions to reduce bias, and whether this repetition might be more effective in changing both bias and biased behavior. When children return to their everyday lives after counter-stereotypical exemplar exposure, they will most likely encounter cultural messages reinforcing their previous biases. As such, it may be essential to provide children with repeated examples of exemplar exposure in order to counteract the messages that maintain bias.

Future research should also examine the efficacy of an alternative strategy to reduce biased behavior: interrupting the effect of intergroup bias on behavior, rather than trying to change the bias itself. This strategy has been successfully employed with adults to prevent intergroup bias from affecting interactions with others (e.g. Devine et al., 2012), as well as to prevent bias from constraining the behavior of members of marginalized groups (e.g. Johns, Inzlicht, & Schmader, 2008). For example, recent work on habit-breaking interventions to reduce implicit racial bias suggests that while this type of intervention might not always change underlying implicit associations, it does successfully reduce participation in biased behavior (Forscher et al., 2017). In the domain of math-gender stereotypes, teaching individuals to reappraise their anxiety has led to successful reduction of stereotype threat effects that lead women to underperform on math assessments (Johns et al., 2008; Jamieson, Mendes, & Nock,
Researchers could make these types of interventions more child-friendly to see whether these strategies successfully mitigate biased behavior in children.

5.2.3  Types of Intergroup Bias

The research presented in this dissertation focuses on two types of bias: racial attitudes and gender stereotypes. Though race and gender are both important social categories within North American society, race is often considered a taboo topic of discussion, while gender is not. Colorblindness is often adopted by individuals who feel uncomfortable discussing race, even though this strategy is not effective in combating racism (Plaut, Thomas, Hurd, & Romano, 2018). In contrast, gender is discussed quite openly, and when gender identity and biological sex correspond, gender differences are often viewed as a product of biology. In general, gender is commonly essentialized as a category, and viewed as having an underlying “essence” that affects individual behavior (Meyer & Gelman, 2016; Rhodes & Gelman, 2009). While past research suggests that race is also essentialized (e.g. Kinzler & Dautel, 2012), gender is arguably essentialized more openly. Thus, gender may be more susceptible to the production of generic language such as “Boys are X” or “Girls do Y”. These types of statements have been linked to further essentialism of the category (e.g. Rhodes, Leslie, & Tworek, 2012; Waxman, 2010; Wodak, Leslie, & Rhodes, 2015). As such, it is possible the gender bias might actually be more difficult to change in children than race bias, because gender stereotypes are more openly encouraged. While both types of bias can be reduced in children immediately after intervention exposure (Block et al. 2018; Gonzalez et al., 2017), gender bias may be harder to change long-
term, as cultural biases may be more prominent within the environment and serve to reinforce previous attitudes and stereotypes.

It is also critical to acknowledge that both race and gender lie on a spectrum (see Dunham & Olson, 2016). Though we use discrete categories to operationalize race and gender in these studies, these categories do not fully capture the complexity of social identity. Many individuals identify as multiracial, non-binary, or transgender, and these more fluid forms of social group identity may significantly impact the malleability of individuals’ intergroup bias. Recent work has shown that transgender children and their siblings are less likely to endorse gender stereotypes (Olson & Enright, 2018). I would speculate that for these children, gender bias is more malleable because gender categories are viewed as less rigid. Similar mechanisms might be at play for children who identify as multiracial, and future research should empirically test these possibilities.

Another important consideration and direction for future research is the role of intersectionality in children’s intergroup bias. In adults, there is evidence that race and gender intersect, and Black men are viewed more negatively than Black women, White men or White women. However, very few studies have looked at children’s bias from the perspective of multiple social categories. To date, one study has found that children have implicit and explicit racial/gender bias that is comparable to that of adults; preschool children associated Black boys more strongly with negativity than Black girls, White girls, or White boys (Perszyk, Lei, Bodenhausen, Richeson, & Waxman, 2019). Understanding the development of intergroup bias from an intersectional perspective is critical to the development of effective interventions, and future work should further investigate the presence and malleability of biases based on multiple social categories.
5.3 Concluding Remarks

In conclusion, the work presented in this dissertation provides us with a better understanding of the nuanced nature of intergroup bias across development and makes a case for the development of interventions to reduce intergroup bias in childhood. There are many more outstanding questions to be answered in this field of research, as well as exciting opportunities to implement practical interventions, like counter-stereotypical exemplar exposure, to change children’s intergroup bias. As our understanding of the development of intergroup bias grows, we will be better able to tailor our interventions for effective change. In the meantime, we can build off the current work to start changing children’s bias through exemplar exposure; the results of this dissertation emphasize the importance of increased diversity and positive, counter-stereotypical depictions of social groups in children’s everyday lives and media consumption. By taking this tangible step, we can start the process of cultivating positive attitudes and stereotypes early in development, before negative biases take their toll on behavior.
References


Appendices

Appendix A  Exemplar Story Text (Study 8 & 9)

Today I am going to tell you three short stories. There are four people who are in each of these stories, so I will tell you a little bit about them first.

Two of these people are named Gina and Gary. Gina and Gary are best friends who do everything together. They are very similar to each other. Gina and Gary are not very kind or friendly, and sometimes do mean things to other people.

The other two people are very different from Gina and Gary. These two people are called Rose and Rudy. Rose and Rudy are best friends who do everything together. They are very similar to each other. Rose and Rudy are both very kind and friendly, and often do nice things for people. Here is the first story about these four people.

Gina and Gary were walking on the street and saw a young girl who had fallen down and hurt her leg. When the young girl saw them she asked for help. “Excuse me, could you please bring me a band-aid for my leg?”

“No way!” said Gina and Gary. Instead of helping the girl, Gina and Gary teased her for falling down and walked away.

A few minutes later, Rose and Rudy walked by the same young girl who had fallen and hurt her leg. When the young girl saw them she asked for help. “Excuse me, could you please bring me a band-aid for my leg?”

“Of course!” said Rose and Rudy. While Rose helped the girl sit up on a bench, Rudy went to the store and got the girl a band-aid.

Can you point to the people who helped the girl in trouble?

Can you point to the people who did not help the girl in trouble?

Here is the second story:

Later that day, Gina and Gary walked by a playground where some teenagers were playing soccer. One of the teenagers kicked the soccer ball outside of the grass, and it rolled onto the street near Gina and Gary. “Hey, could you toss that ball over here?” said one of the teenagers. “No way!” said Gina and Gary. Instead of giving the ball back, Gina grabbed the ball and tossed it to Gary. They laughed and ran away, taking the ball with them.

A while later, Rose and Rudy walked by the same playground where the same teenagers were playing soccer with a new ball. Once again, one of the teenagers kicked the soccer ball outside of
the grass, and it rolled onto the street near Rose and Rudy. “Hey, could you toss that ball over here?” said one of the teenagers.

“Of course!” said Rose and Rudy. Rose grabbed the ball and tossed it to Rudy. He threw it back toward the teenagers, and they were able to play soccer again.

Can you point to the people who gave the teenagers back their soccer ball?

Can you point to the people who took the teenagers’ soccer ball?

Okay, here is the last story:

At the end of the day, Gina and Gary were walking home, when they saw their neighbor Mr. Smith taking groceries out of his car. He was holding a big bag of groceries, and walked away from the car, leaving a bunch of other groceries that he couldn’t carry by himself.

“Hey, let’s take some of those groceries for ourselves,” said Gina. “Good idea, I’m hungry, and I don’t feel like buying my own food,” said Gary. They each grabbed a bag of groceries and ran away. Once they were far away, they took the food for themselves and went home.

A few minutes later, Rose and Rudy walked by Mr. Smith taking groceries out of his car. He was carrying another big bag of groceries, and walked away from the car, leaving a bunch of other groceries that he couldn’t carry by himself.

“Hi Mr. Smith! How are you doing today?” asked Rose. “Can we help you carry some of those groceries?” said Rudy. Mr. Smith nodded, and Rose and Rudy each grabbed a bag of groceries to carry. After helping Mr. Smith, they said goodbye and went home.

Can you point to the people who helped Mr. Smith with his groceries?

Can you point to the people who took some of Mr. Smith’s groceries?