



# Gender and Race in Children's Picture Books: A Tragedy in Three Studies

## Citation

McIntyre, Joseph C. 2017. Gender and Race in Children's Picture Books: A Tragedy in Three Studies. Doctoral dissertation, Harvard Graduate School of Education.

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:33052849>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

**Gender and race in children's picture books: A tragedy in three studies**

**Joseph Clark McIntyre**

**Stephanie Jones**

**James Kim**

**Luke Miratrix**

**A Thesis Presented to the Faculty of the Graduate School of Education of Harvard  
University in Partial Fulfillment of the Requirements for the Degree of Doctor of  
Education**

**2017**

© 2017  
**Joe McIntyre**  
**All Rights Reserved**

## Contents

Abstract .....	ii
Chapter 1 – Introduction .....	1
Chapter 2 – Gender in children’s picture books: The impact of parental decisions .....	16
Chapter 3 – Gender in children’s library checkouts: The impact of children’s decisions .....	57
Chapter 4 – Gender in assigned reading: The impact of teachers’ decisions .....	94
Bibliography .....	135

## **Abstract**

Studies of children's picture books have demonstrated that male central characters significantly outnumber female central characters in a wide range of different books. However, these studies have all been limited in the inferences they could make about books that children actually read due to their approaches to sampling. In this dissertation I adopt a survey sampling perspective to estimate the ratio of male to female central characters in children's picture books actually encountered by children. I then extend my results by use library checkout data to estimate the ration separately for girls and boys and to estimate developmental tendencies for the ratio. Finally, I use data from a randomized controlled trial to demonstrate that teacher book selections fail to improve the ratio of male to female central characters encountered by children.

# Chapter 1 – Introduction

Representation matters<sup>1</sup>. It is important for people to see characters who are like themselves in the books that they read, the movies that they watch, and the games that they play. It is also important, though less frequently noted, for people to see those who are *not* like themselves in the cultural artifacts with which they interact. This latter point is especially true of members of socially dominant groups, in this case people who are White, male, and especially people who are both White and male. In this dissertation, I take it for granted that it is important for people, particularly children, to encounter cultural artifacts which represent characters of various races and genders. It would be a bad thing, for example, for children not to encounter books with female characters, or to encounter drastically more books with male characters than female characters. Various literature that I cite in this dissertation, discussed in the individual studies in subsequent chapters, elaborates on the reasons that representation is important. However, I consider the case straightforward and intuitive enough that I do not need to make it here. This dissertation rests on with the stipulation that children should encounter books about male and female characters in roughly equal proportions, and that they should encounter characters of different races at rates roughly equal to their share of the population.

## *The state of representation*

---

<sup>1</sup> Throughout this dissertation, I consider representation based on gender and race. This should not be construed as dismissing other dimensions of identity, such as socio-economic status, religion, nationality, sexual orientation, interests, personality, region of the country, etc... However, in the cultural artifacts I am considering, namely children's books, the race and gender of the characters are far easier to measure than other aspects. Additionally, the literature with which this dissertation engages is primarily concerned with representation of gender, making that dimension particularly important to me. Finally, without creating a hierarchy of identity, race and gender are clearly hugely important in how we see ourselves and how others see us.

Great strides have been made in recent years in increasing the representation of members of socially non-dominant groups<sup>2</sup>. Black women have leading roles in several new popular television shows such as *Scandal* and *How to Get Away with Murder*. Children's cartoons routinely include female and non-white characters, even if White, male characters are still the central characters in most programs. People's awareness of the importance of representation have been raised by campaigns criticizing the absence of actors of color as nominees for acting awards (Keegan & Andersen, 2016). Activists have forcefully criticized the practice of white-washing; i.e., casting White actors to play characters which were originally intended as non-White (e.g., Seltzer, 2011).

Those who have historically been most overrepresented are sometimes threatened by the rebalancing which promotes the inclusion of other groups in lieu of White men and boys. As feminists have campaigned for more movies featuring women in a larger variety of roles, all-female casts have sometimes proved controversial, as in the recent *Ghostbusters* remake. The movie was viciously criticized by self-described men's rights activists for including only female protagonists (Hornaday, 2016), and the sole Black lead was targeted for racist and misogynistic harassment (Rogers, 2016). Videogame players have reacted aggressively to suggestions that the games they play underrepresent women, and portray the women that they do include in harmful, stereotypical ways. After the cultural critic Anita Sarkeesian identified sexist tropes in videogames, she was forced to cancel speaking events when her opponents threatened to carry out attacks at those talks (Collins, 2014). Some science fiction readers, concerned that the prestigious Hugo

---

<sup>2</sup> Here and elsewhere I identify White people and men as members of socially dominant groups, with White men being especially dominant. "Socially dominant" indicates that members of this group tend to occupy positions of power and authority.

science fiction book awards have been too concerned with issues of representation, have banded together to vote in works by White, male authors telling the stories of White, male characters (Schaub, 2016). Fox news anchor Megyn Kelly declared that “Santa Claus is White,” an ahistorical assertion about an imaginary character (the character of Santa Claus is thought to be based on a Turkish bishop, St. Nicholas), because she was concerned about activists advocating representations of Santa as Black (Feldman, 2013). In 2016, when a Black actor portrayed Santa at the Mall of America, there was substantial online backlash and a local newspaper was forced to close the comments section on its article (Vezner, 2016).

Popular culture has paid somewhat less attention to issues of representation in literature, excepting the battles around the Hugo awards (Schaub, 2016). There are a few reasons for this. First, the audience reached by any book is much smaller than the audience reached by a television show or movie. This makes calling attention to underrepresentation of certain groups in literature less effective than doing so in movies or television. At the same time, the huge volume of books which are published every year makes it difficult to determine the state of representation in literature; it is easy to take a sample of all television shows on major networks, or of all movies with a wide release made in a given year. It is much harder to take a sample of all books produced, because no easily accessible sampling frame exists. It is also straightforward to quantify the popularity of television shows and movies, by measuring their Nielsen ratings and ticket sales, respectively; in contrast, it is exceedingly difficult to measure how popular books are, and therefore how many readers they reach, let alone how frequently each book is read. Finally, racial representation can be difficult to study in literature since the races of

the characters are not always clear. Character races are frequently not made explicit, making it conceptually challenging to identify the races of the characters. Tellingly, this has been an issue with some movie adaptations of popular books. When *The Hunger Games* portrayed a popular character, Rue, as Black, there was an immediate outcry from fans who preferred to think of her as White (Stewart, 2012). Interestingly, the book explicitly described Rue as Black, but this failed to prevent some fans from reading her as White. At the same time, the movie cast Jennifer Lawrence, a White actress, as the lead character, even though the source material identified that character as non-White (Seltzer, 2011).

In this dissertation, I consider issues of representation, particularly gender-based, but also race-based, representation, in children's literature. In the introduction, I begin by defining representation and identifying the kind of representation that I intend to consider. I then explain why I am focusing on children's books. Finally, I briefly motivate and describe my three studies.

### *Representation*

Representation can be conceptualized in (at least) two different ways. The first, which I label *quality of representation*, is the way in which members of various groups are depicted in books. Quality of representation has changed dramatically over time. Characterizations which were once commonplace, such as Black characters being depicted as unintelligent and submissive, or female characters as helpless and needy, have largely been rejected as racist and sexist. These representations may continue to live on in more subtle forms, but society has rejected their most overt forms.

A great deal of work has examined the quality of representation of non-White and female characters in children's literature (see, e.g., Turner-Bowker, 1996; Tepper & Cassidy, 1999; Anderson & Hamilton, 2005). Quality of representation is clearly an important issue. Encountering racist and/or sexist stereotypes in the books they read is likely to have a harmful impact on children in terms of the ways in which they themselves think about people of different races or the opposite sex, or in how they think about themselves. As a concrete example, girls and boys who rarely or never read books with warm, competent father figures may build stereotypes of men as incapable of fulfilling crucial aspects of the role of parent (Anderson & Hamilton). This could be harmful for boys in that it could steer them away from an important and fulfilling role. It could also be harmful for girls in that it could lead them to shoulder an increased proportion of the work of care-giving as both children and adults. However, the ways in which characters of various races and genders are depicted is not the focus of the current set of studies.

The second form of representation, which I label *quantity of representation*, is the prevalence of representation of different identities in books. Here I focus on the inclusion of members of different groups as *central characters*, i.e., the characters that the book is about.

The importance of quantity of representation is less straightforward than that of quality of representation. It is obviously a problem if children's literature portrays characters in ways which are racist or sexist. When a recent book focused on slaves at George Washington's plantation cheerfully baking a cake for their master's birthday, people were quick to identify the harm done by portraying Black people as happily

enslaved (Peralta, 2016). The harms associated with making most characters in books White and male is less clear. However, children learn about the society around them and what it values in part from drawing inferences from the books that they read. When children are consistently exposed to White, male characters in their reading, and not to other sorts of characters, they may assume that these characters are somehow more interesting or more valuable than non-White, female characters. This is especially true for children who are themselves White males, since they are likely to have had fewer personal experiences to show them the importance of non-White people and girls and women.

*Individual books v.s. books as a whole*

Before continuing, it is important to explicitly consider the issue of representation in specific works of literature versus representation in works of literature as a whole. Representations in a specific work of literature can be harmful if the representation is pernicious. However, this is distinct from the issue of whether works of literature as a whole contain *sufficient* representation of various groups defined by race and gender. Quantity of representation is best conceptualized as a feature of a set of books, not of any individual book. It is not harmful that some works of literature have White, male protagonists, or even that some works of literature have only White, male characters<sup>3</sup>. Instead, it is harmful that a large majority of works of literature have White, male protagonists. The harm is not a result of any one book, but rather of a literary ecosystem which is consistently more likely to feature White boys and men than any other group<sup>4</sup>.

---

<sup>3</sup> Although the failure of a book with a large cast of characters to include any non-White or female characters could be problematic. However, this is an issue for another study.

<sup>4</sup> A more nuanced approach might be concerned not with books as a whole, but rather with the distribution of the gender and race of central characters in the books that individual children encounter. For example, it

### *Children's books*

This dissertation focuses on quantity of representation in children's books. In part this is because it engages in conversation with previous studies of gender representation in children's picture books. However, there are substantive reasons to be especially interested in representation in this setting. Childhood is a time of learning about the world. One thing children learn is how society is structured, and which groups are privileged over others. If the books that children read underrepresent certain groups, this will provide children with ample evidence that these groups are less worthy of consideration.

Literature is also different from other cultural artifacts such as television and cinema. First, reading books requires more cognitive effort than watching movies or television. This additional effort makes it even easier to assimilate the implicit messages books send about various groups. Books are also valued more than television or movies as works of art, especially by parents. Parents frequently encourage their children to read more, even as they set limits for how much television and movies may be watched. Books continue to be valued throughout school, even as films and television are considered less valuable forms of entertainment. As a result, the messages books send may be considered more authoritative than the ones sent by television or movies, more reflective of how society aspires to see itself. Finally, unlike other media, many books given the reader total access to the minds and lives of the characters. When we watch

---

might be a good thing if half of the characters in children's books were female. However, if this were true because boys only read books with male characters and girls only read books with female characters, this would color our understanding of the equality between female and male central characters. Two of the studies in this dissertation examine the distribution of character gender at the level of the reader, but the first does not in part because it is far easier to measure the proportion of characters who are female in all books than the distribution of this proportion across children.

television, we observe characters who are external to us. We can draw inferences about their thoughts and motivations, but no more so than with any other person we observe. When we read, we step into the characters. We have total access to their inner lives. In a sense, we become those characters, and in doing so give them some measure of our own humanity. This is why better representation in literature has the potential to be so powerful. Enabling children, and especially members of socially dominant groups, to see themselves in socially non-dominant groups is a potential lever for promoting their appreciation and understanding of these groups.

### *The studies*

This dissertation comprises three studies, all organized around the theme of quantity of representation by race and gender in children's books. The first study examines the distribution of race and gender in central characters among popular books sold nationally. The second and third studies consider books for older children, and are restricted to books read by children at a set of elementary schools in North Carolina. They examine the distribution of the proportion of female central characters across children. Conceptually, the studies examine three different sources of books for children: older relatives, self-directed choices, and teachers.

### *Study 1*

In my first study, I build on literature which finds that children's picture books are more likely to include male central characters than female. I identify a key flaw in the sampling decisions of these previous studies which weighted each book title equally despite some books being encountered by far more children than others. I show that the overrepresentation of male central characters is even more severe in the books that

children read; previous studies had underestimated the ratio of male to female central characters in books children read by failing to account for the fact that books with male central characters are read more often than books with female central characters. I also demonstrate that the books that children read are far more likely to contain White central characters than non-White.

### *Study 2*

In my second study, I examine the genders and races of central characters in books that children in second through fifth grade select from a public-school library. I provide evidence that, although children in this sample are predominantly Black and Latinx, the majority of books that they select feature White central characters. I also show that boys' selections contain far more male central characters than female, while girls' selections are roughly balanced between female and male central characters. Finally, I show that children's preferences for characters of the same gender as themselves tends to grow over time. I connect this to the first study to demonstrate that books with male central characters continue to dominate the books children read even as they age, but I also demonstrate that this imbalance is far larger for the books boys read than the books girls read. I also tie this finding to previous results from the field of gender development which show that boys have more rigid stereotypes than girls about what constitutes gender-appropriate behavior.

### *Study 3*

In the third study, I consider the role of teachers in the overrepresentation of White, male characters in the books children encounter. Previous studies have provided some evidence that teachers prefer to use books with White, male characters in their classes

over books with non-White or female central characters. In the third study, I use data from Project READS to show that teachers have a slight tendency to select books with male central characters for their male students. More importantly, I demonstrate that teachers do not act to correct the overrepresentation of the socially dominant group. This provides evidence that teachers in the schools sampled may not have dramatically increased the overrepresentation of boys and men in the books that children encounter, but that they also do not act against that overrepresentation.

### *Coding*

All the studies described herein require me to code books according to the race and gender of the central character. Here I briefly describe this process, noting any differences across the studies. I do not re-describe this process in any detail in subsequent chapters.

For each book sampled, I began by identifying the central character or characters of the books. The concept of central character is not explicitly defined in the literature on which I draw, but it is like the idea of a protagonist. The central character of a book is the person whom the book is about. A handful of books have more than one central character if there are multiple characters, all of whom are equally central to the narrative. For example, many of the Berenstain Bears books feature Brother and Sister Bear equally, and for these books both serve as central characters. However, the fact that a character has an important role in a story does not mean that they are a central character. In the recent Disney movie *Frozen*, Anna, the younger sister, is the central character. There are other characters with major roles, including her sister Elsa and her companion Kristof, but the story mostly revolves around Anna and her quest to rescue her sister and her

kingdom. Similarly, most Berenstain Bear books focus on one of the characters, usually Brother or Sister Bear, with the others temporarily relegated to supporting roles. If a book had no central character, either because it had no characters or because none of the characters were central to the narrative, I coded the book as having no central character. Most analyses omitted these texts. Having identified the central character, I then coded each book according to the gender and race or ethnicity of the central character.

I identified the central characters of the texts and their gender and race or ethnicity by accessing the amazon.com and books.google.com pages associated with the book, and using descriptions, illustrations, and excerpts from the books as clues.

For example, the book “The Wooden Mile”, by Chris Mould, was sampled in the second study. The cover illustration for the book appears to show a zombie and a wolf and does not conclusively indicate the gender of the main character. The description of the book reads

“Something Wickedly Weird is most definitely here! Crampton Rock seems like a lovely seaside town...at least until dark. When eleven-year-old Stanley Buggles inherits a house from a mysterious uncle he didn't know he had, he also inherits a mystery and some strange and sinister new neighbors. The questions begin to pile up: Why are all the dogs in town three-legged? Why is no one on the streets after dark? Is it true that the man who runs the candy shop is a werewolf? And why do those shoemakers look an awful lot like pirates? With the help of Mrs. Carelli, a housekeeper, and a talking stuffed fish, Stanley begins to unravel the mysteries that haunt his great-uncle's death and have set their sights on him. A thrilling, spooky, and funny read, and the first installment of a kid-pleasing new series.”

Although this description mentions both a male character, Stanley Buggles, a female character, Mrs. Carelli, and a character of indeterminate gender, the talking fish<sup>5</sup>, the structure of the description makes it clear that Stanley is the central character. Stanley is the first character introduced and the narrative is presented as the adventures of Stanley while accompanied by the other characters. The fact that this is a children's book and Stanley is the only child character reinforces his identity as the central character.

In contrast, the book "Weirdo Hallowe'en", by R.L. Stine, is described by the following text.

"Chris and his twin, Meg, have a problem. It started the night before Halloween when the twins found a gang of bullies hassling someone in costume. The twins rescued the kid-only to find that it was a bald, pink alien who promptly swore allegiance to them. "Weirdo" is smelly and gross with a lot of disgusting habits, and the twins can't shake him. Little do they realize that their troubles are only beginning. Weirdo<sup>6</sup>'s invited hundreds of aliens to town on the 31st for a full-scale invasion-and to come live in the twins' house! Will the HorrorLand souvenir save-or harm-them?<sup>7</sup>".

Although the description opens with Chris, who is male, there is not enough evidence to decide which twin, if either one, is the central character. However, in the excerpt of the book provided by amazon.com, it is clear that Meg is the central character, and Chris is a secondary character who exists mostly to irritate Meg. The book is written in first-person

---

<sup>5</sup> In theory, the gender of the fish is not specified. In practice, non-human characters included in books for comic relief are almost always male. Non-human animals which serve as romantic foils for the central character (e.g., the female fox in the Fox and the Hound) are typically female. Dory in Finding Nemo is a notable exception.

<sup>6</sup> Weirdo is also a non-human character included for comic relief and is, of course, male.

<sup>7</sup> Gentle reader, fear not. Charles and Meg both escape without major harm.

perspective, with Meg as the narrator. She is the driver of the action, and Chris is along for the ride.

These two texts were selected to demonstrate some of the more challenging texts to code. The vast majority of texts were far simpler. For example, it is quite easy to determine that a book about Snow White has a female human adult as a central character. Having already read “Where the Wild Things Are”, I was able to code the book as having a male human child as the central character without looking it up online. Similarly, after coding a handful of books from “The Magic Treehouse” series, it was clear that all of the books had two central characters, a male and a female human child. Any book in the Pinkalicious series has a female human child central character. Ultimately, very few titles required a great deal of effort to code.

Possible central-character gender codes included female, male, both, or neither. For most analyses of the gender of the central character, I set aside all books with neither female nor male central characters. These texts were generally informational texts, but also included some narrative texts with non-gendered, non-human animals as central characters or narrative texts with no central characters.

If an informational text without a central character focused exclusively, or almost exclusively on people of a single gender, I assigned that as the gender of the central character. For example, a book about an American football team would be coded as having a male central character. The same would be true of a book about professional basketball if all, or almost all of the people discussed in the book were male. This would be true even if the book had a chapter on women’s basketball<sup>8</sup>. However, if a book on

---

<sup>8</sup> Including a single chapter on women’s basketball in a book focused on male players clearly signals that actual basketball is played by men, and women are just a sidenote.

basketball discussed basketball skills without focusing on players of either gender, I would code it as having neither a female nor a male central character. Alternately, if a book on basketball focused on player biographies, but gave approximately equal attention to female and male players, I would code it as having both female and male central characters. If any book had more than one central character and more than two-thirds of those characters were of a single gender, I coded the book according to the dominant gender.

Possible race and ethnicity codes included African-American, Asian-American, Latino/a, Middle-Eastern, Native-American, White, Multiple, and None. For most analyses of the race and ethnicity of the central character I set aside all books which were coded as None. For books with non-human central characters, if the central character was a pet, possession, or companion of a human, I identified the central character with the race of the human character. As an example, in the book “Stuart Little”, the central character is a mouse, and is therefore raceless. However, Stuart lives in a White household, and almost everyone he encounters in the book is White. I coded the race of the central character in this book as White, because Stuart was most closely affiliated with White characters. In contrast, in the book “Goodnight Moon”, all the characters are rabbits, kittens, and mice. There are no characters in this book who have a recognizable race, so I coded this book as having a central character with no race.

For many books, it was impossible to identify the race of the central character because there were no illustrations of the central characters and text descriptions did not label them with their races. Future research might determine if and how children assign race to characters when the character’s race is ambiguous or undefined. For example,

children might identify non-human animal characters as White if they are depicted as having have blue eyes, or if they are given stereotypically White names. Alternately, perhaps children assume that characters share their own race unless given explicit information to the contrary. However, for this analysis I labeled the race of the central character in these cases as None. For most analyses, I dichotomized race as either White or non-White, treating books with central characters of multiple different races as non-White<sup>9</sup>. The sole exception was the first study. In the first study, most non-White characters were Disney princesses and fairies who appeared as central characters in individual stories in books with White characters. It seemed unreasonable to label a book with ten stories about White princesses and two stories about non-White princesses as having a non-White central character. In the first study, I assigned each book a fractional value for race, equal to the proportion of central characters who were White. This was not necessary in the second or third studies which included few or no collections of stories. For the books in these studies it made more sense to count any book with at least one non-White central character as having non-White central characters.

---

<sup>9</sup> This decision may seem questionable. I did so after observing how few of the books had non-White central characters. In this context, I believed that encountering a book with any non-White central characters would be most similar to encountering a book with only non-White central characters. Even after this generous labeling of books as having non-White central characters the vast majority of books featured White central characters.

## **Chapter 2 – Gender in children’s picture books: The impact of parental decisions**

### **Abstract**

Previous studies have demonstrated that male central characters outnumber female central characters in children’s picture books. However, these studies all suffer from a common methodological flaw in that they fail to weight the books by their popularity. In this paper, I develop a new dataset and a new approach to sampling which offers an estimate of a more interesting quantity, namely the ratio of male to female central characters in books that children actually encounter. I also extend previous results by considering racial representation, and whether gender representation differs by character race. Finally, I demonstrate how this approach can be used to better investigate the quality of representation in the books that children encounter.

Reading books with their parents is one of the first ways in which children are exposed to the world around them, and through which they learn the expectations that society has for them. In particular, through encounters with the written word, children learn about gender and about how the two genders are hierarchically arrayed<sup>10</sup>. One way that children learn about gender is by seeing whose stories, those of women and girls and those of boys and

---

<sup>10</sup> Recent scholarship has encouraged us to move beyond a gender binary wherein all persons can be unambiguously categorized as either female or male. While this is a laudable goal, our society and language still classify persons according to those two genders, and it is likely that children do as well. Thus, this paper will treat gender as a binary without endorsing the normative proposition that gender ought to be a binary system. As a practical note, all books analyzed in this paper also treat gender as a binary construct.

men, are told. If children consistently encounter books which are about male characters, this may lead them to see boys and men as more central to society than girls and women.

In this analysis, I examine the ratio of male to female central characters in children's picture books. Other articles on gender representation in children's picture books have made the case that it is harmful to children to encounter stories which predominantly feature male central characters (see, e.g., McCabe, Fairchild, Grauerholz, Pescolido, & Tope, 2011, or Hamilton, Anderson, Broaddus & Young, 2006). In this paper, I will take it for granted that it is bad when children primarily encounter books which feature characters of a single gender. Specifically, it is bad when children primarily encounter books that feature male central characters. Exposing children to texts which are disproportionately about male central characters reinforces a gendered system in which men and boys are considered more central to society than women and girls.

My goal is simply to improve on existing estimates of the disparity in representation by using an approach to sampling books which better captures the phenomenon of interest, namely the ratio of male to female central characters in books to which children are exposed. I do not present evidence that this exposure causes harm, and in the discussion I outline why this may not be feasible. However, it seems uncontroversial to claim that something is wrong when very few books that children encounter are about girls and women.

### *Results from previous studies*

Previous studies of gender representation have consistently found that males outnumber females as central characters. This is true of various collections of books,

including Caldecott Medal winning books (Hamilton, Anderson, Broaddus & Young, 2006; Kortenhuis & Demarest, 1993), popular books (Hamilton, Anderson, Broaddus & Young; Tepper & Cassidy, 1999), books in series (Poarch & Monk-Turner, 2001), books sampled from library holdings (Kortenhuis & Demarest), and books published in the Children's Catalog (McCabe, Fairchild, Grauerholz, Pescolido, & Tope, 2011).

Studies of recently published books have tended to find less disparity in representation. McCabe and her coauthors (2011) report that in books published in the 1990s, the ratio of male to female central characters in books in their sample approached 1.2:1, down from 2.5:1 in books published in the 1960s. The authors trace a rising and falling pattern to the ratio, which was close to 1.2:1 in books published in the 1910s, then rose to a peak from the 1930s to the 1960s, before falling from the 1970s through the 1990s. They link this pattern to waves of feminist activity, which peaked in the 1910s then subsided until the end of the 1960s and beginning of the 1970s. There is reason to believe, then, that publishers are publishing more books with female central characters than has been the case in previous decades, or at least that this was so through the 1990s.

#### *Books published v.s. books encountered*

Although the studies cited above provide us with valuable information about the prevalence of female central characters in award-winning books, or in published books, they shed little light on how frequently children encounter these female characters. If our interest is in determining if children primarily read books with male central characters, then a different approach to sampling is required.

Many of the studies discussed above estimate proportions of characters who are female and male by sampling from well-known book series, including Caldecott Medal winners and runners up (Kortenhaust & Demarest, 1993; McCabe, Fairchild, Grauerholz, Pescosolido, & Tope, 2011; Anderson & Hamilton, 2005; Hamilton, Anderson, Broaddus, & Young, 2006; Turner-Bowker, 1996), Notable Books for Children (Gooden & Gooden, 2001), or Little Golden Books (McCabe, Fairchild, Grauerholz, Pescosolido, & Tope). One problem with this approach, which is acknowledged by the authors, is that award-winning books may differ systematically from other books, and so the population about which inferences may be made is restricted.

Other studies use books sampled from library holdings (Poarch & Monk-Turner, 2001; Tognoli, Pullen, & Lieber, 1994), lists of best-sellers (Anderson & Hamilton; Hamilton, Anderson, Broaddus, & Young), or books which parents report reading to their children. Cassidy and her co-authors (1998) created a sample of books by asking parents of 47 children, aged three to six, to report the books that they read to their children over the course of a single week. Tepper and Cassidy (1999) found that the books in this sample tended to have male titles and central characters. However, the sample of parents from which the original study identified commonly-read books was predominantly and was drawn from suburbs of two large Eastern cities, and is almost certainly not representative of American parents as a whole.

All of these studies take samples of book titles and in doing so treat individual books as the level of analysis. Although this is a reasonable approach, the ratio of male to female central characters in books published may be very different from the ratio in books that children actually encounter. As a thought experiment, suppose that 50% of all

books published have a female central character, and 50% have a male. In this case, the ratio of male to female central characters in published books would be 1:1, or perfect parity. However, suppose further that an average of 100 children read each book with a female central character while an average of 1,000 read each book with a male central character. Then the ratio of male to female central characters that children encounter will be 10:1, because far more children encounter each book with a male central character than encounter each book with a female central character. In this hypothetical scenario, the population of titles is evenly split between books with female and male central characters, and studies which randomly sampled book titles would reveal no bias in representation. However the books which children actually encounter would almost exclusively feature male central characters. Figure 1 represents this possible universe. In the left-hand image, books appear to be equally balanced between male and female central characters. However, the right-hand side weights each book by its readership, revealing a fictional universe which is dominated by hugely popular books with male central characters.

There are reasons to believe that this hypothetical scenario is at least partially correct. Other data suggest that parents tend to read books with male central characters to both girls and boys, but to read books with female central characters mostly to girls (McIntyre, 2017). This would result in titles with male central characters being far more popular than titles with female central characters, and therefore in children being much more likely to read books about male central characters.

*A different approach to sampling*

An approach better aligned with the goal of understanding the ratio of male to female characters that children encounter is to treat an individual encounter with a book as the unit of analysis<sup>11</sup>. One way to conceptualize an encounter with a book would be as an individual experience of reading a book. Another way would be as the first time a child reads a given book. Either conceptualization, or possibly others, could be relevant depending on the mechanism by which encounters with books are thought to shape children's understandings of gender. Regardless of how a child-book encounter is defined, a researcher could attempt to estimate the ratio of male to female central characters in books that children encounter by taking a random sample of child-book encounters, coding the books involved in those encounters as having either a female or male central character, and using this sample to estimate the ratio.

This approach presents two key challenges. The first is that it is much more challenging to take a sample of encounters than of titles. Sampling frames of book titles are readily available, though relying on a sampling frame which only includes award winning books runs the risk that these books may be very different from other books. Taking a simple random sample, or even a census, of these books is trivially simple. In contrast, there is no handy sampling frame of child-book encounters from which to sample.

The second challenge is that a random sample of encounters may result in more unique books than it is feasible to code. The major cost to the researcher in estimating the

---

<sup>11</sup> This could also be problematic if some children encounter books far more frequently than others. For example, suppose that girls encounter far more books than boys, and these books are balanced with respect to central character gender. In contrast, suppose that boys only encounter books about boys. Then on average the ratio of male to female characters encountered might be close to 1:1, but 50% of all children would only read about boys, while 50% read almost equally about girls and boys. In this case, we might be better off treating individual children as the unit of analysis. Studies two and three attempt to do so.

ratio of male to female central characters is the amount of time required to code books according to the genders of the central characters.<sup>12</sup> A simple random sample of child-book encounters could easily result in thousands of unique books which need to be coded according to the central character gender or some other feature. Doing so could be extremely challenging and time consuming.

Design-based approaches to inference, or survey statistics, offers a solution to both challenges. Survey statistics are generally interested in estimating population-level totals and averages, for example the mean income of all people in America. Simple random samples, in which each possible sample of Americans is equally likely to be drawn, are infeasible due to the associated expense and logistical challenges. A more reasonable approach is to take a weighted sample of higher-order units, such as census blocks, estimate the mean income in that unit, and then use the known probabilities of selecting each unit to generate an unbiased estimate of the population parameter of interest. We can improve our efficiency if we have access to relevant information about the sampling blocks, such as the number of people in each block.

In the current context, a similar approach could work as follows. Suppose that we have access to a sampling frame of book titles, and for each title we know how often children encounter that title. This is much more reasonable than asking for a sampling frame of all child-book encounters. We can take a random sample of titles from the frame, either a simple random sample or, more efficiently, a sample where the probability of selecting a given book is proportional to how frequently it is encountered. This will

---

<sup>12</sup> This is especially costly when estimating other gendered characteristics which require carefully reading books, such as whether male characters in the book are presented as more agentic than female characters, or counting the number of male and female characters in illustrations in the books.

make it more likely that we will draw more popular books, which are more important for estimating the ratio in the population. We can restrict our sample size to a number which is reasonable to code; if our coding scheme is elaborate and coding books is relatively expensive we may need to take a smaller sample than for a simple coding scheme. We can code these books, and then use the known probability of selecting each book to obtain unbiased estimates of the number of encounters with books which involve a male central character and the number of encounters which involve a female central character. We can use these totals to estimate the ratio of male to female characters in encounters with books. We can also obtain confidence intervals for the ratio and conduct significance tests.

It is challenging to find a sampling frame which lists all children's picture books and the associated number of encounters for each book. However, a reasonable proxy for encounters with books would be the number of book titles sold in a given year. This should be roughly proportional to the number of times each title is encountered by children, and would be much easier to obtain than data about actual encounters. Alternately, surveying parents as to which books they read to their child in a given week could be used to construct an approximation to the sampling frame of interest, though it would be essential to administer that survey over the course of an entire year; my analyses suggest that the ratio of male to female central characters in books sold varies systematically over the course of the year.

In this analysis, I use sales data for children's picture books to attempt to approximate the number of encounters children have with male and female central

characters in highly popular children's picture books. In the discussion, I point out how survey data could be used to improve these estimates.

### *Race*

Although the ratio of male to female characters has been estimated in multiple different studies, and this ratio has been estimated separately for child characters, adult characters, and non-human animal central characters in at least one (McCabe, Grauerholz, Pescolido, & Tope, 2011), to my knowledge no study has attempted to estimate separate ratios of male to female central characters by character race. This is an important omission. Race and gender interact powerfully; gender does not function in the same way across racial groups. Equivalently, race does not function in the same way across genders. Much modern scholarship on gender insists on the importance of understanding the different experiences of women of different races (Cho, Crenshaw, & McCall, 2013). A Black girl may not see herself represented in a White female central character, just as a White girl may not see herself represented in a White male central character. Therefore, it is important to try to estimate the ratio of male to female central characters as a function of character race, as well as for characters overall.

### *Research questions*

In this study, I address two key research questions. First, what is the ratio of male to female central characters in highly popular books that children purchase? Second, does the ratio differ according to the races of the characters?

I hypothesize that the ratio in books sold will be larger than the ratio in book titles, because books with male central characters will tend to sell more copies than books with female central characters. I have no hypothesis about the different ratios for central characters who are of various races.

## **METHODS**

### **Data**

Each week, *The Publisher's Weekly*, a trade magazine published for people in the book industry, reports the week's top 25 best-selling books in several different categories. One category that the site tracks is children's picture books. For each title, the site reports the number of copies it sold during that week, and the number of copies it sold during the year to date. Records go back to the week of June 6<sup>th</sup>, 2012, though in 2012 the sales, year to date, record sales from January through that date. The present study uses records from the beginning date until January 11<sup>th</sup>, 2016, to create a dataset of popular children's books<sup>13</sup>. Over the span of the study, which covers 190 weeks<sup>14</sup>, 340 books have appeared on the bestseller list at least once.

### **Variables and Coding**

I coded all 340 books as having a male or female central character. I also coded all books according to the race or ethnicity of the central character. Detailed descriptions of the coding process are given in introduction of this dissertation. Note that, given the almost total lack of books with non-White central characters, I assigned race codes equal to the

---

<sup>13</sup> This is the last week for which sales from 2015 are reported; sales appear to be reported roughly two to three weeks after they are recorded.

<sup>14</sup> *The Publisher's Weekly* did not report bestsellers for the week of October 6<sup>th</sup>, 2014. Two books on the list, *An Abundance of Katherines* and *Hush, Hush*, were not children's picture books and were not included in my analyses. They appear to have been listed mistakenly.

proportion of characters in each book who were of various races. This was especially appropriate in the first study in which non-White central characters tended to be Disney princesses and fairies in story collections which also included White central characters.

### **Analysis**

I calculated the ratio of male to female central characters, the ratio of White to non-White central characters, and the ratio of male to female characters for White and non-White characters in books as a whole, unweighted by popularity. I then calculated the same at the level of individual purchases. For example, the book *How the Grinch Stole Christmas*, which has a male central character, sold 92,091 copies in 2012, 125,558 copies in 2013, 146,249 copies in 2014, and 163,396 copies in 2015. As a result, it contributes 527,294 to the weighted total number of male central characters, but only a single observation to the unweighted total.

Given access to full sales data from 2012 to 2015 it would be possible to directly calculate these ratios directly and there would be no need for inferential statistics. However, because of the way that *The Publisher's Weekly* presents sales data, it is only possible to view yearly sales for a book up to the last week in which it appears on a bestseller list. In 2013 the book *Fancy Nancy: Heart to Heart*, which is a Valentine's Day themed book, sold 46,617 copies by the week of February 25<sup>th</sup>. For the rest of the year, the book never appeared on the bestseller list. I can be certain that the book sold at least 46,617 copies in 2013, and I can establish an upper bound on the number of copies sold by noting that in every week after that one, the book sold too few copies to be included in the bestseller list. However, this leaves an extremely wide range of possible sales, between 46,617 and 227,716. The issue is much less severe for a book like *Princess*

*Bedtime Stories*, which in 2013 last appears on the bestseller list on December 9<sup>th</sup>, almost the last day of the year. This book sold between 29,809 and 61,742 copies in that year.

I addressed this uncertainty by using a variety of methods to impute missing book sales. My main approach entailed doing the following: I assumed that in each week in which a book did not appear on the top-25 bestseller list, the number of copies it sold followed a Poisson distribution with a constant mean. I estimated this mean and uncertainty estimates, treating it as a Bayesian posterior distribution for the mean. I then took draws from the posterior and used these to simulate book sales in weeks after the last week the book was observed in each given year. This approach basically assumes that the average number of copies a book sold in weeks it was unobserved before the last date that it was observed are a good proxy for the average number of copies it sold in weeks after the last date it was observed. I used this approach to obtain point estimates and uncertainty intervals. I used the other imputation approaches as sensitivity checks.

## **Findings**

Each book appears on the list an average of 13.5 times, though the median number of appearances is only 5. A handful of extremely popular books, such as *The Very Hungry Caterpillar* and *Goodnight Moon* appear in almost every week. Figure 2 shows a histogram of the number of weeks each book appeared on the bestseller list. The most popular book in the sample sold 1.76 million copies over the entire study, while the least popular sold only 2,900. In fact, the most popular book sold more copies than the least popular 30% of the sample. Figure 3 shows a histogram of the total books sales over the course of the study. This suggests the importance of weighting each book by how frequently it is encountered by children. Even in this sample of books which are

identified as extremely popular, some books are vastly more popular than other, and will therefore be encountered by more children.

Before weighting, books in the dataset have slightly more male central characters than female. Overall, 80% of books have a central character with an identifiable gender. Of these, 25% of books have a female central character, 69% have a male central character, and 6% have both a female and a male central character. Counting books with both female and male central characters as contributing a half of a female central character and a half of a male, 72% of all central characters are male and only 28% are female. The ratio of male to female central characters in all bestselling books is approximately 2.5:1. See Table 1 for complete results.

Turning to race, 48% of the books which appear in the dataset have central characters with identifiable races. Only 2% of central characters with an identifiable race are Black (Tiana and Doc McStuffins) and 2% are Latina (Dora the Explorer). The other 96% are White. There are vanishingly few central characters who are Asian (Mulan), Middle-Eastern (Jasmine), or Native-American (Pocahontas). Given the tiny numbers of non-White characters, I will henceforth treat race as a dichotomy, with central characters identified as either White or non-White. The ratio of White to non-White central characters is 24:1.

In this dataset, all of the non-White central characters are female, meaning that, for non-White central characters, the ratio of male to female characters is 0. This is surprising, but is driven by the fact that, with a single exception, the only non-White central characters in the dataset are Disney princesses, Disney fairies, Disney's Doc

McStuffins, and Dora the Explorer<sup>15</sup>. The non-White Disney princesses and fairies only appear in large collections of stories in which the majority of characters are White<sup>16</sup>.

For books with White central characters, 24% have a female central character, 69% have a male central character, and 6% have both female and male central characters. This means that in books with White central characters, 72% of central characters are male and 28% are female, for a ratio of 2.6:1. The fact that this is almost identical to the ratio in all books is due to the fact that so few central characters are non-White.

The previous results are not weighted by the number of times each book was sold over the period; weighting dramatically increases the overrepresentation of male central characters. Using my preferred approach to imputation, I estimate that 19% of books sold had a female central character, 77% had a male central character, and only 3% had both female and male central characters. Thus 21% of central characters in all books sold were female, while 79% were male. This gives a ratio of male central characters to female of close to 3.8:1 in popular books sold during this period<sup>17</sup>.

The ratio of male to female characters in popular books sold is substantially larger than the ratio in popular books as a whole. The reason for this is that books with male

---

<sup>15</sup> The sole exception is *Rad American Women A-Z: Rebels, Trailblazers, and Visionaries Who Shaped Our History . . . and Our Future!*, which is an alphabet books featuring notable American women of many different races and ethnicities. This book is strikingly different from other books with non-White central characters,

<sup>16</sup> Despite the fact that most Disney princesses are White, the company appears to see the non-White princesses, especially Tiana, who is Black, as important parts of their collection. Most blurbs for collections of stories about Disney princesses highlight the fact that the books include Tiana, and she frequently appears in the title illustrations.

<sup>17</sup> Many other authors prefer to allow each both with both male and female central characters to contribute one female and one male central character. I believe that allowing each one to contribute half of a female and half of a male central characters is more sensible, since a book with a single female central character presumably signals the importance of girls and women more powerfully than a book with both a female and a male central character. However, using the common approach described above, the ratio of male to female characters declines only slightly to 3.6:1. Dropping books with both female and male central characters from the analysis completely increases the ratio to 4.0:1.

central characters tend to sell more copies than books with female central characters. Over the period studied, books with both female and male central characters sold an average of 110,000 copies per book, books with female central characters sold an average of 80,000 copies per book, and books with male central characters an average of 140,000 copies per book. Not only do most of the popular books feature male central characters, those which do sell far more copies than books which feature female central characters. An analysis which samples book titles will miss the disparity in average copies sold and will thus underestimate the ratio of male to female central characters in books which children encounter.

After weighting the races of the central characters by the number of sales each book had, I found that 98% of central characters in books sold were White, while only 2% were non-White, giving a ratio of White to non-White characters in books sold of 49:1. This is only two percentage points different than the unweighted proportions, but represents a halving of the proportion of non-White central characters in books sold. Once again, this is attributable to the fact that books with White central characters tend to outsell books with non-White central characters. Books with White central characters were estimated to have sold an average of 123,000 copies, while books with non-White central characters were only estimated to have sold an average of 55,000 copies. The already huge imbalance between White and non-White central characters becomes even larger when accounting for differences in sales.

Since there were no non-White male characters, the ratio of male to female non-White characters is unchanged by weighting books by the number of copies sold. For books with White central characters, 24% had a female central character, 75% had a male

central character, and 1% had both a female and a male central character. Overall 25% of central characters in books sold were female while 75% were male. This gives a ratio of male to female White central characters in books sold of close to 3.1:1.

As an exploratory analysis, I stratified books according to the years in which they were first published, then estimated the weighted proportions and ratios within these strata (for several strata it was impossible to estimate ratios since no books with female central characters from these decades appeared on the lists). For simplicity, this analysis ignores the handful of books with both male and female central characters. Table 2 shows the results by decade, while Figure 4 provides a graphical display. Although the ratio of male characters to female in books encountered is lower in recently published books than in older books, it remains quite high; there is little evidence that the trend towards parity in books published is leading to parity in books encountered.

## **DISCUSSION**

Using national sales data for highly popular children's picture books, I found that children tend to encounter books with male central characters, and that the ratio of male central characters to female is greater after weighting for the popularity of the books. This confirmed my first hypothesis. I also confirmed my second hypothesis, finding evidence that the ratio of male to female central characters is very different for White central characters than for non-White, due to a total absence of non-White central characters.

*Gender representation in the books children read*

Although studies have found that the ratio of male to female central characters in recently published books has approached parity, my results suggest that popular books which are purchased for children, and which children presumably actually encounter, continue to have far more male central characters than female. In fact, the ratio of male to female central characters in popular books sold is substantially higher than the ratio estimated in any of the preceding studies. Figure 5 and Table 3 situate these results, both weighted and unweighted, in the results from previous studies. This in no way suggests that previous studies have been incorrect, only that they are concerned with estimating something different. However, insofar as researchers are interested in the ways that books shape children's understanding of gender, it is important to realize that the books children encounter may drastically overrepresent male characters, even if published books do not.

There are two explanations for the difference between these results and those other studies. First, although recently published books may be almost equally likely to have a female central character as a male, parents continue to purchase older texts which are much more likely to feature male characters. Studies which focus on the year in which a book was published may miss the fact that the books children encounter were often written long before the current year. The books in my dataset included titles from every decade from the 1950's to the present. In selecting books to purchase for children, adults are likely to draw on "canonical" books which they themselves remember reading as children. Dr. Seuss books remain staples in children's libraries, as do the works of Maurice Sendak. However, these books are also far more likely to feature male characters than more recent titles.

Something more seems to be at work here than the fact that books published in past decades are more likely to feature male characters than recently published books. Although many books published in the past have featured female central and title characters, very few of them remain popular today. The oldest four books with female central characters were published in 1964, 1967, 1976, and 1978. Aside from those, all others were published in after the year 2000. In contrast, one book with a male central character was published in the 1900s, five were in the 1940s, four in the 1950s, 12 in the 1960's, seven in the 1970s, five in the 1980s, and eight in the 1990s. Books about boys and men may be more likely than books about girls and women to remain popular over time, continuing to be used for decades after they are published.

This is an important result because books which remain popular over extended periods of time are likely to have a greater influence on children's perceptions of men and women than books which do not. *Green Eggs and Ham* has been read to millions of children and will probably continue to be read to millions more. Children will read and discuss this book with their parents and grandparents, and will likely read it to their children and grandchildren. Children and other adults will expect other people in society to be familiar with the book and its characters. The fact that these discussions will revolve around two distinctly male characters will serve to create a shared fictional universe which is slightly more male. Even if *Pinkalicious: The Princess of Pink Slumber Party* is popular right now, it will not have a lasting impact on society or on children's perceptions of the relative importance of women and men unless it remains influential.

Another possible reason for the dearth of older books with female central characters is that parents no longer consider female characters from the past to be

appropriate or interesting. Given the extremely stereotypical ways in which girls and women have historically been depicted, the fact that most female characters appear in recently written books could be seen as a rejection on the part of parents of the stereotypical representations of girls and women contained in older books. As important as increasing the quantity of representation of women and girls in children's literature may be, parents may not be willing to do so at the cost of exposing their children to damaging sexist stereotypes, or to dull characters and plots. However, if this is the reason that older books with female characters are no longer popular, parents have not replaced them with more modern books with female central characters.

The second reason for the discrepancy between the male to female ratio calculated in this study and results from previous studies is that books with male central characters are far more popular than books with female central characters, regardless of when they were published. On average, each book with a male central character sold almost twice as many copies as each book with a female central character. In treating book titles as the level of analysis rather than book sales, researchers have obscured these differences and have understated the extent to which children's reading environments are male dominated. The reason the books with male characters are more popular than books with female characters is not clear. One possibility is that parents believe that books about girls are appropriate only for girls, while books about boys are thought to have universal appeal.

This also highlights the fact that parity in the rates at which female and male characters are represented in published books will not necessarily result in parity in the rates at which children encounter female and male characters. If books with male central

characters remain more popular than books with female central characters, children will encounter more male central characters than female, even if newly published books have equal rates of female and male central characters. There is only so much that publishers can do to achieve gender parity in representation in books, if parents continue to preferentially purchase books with male central characters.

The exploratory results from this study suggest that this may be the case. Although recently published books tend to feature male and female central characters at similar rates, even among recently published books far more copies of books with male central characters are sold than copies of books with female central characters.

#### *Kinds of representation of female central characters*

Although the focus of this study is on the rates at which children's picture books feature female and male central characters, the ways in which those central characters are represented is also important. It is beyond the scope of this research to examine in detail how female and male central characters are portrayed in these books, but one result is worth mentioning. Female central characters tend to be portrayed in extremely restrictive, stereotypical ways. Of the female central characters in this dataset, a full 20% are princesses, and an additional 5% are fairies. Representing female characters as princesses or fairies is not problematic in and of itself. Although they represent an oppressive social system, individual princesses have many positive characteristics, including compassion, courage, and ice powers; despite their reduced stature and inappropriately sexualized clothing, fairies are powerful, kind, and ardent environmentalists. However, it is limiting for children's understandings of the roles that women and girls can occupy to only or

almost only encounter girls in these roles. Very few actual women or girls are princesses, and none are fairies, and neither fairies nor princesses play an especially large role in modern society. The “specialness” of princesses and fairies is a function of a stable status. Princesses are interesting because they are the daughters of queens and kings, or because they are the spouses of princes. Fairies are interesting because of their magical powers and diminutive sizes. Neither of these are realistic roles for most girls.

In contrast, male central characters are represented in extremely diverse ways in popular books, from superheroes to cats to Santa Claus. Many male central characters are simply unremarkable human children, which is less common for female central characters. It may not be reasonable for boys to aspire to become Santa Claus, but it is easy to imagine themselves as the boy in “The Cat in the Hat” or “The Polar Express”. While there are several popular books with female central characters in which those characters are neither princesses nor fairies, books with princesses and fairies as central characters represent a large proportion of all books with female central characters.

Previous literature has not examined the roles occupied by female and male central characters using a quantitative approach.

### *Race*

The results for race are even stronger. There are almost no non-White central characters in popular children’s books. Furthermore, as with the ratio of male to female central characters, the ratio of White to non-White central characters grows tremendously after accounting for the different numbers of books sold. Most books don’t feature non-White

central characters, and the books that do feature non-White central characters are far less popular than those that do not.

The easiest explanation for this disparity is that most books are purchased by White parents for their White children, and these parents select books which are centered on White characters. A majority of young children, close to 67%, are White, and most of these White children are non-Hispanic Whites. White parents have more disposable income than parents of other races aside from Asian-American parents, and presumably purchase more books than parents of other races. If White parents only, or almost only buy books with White central characters, this would explain why almost all the most popular books feature White central characters. Although Black parents might purchase books with Black central characters, none of these would become popular enough to make the list of the top 25 bestsellers in any given week because they would be swamped by books with White central characters. Clearly it would be unfortunate if White children did not encounter books with non-White central characters especially given the power and wealth White people control in the United States. However, the harm would be different than if neither White nor non-White children encountered non-White central characters. Because there is no way of knowing the races of the parents who purchased the various books, nor of determining how many less-popular books focus on non-White central characters, this explanation must remain speculative.

Another possible explanation is that highly popular older books, such as those by Dr. Seuss or Maurice Sendak, only include White characters. As with female central characters, it may be that books with non-White central characters have a difficult time entering the canon of children's literature. Alternately, perhaps older books represent

non-White characters in stereotypical ways which are no longer considered acceptable, and so only newer books with non-White characters are popular. Since the very bestselling books tend to be older, this would partially account for the imbalance in rates of representation between non-White and White characters. However, this cannot be the whole explanation. Even restricting the sample to books published after the year 2000, 93% of book titles with a main character of an identifiable race have a White main character. After weighting by numbers of books sold, this proportion jumps to 97%. Even in recently published books, non-White characters are dramatically underrepresented. Unfortunately the representation of non-White central characters in these books is so low that it is impossible to conduct a more detailed analysis.

### *Race and Gender*

Even more surprising is the difference in the ratio of male to female central characters between characters who are White and characters who are non-White. Although for books with White characters, male central characters greatly outnumber female central characters, for books with non-White central characters the reverse is true. This is largely attributable to the fact that most non-White characters appear in books by Disney. Most of Disney's human central characters are female, and Disney's collections include a substantial number of non-White characters<sup>18</sup>. If there were a popular line of books featuring Disney princes as central characters, it would also include non-White

---

<sup>18</sup> Since, with few exception, the princesses and fairies are ostensibly from worlds which are unlike our own, it is somewhat awkward to label them as White, Black, or another race. Race is inextricably linked to the historical experiences of people in our actual world, which is one there was some controversy when Disney labeled princess Sofia their first Latina princess. However, it is reasonable to expect that children will understand a character like Jasmine to be Arabic, even though there is no actual kingdom of Agrabah in the Arabic peninsula.

male characters. Without the Disney books, there would be almost no representation of non-White female central characters either. It is important to note that these fairy and princess characters appear in compilations of stories which always include and frequently feature White characters. For example, although the Disney fairies books have stories which feature non-White fairies, the central character of the entire series is Tinkerbell, who is White. It is possible that, although these books contain non-White princesses and fairies, children mostly read the stories with White characters, or that children understand the books to be about the White characters.

However, there are also a handful of standalone non-White female central characters, including Doc McStuffins, Dora the Explorer, and a book of “Rad American Women” which includes multiple non-White American women who are notable for being rad, but there are no non-White male central characters. The reason for this absence is not immediately clear. However it is striking that non-White boys and man are completely absent from the most popular books. Dora the Explorer has a male equivalent, Diego, but his books are not popular enough to appear in these lists. There are other non-White male characters in children’s media, but these characters are also not popular enough for books featuring them as central characters to appear in the lists of bestsellers. This may contribute to our understanding of the intersection between race and gender in young children. Perhaps parents, and especially White parents, are more comfortable with exposing their young children to non-White female characters than to non-White male characters. Alternately, perhaps non-White female characters are attractive to publishers because they do double duty, increasing the representation of both non-White and female characters with a single character.

Non-white boys are rendered almost completely invisible in popular children's books. The same is also true of non-White girls, though their absence is not quite as total. This evidence suggests that most children will never, or almost never, encounter picture books which are about non-White characters, at least in the popular books that they read.

## **Limitations**

### *Popular books v.s. all books*

A limitation of this approach is that it focuses exclusively on books which are popular enough to be one of the top 25 bestsellers in a given week. These books may not be representative of all the books that children encounter. As an extreme example, if there are many, many more books about girls than about boys, but individual books about girls tend to be less popular than individual books about boys, then it may be the case that the majority of popular books were about boys, even though most children's books purchased featured female main characters. More plausibly, it could be that the ratio of male to female characters in all books sold is lower than the ratio of male to female characters in extremely popular books sold because there are many books about female central characters which never sell enough copies to secure a place on the list of most popular books.

A future study could use data from *The Publisher's Weekly* and some larger source of book sales, such as amazon.com, to create a link between average Amazon rank across the course of a year and total yearly sales of a book. Having established a linking function, it would be possible to randomly sample from amazon.com's list of books, which includes both highly popular and less popular titles. Taking samples of books with

the probability of selecting a book proportional to the number of estimated sales of that book would enable a researcher to efficiently estimate the ratio of male to female characters in all children's picture books purchased in a given year, without restricting the population to extremely popular books as the present paper does. In doing so, it would be important to average sales ranks across a year, because there is marked seasonal variability in the ratio of male to female characters in books sold. Books sold around Valentine's Day and Christmas, for example, tend to have far more female central characters than books sold at other times of the year. For the present study, I address this concern by restricting my inferences to the population of books which are extremely popular.

However, even if the characters children encounter are more likely to be female than my findings suggest, there are additional reasons to be especially concerned with the imbalance in the most popular books. Because these books are the most popular, they are likely to have an outsized impact on children encountering them. These are the books that children are likely to talk about with other children, to see in their friends' homes, and to watch as movies. These books will be seen as universal, while other, less popular books remain particular. As a result, children may be more likely to assimilate the lessons they receive from the most popular books than from the less popular books; the more popular books may have a stronger influence on the development of children's understandings of society and its values.

A similar argument is advanced by researchers who sample from Caldecott medal winners. Some authors claim that the prominence of Caldecott medal winners suggests that these books will be representative of other books children will encounter in their

reading (Crabb & Bielawski, 1994). Others argue that Caldecott medal winners have outsize influence on society because of the prestige of the award (Weitzman, Eifler, Hokada, & Ross, 1974). However, it seems more likely that young children will be more influenced by the frequency with which they encounter rather than the awards a book has received. Children are likely to notice that they see a book in multiple different locations and that their friends, parents, and other adults are all aware of a book. Books which they see in many different places, i.e., highly popular books, will be more influential on how they see the world than books which they only see at home. It is less clear that children will be aware of how prestigious a Caldecott medal is, or that they will assign more weight to books that have been honored in that way.

*Book encounters as a whole v.s. books encountered by individual children*

Another drawback to my approach is that I can only observe overall book sales. It is impossible to know the ratios of male to female central characters in books that individual children encounter. On the whole, approximately one out of five central characters is female, but individual children may read books that include substantially fewer or more female characters.

Given the huge disparity in the proportions of central characters who are female and male, it is likely that most children primarily encounter male central characters in picture books. However, if parents are disproportionately likely to purchase books with female central characters for their daughters, then the books boys encounter may be even more unbalanced. This seems likely since so many of the female central characters in these books are princesses or fairies, and it seems somewhat unlikely that most parents

read these books to their male children. In fact, it seems plausible that most boys almost never encounter books which are about a female character. This is extremely problematic. Consistently and exclusively encountering male protagonists is likely to send a strong signal to boys that their own stories are more important than those of girls. Given the disproportionate power that men exercise in American society, it is especially harmful for boys to be influenced to see girls and women as less important than boys and men. My data do not allow me to estimate the variability in the ratio of male to female characters which children encounter, but doing so is an important topic for future research.

#### *Books purchased v.s. books encountered*

An assumption which I make is that the books which children encounter are similar to the books that are purchased. This may not be the case. To begin with, not all children's picture books are purchased for children. Many copies of Dr. Seuss's *Oh the Places You'll Go!* are given to students graduating from high school or college and are never read to children. Additionally, children routinely encounter books which are not purchased and therefore will not contribute to my estimates of how many copies each book sold. If parents disproportionately check out books with female main characters from the library and disproportionately purchase books about boys, my estimate will tend to inflate the true ratio of male to female main characters. I have no reason to believe that this is the case, but it remains a possibility.

More importantly, book purchases can be thought of as partially representative of children's home libraries. However, children do not encounter every book in their library at the same rate; some will be more popular and other less. It is possible that, although a

large majority of books that children have at home feature male main characters, children prefer to read the books which feature female main characters. In this case, it could be that children tend to encounter female and male characters at roughly the same rate. However, once again, there is no reason to expect this. If anything, it seems likely that the sorts of books which are most popular overall, i.e., books about boys and men, are also most frequently read. In this case the 3.8:1 ratio of male to female main characters in books sold may actually represent an underestimate of the ratio in books read.

Future research into gender in children's books could attempt to sample books in proportion to the rates that children encounter them by using a survey of parents to determine which books they read to their children, and sampling titles in with probability proportional to how frequently they are read. Cassidy and her co-authors (1998) obtained a sample with similar properties by asking parents to identify books that they had read in the previous week. However, their sample of parents only included suburban White mothers of children at two daycare centers, and so their result cannot be considered representative of the country as a whole. They also did not weight their analyses by how frequently each book was read which makes the parameters they estimated less interesting and difficult to interpret.

Another project might, over the course of a year, use a telephone survey or an online survey to ask parents to list all the books they read to their children over the previous week, and how frequently they read each book. This could be used to construct a dataset of book encounters. It would then be possible to estimate the gendered characteristics of books which children actually encounter, and to estimate how those

characteristics differ according to child and parent gender, race, socioeconomic status, and other variables.

*The effects of disproportionate gender representation*

Ultimately, the importance of this line of research is a function of the effect that encountering literature with disproportionately few female and non-White characters is likely to have on children. However, to my knowledge there is no literature which has attempted to estimate this effect. In fact, under the potential outcomes framework, the dominant framework for causal inference in the social sciences, it is impossible to define the causal effect of continuous exposure to books with male central characters since there is no intervention to be studied.

We might approach an estimate of the causal effect of constant exposure to books with male central characters by imagining how a child's conception of gender would differ according to the ratio of male to female central characters that she or he encountered. Under one condition, which we might label treatment, we could imagine a child who was exposed to books with an approximately equal balance of female and male central characters. Under the other condition, which we might label control, we could imagine a child who was exposed to books which predominantly featured male central characters. However, in addition to being unfeasible, these children would continue to live in a society in which children picture books as a whole overrepresented male central characters; in this context, changing the books encountered by a single child would likely be a poor estimate of the overall effects of the imbalance. However, conducting studies in

which children were exposed to more books with female central characters might at least shed some light on the effects of the overrepresentation of male central characters.

*Extensions and directions for future research*

In this study I have only estimated ratios of male to female central characters. However, the approach I use could be easily extended to estimating other gendered characteristics of children's picture books. For example, Hamilton, Anderson, Broaddus and Young (2006) explored the question of whether male characters were presented as more active than female characters in recent bestsellers or Caldecott medal winners. Counter to their expectations, they found no evidence that this was the case. However, this analysis was conducted at the level of books titles, without accounting for the fact that some books are drastically more popular than others. If parents disproportionately purchase books which depict female characters as less active than male characters, then the books that children actually encounter in their reading may reinforce this gendered stereotype. Ultimately, the effect of stereotypical portrayals of female and male characters on children will depend on how those characters are portrayed in books that children actually encounter.

The approach to estimation that I describe in this article, or the dataset that I have developed, could be used to improve our understanding of how passive the female characters that children encounter are. It would be possible, for example, to code each character in each book according to her or his passivity, and then estimate how passive female and male characters are in the average book sold. Doing so would give a better picture of the sorts of characters that children will actually encounter, and how those characters might impact the ways that children think about gender.

## **Conclusion**

Insofar as the most popular books represent the books which children encounter most, the books which children read present an extraordinarily male, white landscape. Other studies have also found that children's picture books overrepresent boys and men while underrepresenting girls and women. However, when we account for the fact that some books are more popular than others, the imbalance grows dramatically. The fact that recently published and award-winning books are doing a better job of foregrounding stories about girls and women is a positive development. However, as long as books about boys and men remain more popular than books about girls and women, children will continue to encounter reading universes which implicitly support the existing social hierarchy.

*Figure 1: A possible universe in which books with male central characters are as common as books with female central characters (left figure), but are far more popular (in the right figure, circle area is proportional to popularity). Books with male central characters are in blue, while books with female central characters are in green.*

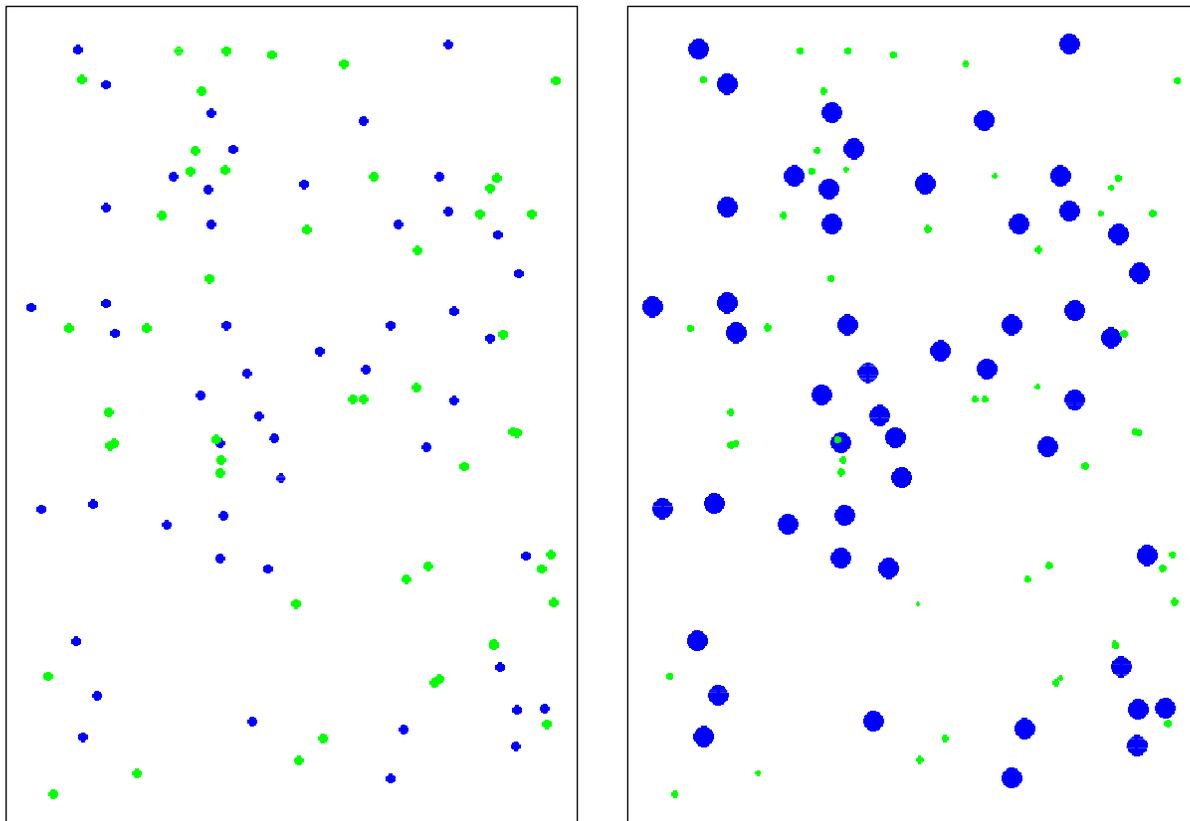
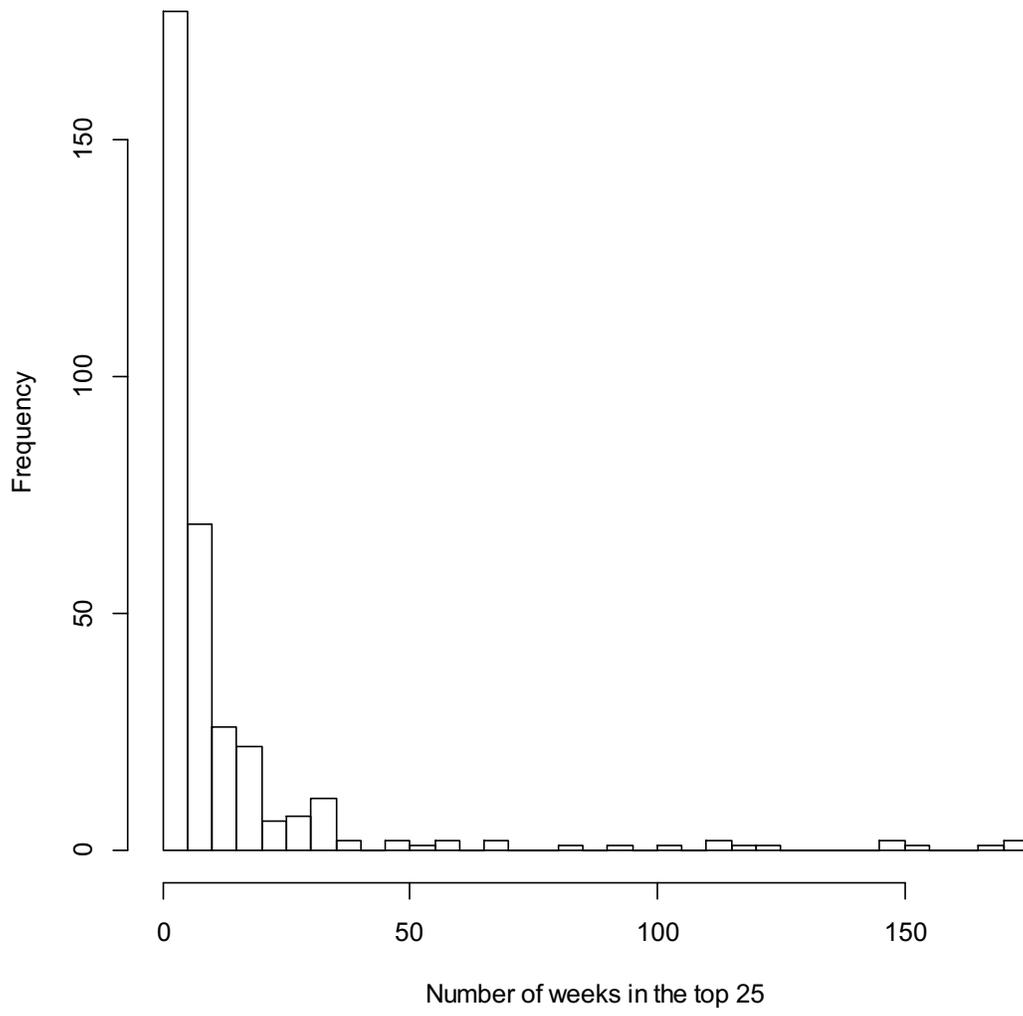
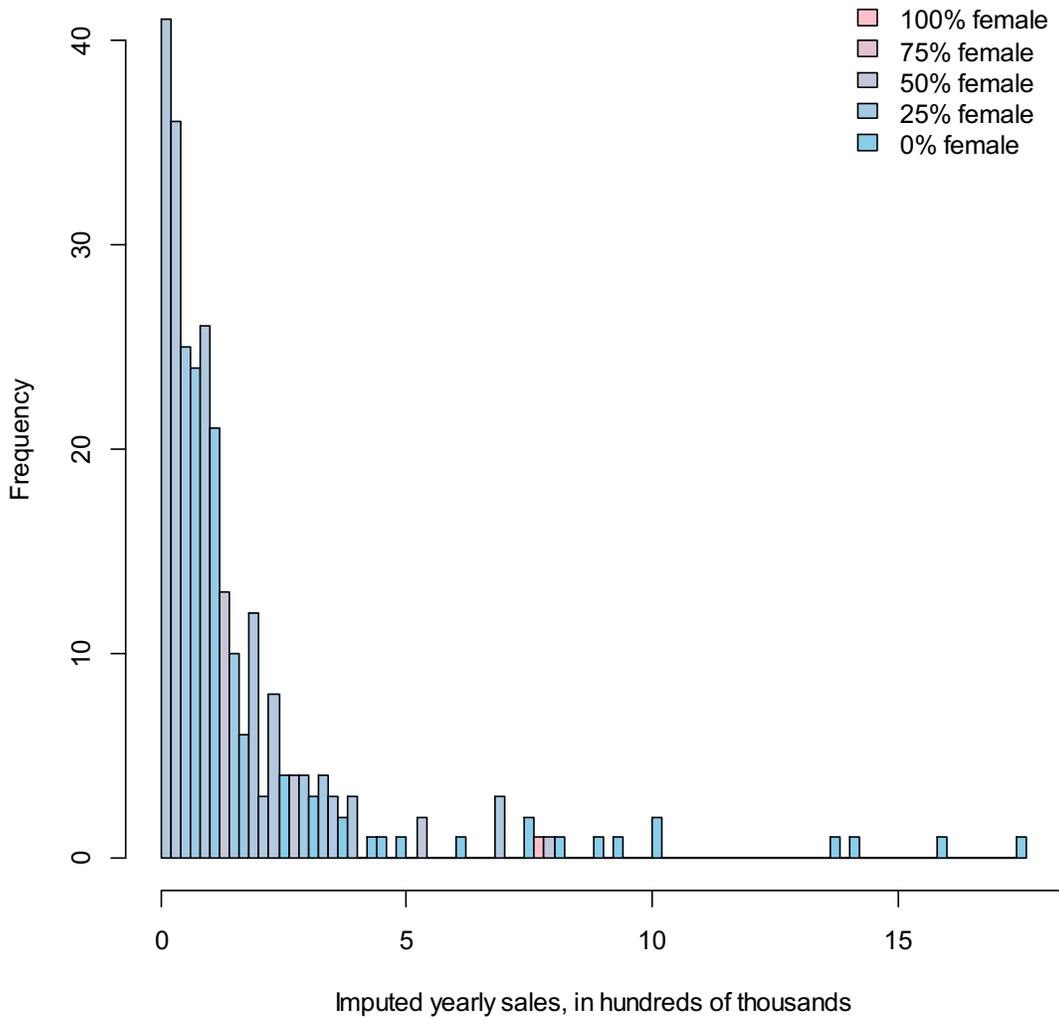


Figure 2: Number of weeks each book in the sample appeared on the bestseller list.



*Figure 3: Number of titles sold per year by each book in the sample. Bars are colored according to the proportion of titles which had a female central character in a very sexist way, with pinker bars indicating more female central characters and bluer bars indicating fewer. The reader will note that the bars representing the most popular books are entirely blue, i.e., all have male central characters, while the bars representing the least popular books have slightly more pink.*



*Figure 4: Total books sold by central character gender and decade published. As before, the color scheme, which uses blue to indicate books with male central characters and pink to indicate books with female central characters, is deplorably sexist.*

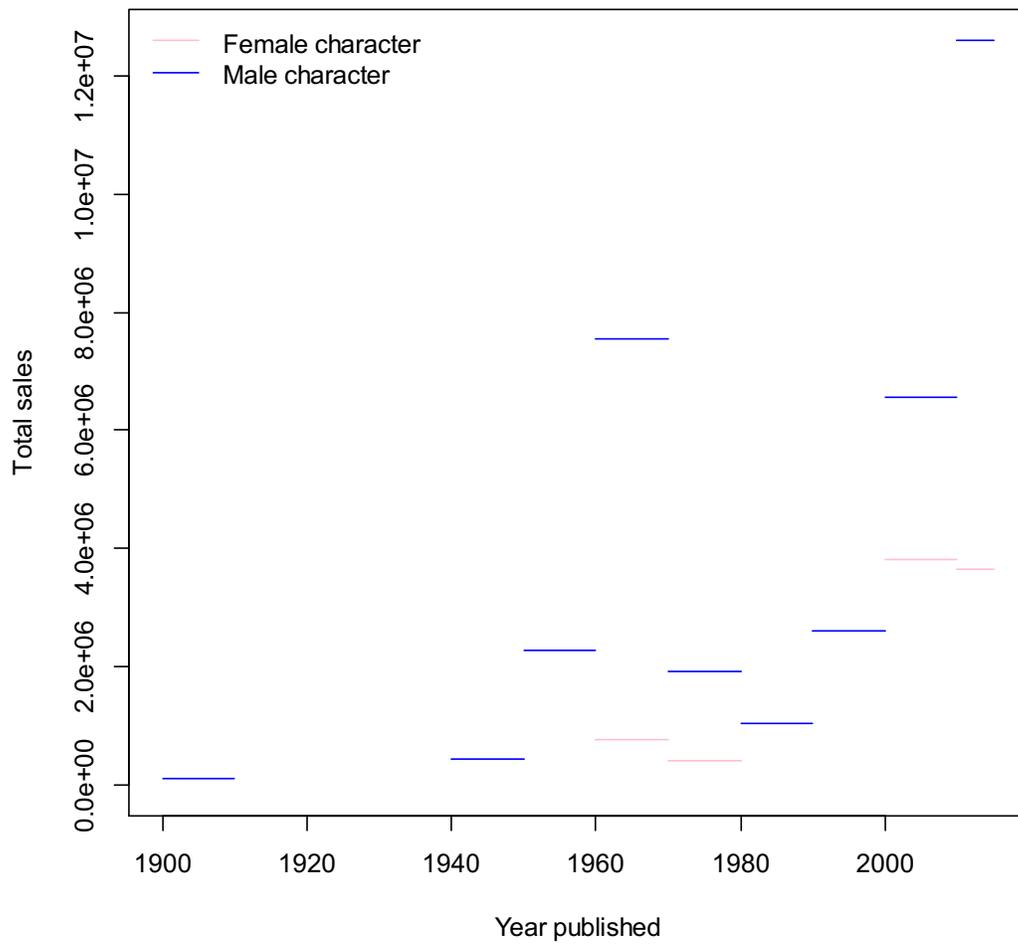
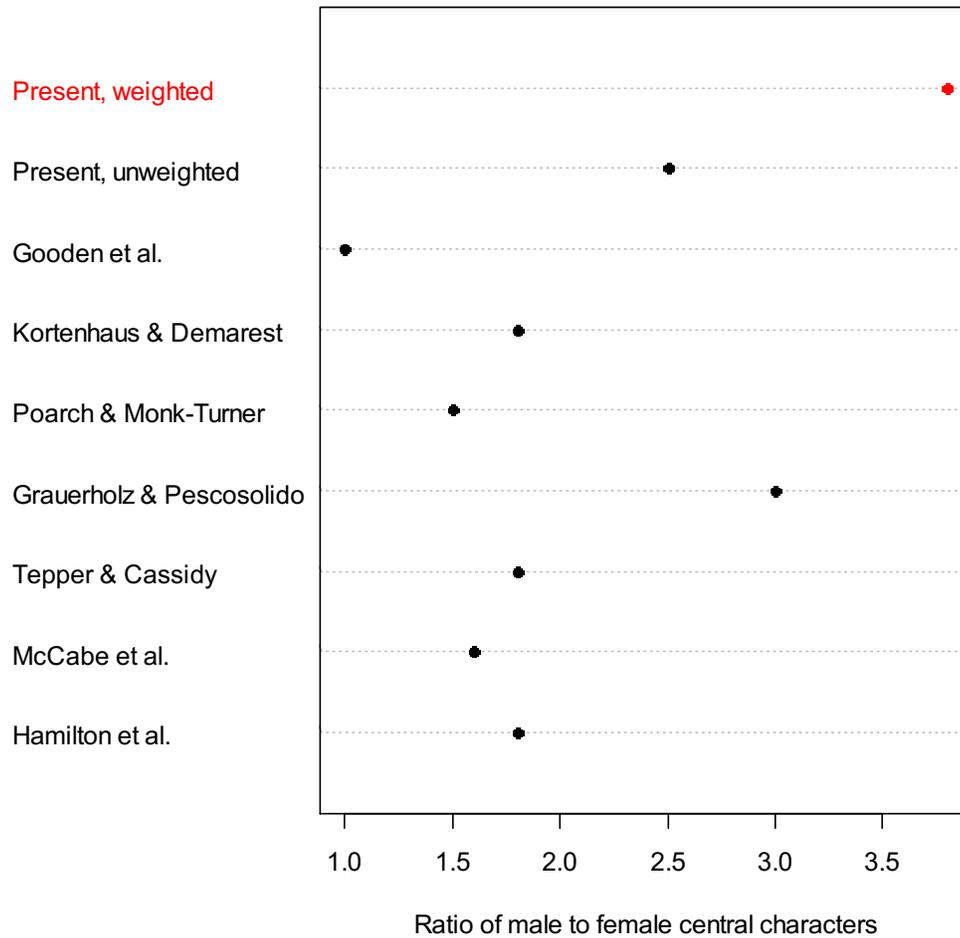


Figure 5: Ratios of male to female central characters in the present study and in previous studies. Although the unweighted estimate from the present study is more or less in line with previous results, the weighted estimate is far larger.



*Table 1: Study results before and after weighting*

<u>Outcomes</u>	<u>Book titles (unweighted)</u>	<u>Books sold (weighted)</u>
Proportion male	.72	.79
Proportion female	.28	.21
Ratio of male to female	2.6:1	3.8:1
Proportion White	.96	.97
Proportion non-White	.04	.03
Ratio of White to non-White	24:1	38:1
Proportion male (White characters)	.73	.75
Proportion female (White characters)	.27	.25
Ratio of male to female (White characters)	2.7:1	3.1:1
Proportion male (non-White characters)	0	0
Proportion female (non-White characters)	1	1
Ratio of male to female (non-White characters)	0:1	0:1

*Table 2: Books sold by gender and ratios of male books to female books*

<u>Decade published</u>	<u>Female total sold</u>	<u>Male total sold</u>	<u>Ratio, male to female</u>
1900		111,000	
1940		432,000	
1950		2,289,000	
1960	775,000	7,539,000	9.73:1
1970	419,000	1,935,000	4.62:1
1980		1,056,000	
1990		2,616,000	
2000	3,827,000	6,562,000	1.71:1
2010	3,662,000	12,605,000	3.44:1

*Table 3: Results from previous study and the current study*

<u>Study</u>	<u>Sample</u>	<u>Years</u>	<u>Ratio of male to female central characters</u>
Hamilton, Anderson, Broaddus & Young	30 Caldecott Medal winners and 155 best-selling books	1995-2001	1.8:1
McCabe, Fairchild, Grauerholz, Pescolido & Tope	5,618 books listed in the Children's Catalog and book series	1900-2000	1.6:1
Tepper & Cassidy	178 books parents reporting reading to their children	NA	1.8:1
Grauerholz & Pescolido	2,216 books in the Children's Catalog	1900-1984	3:1
Poarch & Monk-Turner	22 books drawn from series	NA	1.5:1
Kortenhaus & Demarest	150 Caldecott medal winners, runners-up, and non-award winning books	1940-1989	1.8:1
Gooden & Gooden, 2001	83 ALA Notable Books for Children	1995-1999	1:1
Present study (unweighted)	340 bestselling books	2013-2015 (read, not published)	2.5:1
Present study (weighted)	340 bestselling books	2013-2015 (read, not published)	3.8:1

## **Chapter 3 – Gender in children’s library checkouts:**

### **The impact of children’s decisions**

#### **Abstract**

The first study provides evidence that the picture books which adults purchase for children predominantly feature White, male central characters. This creates a fictional universe for young children which revolves around stories about boys and men, in which girls and women are either absent or assigned to supporting roles. However, this leaves open the question of whether this ratio differs for individual children, and whether it differs between girls and boys. In this study I provide evidence that young children prefer to read books with central characters of the same sex as themselves, that this preference is vastly stronger for boys than for girls, and that the difference in preferences between girls and boys grows from 2<sup>nd</sup> grade through 5<sup>th</sup> grade.

#### **Background**

Previous research, including the first study in this dissertation, has demonstrated that children’s picture books are far more likely to have male central characters than female (among many others, McCabe, Fairchild, Grauerholz, Pescolido, & Tope, 2011; Hamilton, Anderson, Broaddus & Young, 2006). This is true at the level of book titles, and it is even truer of the books that children encounter; a huge majority of children’s picture books purchased have male central characters. The first literature that children read, which offers them an important window on the society in which they live, tell the stories of boys and men.

Children's picture books are an important subject to study, since they represent the first books that children encounter and convey an implicit endorsement from children's parents. Most children will encounter these books while reading with parents. Reading these books will be a set of formative childhood experiences. Children will listen carefully and remember these books fondly. If they become parents, they will likely return to the books they first encountered as children when selecting reading material for their own children.

The gendered composition of these books is also important because children encounter them when they are first beginning to form their ideas about gender (Martin, Ruble, & Szkrybalo, 2002; Martin & Ruble, 2004; Ruble, et al., 2007). From age one to six, children progressively build their understandings of gender. They acquire the ability to identify genders using perceptual clues. They develop the vocabulary to correctly label gender. By around three years of age, children can correctly identify their own genders. Between the ages of three and six, children come to identify gender as a fixed, immutable characteristic. This development of gender-based concepts and schema all occurs within an environment which is permeated by books telling stories about male characters instead of female characters. Given this reading environment, it is not surprising that as early as age five, children tend to ascribe assertive traits to males, and affiliative traits to females (Best & Thomas, 2004); in the books they encounter, men and boys drive the stories, while women and girls provide support.

However, only a small fraction of the books a person encounters over the course of childhood, let alone a lifetime, will be children's picture books. A far greater proportion will be self-selected. To measure the proportion of central characters that

children encounter who are female, it is important to move beyond the children's picture books that they encounter, into the books that they select for themselves.

Furthermore, the development of ideas about gender does not stop at the end of early childhood. Children continue to build their ideas about gender, gender roles, and gender stereotypes throughout middle-childhood (Katz & Ksansnak, 1994; Signorella, Bigler, & Liben, 1993). Studying the books that children read as they move through middle-childhood can help us to better understand the literary environment that children experience as their understandings of gender and gender roles change during this period.

Books read by older children are also reflective of how those children already think about gender. Young children, who rely on parents to select books for them to read, have relatively little input in the books that they encounter. However, by middle childhood, here defined as ages eight through 11, children are capable of identifying stories that they enjoy, and seeking them out. Therefore, the books that children read during this period are likely to better reflect their own interests and beliefs than the books purchased for them when they were younger. Many studies of gender in children's books are based on a concern that the way gender is portrayed in those picture books will shape how children think about gender throughout life. If this is the case, then it may be reflected in the books that children read as they age.

A large majority of the children's picture books that children encounter contain male central characters. A concern motivating this dissertation and other research into gender in children's literature is that this imbalance will cause children to see boys and men as more important than girls and women, and stories about boys and men more worthy of being told than those about girls and women. If this is the case, then it should

be expected that children choose to consume more books with male central characters once they have more control over their reading. This would also be concerning, since it could suggest that the gender imbalance in children's picture books is self-perpetuating; encountering more male central characters than female leads children to view male central characters as more important than female. This, in turn, leads them to seek out more books with male central characters than female, continuing a cycle whereby the experiences and stories of women and girls are considered less valuable and interesting than those of men and boys.

#### *Individual-level data*

A limitation to studies estimating the prevalence of male central characters in children's picture books, including the first study in this dissertation, is that they focus on the marginal distribution of books or book sales, without examining how these are distributed within readers. Studies have convincingly demonstrated that, on the whole, male central characters are more common than female central characters. However, while this is true on average, it is not known if this is true for subgroups of children. Specifically, it is unclear whether girls and boys are equally likely to be exposed to books with male central characters, or whether girls encounter more female central characters, while boys encounter almost none.

Common stereotypes assume that books with male central characters are appropriate for all children, while books with female central characters are primarily intended for girls. If parents select books using these assumptions, it is possible that the imbalance in the ratio of male to female central characters in highly popular children's books stems from the fact that parents exclusively purchase books with male central

characters for their boys but purchase more balanced books for their girls. This could explain why the books with female central characters seem so strongly sex-typed (with most central characters being princesses, fairies, or, occasionally, princess-fairies), while male central characters represent a wider range of roles. If books with female central characters are only intended to appeal to girls, while books with male central characters are meant to reach a wider audience, we might expect female central characters to be extremely feminine, while male central characters are less constrained.

Unfortunately, it is impossible to test this hypothesis with aggregate data since the gender of the child for whom any given book was purchased cannot be determined. Similarly, previous studies cannot estimate whether the proportion of central characters who are male in the books children read change as children age because they do not include individual children in their datasets.

In an ideal world, we might hope that both boys and girls would encounter books with both female and male central characters in roughly equal proportions. Perhaps children should encounter a slighter higher proportion of books with characters of their own gender because reading about children like them helps girls and boys to better recognize and understand their own experiences. Alternately, perhaps it would be better for children to encounter a slightly higher proportion of children of different genders than their own because these characters would offer them a crucial insight into the experiences of their peers who are different than themselves. Regardless, it is important to understand how the proportion of female characters encountered by children differs across children, and especially how it differs between girls and boys. However, current estimates and

approaches to estimation make it impossible to determine the proportion of central characters in children's books who are female separately for female and male readers.

*Developmental change in the quantity representation of female and male characters*

It would also be a mistake to assume that the ratio of male to female characters that children encounter is static over the course of children's development. As children develop and mature, the books that they enjoy reading change. At the same time, it is reasonable to expect that the representation of female and male characters in books that children read will also change. To begin with, there is evidence that middle childhood is a time when children's conception of gender and gender roles are becoming more and more solidified. This is thought to be especially true of boys, who maintain relatively rigid conceptions of gender relative to girls (Signorella, Bigler, & Liben, 1997; Katz & Ksansnak, 1994; Huston, 1985). Boys also tend to be more rigid in conforming to gender role stereotypes than girls (Levy, Taylor, & Gelman, 1995; Galambos, Almeida, & Peterson, 1990). One possible reason for this difference in developmental trajectories between girls and boys is that, while both girls and boys perceive pressure to behave in ways which are appropriate to their genders, masculine for boys and feminine for girls, both girls and boys recognize that boys, men, and masculinity are all considered more valuable than girls, women, and femininity (Coyle, Fulcher, & Trübutschek, 2016).

These different developmental trajectories raise the possibility that children's reading patterns develop similarly. Perhaps both girls and boys experience increasing preferences to read books about characters who are similar to them in terms of gender, but both girls and boys also recognize that stories about boys are more valued than stories about girls. This might lead boys to read books which disproportionately feature male

central characters, because these books both feature the dominant gender and feature characters who are similar to the boys themselves. In contrast, girls might select books which are more balanced in terms of central character gender, since their desire to read about characters similar to themselves is partially offset by their desire to read about boys. However, since no studies have estimated the distribution of character gender in books that children read, or how this changes as children mature, and whether the trajectories differ between girls and boys, this remains speculative.

*The importance of books for children's understanding of gender*

It would be incorrect to suggest that books are solely responsible for providing children with information about gender. Children receive messages about gender and the relative value of the genders from many different sources. Books are only one of these, and must compete with movies, music, television, and other media. Despite this, books have the potential to be uniquely influential in how children think about gender, and in particular how they think about members of the opposite gender. To begin with, reading a book entails a deeper level of interaction than watching a TV show, movie, or music video. This heightened focus is likely to cause children to be more influenced by the books that they read than by other forms of media. This is especially true since books are given a special status in society. They are privileged above other forms of entertainment, and so children may be especially receptive to their messages.

Additionally, books, unlike other forms of entertainment, give the reader direct access to the mental states of at least some of the characters. In a sense, the reader actually becomes the character when reading a book, privy to the character's thoughts and motivations. In other forms of entertainment there remains a strong divide between

the viewer and the characters; the viewer can guess at what the characters are thinking and feeling, but cannot directly observe. This is less true of children's picture books which tend not to explore the thoughts and feelings of their characters. However, the books that children in middle childhood read, for example novels by R.L. Stine and the *Magic Treehouse* series, actually make the thoughts and emotions of their characters central to the narrative, requiring children to actually understand what the central characters are thinking and why. In a sense, the reader is encouraged or required to adopt the perspective of the central characters, at least on some level.

Recent studies have found that reading fiction tends to promote empathy in the reader, which may be at least partially attributable to identification that books create between the reader and the characters (Johnson, 2012; Johnson, Cushman, Borden, & McCune, 2013; Bal & Veltkamp, 2013; Djikic, Oatley, & Moldoveanu, 2013). Thus it is reasonable to expect that reading books about male characters will help children to better empathize with and take the perspectives of boys and men. At the same time, reading books about girls and women should help children to better empathize with and take the perspectives of girls and women.

### *Books and race*

Many of the same questions and hypotheses arise when considering character race. Although less research has been done on the distribution of race in characters in children's books, the first study in this dissertation gave evidence that non-White characters are so drastically underrepresented that it is sensible to speak of non-White characters, rather than characters of specific non-White races. That is, children's picture books contain such inadequate representation of characters who are not white that it is

more reasonable to speak of the proportion of central character who are non-White, than of the proportion who are Black, Hispanic, or any other race.

As with gender, it seems possible that children of different races will tend to encounter books with characters of the same race as themselves, either because their parents purchase these books for them, or because, when they reach an appropriate age, they seek those books out for themselves. Also as with gender, it is reasonable to expect that this tendency becomes more pronounced, especially for non-White children, as they age and begin to acquire a deeper understanding of race and a stronger affiliation with their racial community (Murray & Mandara, 2002). Once again, the first study does not allow us to see the distribution of race in books encountered by individual children, and therefore makes it impossible to determine if children of different races encounter non-white characters at different rates, or if this distribution changes as children age.

These are important questions. As discussed before, representation of female characters is important, as it shows introduces children to the stories, experiences, and inner worlds of girls and women, and sends the messages that these stories, experiences, and inner worlds are valuable and worth attending to. Encountering books about non-White characters has the potential to do the same thing for children's understandings of people of different races<sup>19</sup>. It is therefore worth determining whether and how frequently children encounter non-White children in books, and whether this changes over time.

### **Research questions**

---

<sup>19</sup> It cannot be stressed enough that in an ideal world children would encounter multiple books featuring characters of all different races, and that it is inadequate to lump all characters who are not White into a single category. However, as mentioned earlier, and as will become clear in this study, in the actual world there are so few central characters that are not White that it is more reasonable to hope that the books children encounter will not *only* include White central characters.

In this study, I sought to address the following research questions:

- 1a) Do children of different genders have different probabilities of checking out library books with female central characters<sup>20</sup>?
- 1b) Does the probability of checking out a library book with a female central character differ between children, conditional on child gender?
- 1c) Does the probability of checking out a library book with a female central character change over time?
- 1d) Does the rate of change over time in the probability of checking out a library book with a female central character differ between girls and boys?
- 2a)-2d), the same, except for the probability of checking out a book with a non-White central character predicted by a child's race.

All questions are asked in the context of the sample, which includes 2<sup>nd</sup> through 5<sup>th</sup> grade students in a school district in North Carolina.

Based on previous literature on gender, I hypothesize that male central characters will outnumber female central characters; that the disparity will larger in books boys read than in books girls read, and that as boys age, they will become more likely to check out books with male central characters. I make no hypotheses regarding overall change over time, or about change in the books girls will read. Based on previous literature on race, I hypothesize that children will disproportionately read books about White central characters, but that non-White students will read fewer books with White central

---

<sup>20</sup> This paper treats gender as a binary construct, an approach which has been problematized in recent works in queer theory (e.g., Butler, 2011; Kerfoot & Knights, 1994). However, while the notion of a fixed, dichotomous gender may be problematic, it is very much the reality of the literature that children encounter, or at least the books that they encounter in the texts that they checked out from their school libraries. It is also the way that children are identified in the administrative records to which I have access.

characters than White students. I also hypothesize that over time, non-White children will read more books with non-White central characters.

## **Method**

### **Participants**

My sample consisted of all transactions made by 2<sup>nd</sup> and 3<sup>rd</sup> graders in several elementary school libraries in a single urban public school district in North Carolina made from the beginning of AY 2012/2013 until December 19<sup>th</sup>, 2014, a total of close to two and a half years. It included a total of 70,640 checkouts. Each row in the dataset listed the date of the checkout, the title, ISBN, and author of the book checked out, and a non-identifying numeric code associated with the student checking out the book. Student-level data included student grade at the beginning of the study, race, and gender. The dataset included checkouts from 1,308 unique students, and checkouts of 17,532 unique titles<sup>21</sup>.

Of the students, 51% were identified as female and the rest as male. 54% of the students were identified as Black, 37% as Hispanic<sup>22</sup>, 4% as White, and 5% as other, a group that was primarily American Indian, but also partially Asian-American. Within each racial group, the gender split was roughly equal to the split in the full sample. However among Black students, girls were overrepresented (55%), while among Hispanic students, boys were (54%). The number of White students and students of other races and ethnicities was too low for the proportions to be meaningfully estimated. A chi-squared test revealed that the differences in proportions was too large to be attributed to

---

<sup>21</sup> This may be an overestimate of the number of unique titles, as different libraries sometimes recorded books using slightly different names. For example, one library might refer to a book as “The adventures of Captain Underpants”, while another referred to the same book as “The adventures of Captain Underpants : an epic novel”. I found and fixed as many of these as I was able to, but some undoubtedly remained.

<sup>22</sup> These analyses all treat “Hispanic” as a race rather than an ethnicity.

chance ( $\chi^2(df = 3) = 10.21, p = .017$ ). Keeping in mind that only students who used the school library appear in the dataset, it is possible that Black boys are relatively less likely than Black girls to access the school library, leading to the difference in proportions.

57% of the students were 2<sup>nd</sup> graders at the beginning of the survey, while 43% were 3<sup>rd</sup> graders. Neither gender proportions nor race and ethnicity proportions appeared to differ by grade. Table 1 presents sample demographic statistics.

Although for reasons of de-identification this dataset does not include information about student SES or language, other analyses of this district suggest that students there are disproportionately lower SES, and that a large proportion of students who are identified as Hispanic are also classified as having limited English proficiency.

## **Variables**

My goal was to classify books according to the gender and race of the central character. These were the primary outcomes of this study. Because there were too many unique titles to code, I took a simple random sample of 1,000 students, and for each student sampled I took a simple random sample of up to 5 of that student's checkouts. I sampled all of a child's books if she or he had five or fewer transactions. This resulted in a dataset of 4,968 transactions completed by 1,000 students, and representing 3,518 unique book titles. See the coding sub-chapter for a description of the coding process.

Predictor variables include student gender, race, grade at the start of the study, and age at the time of a given transaction. The deidentification process stripped student ages from the dataset. However, I was able to create a proxy for age by combining the grade variable and the checkout date. I assigned an age of seven to all students who were

in 2<sup>nd</sup> grade at the start of the study, and an age of eight to all students who were in third grade. I then added the fraction of a year which had elapsed since the beginning of the study, August 27<sup>th</sup>, 2012, and the transaction. For example, suppose a student began the study in 2<sup>nd</sup> grade, and checked out a book at the end of February, 2013. This student's imputed age would be 7.5, i.e. 7 for being in second grade at the beginning of the study, and an additional 0.5 for the half of a year which had elapsed between August and February. Although this imputed age is not exact, it is a useful proxy, and differences in age within a student (i.e. the amount of time elapsed between a pair of transactions, or between the checkout and the beginning of the study) are measured exactly.

### **Analytic plan**

I began by generating simple descriptive statistics for both students and books in my sample. I then fit a series of logistic regression mixed models predicting the probability of a student checking out a book including a female central character<sup>23</sup>. To answer my research questions regarding gender, I fit the logistic mixed regression model

$$\begin{aligned} \text{logit}(\Pr(\text{BOOKGENDER}_{ij} = \text{Female or Both})) \\ = \beta_0 + \beta_1 \text{MALE}_i + \beta_2 \text{AGE}_{ij} + \beta_3 \text{MALE}_i \times \text{AGE}_{ij} + u_{0i} + u_{2i} \text{AGE}_{ij} \end{aligned}$$

where  $\text{BOOKGENDER}_{ij}$  indicates the gender of the central character of book  $j$  read by student  $i$ ,  $\text{MALE}_i$  is an indicator variable identifying whether student  $i$  is male,  $\text{AGE}_{ij}$  represents the age of child  $i$  when she or he checked out book  $j$  and has been scale so that  $\text{AGE} = 0$  corresponds to the mean age of children in the dataset, roughly 8 and a half

---

<sup>23</sup> As with race and ethnicity, I treated books with both female and male central characters as having a female central character. This was motivated by the fact that relatively few books had only a female central character.

years,  $u_{0i}$  is a random intercept offset for student  $i$ <sup>24</sup> and  $u_{2i}$  is a random slope offset for student  $i$ . Here the coefficient  $\beta_1$  represents the difference in the probability of checking out a book with a female central character between a typical boy and a typical girl at mean age,  $\beta_2$  represents the association between student age and the outcome for girls, and  $\beta_3$  is the difference in this association between girls and boys<sup>25</sup>. The variance of the random intercept offset roughly represents the student-level variability in the probability of accessing books with female central characters, conditional on student gender.

I then fit a parallel model for the probability of a book having a White central character, and replacing student gender with student race, with values of Black, Hispanic, White, and other, as a predictor.

## Results

I begin with basic descriptive analyses before turning to the results of the models.

Of the books checked out, 76% had a central character with an identifiable gender. The majority of the 24% that did not were science books or technical books teaching children to draw, an extremely popular category. A full 58% of the checkouts with central characters with an identifiable gender had only a male central character. A further 34% had only a female central character, and 8% had both female and male central characters. 60% of books checked out had a central character of an identifiable race. Of these, 81% had only a White central character, 10% had only Black central

---

<sup>24</sup> I fit all models using R, version 3.2.3 (R Core Team, 2015). I fit the logistic regression models using functions provided by the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). When models failed to converge using lme4, I used functions provided by the ordinal package (Christensen, 2015) to fit similar models. For all model parameters, I present confidence intervals obtained by profile maximum likelihood methods. P-values are obtained using likelihood-ratio tests.

<sup>25</sup> I also fit and present (in a table) preliminary models which include only gender or only the main effects of gender and age, but I focus my discussion on the final model.

characters, 2% had only Latino/a central characters, 1% each had only Asian/Asian-American or an American-Indian central characters, and 8% had central characters of multiple races<sup>26</sup>. For both books with White central characters and books with non-White central characters, 56% of books had only a male central character.

Boys tended to read books with male central characters. 77% of check-outs to boys had a male central character. Only 15% had a female central character and 8% had both female and male central characters. Girls were far more balanced in their selections. 51% of the checkouts to girls had a female central character, 42% had a male central character, and 7% had both female and male central characters. Framed slightly differently, 49% of the books girls checked out had at least one male central character, while only 23% of the books boys checked out had any female central characters. Table 2 shows these sample statistics. A plurality of boys did not check out a single book with a female central character. Figure 1 shows histograms of the proportion of books checked out by individual students with at least one female central character and the proportion with at least one male central character, broken out by child gender.

The main race of the character, conditional on the race of the student, was slightly more complex. For Latino/a, White, and other non-Black students, around 85% of the checkouts with a central character of an identifiable race had a White central character, with around 5% Black and 5% central characters of multiple races. Hispanic students were slightly more likely than others to check out books with Latino/a characters, but only 3% of their checkouts had Latino/a characters. For black students, 78% of the

---

<sup>26</sup> Most of these books were about sports teams or players, and focused on Black and White athletes. This is also true of many of the books with non-White characters. Restricting the sample to fiction books, 87% of characters are White.

checkouts with a central character of an identifiable race had a White central character, 13% had a Black central character, and 6% had central characters of multiple different races. Although checkouts by Black students were overwhelmingly White, they were less so than for students of other races. In a model regressing an indicator variable for whether a central character was Black on an indicator for whether a student was Black with a random student intercept, the race coefficient was positive and statistically significant ( $\hat{\beta}_{BLACK} = 1.04, p < .001$ ).

The first fitted model was

$$\begin{aligned} \text{logit}(\text{Pr}(\text{BOOKGENDER}_{ij} = \textit{Female or Both})) \\ = 0.33 - 1.63\textit{MALE}_i + 0.10\textit{AGE}_{ij} - 0.24\textit{MALE}_i \times \textit{AGE}_{ij} \end{aligned}$$

The intercept estimates the log-odds in favor of a typical girl of mean grade checking out a book with at least one female central character. Girls of mean age are slightly more likely than not to check out books with female central characters. The coefficient of *MALE* estimates the difference in log-odds of checking out a book with a female central character between boys and girls of mean age; at this age, boys are substantially less likely to check out books with female central characters than girls are, and far less likely than not to check out a book with any female central characters. The coefficient of age indicates the change in log-odds of the probability of checking out a book with a female central character associated with a change in age of one year for girls in the sample. For girls, an additional year of age is associated with a slightly higher probability of checking out a book with a female central character. Finally, the coefficient of the age-by-gender interaction represents the difference in the association between age and the probability of checking out a book with a female central character between boys

and girls. For boys, each additional year of age is associated with a lower probability of checking out a book with at least one female central character. Table 3 shows the fitted model, as well as models including only gender and only the main effects of gender and age as predictors. Table 4 shows parallel models predicting the probability of checking out a book with at least one male central character; they are similar to the models described in model 3, except that books with both male and female central characters are counted as books with male central characters. Figure 2 displays the model predicted probabilities of checking out a book with at least one female central character for a typical girl or boy of various ages. Figure 3 displays the model predicted probabilities of checking out a book with at least one male central character. The gendered divergence in preference for books with female central characters is clear from the plots; by 5<sup>th</sup> grade there is almost no overlap in the proportions of books with female central characters that girls and boys check out.

Turning to race, the fitted model is

$$\begin{aligned} \text{logit}(\text{Pr}(BOOKRACE_{ij} = White)) \\ &= 1.42 + 0.65HISPANIC_i + 0.88WHITE_i + 0.27OTHER_i + 0.13AGE_{ij} \\ &+ 0.00HISPANIC_i \times AGE_{ij} + 0.12WHITE_i \times AGE_{ij} \\ &- 0.03OTHER_i \times AGE_{ij} \end{aligned}$$

where  $HISPANIC_i$ ,  $WHITE_i$ , and  $OTHER_i$  are indicator variables for student race<sup>27</sup>, age is defined as before, and the interaction is similar. From this model, there is evidence that Hispanic and White students of mean age read more books with White central characters

---

<sup>27</sup> This model treats Black students as the reference category. This decision was made because Black students make up a slight plurality of the sample, and because initial analyses suggested that Black students are the least likely to read books with White central characters.

than do Black students of mean age. There is no evidence of an association between student age and the probability of checking out a book with a White central character, although a simpler model with no race by age interactions finds a weak positive association. Table 5 presents the final model as well as simpler models with only student race or the main effects of age and race as predictors. Table 6 presents parallel models predicting the probability of checking out books with Black central characters, which decreases over time for students of all races.

## **Discussion**

This study extended results from previous literature which had found that children's picture books tend to overrepresent White male central characters. Using a longitudinal dataset containing children's library book checkouts over a period of two-and-a-half years, it provides evidence that children check out more books with male central characters than female central characters, that this tendency is stronger for boys than for girls who check out books with an approximate balance between male and female central characters, and that the tendency to check out characters of the same gender as the reader tends to grow as children age. All three research hypotheses regarding gender were confirmed. There is also evidence that children preferentially check out books with White central characters, and that this preference is especially strong for White students. However, there was no evidence that Black and Hispanic children checked out fewer books with White central characters over time.

These results are consistent with previous estimates, which showed that male central characters outnumber female central characters in children's picture books, with theories of gender which suggest that reading cross-gender books is more threatening for

boys than for girls (Dutro, 2001; Dutro, 2002), and with developmental studies which have demonstrated that boys develop more rigid concepts of gender-appropriate behavior than girls (Katz & Ksansnak, 1994). However, it also extends previous studies on children's picture books by directly estimating the ratio of male to female central characters within children, by estimating changes in this ratio over time, and by using more plausible measures of children's reading.

*Child gender and character gender*

The first hypothesis, that children would be more likely to check out books with male central characters than female, was confirmed. This result extends previous estimates.

The ratio of male to female central characters estimated in this study, approximately 1.7:1, is consistent with previous estimates, though far lower than the ratio calculated in the first study. The reasons for this discrepancy are unclear. The first study used national data while the second study only uses data from a single school district in a single state; perhaps children in North Carolina are different from children around the country.

Additionally, while the first sample draws on children's picture books which are presumably purchased by parents and read to young children, the second study uses books selected by grade school children; perhaps children of this age select different books than the parents of young children select for them. Finally, the first study relies on data which are only loosely connected to the actual reading behavior, and are restricted to only the most popular books sold in the country; perhaps the difference in the ratios estimated is due to methodological flaws in the first study, which are corrected by the second study.

Also as hypothesized, boys were less likely than girls to read books with female characters. A full 77% of the books boys checked out had only male central characters. In a sense, this is an understatement of the preference boys showed to books about male characters: only 14% of the books boys checked out had *only* female central characters, while 8% had both male and female central characters. Previous research has documented a tendency on the part of boys to identify male characters as central to a text, even if those characters are “objectively” peripheral (Segel, 1986; Davies, 2003). Therefore, in their own perceptions, the books boys read may be even more masculine than this analysis suggests; boys might have identified the male characters in the books with both male and female central characters as the true central characters, while identifying the female central characters as peripheral. In contrast, girls’ selections are much more balanced. Although girls read more books with female central characters than male, 50% of books checked out by girls had at least one male central character, and a full 41% had only a male central character. It seems quite clear that girls are more willing to read books about boys than boys are to read books about girls.

Students’ preferences are not monolithic, and there is a great deal of variability in the rates at which girls and boys encounter books about characters of different genders. However, the gender gap is extremely large relative to these individual differences, and even individual boys are extremely unlikely to select books which are balanced with respect to central character gender. Figure 4 displays a histogram of the model implied probabilities of checking out books with male or female central characters, broken down by student gender. It is striking that none of the more than 450 boys in the sample is estimated to have even a 50% chance of selecting a book with any female central

characters, while all boys are predicted to have over a 70% chance of selecting a book with at least one male central character. In contrast, most girls have close to a 50% chance of selecting a book with a male central character, and some girls even have up to a 60% chance of selecting a book without a single female central character.

These results are consistent with those reported by Dutro (2001; 2002; see also McCreary 1994) who observed that 5<sup>th</sup> grade boys were extremely reluctant to be observed reading books about girls, while girls of the same age were much more comfortable being observed reading books about boys. Dutro develops the concept of a gendered boundary which is permeable to girls, but less so for boys; crossing that boundary exposes a boy to ridicule and loss of stature for embracing something feminine. This is an important reminder that gender not only divides children, it also helps to rank them. Crossing the gender boundary is difficult for boys because masculinity is prized and femininity is devalued (Coyle, Fulcher, & Trübutschek, 2016). It is easier for girls because in reading books about boys, they are conforming to societal norms that boys are more important and worth reading about. Boys who read about girls are doing the opposite, and doing so is more dangerous for them. Supporting this hypothesis, Dutro also finds that boys whose status and masculinity are already secure, such as star athletes, are more willing to cross the gendered boundary than boys whose statuses are less secure. This dataset lacked covariate information on the social statuses of children checking out library books, but future quantitative work could consider this question.

The finding is also consistent with work on the development of gender, which finds that boys tend to form more rigid conceptions of gender than girls, and to conform more closely to stereotypes about gender (Signorella, Bigler, & Liben, 1997; Katz &

Ksansnak, 1994; Huston, 1985). Girls in this sample are willing to treat gender as flexible by reading stories which focus on the experiences of boys. In contrast, the majority of boys only, or almost only, read books with male central characters.

### *Gender and age*

The differences in the rates at which boys and girls check out books with male and female central characters are not static. As they age, boys become even less likely to check out books with female central characters, while girls become more likely. Figure 2 displays this trend. At a time when girls and boys are beginning to form romantic attachments, most frequently with members of the opposite sex, they are also checking out fewer and fewer books which actually depict members of the opposite sex. This may be harmful to children, as one benefit of reading, and particularly of reading fiction, is thought to be the promotion of empathy (Johnson, 2012; Johnson, Cushman, Borden, & McCune, 2013; Bal & Veltkamp, 2013; Djikic, Oatley, & Moldoveanu, 2013); if children are reading less about members of the opposite sex, they may also be developing less empathy for members of the opposite sex. At the very least, it signals a troubling lack of interest in the perspectives and stories of members of the opposite sex on the part of boys.

### *Reading and the development of empathy*

A great deal of scholarship has been concerned with the importance of children, especially children from marginalized groups, seeing themselves reflected in cultural artifacts that they encounter. It is important for girls and boys to encounter characters in whom they can see themselves (Disch, 2016; Pho, Chorn, Rustad, Kwaymullina, & Hartness, 2015; Kidd 2015), including characters of the same gender as they are. Seeing themselves in cultural artifacts assures girls that their stories and experiences are valuable

and worthy of attention. They convey to girls that they have a place in books. Insofar as library checkouts represent the books that children read, it appears that most children in this school district encountered children of the same gender as they are in the books that they read<sup>28</sup>.

Perhaps equally important, though, is that children see those who are *not* like them reflected in the cultural artifacts they encounter. If books are a safe, but important way to explore the world and expand their perspectives (Fox, 1993), then it is important for children to encounter works which allow them to explore perspectives which are different from their own. Reading books gives children direct access to the thoughts of the characters, especially the central characters. It encourages them to understand and engage with those thoughts, and to recognize the thinker as important. Girls and boys who read stories centered on the experiences of members of the opposite sex are likely to have an easier time taking each other's perspectives.

In this dataset, girls are far more likely than boys to encounter members of the opposite sex in their reading. Couple this with the disproportionate number of male central characters that girls have probably encountered in children's picture books, not to mention the many movies and television shows centered on male central characters, and it seems clear that most girls will have had a great deal of experience with male central characters. The books they encounter in middle childhood will extend those experiences

---

<sup>28</sup> The same is not true of other characteristics of children. Although only 4% of students were identified as White, in the checked out books 81% of the central characters with an identifiable race were White, and only 2% were Latino/a despite 37% of students in the sample being Hispanic. No central characters were involved in same-sex relationships in the texts of the books although a number of central characters were involved in romantic relationships with members of the opposite sex. Most, if not all, central characters lived in stable two-parent families with married opposite-sex parents. There are no data available which speak to the proportion of children in this sample who live in households headed by a single parent, but it seems likely that this is true of at least some families. Clearly there are important aspects of many children's lives which are not represented in the books they check out from the library.

by requiring them to identify with those male central characters and to take their perspectives. Girls will be required to engage directly with the thoughts, perspectives, and experiences of boys and men. In contrast, many boys will not encounter female central characters. In addition to a dearth of female central characters in children's picture books, as well as movies and television, many boys will never or almost never check out books with female central characters. The literature these boys read will not require them to engage with the thoughts, perspectives, and experiences of girls or women. This seems likely to hamper the development of empathy for girls and the ability to take their perspectives.

### *Race*

As hypothesized and consistent with the first study, children were far more likely to check out books with White central characters than non-White. The ratio of White to non-White central characters observed in this study is roughly 4:1, which is much lower than the ratio estimated in the first study of close to 40:1.

### *Limitations*

My goal has been to estimate the proportion of books with female and male central characters, and with White and non-White central characters, that children read.

However, these data fail to estimate that quantity in two ways.

The first problem is that these data can only be used to estimate the proportion of female and male characters in books that children check out. These are not necessarily the same as the books that they read. For one thing, children may not read all of the books that they check out. If boys disproportionately fail to read books that they check out which have male central characters, then my analyses may overstate the extent to which

boys favor books with male central characters over books with female central characters. More likely, if boys mistakenly check out certain books believing that they have male central characters and then fail to read them when they realize that this is not the case, my results may be too modest.

Additionally, children have access to books other than the ones they check out from the school library, and school library books may only represent a small fraction of the books that children read. Boys may have books at home, or books that they check out from other libraries, with female central characters. If these books have more female central characters than the ones they check out from the school library, my analyses may overstate these boys' preference for books with male central characters. This is a very real possibility. In an examination of the behavior of girls and boys in a classroom that she studied, Dutro (2001; 2002) demonstrated that one way in which 5<sup>th</sup> grade boys perform masculinity is by loudly and publicly disparaging books with female central characters. She interpreted these performances as evidence not that boys actually disliked reading books about girls, but observed that the devaluation of the feminine helped the boys to establish their own masculinity more firmly. Selecting a book from a school library is a public action which could expose a student to ridicule if the selection is judged inappropriate by his peers. This might explain how reluctant boys in this sample were to select books with female central characters. It is less certain how these boys would behave in a more private setting, and it is possible that they have access to, and choose to read, books with female central character at home.

The second issue is that children in this sample come from a single school district and may not be representative of other children in the US. This limitation is common to

much social science research, but should always be highlighted. This is especially true when doing research on gender; attitudes towards gender vary widely across the country. The way that Black and Latino/a/x children in this particular district in North Carolina think about reading books with female characters may be different from how other children in the country do.

There are two additional limitations which do not directly threaten the estimation of the proportion of interest, but do change the interpretation that I give to that proportion. To begin with, it is important to note that I have no data for the proportion of books in the library as a whole which have male or female central characters. In this study, I write about the preference which boys show for books with male central characters. However, the way that I understand this preference depends intimately on the ratio of male to female central characters in books in the library holding. For example, if 80% of books in the library include a male central character, it may be that boys are simply picking books off of the shelves at random without thought to the genders of those characters. This would also suggest that girls were going out of their ways to find books which were approximately balanced in terms of character gender.

This situation would still be somewhat problematic. Even if boys were not intentionally sexist in their book selections, they would be confronted with an extremely sexist collection of library books; they would not be exposed to stories which centered on the experiences of girls and women. However, this would have different ramifications for our understandings of how that imbalance came to be, and for how it could be addressed. In this case, it would seem imperative to ensure that library collections had as many books with female central characters as male, which might naturally remedy the

imbalance in the books that children actually encountered. In contrast, if library collections are more balanced and boys are seeking out books with male central characters, it would be more important to find ways to persuade boys to be interested in stories about girls, and to feel comfortable in doing so. It seems most likely that the truth is somewhere in between: the libraries most likely have more male central characters than female, and both boys and girls most likely selectively check out texts with central characters of their own genders.

Finally, some significant proportion of books that children checked out may have been assigned by their teachers or otherwise required by some other external agency. It could be that the ratios of interest are different in the subset of books that children selected for themselves. Once again, this would not change the proportion being estimated and would still be problematic, because regardless of the reason, boys continue to encounter far more stories about themselves than about girls. However, it would once again change our understanding of how the imbalance came to be, and of how to address it.

It should be noted that this explanation cannot account for most of the imbalance, since girls and boys are enrolled in the same classes and are presumably being assigned substantially similar reading material. Assuming that teachers assign similar reading to male and female students, removing the teacher-assigned books would actually increase the disparity between girls and boys in child-selected books.

## **Conclusion**

Reading is a mechanism by which children and adults can be encouraged to understand the perspectives of those who are not like them, including members of the

opposite gender. For most girls, it appears that the books they check out are at least exposing them to the experiences of boys, who are the main characters of a large proportion of the books that girls check out from their school library. In contrast, most boys check out almost no books with female central characters, meaning that boys get very little exposure to the experiences of girls in their reading. Boys and men frequently complain that girls and women are difficult for them to understand, but a good first step to understanding might be to actively seek out the perspectives of girls and women which are presented in books.

Table 1: Demographic characteristics of students (n = 982) included in the sample

<u>Category</u>		<u>Percent</u>
<i>Gender</i>	<i>Female</i>	51%
	<i>Male</i>	49%
<i>Race</i>	<i>Black</i>	39%
	<i>Latinx</i>	54%
	<i>White</i>	3%
	<i>Other</i>	4%
<i>Grade (at the beginning of the study)</i>	<i>2<sup>nd</sup></i>	57%
	<i>3<sup>rd</sup></i>	43%

Table 2: Proportions of books checked out with central characters of given genders

	<u>Female only</u>	<u>Male only</u>	<u>Male and female</u>
<i>Boys</i>	14%	77%	8%
<i>Girls</i>	51%	42%	7%

*Table 3: A taxonomy of models predicting the log-odds ratio of a child checking out a book with at least one female central character. In these models age has been scaled so that the mean is 0 and the standard deviation is 1. The actual mean age is around 8.5 years old, while the actual standard deviation is around 10 months, or almost one year.*

	<b>Gender only</b>	<b>Gender and age</b>	<b>Gender interacting with age</b>
Male	-1.68 <sup>***</sup>	-1.66 <sup>***</sup>	-1.68 <sup>***</sup>
	(0.09)	(0.09)	(0.09)
Age (scaled)		0.03	0.14 <sup>**</sup>
		(0.04)	(0.05)
MaleXAge (scaled)			-0.30 <sup>**</sup>
			(0.09)
Log Likelihood	-2290.28	-2288.69	-2283.31
AIC	4586.57	4589.38	4580.63
BIC	4605.27	4626.79	4624.26
Num. obs.	3766	3765	3765
Groups (studentid)	982	982	982
Variance: studentid: (Intercept)	0.41	0.40	0.39
Variance: studentid: scale_age		0.01	0.03

\*\*\* p < 0.001, \*\* p < 0.01, \* p < 0.05

*Table 4: A taxonomy of models predicting the log-odds ratio of a child checking out a book with at least one male central character. In these models age has been scaled so that the mean is 0 and the standard deviation is 1. The actual mean age is around 8.5 years old, while the actual standard deviation is around 10 months, or almost one year.*

	<b>Gender only</b>	<b>Gender and age</b>	<b>Gender interacting with age</b>
Male	1.94 <sup>***</sup>	1.91 <sup>***</sup>	1.93 <sup>***</sup>
	(0.10)	(0.10)	(0.10)
Age (scaled)		-0.04	-0.13 <sup>*</sup>
		(0.04)	(0.05)
MaleXAge (scaled)			0.34 <sup>***</sup>
			(0.10)
Log Likelihood	-2104.99	-2102.90	-2097.27
AIC	4215.98	4217.80	4208.55
BIC	4234.68	4255.20	4252.18
Num. obs.	3766	3765	3765
Groups (studentid)	982	982	982
Variance: studentid: (Intercept)	0.38	0.37	0.35
Variance: studentid: scale_age		0.01	0.05

\*\*\* p < 0.001, \*\* p < 0.01, \* p < 0.05

Statistical models

Table 5: A taxonomy of models predicting the log-odds ratio of a child checking out a book with at least one White central character. In these models age has been scaled so that the mean is 0 and the standard deviation is 1. The actual mean age is around 8.5 years old, while the actual standard deviation is around 10 months, or almost one year.

	<b>Race only</b>	<b>Race and age</b>	<b>Race interacting with age</b>
Hispanic	0.65*** (0.13)	0.65*** (0.13)	0.65*** (0.13)
White	0.80* (0.35)	0.83* (0.35)	0.88* (0.38)
Other	0.27 (0.29)	0.27 (0.29)	0.27 (0.29)
Age (scaled)		0.15* (0.08)	0.15 (0.09)
HispanicXAge (scaled)			0.00 (0.12)
WhiteXAge (scaled)			0.12 (0.39)
OtherXAge (scaled)			-0.03 (0.30)
Log Likelihood	-1393.92	-1391.31	-1391.26
AIC	2797.84	2798.63	2804.52
BIC	2827.81	2846.59	2870.46
Num. obs.	2966	2965	2965
Groups (studentid)	954	954	954
Variance: studentid: (Intercept)	0.64	0.64	0.64
Variance: studentid: scale_age		0.02	0.01

\*\*\* p < 0.001, \*\* p < 0.01, \* p < 0.05

*Table 6: A taxonomy of models predicting the log-odds ratio of a child checking out a book with at least one Black central character. In these models age has been scaled so that the mean is 0 and the standard deviation is 1. The actual mean age is around 8.5 years old, while the actual standard deviation is around 10 months, or almost one year.*

	<b>Race only</b>	<b>Race and age</b>	<b>Race interacting with age</b>
Hispanic	-1.05***	-1.08***	-1.07***
	(0.19)	(0.19)	(0.19)
White	-0.82	-0.85	-0.73
	(0.47)	(0.46)	(0.49)
Other	-0.57	-0.58	-0.54
	(0.41)	(0.41)	(0.42)
Age (scaled)		-0.28*	-0.33*
		(0.14)	(0.15)
HispanicXAge (scaled)			0.16
			(0.19)
WhiteXAge (scaled)			0.35
			(0.53)
OtherXAge (scaled)			0.32
			(0.44)
Log Likelihood	-896.92	-893.66	-892.90
AIC	1803.84	1803.32	1807.81
BIC	1833.82	1851.28	1873.75
Num. obs.	2966	2965	2965
Groups (studentid)	954	954	954
Variance: studentid: (Intercept)	1.14	1.20	1.19
Variance: studentid: scale_age		0.21	0.19

\*\*\* p < 0.001, \*\* p < 0.01, \* p < 0.05

Figure 1: Observed distribution of central character gender in books selected by children in the sample, broken out by child gender. Note that a child checking out five books with central characters with identifiable genders must have a proportion of 0, .2, .4, .6, or .8, which leads to some chunkiness in the histograms.

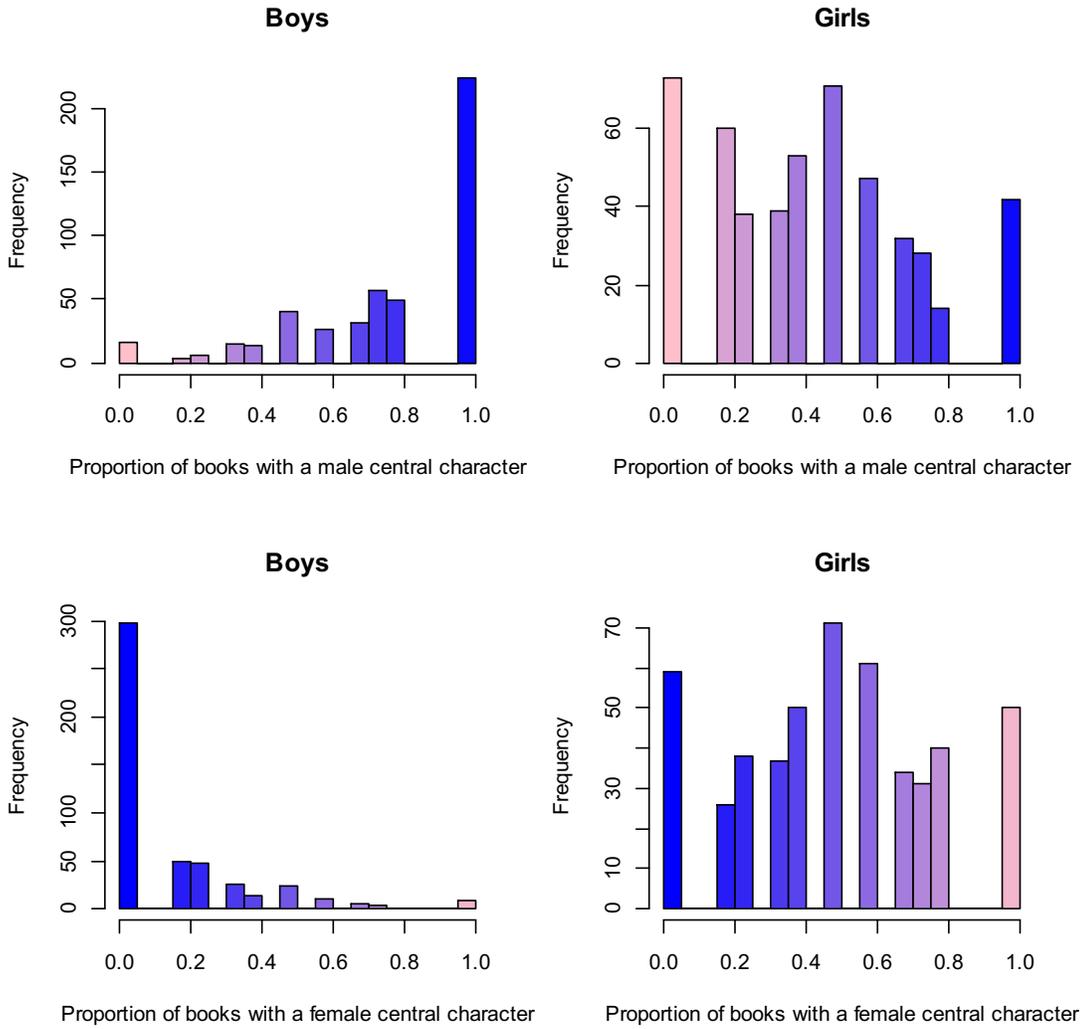


Figure 2: Model implied probability of a child checking out a book with at least one female central character, conditional on child age and gender. The dashed lines indicate model implied probabilities for children with random intercepts at the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles of the empirical distribution.

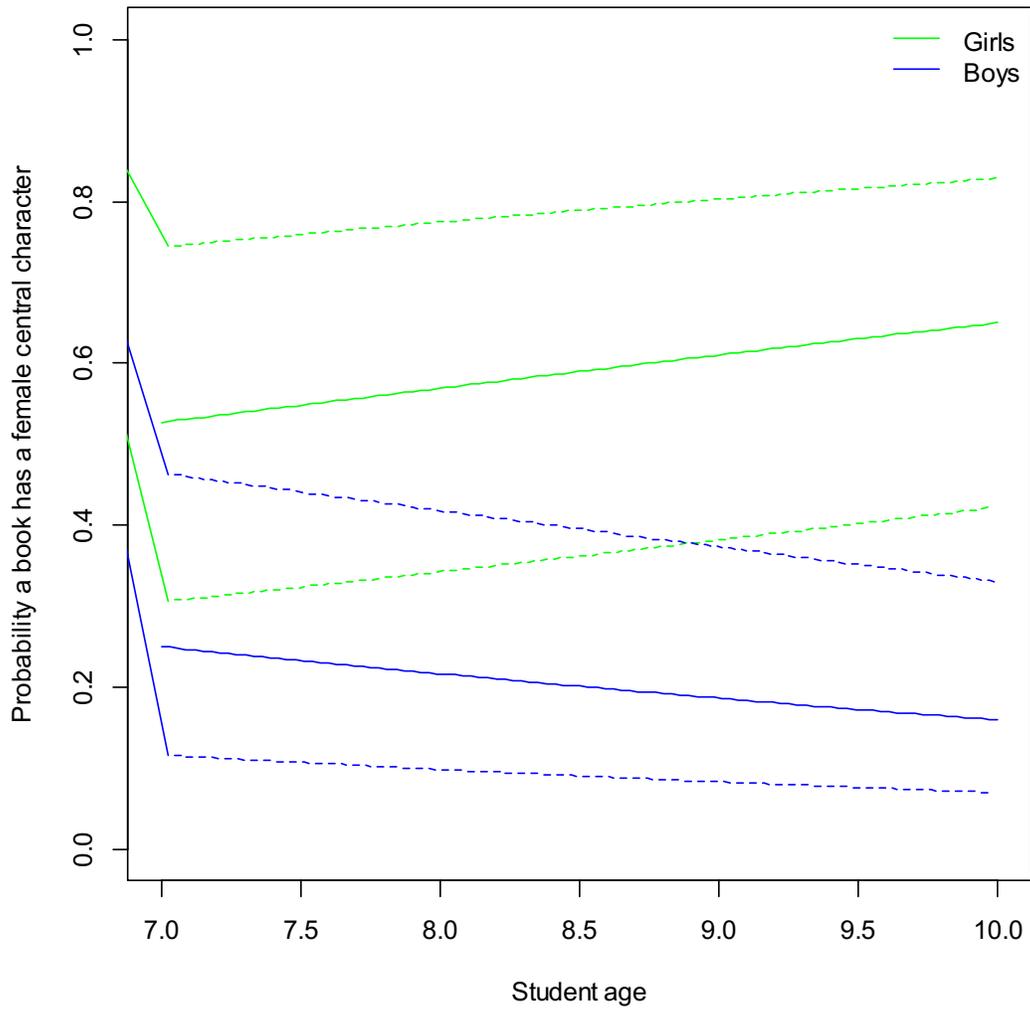


Figure 3: Model implied probability of a child checking out a book with at least one male central character, conditional on child age and gender. The dashed lines indicate model implied probabilities for children with random intercepts at the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles of the empirical distribution.

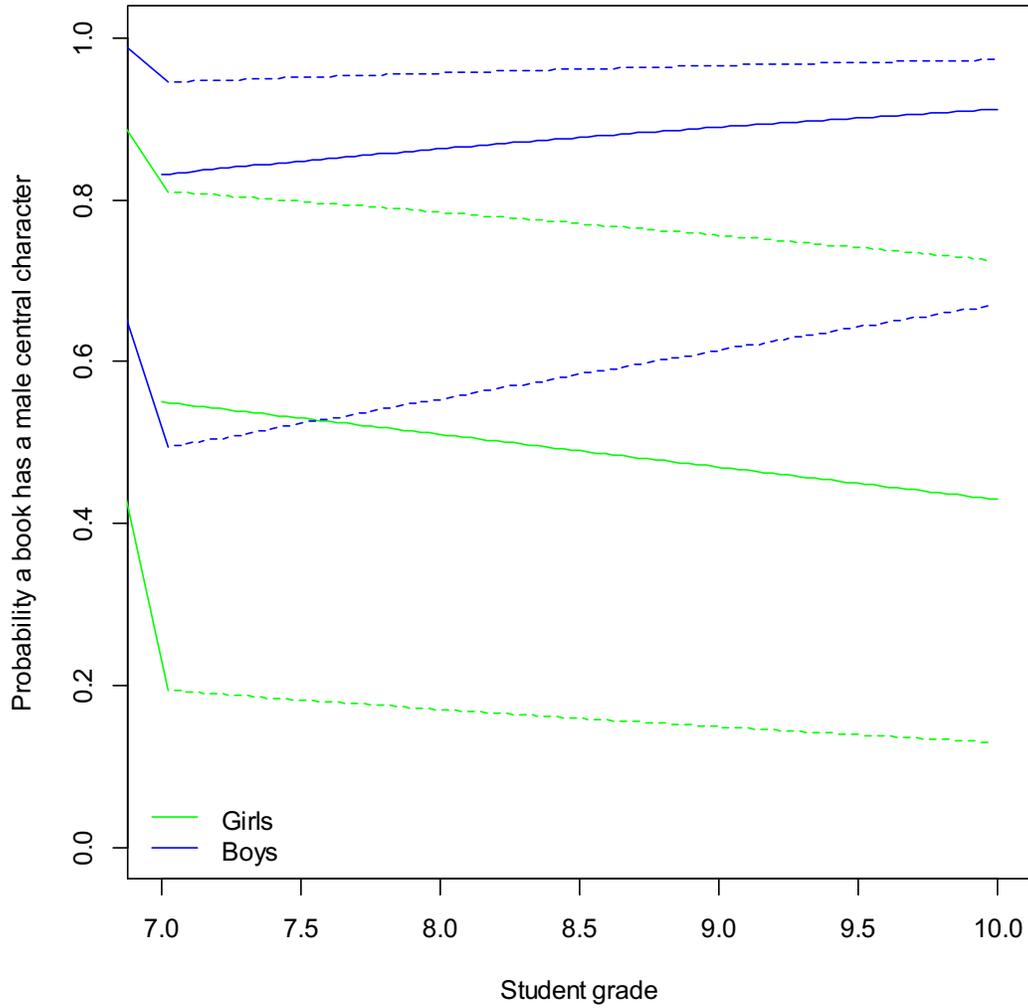
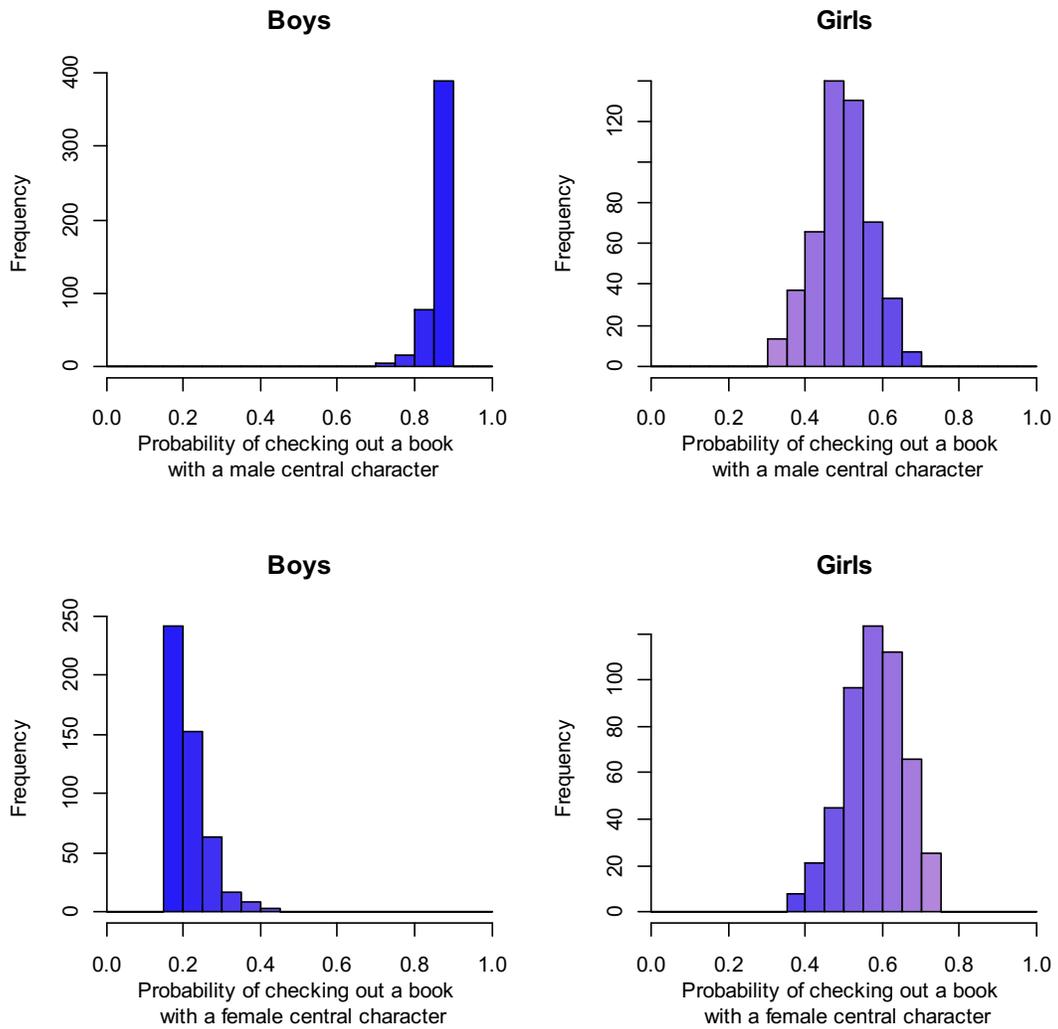


Figure 4: Model estimated probabilities of selecting books by central character gender by children in the sample, broken out by child gender.



## **Chapter 4 – Gender in assigned reading: The impact of teachers’ decisions**

### **Abstract**

Theorists have developed the concept of a “selective tradition”, which privileges the perspectives and stories of members of dominant social groups. Empirical research has demonstrated the existence of the selective tradition in children’s literature. Children’s books which are identified as worthy of esteem tend to tell stories about White male characters. Studies have shown that teachers tend to use books written about White male characters in their classrooms. The current study builds on previous work by considering how teacher book selections can increase or decrease the representation of socially dominant groups in books that children encounter. It improves on previous studies by examining teacher decisions in an ecologically valid setting and considering how teacher decisions about which books to highlight vary with child gender. I find mixed evidence of the maintenance of a selective tradition via teacher book selections. On the one hand, teachers do little to privilege books about White male characters; on the other hand, they do not act to ensure that children’s books represent members of non-dominant groups. I also show that teacher decisions tend to increase the representation of male characters for boys, while increasing the representation of female characters for girls.

Books are an important medium through which children learn about their society and what their society values. In addition to the explicit messages about values that books contain, children can draw inferences about the sorts of things and types of people which

are considered important by observing what tends to be written about. This tendency to privilege stories about socially dominant groups, in the current studies defined as people who are White and male<sup>29</sup>, is referred to as a “selective tradition” (Taxel, 1981). The selective tradition identifies groups which are favored by society as more legitimate subjects of stories than members of less favored groups. It also privileges the authorial voices of members of socially dominant groups, considering books written by White male authors as more valuable than books written by members of other groups. The current studies do not consider this aspect of the selective tradition, focusing on representation in characters, rather than representation in authors. A substantial minority of Americans are non-White, and a majority are female, but the selective tradition treats their stories as less important than stories about White male characters.

This is subtly different from claiming that books overrepresent central characters as White and male. The selective tradition highlights not simply overrepresentation, but overrepresentation in books that “matter”. Children’s literature is filled with books which are published, read by a handful of children, and then disappear without making any mark on the wider culture. In contrast, a handful of books, such as *Goodnight Moon*, or *The Cat in the Hat*, persist for much longer periods of time and become hugely influential. These texts are encountered by generations of children and their parents. They become canonical instances of children’s literature, defining what children’s literature is, and what it should aspire to be. They are made into movies and reprinted again and again.

---

<sup>29</sup> Social dominance is associated with characteristics other than race and gender. Heterosexuals, wealthy people, and Christians are all dominant groups in American society. However, character race and gender are fairly easy to decode in children’s books, while sexual orientation, socioeconomic status, and religion are far harder to identify. Future studies might look at the representation of other privileged groups in children’s literature.

Children are likely to remember these books as they age, because of the books' outsize influence and longevity.

To a certain extent, samples of popular books, as in the first study of this dissertation, capture books that are identified as canonical. Many of the most popular books are classics of children's literature which have an outsize influence on the wider society. On the other hand, it also included many books which were popular in the moment, but may not ultimately enter the canon of children's literature. For example, books from the *Fancy Nancy* and *Pinkalicious* series make several appearances in the list of the most popular children's books, especially near holidays. *Fancy Nancy* and *Pinkalicious* books both entered the list of popular children's books near Valentine's day with Valentine's Day themed offerings. However, it is much less clear that these texts have entered the canon of children's literature. It seems much more likely that they will quickly disappear from the lists of popular books, never to return.

Although this is the first chapter in which I explicitly invoke the term "selective tradition", the first study can easily be framed as an examination of the role of the selective tradition in purchases of children's picture books. The second study can be viewed as providing evidence that the selective tradition perpetuates itself in the books that boys check out from school libraries. In the third study, however, I explicitly examine how teacher decisions can work to maintain the selective tradition by exposing children to books which predominantly feature White male central characters and exclude female and non-White central characters.

It is important to note that the overrepresentation of socially dominant groups which is characteristic of the selective tradition, is not a feature of any particular book.

Although it is possible for individual texts to represent female and non-White characters in ways that are harmful, the harms of the selective tradition are more subtle. The selective tradition operates when the majority of books a child encounters, or the majority of books which are held up as especially valuable and worthy of attention, feature members of socially dominant groups. There is nothing wrong with the fact that certain books tell stories about White male characters. Indeed, it would be harmful if there were no books with White male central characters. The harm comes when a large majority of important books feature these characters.

### **Background**

Two previous studies have claimed that teachers play a role in perpetuating the selective tradition. In the first, Luke, Cooke, and Luke (1986) asked 54 preservice primary school teachers in Queensland Australia to select a book that they liked and believed that primary school children would like and benefit from. They found that 74% of the books identified featured male central characters, while only 10% featured female central characters. Additionally, they found that only slightly under 4% of the selections featured non-White central characters, both of whom were male. Interestingly, the only teachers who identified books

This study provided powerful evidence that the characters whose stories were told by the books that these preservice teachers considered interesting and appropriate for use in the classroom were almost exclusively White, and predominantly male. It is easy to imagine that the preferences these teachers showed for books about White, male characters could lead them to foreground these works for their students. Teachers might

be more likely to use these books in class, to assign them to children, or to recommend them to children looking for books.

However, the study was focused on the perception of preservice teachers and the need for training in developing inclusive reading lists, and did not demonstrate that teacher preferences actually resulted in children being exposed to books which predominantly featured White, male characters. This problem is especially acute since teacher preferences for books with White, male characters might change after experience teaching. It is also unclear if teacher's actual behaviors would mirror their responses to a thought-experiment. It could be that when teachers are asked to identify a good book they immediately think of books with White, male characters, but when they need to actually select books for use in the classroom or to give to children their selections are different.

In the second study, Jipson and Paley (1991) attempted to move beyond these limitations. They asked 55 elementary school teachers from Massachusetts, Wisconsin, and Oregon to identify three books that they had used in their classrooms in the past year. They found that 65% of the books identified featured male central characters, while only 6% featured characters from "North American minority cultures"; it is unclear if other texts included other non-White central characters, though arguably it is more important for children to be exposed to characters who represent the diversity in their own societies rather than the diversity which exists around the world. Only a handful of teachers explicitly cited the gender or race of the central character as a reason for assigning books and mostly to explain the inclusion of a book with a female or non-White protagonist, suggesting that most teachers did not see the inclusion of primarily White, male stories as requiring a gender- or race-based explanation. Stories about non-White and female

characters are justified because of the representation they give to certain groups; the far more common stories about White, male characters are justified on non-racial, non-gendered grounds.

Jipson and Paley (1991) demonstrated that teachers were actually more likely to report using books with White, male central characters in their classrooms. Teachers not only talk the talk by being more likely to identify these books as useful for their classes, they walk the walk by being more likely to use them. One potential limitation in the inferences supported by the study is that teachers were asked to report any three books that they had used in the past year. Since teachers presumably used more than three books on average, this means that their results capture not only teacher tendencies to use books with White, male central characters, they also capture teacher tendencies to report using these books. It is possible, for example, that teachers are more likely to report the books that they have used which have non-White or female central characters, then the books that they use in class may be even more dominated by white, Male characters than the books that they report having used.

#### *Open questions*

Although these studies demonstrate that teachers in two populations in the late 1980's and early 1990's privileged stories about White, male characters it is reasonable to wonder if this is still the case. Over the intervening 25 years, awareness of the importance of proportional representation has grown. Feminist and critical race theoretical critiques of underrepresentation of female and non-White characters in the media have led to a greater appreciation of the harms of underrepresentation of socially disadvantaged

groups. *Frozen* was a huge success despite featuring two female central characters<sup>30</sup>.

Television shows such as *Scandal* and *How to Get Away with Murder* feature Black women as their protagonists. Video games have begun to include more female characters in roles beyond sex objects<sup>31</sup>. Given the progress which has been made in other areas, it is reasonable to wonder if teachers are more likely to value books with non-White or female central characters.

On the other hand, the National Educator's Association's (NEA's) list of the top 100 children's books, based on a survey administered to NEA members, includes substantially more books with male central characters than female, and far, far more books with White central characters than non-White. One of the key reasons for this disparity appears to be a large number of older books which feature White, male central characters, such as *The Hobbit*<sup>32</sup>, *The Cat in the Hat*, and *Harold and the Purple Crayon*. As I showed in the first study, many older books, which predominantly feature White, male central characters, are hugely popular, especially compared to newer books which are better at including diverse representations of gender and race. Although society may be more aware of the importance of representing non-dominant social groups, this awareness may not be enough to overcome the powerful inertia privileging books which teachers remember from their own childhoods, and which overrepresent White, male characters.

---

<sup>30</sup> Of course literally every other character was male, so *Frozen* can hardly be accused of failing to include male characters.

<sup>31</sup> Or, if they treat women as sex objects, they feel compelled to come with implausible backstories which explain the objectification as in the case of Quiet in *Metal Gear Solid V*; the character is a soldier who wears a bikini top and ripped pants because she needs her skin exposed to carry out photosynthesis.

<sup>32</sup> Hobbits, Elves, and Dwarfs are all White, as are all Men who didn't fall under the sway of Sauron. On the other hand, Orcs and corrupted Men map nicely onto stereotypes about Black people.

Another open question is whether there are additional mechanisms through which teachers maintain the selective tradition by exposing children to books which privilege the stories of the socially dominant group. Selecting books to be read in class is clearly a powerful way for teachers to do so, but teachers also influence the books that children encounter by assigning books and making recommendations to individual students. Regardless of what books a teacher reads in class, she or he is also responsible for identifying books which are appropriate for individual students. It may be that the books which teachers identify as appropriate for individual students differ from the books which they identify as appropriate to use in class.

Classroom reading is, by definition, selected independent of the characteristics of individual students since the books are read to the entire class. However, teachers can recommend and select books for individual students as well. In this case, teachers may use characteristics of specific children, including their genders and races, in order to select books. This opens the possibility that teachers are in fact creating multiple different selective traditions. Although the concept of the selective tradition refers to a cultural privileging of certain voices and perspectives, not every student experiences culture in the same way.

#### *Research questions*

In this study, I address the following questions:

- 1) Do teachers show a preference for selecting books with male central characters for children to read over the summer?
- 2) Do teachers show a preference for selecting books with White central characters for children to read over the summer?

3) Does the preference for male central characters differ by the race of the central character?

4) Does this preference vary according to the gender of the student?

The hypothesis of the selective tradition predicts that teachers will disproportionately assign books with White male central characters to children. It does not make predictions for the third or fourth research questions.

## **Method**

### **Participants**

My sample consisted of 583 3<sup>rd</sup> grade students in 34 classrooms in 11 schools in North Carolina. The sample was 51% female and 49% male. Of the students, 43% were identified as Black, 29% as Hispanic, 17% as White, and 11% as another race. Given other information about the schools, it is likely that most of the 11% were American-Indian and Asian or Asian-American. Most students who attended these schools receive free or reduced price lunches.

The sample was originally created by project READS, an intervention designed to combat summer reading slump by teaching students effective reading comprehension strategies and providing them access to high-quality books. One aspect of the READS intervention is that students are provided with 10 books over the summer. Two of these books, called lesson books, are assigned to the entire class. The other eight are algorithmically matched to students based on their reading comprehension level as assessed by the Iowa Test of Basic Skills reading comprehension subscale (converted into Lexile scores), and based on their expressed preferences for different sorts of books. All

books are selected from the READS library, a set of 471 texts identified as both high quality and interesting to children.

Student preference for different sorts of books was determined by administering a short survey in the spring. On the survey, students were asked to report how much they enjoyed reading each of eighteen different categories of books. Categories included, e.g., “books about girls my age”, “books about boys my age”, “sports biographies”, “books about science and facts”, etc.... Each type of book was accompanied by pictures of the covers of representative texts (e.g., the “books about girls my age” option was accompanied by a picture of a book about a young girl). Students reported how much they enjoyed reading that sort of book on a scale from one frowny face through two smiley faces.

Lexiles for the books were computed by MetaMetrics using their proprietary formula. Book categories were assigned by the READS team by performing content analyses on the books. The author had no input or control over any of these variables.

Books were matched to students by finding eight books from the READS library, which best matched student preferences while remaining close to their measured reading ability. Books could be matched to students as long as they were no more than 50 Lexiles above the students’ measured reading comprehension, and no more than 100 Lexiles below.

The students in this sample were enrolled into the alternative treatment wing, in which teachers were permitted to alter the basic treatment. One common way in which teachers adapted the READS intervention was by changing the books which students were sent over the summer. Teachers in this condition were sent a list of the ten books

that would be sent to each child, i.e., the two lesson books and the eight matched books. They were also sent ten additional books for each child which were matched in the same way; these ten alternate titles were the next ten best matches. Teachers were permitted to replace any of the lesson books or matched books with any of the ten alternate books<sup>33</sup>. This permitted teachers to use their rich knowledge of their students and their students' interests to select the most appropriate books to send over the summer.

Several teachers decided to swap out both lesson books and replace them with their own selections. Because one lesson book, which was assigned to all students by default, had a female central character and the other had neither a female nor a male central character, including the lesson books in my analyses inflated the number of initially assigned books with a female central character, the number of books switched out with a female central character, and teacher's overall tendency to switch out books. However, switching out a lesson book indicated something different than switching out a matched book; there was no reason to think that switching out the lesson books indicated a preference for books without female characters, even though one of the lesson books had a female central character. Therefore, I made the decision to eliminate lesson books from the analysis completely.

All teachers were women, and no other teacher-level data are available. I dropped all teachers from the dataset who made no substitutions, taking this as evidence that the teacher was not actually considering the possible switches.

---

<sup>33</sup> Teacher selections were slightly constrained by a requirement that at least two books be informational, rather than narrative, texts. Teachers also had the opportunity to replace the entire set of matched books with a new set of either lower or higher Lexile; essentially, teachers could determine based on their own knowledge of and experience with their students that their students' true reading comprehension ability was either lower or higher than the ITBS-based estimate, and request a new set of matches based on their personal perceptions.

## **Variables and coding**

*Student gender:* Student gender was reported to project READS by the schools. Students were identified as either female or male.

*Student race:* Student race was also reported by the schools. Students were identified as African-American, Hispanic, White, or Other.

*Central character gender and race:* I describe the coding of central character gender and race in the introduction.

## **Analytic plan**

I began by examining the proportions of central characters who were female and male in 1) the set of all books which were eligible to be assigned to children, 2) books that were originally matched to children, 3) books which were not originally matched to children, 4) books which were switched out from children's reading lists (i.e., replaced by other books), 5) books which were switched in to children's reading lists (i.e., selected to replace other books), and 6) books which were ultimately assigned to children. I also examined the same proportions, broken down by child gender.

I then calculated the difference in the proportion of central characters who were female between 1) all library books and books which were matched *or* held in reserve, 2) books which were originally matched to students and books which switched out, 3) books which were originally held in reserve and books which were switched in, and 4) books which were ultimately assigned to children and books which were initially assigned. The first difference represents the tendency of the algorithm matching books to children to over- or under-represent books with female central characters. The second difference represents teachers' tendencies to disproportionately switch out assigned books with

female central characters. The third difference represents teachers' tendencies to disproportionately switch in reserved books with female central characters (or with male central characters, as the case may be). The fourth difference represents the overall effect that teachers had on the percentage of central characters who were female in children's assigned book lists.

I then fit a series of mixed-effects logistic regression models to assess the statistical significance of the differences I found from comparing proportions. The first model predicted teachers switching out a book based on the gender of the main character of the book. It took the form

$$\log\left(\frac{PR(SWITCHOUT_{ijkl})}{1 - PR(SWITCHOUT_{ijkl})}\right) = \mu + \alpha_j + \gamma_k + \nu_l + \beta_1 FC_i,$$

where  $i$  indexes books,  $j$  indexes students,  $k$  indexes teachers, and  $l$  indexes schools, and  $FC_i$  indicates whether book  $i$  had a female central character. I fit this model to a subset of the data comprising all books initially matched to students with central characters who were either female or male.

Here  $\mu$  is roughly equivalent to the probability of switching out a book with a male central character for a typical student with a typical teacher in a typical school. The various  $\alpha$ ,  $\gamma$ , and  $\nu$  terms are random student-, teacher-, and school-level random offsets, and are assumed to follow mutually independent normal distributions with means 0 and variances estimated from the data. The term  $\beta_1$  is roughly equivalent to the difference between the probability of switching out a book with a female central character and the probability of switching out a book with a male central character. This is the key parameter of interest for this study. I also considered models which allowed  $\beta_1$  to vary across students, teachers, and schools.

Next I fit models which added an indicator for students being female, and an interaction between student gender and character gender to determine if the probability of switch out a book with a female central character differed according to student gender.

I then conducted similar analyses to predict the probability of switching in a book initially assigned to reserve, using a parallel series of models. I fit these models to a subset of the data comprising all books initially assigned to reserve which had central characters who were either female or male.

Finally, for each student I calculated the difference in the proportion of books which she or he was ultimately assigned which had a female central character and the proportion of books she or he was initially assigned. I labeled this difference  $\Delta^{Female}$ ; positive values indicated that a greater proportion of books that a student was ultimately assigned had female central characters than the proportion of books which the student was initially assigned. I then fit a mixed-effects linear regression model,

$$\Delta_{jkl}^{Female} = \beta_0 + \beta_1 STF M_j + \gamma_k + \nu_l + \varepsilon_j.$$

where  $STF M_j$  is an indicator for a student being female. Here the goal of inference was estimating  $\beta_0$ , which represented the change in proportion of central characters who were female for male students,  $\beta_0 + \beta_1$ , which represented the change for female students, and  $\beta_1$ , which represented the difference in those changes. However, as before, I was also interested in estimating the variance of  $\gamma$  and  $\nu$ , and I fit additional models which allowed  $\beta_1$  to vary within teachers and schools<sup>34</sup>.

---

<sup>34</sup> The advantage of fitting the previous models in addition to this final one is two-fold. First, they increase the power of the analysis by conditioning on the number of books with female central characters initially assigned to the student or initially put in reserve. Second, they help to understand the process by which the overall composition of the student reading lists changes; it is useful to estimate both the probability of switching books in and of switching books out, even if the ultimate goal is to see how books lists changed overall.

I then replicated all analyses using character race as a predictor. For these analyses I did not interact character race with any student variables. Theoretically it would have made sense to expect that character race and student race might have interacted but with so few white students, I elected not to test this hypothesis.

Finally, I refit all models using character gender, character race, student gender, and the interactions between character gender and character race, and between character gender and student gender as predictors.

## **Results**

I begin with brief descriptive results before discussing the analyses.

### *Descriptive results*

#### *Book switching*

Teachers in the sample actively engaged in switching books. On average, teachers switched out almost 15% of the assigned books<sup>35</sup>. However, there was substantial teacher-level variability in the tendency to alter the book lists. A handful of teachers switched between 20% and 25% of the assigned books, while around the same number switched between 1% and 10%. Most teachers fell between these two extremes. There was also a great deal of variability at the student level. The 25<sup>th</sup> percentile of the percentage of books switched for students was 0%, the median was 25%, and the 75<sup>th</sup> percentile was 50%. A handful of students had over 60% of their lists switched.

#### *Central character gender*

---

<sup>35</sup> Recall that the sample was constructed by dropping all teachers who failed to switch any books, as they seemed to have opted out of participating in the intervention.

Books in the READs library tended to have male central characters. Of the books, 46% had only male central characters, 28% had only female central characters, 4% had both female and male central characters, and 21% had neither female nor male central characters. Considering only books with central characters with identifiable genders, 59% had only male central characters, 36% had only female central characters, and 6% had both female and male central characters. Given how few books had both female and male central characters, and the complications they pose for my models, I set them aside for the rest of this analysis. This ratio of male to female central characters, 1.64, is markedly lower than studies of children's picture books have estimated, and substantially lower than the ratio in the books that children check out from their school libraries, presented in the second study. However, it is still quite far from 1:1, or parity, and strongly favors books with male central characters.

The books which were matched to children were also imbalanced along central character gender. For female students, 40% of the initial matches had female central characters, while 60% had male. For male students the balance was markedly worse; 26% of books initially assigned to boys had female central characters, while 74% had male. The books which were initially held in reserve had a similar distribution of character gender. Table 2 shows the proportion of books with male and female central characters, by student gender, at various stages of the book assignment process.

Teachers tended to switch in books with female central characters for girls and with male central characters for boys. For girls, 41% of the books that teachers switched in had female central characters. For boys, this was true of only 17% of books. In

contrast, teachers tended to switch out books with female central characters in the same proportions for both girls and boys.

Overall, the percentage of central characters in assigned books who were female did not change much. For girls, there was a tiny increase, from 40% to 41%. For boys, there was a decrease, from 26% to 23%.

#### *Central character race*

The READS library skewed heavily white. Among books with central characters with identifiable races, a full 82% were white. Among books initially matched to children, 84% had white central characters. The same was true of books initially held in reserve, as well as books switched in and out. Teacher actions only slightly changed the proportion of central characters who were white, increasing it from 84% in initially matched books to 85% in the final reading lists. On the other hand, given how few non-white characters were initially assigned to students, this constitutes a 7% reduction in the proportion of non-white central characters that children would have received if not for teacher selections. Table 3 displays the proportions of books with White and non-White central characters at various stages of the book assignment process.

#### *Central character gender and race*

There was some indication that character gender and race interacted in books in the READS library. In books with a non-White central character, 49% of central characters were female. In books with a White central character, only 37% of central characters were female. However, possibly due to the small number of books with non-White central characters, this difference was not statistically significant ( $\chi^2(df = 1) =$

2.80,  $p = .094$ )<sup>36</sup>. Overall, 10% of the books in the library had a non-White female central character, 10% had a non-White male, 30% had a White female, and 50% had a White male. Table 4 shows the proportions of books with female and male White and non-White central characters broken down by student gender at various stages of the book assignment process.

The matching process increased the proportion of books with White male central characters at the expense of all other groups. Overall, 8% of books initially assigned to children had a non-White female central character, 8% had a non-White male, 24% had a White female, and 59% had a White male. The same was approximately true of books initially held in reserve.

Books switched in and books switched out had approximately the same gender breakdown for books with White central characters. This is counter to the first hypothesis, which predicted that teacher decisions would increase the representation of White male central character. However, consistent with the hypothesis, books with non-White male central characters were disproportionately likely to be switched in while books with non-White female central characters were disproportionately likely to be switched out. Proportions in the final set of books assigned to students was similar to those in the initial set, suggesting that teacher actions had little impact on the overall distribution of race and gender in books assigned to children.

---

<sup>36</sup> A statistically significant result would be difficult to interpret, as it might point to the mechanism by which books are written and published, or to the mechanism by which READS selected books for its library. However, this discrepancy is noteworthy since it is directly in opposition to the pattern observed in study 1, in which there were no non-White male central characters, but far more White male central characters than White female.

These proportions obscure differences between female and male students. Those differences will be explored in the modeling section.

### *Models*

#### *Probability of switching out*

Results for all of the models are presented in Tables 8, 9 and 10. Model implied probability of switching in or out, or of entering the final book list, a more interpretable metric, is presented in Tables 5, 6, and 7. Figures 1, 2, and 3 display the model-implied probability of switching books in or out, and of a book entering the final book list based on book and child characteristics. I present only the results from the most complex models in text. The first model predicted the probability of teachers switching out assigned books, conceptually the first teacher-driven decision in assembling the final book list. The fitted model was

$$\begin{aligned} \log\left(\frac{PR(SWITCHOUT_{ijkl})}{1 - PR(SWITCHOUT_{ijkl})}\right) \\ = -2.22 + 0.70CHFM_i - 0.09CHNW_i + 0.98CHFM_i \times CHNW_i \\ + 0.41STFM_j - 1.46STFM_j \times CHFM_i \end{aligned}$$

Where *CHFM* is an indicator for a female character, *CHNW* is an indicator for a non-white character, and *STFM* is an indicator for a female student. In this model, *i*, *j*, *k*, and *l* index books, students, teachers, and schools, respectively. All of the estimated coefficients were statistically significant, with the exception of the coefficient of *CHNW*. Because of the presence of multiple interactions, the model coefficients are difficult to interpret. However, Figures 1 and 2 display the model-implied probabilities of switch out books based on student and character gender (Figure 1), and based on student and character gender and character race (Figure 2). Tables 5, 6, and 7 also display these

probabilities. Especially noteworthy is the strong tendency teachers show to switch out books with non-White female characters that are initially assigned to boys (35%). Other probabilities range from a low of around 6% (switching out books with White female central characters for girls or books with non-White male central characters for boys) to a high of around 17% (switching out books with non-White female central characters for girls or books with White female central characters for boys).

Teachers tended to switch out books with central characters with genders different from the children to whom they were assigned, though this difference was far greater for boys than for girls. Teachers were far more likely to switch out books with female central characters that were assigned to boys (20%), than books with male central characters (10%). They showed far less differences with girls, switching out roughly the same proportion of books with female central characters (11%) as with male central characters (13%). Overall, teachers were more likely to switch out books with non-White central characters (17%) than with White central characters (11%).

*Probability of switching in*

The next set of fitted models predicted the probability of teachers switching out books based on character and student characteristics. The final fitted model was

$$\begin{aligned} \log\left(\frac{PR(SWITCHIN_{ijkl})}{1 - PR(SWITCHN_{ijkl})}\right) \\ = -1.04 - 0.27CHFM_i + 0.34CHNW_i - 0.84CHFM_i \times CHNW_i \\ - 0.22STFM_j + 1.14STFM_j \times CHFM_i \end{aligned}$$

All of the model parameters are statistically significant except for the coefficient of the interaction between character race and gender. Once more the model is difficult to interpret given the interactions. Figures 3 and 4 and Tables 5, 6, and 7 display the model

estimated probabilities of switching in a book conditional on student and central character gender and race. Most of the probabilities of switching in are between 20% and 30%, but the probability of switching in a book about a female central character, either non-White or White, for a male student is substantially lower, at close to 14%.

Overall, teachers were more likely to switch in books with central characters of the same genders as the children, though the gap was wider for girls than for boys. For boys, teachers were more likely to switch in books with male central characters (27%) than with female central characters (15%). Once again, for girls, the probability of switching in books with male central characters (22%) was fairly close to the probability of switching in books with female central characters (27%). Teachers were slightly more likely to switch in books with non-White central characters (27%) than books with White central characters (23%).

*Probability of entering the final list*

The final set of fitted models predicted the probability of a matched book entering students final book lists. The most complex fitted model was

$$\begin{aligned} \log\left(\frac{PR(INFINAL_{ijkl})}{1 - PR(INFINAL_{ijkl})}\right) \\ = 0.17 - 0.45FM_i + 0.09CHNW_i - 0.20CHFM_i \times CHNW_i - 0.24STFM_j \\ + 0.78STFM_j \times CHFM_i \end{aligned}$$

All coefficients are statistically significant except for the coefficients of *CHNW* and of *CHFM* × *CHANW*. Figures 4 and 5 displays the model estimated probabilities of books matched books ultimately being assigned to students. The most striking probabilities are those of books assigned to boys with either White female (43%) or non-White female (40%) central characters ultimately ending up on the final reading list. The probabilities

of a male character entering a boy's list or a female character entering a girl's list were all close to 55%. On the whole, teachers actions made it more likely that girls would encounter books with female central characters and boys would encounter books with male central characters.

#### *Total changes in book lists*

The fitted model estimating the change in the proportion of female central characters due to teacher interventions was

$$\Delta_{jkl}^{Female} = -.03 + .04STFM_j$$

Both coefficients were statistically significant. Teacher actions decreased the number of female central characters encountered by male students. The evidence that teacher actions increased the number of female central characters in books encountered by female students (i.e.,  $H_0: \beta_0 + \beta_1 > 0$ ) was not statistically significant ( $\hat{\beta}_0 + \hat{\beta}_1 = .02, t_{28} = 1.87, p = 0.073$ ).

#### **Discussion**

Counter to previous research on the role of teachers in perpetuating the selective tradition (Jipson & Paley, 1991; Luke, Cooke, & Luke, 1986), this study finds no evidence for its primary hypothesis, that teacher book recommendations systematically privilege books with male central characters. Instead, it finds that teachers' decisions marginally increase the proportion of male central characters that boys encounter, but this effect is extremely weak, especially compared to the huge imbalances reported in previous literature. There is evidence to support the secondary hypotheses, that teachers preferentially switch out books with female central characters and switch in books with male central character when choosing books for boys. There was also evidence that

teacher systematically switch in books with female central characters when choosing books for girls, but no evidence that teachers systematically switched out books with male central characters for girls. This is less consistent with the hypothesis of the selective tradition and more consistent with the hypothesis that teachers expect books with male characters to be more appealing to boys and books with female character to be more appealing to girls.

The exploratory hypothesis, that rates of switching in would vary by the races of the characters, was confirmed, though the patterns of switching were complex and difficult to interpret. One interesting result was that teachers were especially likely to switch out books with non-white female central characters for both female and male students.

#### *Divergences from previous literature*

Previous studies of teacher behavior have found that teachers tend to select books with male central characters when deciding what books to read to their classes. These studies tie teacher decisions to the selective tradition in children's literature, a tradition which elevates books by and about privileged members of society, namely White men and boys, at the expense of others. Luke, Cooke, and Luke focus on preservice teachers and argue for preservice training to help teachers select more representative texts for use in class. Jipson and Paley demonstrate that the preference for using books with White male central characters persists in experienced teachers.

However, these studies have only examined the decisions teachers made in selecting books for classroom use. While the books that teachers use in their classrooms send a powerful signal regarding what sorts of characters are worth paying attention to,

this is only one mechanism by which teachers expose children to books. Teachers also have the ability to direct individual children to particular books which they think are well-suited for those children. While book recommendations lack the public, performative nature of books used in class, they may be more influential on children's later reading because they are tailored to particular children, and therefore more likely to shape the books that children select for themselves.

There are reasons to expect that the books which teachers select for children may be different from books that they use in class in terms of central character gender and race. As the second study in this dissertation demonstrates, both girls and boys seem to independently read books with male central characters, while very few boys select books about female central characters. In choosing a book which would be acceptable to all students in the classroom, a teacher might feel constrained to select a book with a male central character; many boys might refuse to read a book with a female central character. However, in selecting books for individual children, a teacher might feel more comfortable recommending books with female central characters to girls or to boys that she<sup>37</sup> knows would be comfortable reading them. This could account for the finding in this study that teachers did not impact the overall proportion of female central characters in books that children encountered: decisions which made boys' book lists more masculine were offset by decisions which made girls' book lists more feminine.

Another key difference is time. Both previous studies were conducted over two decades before the present study. Although few scholars of gender would argue that women have achieved equal representation with men over that time (and this dissertation

---

<sup>37</sup> All teachers in the sample, as well as the vast majority of elementary school teachers around the country, are female, so I use feminine pronouns to refer to these teachers.

demonstrates that they have not in the realm of children's literature), it may be that teachers are more comfortable in asking boys to read books with female central characters than they would have been in the past. Perhaps present-day teachers are better able to see the universal appeal of books with female central characters.

It is also worth keeping in mind that the book assignment mechanism presented teachers with book lists which were already highly imbalanced in terms of central character gender, especially for boys. 74% of books assigned to boys had male central characters, while only 26% had female central characters. Teacher actions slightly increased the percentage of books with male central characters to 77%, a small difference, but a notable one in the context of a highly imbalanced starting distribution. Even in books assigned to girls, 68% of the matched books had male central characters, and teacher decisions only decreased that proportion to 67%. Although in this context teacher actions only marginally increased the proportion of male central characters encountered by boys, it is less clear what they might have done had the book assignment process assigned more books with female central characters to male students.

### *Race*

Previous studies found that teacher book selections consistently privileged books about White central characters. However, teacher book selections showed a different pattern. Teacher decisions ultimately did not change the overall distribution of character race in the books children encountered; books with non-White central characters had the same probability of entering children's final lists as books with White central characters, although teachers were far more likely to switch out books with non-White female central

characters than books with White female central characters or books with non-White male central characters.

The reason for this divergence from previous literature is not entirely clear. One possible explanation is that the majority of students in this sample (83%) were non-White. Previous literature has not examined how the composition of classrooms impacts the ways in which teachers further the selective tradition. However, it is reasonable to expect that teachers in classrooms which are predominantly non-White would feel more comfortable in using texts which foreground non-White characters. Alternately, perhaps the selective tradition itself is changing to give more prominence to non-White characters. Finally, as with gender, the initial matches drastically overrepresented White central characters; 84% of books initially matched to children had White central characters. Teacher decisions must be understood in this context.

#### *Teachers decisions in context*

It is helpful to take several steps back from the focus of this study, i.e., the role of teacher decisions in perpetuating the selective tradition, in order to better understand the sequence of steps that led to such dramatic overrepresentation of books about White, male central characters. In this context, we can see the books' failure to adequately represent the world that children inhabit in terms of gender and race as arising through four processes. First, children's books as a whole over-represent stories about members of socially dominant groups, in this case White male characters. Studies have demonstrated that this is true, at least for children's picture books.

Second, books about dominant groups may be more likely to be identified as high quality than books about non-dominant groups. The library used in this study was

developed to offer children access to high quality books that they are likely to read and enjoy. Experts in children's literacy worked hard to identify books which would be interesting and accessible to students with a wide range of reading comprehension abilities. However, 62% of the books in the READS library have male central characters, 80% have White central characters, and 50% have White male central characters. It is unclear how this compares to the set of all possible books which could have been selected, but it is true that the library which was created by the READS team with high-quality texts drastically over-represents dominant social groups.

Third, books in the READS library which are about dominant groups are substantially more likely to be matched to students than books about non-dominant groups. 62% of the books in the READS library had male central characters, but 68% of the books matched to children did. 80% had White central characters, but this was true of 84% of the books matched to children. 50% of books in the READS library had central characters who were both White and male, but 59% of books assigned to children did. This was especially true for books assigned to boys. A full 67% of books matched to boys had White male central characters, while only 5% had non-White central characters and 19% had White female central characters. In the READS library, in contrast, 10% of central characters were non-White females, and 30% were White females. The reason for the disproportionate assignment of books with White male central characters is not clear, but may stem from the fact that books which were well-matched to children's interests, including humor and adventure books, were the most likely of all books to have White male central characters. Books dealing with cultural issues, families, and friendships

included more non-White and female central characters, but these categories were less popular with children.

All three of these processes, the disproportionate representation of socially dominant groups in children's literature as a whole, the disproportionate representation in the library of books, and the disproportionate assignment of books to children resulted in highly imbalanced reading lists before teachers even encountered them. The books available to students already featured far more White male central characters than any other group. The imbalance was already so severe for boys that even if teachers had created reading lists with as many female central characters as possible, treating that as the only consideration in assigning books, 50% of the central characters in books assigned to boys would still have been male.

The fourth mechanism by which socially dominant groups come to be overrepresented in book lists, teacher selection, should be understood in this context. Teachers would have been hard-pressed to craft books lists which gave fair representation to girls, women, and non-White people, even if this had been their key goal. On the other hand, after seeing the books to which children were originally matched and the books held in reserve, teachers had the opportunity to make children's reading lists either more or less reflective of the diversity of the society in which they live, but overall, teacher decisions led to almost no change in the overall representation of female or male non-White or White central characters.

### *Limitations*

The design of this study offers unusually strong internal validity, since it is possible to observe both the books children would have been assigned without teacher

intervention and the books children were actually assigned. The study design also ensure that books are already well-matched to children in terms of their reading level and the reading interests that they have expressed. This makes it more plausible that teacher decisions to switch books in or out were based on central character gender rather than other aspects of the books. Finally, the study permits me to directly observe all of the possible books that could be assigned, which makes it possible to observe book lists in the context of all possible lists; previous studies of the selective tradition have not been able to identify the set of books teachers were choosing from, which made it impossible to determine whether the books teachers selected featured more or fewer female central characters than the books available to them.

However, there remain threats to the inferences I draw. The conclusions are most solid when understood as descriptive. It is the case that teacher decisions led boys to encounter more male central characters than they would have based solely on their preferences and reading levels. It is also the case that teachers assigned reading lists to children which contained far more male central characters than female. However, it is less certain that this effect is due to teachers intentionally selecting books with male central characters for male readers. It is possible that teachers were instead responding to other characteristics of the books, correlated with but not identical to central character gender in making their decisions. This would not change the fact that teacher decisions increased the overrepresentation of male central characters in books assigned to boys, but it would change our understanding of those decisions.

External validity is harder to establish. This study focuses on a small set of teachers in a set of schools in a single southern state. While there is no clear reason why

teachers in other schools should choose books differently, there is also no compelling reason why they should make similar choices. This analysis is best thought of as an exploratory attempt to detect the selective tradition at work in teachers' book recommendations in a single setting. The findings from this survey are at least partially consistent with research done in other settings. Teachers in this sample did not greatly increase the overrepresentation of male central characters in books that children read. However, just as Jipson and Paley (1991), and Luke, Cooke, and Luke (1986) found in very different settings, they were comfortable assigning books to children which contained far more male central characters than female.

### **Conclusion**

Teacher decisions made a tiny contribution to the severe overrepresentation of White, male characters in the books children in this sample received for summer reading. However, much more important, teachers failed to act to correct the imbalance that existed in the books which were initially assigned. Had teachers shown no preference for books with White, male central characters, these books would still have constituted a huge proportion of the books assigned to children, and an overwhelming proportion of books assigned to boys. Teacher decisions actually worsened the situation for boys, even as they slightly improved the situation for girls. A key takeaway might be that achieving proportional representation of gender and race in books that children encounter requires teachers to go beyond a commitment to neutrality in selecting books, to actively seeking out books with non-White and female characters for children, and especially boys, to read. When the deck is already stacked so badly in favor of socially dominant groups, ignoring race and gender won't solve anything.

Table 1: Student demographic information (n = 583).

<u>Category</u>			
Gender		Female	51%
		Male	49%
Race		Black	43%
		Hispanic	29%
		White	17%
		Other	11%

Table 2: Proportions of books with female and male central characters, by student gender, at various points of the book assignment process.

<u>Child gender</u>	<u>Central character gender</u>	<u>Library (353 books)</u>	<u>Matched (7346 books)</u>	<u>Assigned (3234 books)</u>	<u>Reserve (4112 books)</u>	<u>Switched in (1072 books)</u>	<u>Switched out (482 books)</u>	<u>Final (2752 books)</u>
Overall	Female	38%	32%	33%	32%	29%	36%	33%
	Male	62%	68%	67%	68%	71%	64%	67%
Girls	Female		38%	40%	37%	41%	35%	41%
	Male		62%	60%	63%	59%	65%	59%
Boys	Female		26%	26%	25%	17%	38%	23%
	Male		74%	74%	75%	83%	62%	77%

Table 3: Proportions of books with White and non-White central characters at various points of the book assignment process.

<u>Central character race</u>	<u>Library (333 books)</u>	<u>Matched (6719 books)</u>	<u>Assigned (2935 books)</u>	<u>Reserve (3784 books)</u>	<u>Switched in (946 books)</u>	<u>Switched out (437 books)</u>	<u>Final (2498 books)</u>
Non-White	20%	16%	16%	16%	19%	22%	15%
White	80%	84%	84%	84%	81%	78%	85%

Table 4: Proportions of books with female and male White and non-White central characters, by student gender, at various points of the book assignment process.

<u>Child gender</u>	<u>Central character gender and race</u>	<u>Library (304 books)</u>	<u>Matched (5869 books)</u>	<u>Switched in (863 books)</u>	<u>Switched out (437 books)</u>	<u>Final (2155 books)</u>
Combined						
	Non-White female	10%	8%	8%	14%	8%
	Non-White male	10%	8%	11%	7%	8%
	White female	30%	24%	21%	21%	25%
	White male	50%	59%	60%	58%	59%
Girls						
	Non-White female		11%	13%	17%	11%
	Non-White male		8%	10%	8%	7%
	White female		29%	29%	16%	32%
	White male		52%	48%	60%	50%
Boys						
	Non-White female		5%	3%	10%	4%
	Non-White male		9%	13%	7%	9%
	White female		19%	11%	26%	18%
	White male		67%	73%	57%	69%

Table 5: Probability of books being switched out, switched in, and entering the final list as a function of child and central character gender.

<u>Child gender</u>	<u>Central character gender</u>	<u>Probability of switching out</u>	<u>Probability of switching in</u>	<u>Probability of entering final list</u>
Girls				
	Female	11%	27%	55%
	Male	13%	22%	44%
Boys				
	Female	20%	15%	50%
	Male	10%	27%	56%

Table 6: Probability of books being switched out, switched in, and entering the final list as a function of central character race.

<u>Central character race</u>	Probability of switching <u>out</u>	Probability of switching <u>in</u>	Probability of entering <u>final list</u>
Non-White	17%	27%	51%
White	11%	23%	51%

Table 7: Probability of books being switched out, switched in, and entering the final list as a function of child and central character gender and central character race.

<u>Child gender</u>	<u>Central character gender and race</u>	Probability of switching <u>out</u>	Probability of switching <u>in</u>	Probability of entering <u>final list</u>
Combined	Non-White female	20%	25%	53%
	Non-White male	11%	30%	49%
	White female	11%	21%	51%
	White male	12%	24%	51%
Girls	Non-White female	16%	29%	56%
	Non-White male	13%	28%	40%
	White female	7%	27%	54%
	White male	14%	21%	43%
Boys	Non-White female	35%	15%	50%
	Non-White male	9%	33%	54%
	White female	18%	13%	48%
	White male	10%	26%	56%

Table 8: A series of hierarchical logistic regression models predicting the probability of switching out assigned books based on book and student characteristics

	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>	<b>Model 4</b>
Intercept	-2.08*** (0.24)	-2.24*** (0.24)	-2.10*** (0.24)	-2.22*** (0.25)
Female character	0.27* (0.11)	0.85*** (0.16)		0.70*** (0.20)
Female student		0.32* (0.13)		0.41** (0.15)
Female characterXFemale student		-1.03*** (0.22)		-1.46*** (0.26)
Non-White character			0.48*** (0.14)	-0.09 (0.23)
Female characterXNon-White character				0.98** (0.32)
Log Likelihood	-1279.97	-1258.06	-1153.88	-976.66
AIC	2569.93	2530.13	2317.75	1971.33
BIC	2600.34	2572.63	2347.67	2023.79
Num. obs.	3234	3204	2935	2513
Groups (reads_id)	583	577	579	571
Groups (t_id)	34	34	34	34
Groups (sch_id)	11	11	11	11
Variance: reads_id: (Intercept)	0.02	0.03	0.13	0.03
Variance: t_id: (Intercept)	0.45	0.47	0.48	0.51
Variance: sch_id: (Intercept)	0.39	0.36	0.39	0.34

\*\*\* p < 0.001, \*\* p < 0.01, \* p < 0.05

Statistical models

Table 9: A series of hierarchical logistic regression models predicting the probability of switching in reserve books based on book and student characteristics

	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>	<b>Model 4</b>
Intercept	-1.12*** (0.17)	-1.02*** (0.17)	-1.23*** (0.15)	-1.04*** (0.16)
Female character	-0.11 (0.08)	-0.69*** (0.14)		-0.85*** (0.17)
Female student		-0.23* (0.09)		-0.27** (0.10)
Female character X Female student		0.96*** (0.17)		1.14*** (0.20)
Non-White character			0.24* (0.10)	0.34* (0.14)
Female character X Non-White Character				-0.22 (0.23)
Log Likelihood	-2241.65	-2199.07	-2040.22	-1786.73
AIC	4493.30	4412.13	4090.43	3591.46
BIC	4524.91	4456.31	4121.62	3646.38
Num. obs.	4112	4066	3784	3300
Groups (reads_id)	582	576	582	574
Groups (t_id)	34	34	34	34
Groups (sch_id)	11	11	11	11
Variance: reads_id: (Intercept)	0.00	0.00	0.00	0.00
Variance: t_id: (Intercept)	0.35	0.35	0.29	0.32
Variance: sch_id: (Intercept)	0.16	0.16	0.12	0.12

\*\*\* p < 0.001, \*\* p < 0.01, \* p < 0.05

Statistical models

Table 10: A series of hierarchical logistic regression models predicting the probability of books entering the final reading list based on book and student characteristics

	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>	<b>Model 4</b>
Intercept	0.10*	0.19***	0.05	0.17**
	(0.04)	(0.05)	(0.04)	(0.05)
Female character	-0.05	-0.43***		-0.45***
	(0.05)	(0.08)		(0.09)
Female student		-0.19**		-0.24***
		(0.06)		(0.06)
Female character X Female student		0.66***		0.78***
		(0.10)		(0.12)
Non-White character			-0.03	0.09
			(0.07)	(0.10)
Female character X Non-White character				-0.20
				(0.15)
Log Likelihood	-5081.61	-5008.84	-4651.99	-4002.46
AIC	10173.22	10031.69	9313.97	8022.93
BIC	10207.73	10079.93	9348.04	8082.94
Num. obs.	7346	7270	6719	5813
Groups (reads_id)	583	577	583	577
Groups (t_id)	34	34	34	34
Groups (sch_id)	11	11	11	11
Variance: reads_id: (Intercept)	0.00	0.00	0.00	0.00
Variance: t_id: (Intercept)	0.01	0.01	0.01	0.01
Variance: sch_id: (Intercept)	0.01	0.01	0.01	0.00

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$

Statistical models

Figure 1: Model-implied probabilities of books being switched out of children's reading lists based on the genders of the central character and of the student.

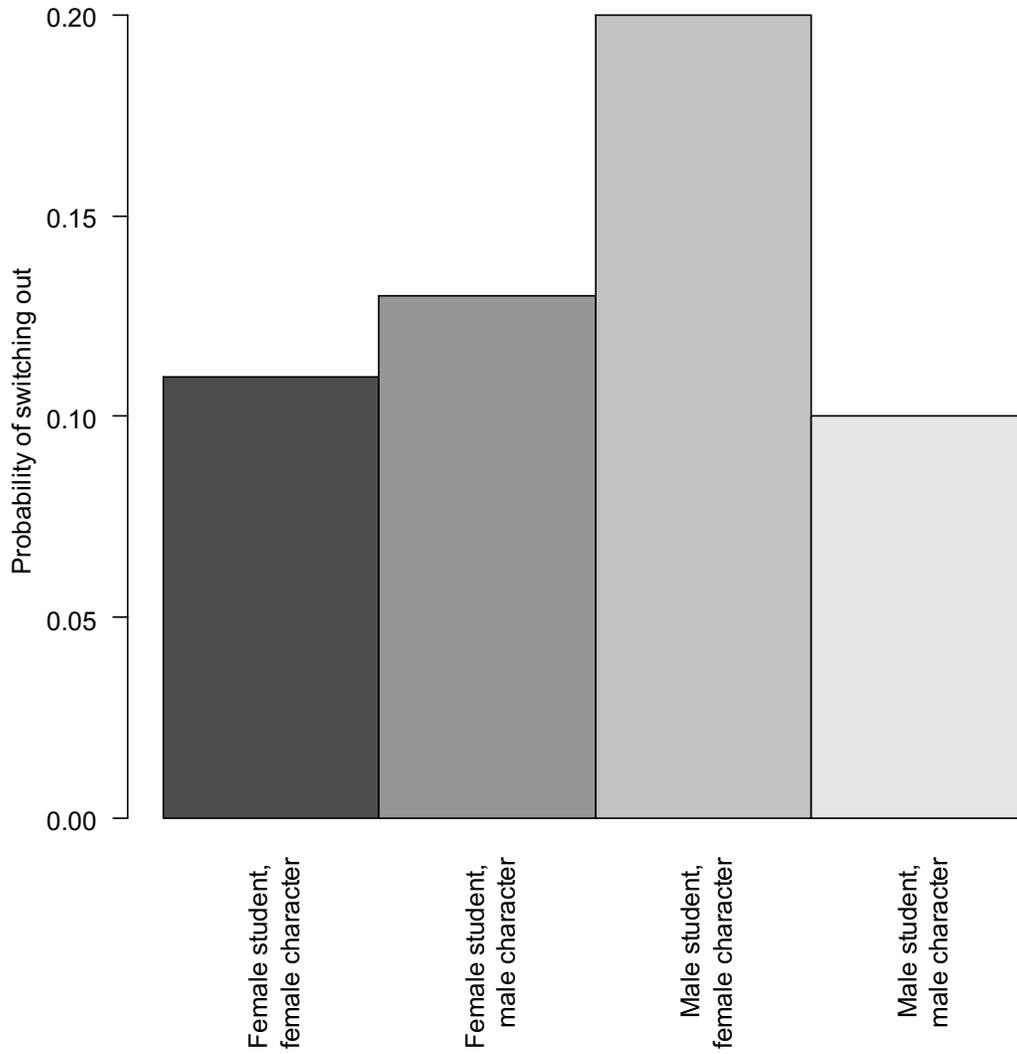


Figure 2: Model-implied probabilities of books being switched out of children’s reading lists based on the genders of the central character and of the student and the race of the central character.

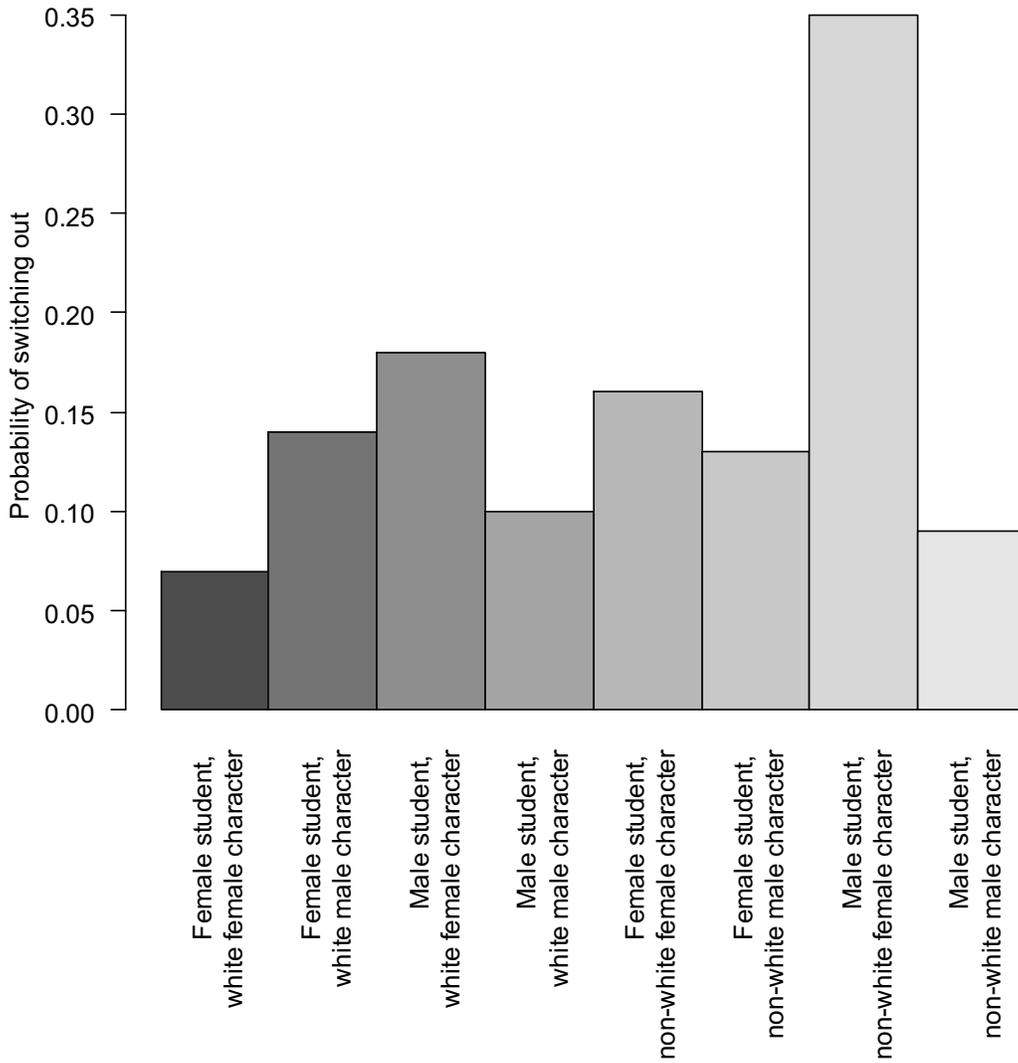


Figure 3: Model-implied probabilities of books being switched into children's reading lists based on the genders of the central character and of the student.

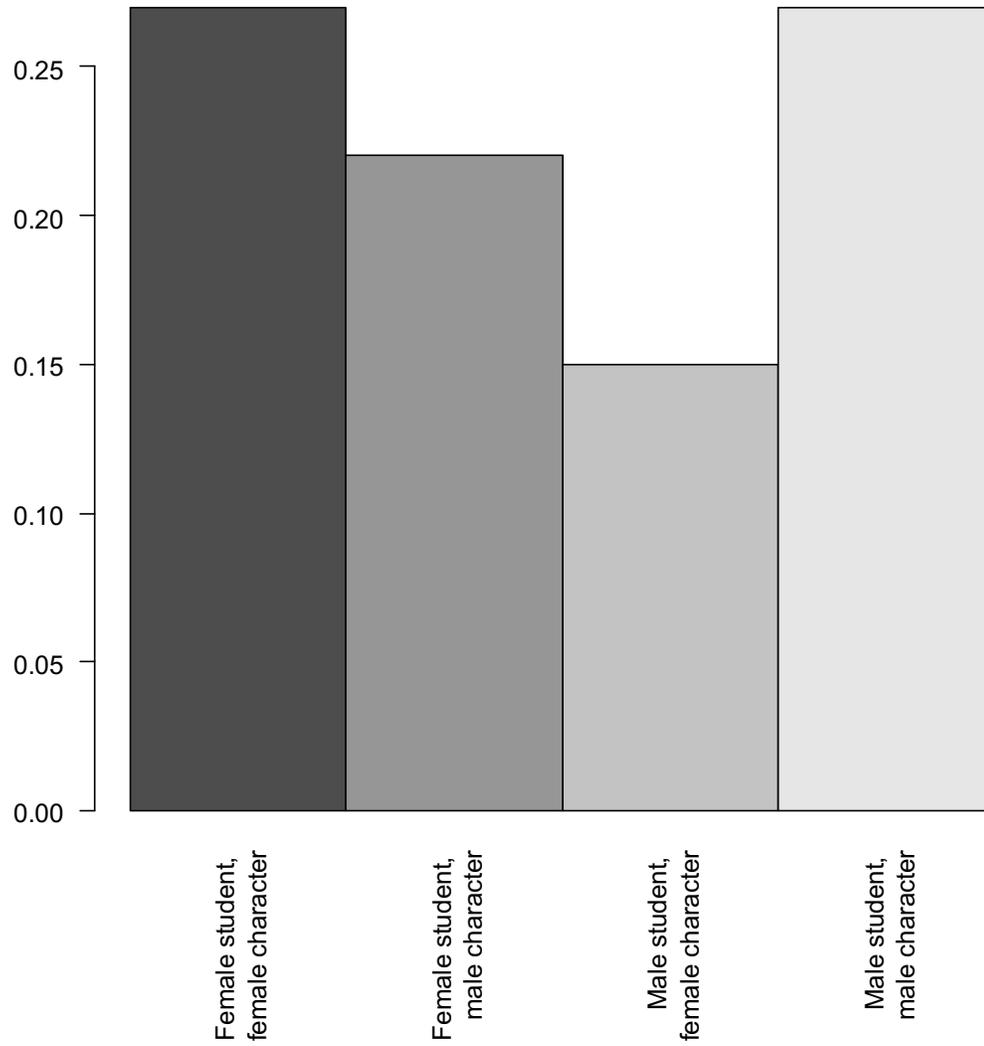


Figure 4: Model-implied probabilities of books being switched into children’s reading lists based on the genders of the central character and of the student and the race of the central character.

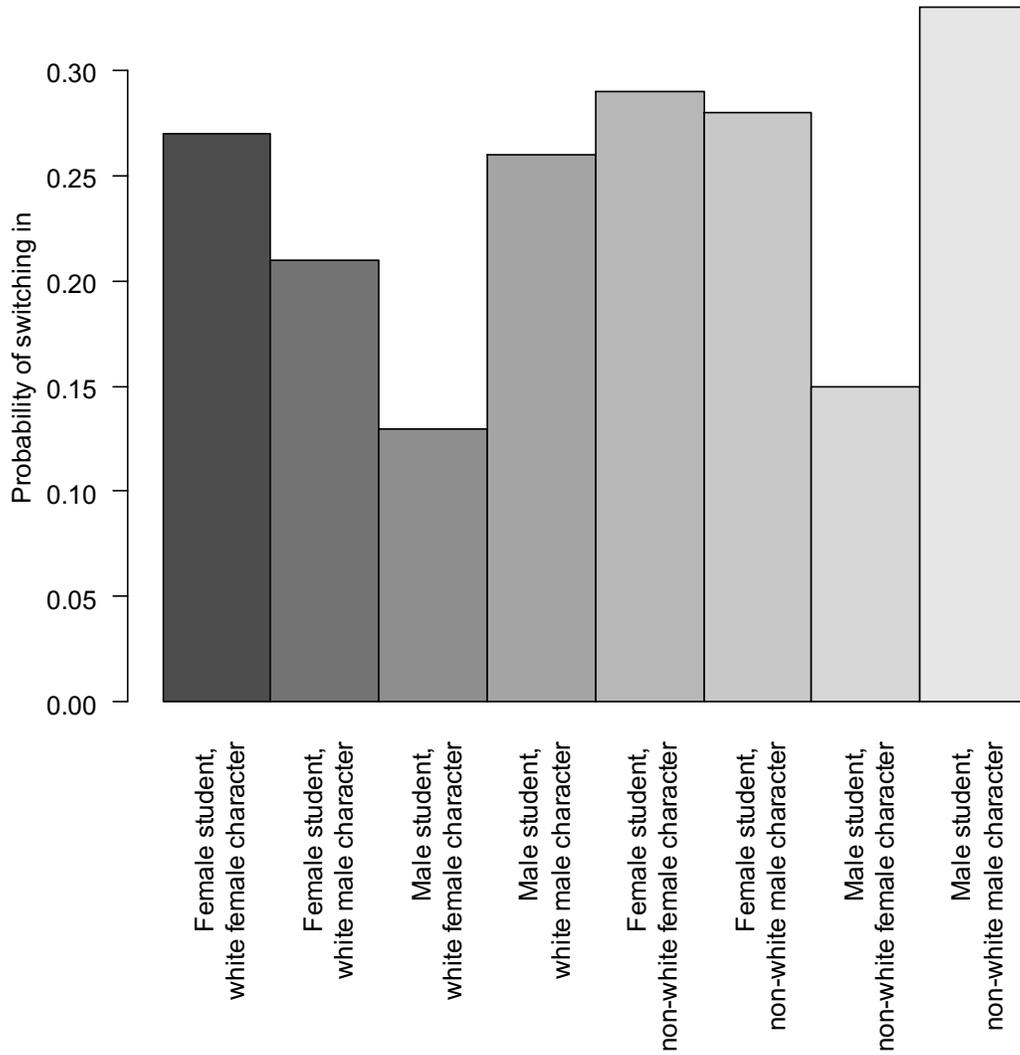
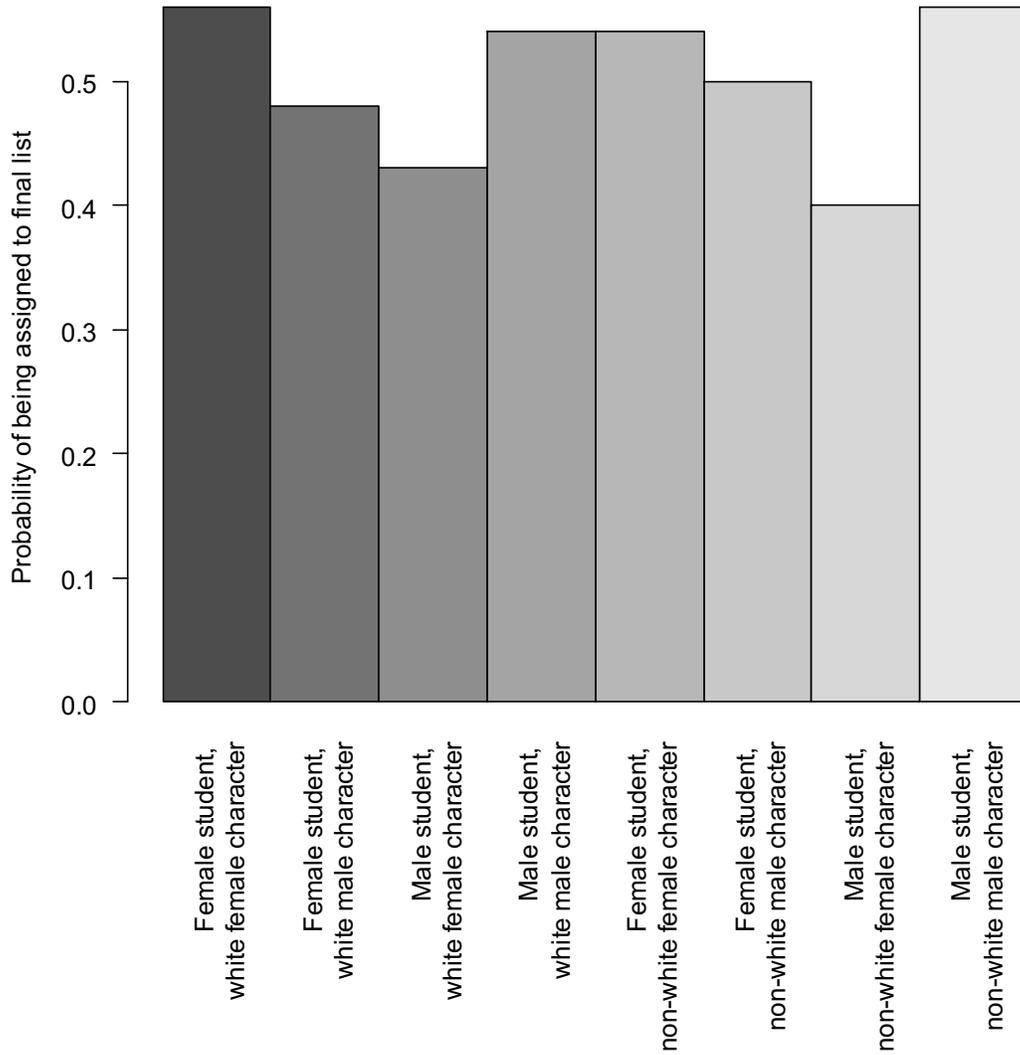


Figure 5: Model-implied probabilities of books ending up in children’s reading lists based on the genders of the central character and of the student and the race of the central character.



## Bibliography

- Anderson, D. A., & Hamilton, M. (2005). Gender role stereotyping of parents in children's picture books: The invisible father. *Sex Roles, 52*(3-4), 145-151.
- Anderson, R. C., Wilson, P. T., & Fielding, L. G. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly, 28*(5), 285-303.
- Bal, P. M., & Veltkamp, M. (2013). How does fiction reading influence empathy? An experimental investigation on the role of emotional transportation. *PloS one, 8*(1), e55341.
- Bates D, Maechler M, Bolker BM and Walker S (2015). "Fitting Linear Mixed-Effects Models using lme4." ArXiv e-print; in press, *Journal of Statistical Software*, <URL: <http://arxiv.org/abs/1406.5823>>.
- Best D. L., & Thomas J. J. (2004). Cultural diversity and cross-cultural perspectives. In A. H. Eagly, A. E. Beall, & R. J. Sternberg (Eds.), *The psychology of gender* (2nd ed., pp. 296–327). New York: Guilford.
- Bleakley, M. E., Westerberg, V., & Hopkins, K. D. (1988). The effect of character sex on story interest and comprehension in children. *American Educational Research Journal, 25*(1), 145-155.
- Butler, J. (2011). *Gender trouble: Feminism and the subversion of identity*. Routledge.
- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics, 90*(3), 414-427.
- Cassidy, K. W., Ball, L. V., Rourke, M. T., Werner, R. S., Feeny, N., Chu, J. Y., Lutz, D. J., & Perkins, A. (1998). Theory of mind concepts in children's literature. *Applied Psycholinguistics, 19*(3), 463-470.
- Chapman, M., Filipenko, M., McTavish, M., & Shapiro, J. (2007). First Graders' Preferences for Narrative and/or Information Books and Perceptions of Other Boys' and Girls' Book Preferences. *Canadian Journal of Education/Revue canadienne de l'éducation, 53*(1), 531-553.
- Chiu, L.-H. (1973). Reading preferences of fourth grade children related to sex and reading ability. *The Journal of Educational Research, 66*(8), 369-373.

- Cho, S., Crenshaw, K. W., & McCall, L. (2013). Toward a field of intersectionality studies: Theory, applications, and praxis. *Signs: Journal of Women in Culture and Society*, 38(4), 785-810.
- Christensen, R. H. B. (2015). ordinal - Regression Models for Ordinal Data. R package version 2015.6-28. <http://www.cran.r-project.org/package=ordinal/>.
- Collins, P. H. (2000). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment* (2nd ed.). NY: Routledge.
- Collins, S. T. (2014, October 17). Anita Sarkeesian on GamerGate: 'We have a problem and we're going to fix this'. *Rolling Stone*. Retrieved from <http://www.rollingstone.com>.
- Connor, J. M., & Serbin, L. A. (1978). Children's responses to stories with male and female characters. *Sex Roles*, 4(5), 637-645.
- Coyle, E. F., Fulcher, M., & Trübtschek, D. (2016). Sissies, mama's boys, and tomboys: Is children's gender nonconformity more acceptable when nonconforming traits are positive?. *Archives of sexual behavior*, 45(7), 1827-1838.
- Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, 1241-1299.
- Davies, B. (2003). *Frogs and snails and feminist tales: Preschool children and gender*. Rev. ed. Cresskill, NY: Hampton Press.
- Davies, J., & Brember, I. (1993). Comics or Stories? Differences in the Reading Attitudes and Habits of Girls and Boys in years 2, 4 and 6. *Gender and Education*, 5(3), 305-320.
- Djikic, M., Oatley, K., & Moldoveanu, M. C. (2013). Reading other minds: Effects of literature on empathy. *Scientific Study of Literature*, 3(1), 28-47
- Disch, L. (2016). Representation. In *The Oxford Handbook of Feminist Theory*.
- Dutro, E. (2001). "But that's a girls' book!" Exploring gender boundaries in children's reading practices. *The Reading Teacher*, 376-384.
- Dutro, E. (2002). "Us boys like to read football and boy stuff": Reading masculinities, performing boyhood. *Journal of Literacy Research*, 34(4), 465-500.
- Feldman, B. (2013, December 12). Megyn Kelly assures everyone that Santa is White even though Santa does not exist. *The Atlantic*. Retrieved from <http://www.theatlantic.com>.

- Fox, M. (1993). Men who weep, boys who dance: The gender agenda between the lines in children's literature. *Language Arts*, 70(2), 84-88.
- Galambos, N. L., Almeida, D. M., & Petersen, A. C. (1990). Masculinity, femininity, and sex role attitudes in early adolescence: Exploring gender intensification. *Child Development*, 61(6), 1905-1914.
- Gambell, T. J., & Hunter, D. M. (1999). Rethinking gender differences in literacy. *Canadian Journal of Education/Revue canadienne de l'education*, 1-16.
- Gooden, A. M., & Gooden, M. A. (2001). Gender representation in notable children's picture books: 1995–1999. *Sex Roles*, 45(1-2), 89-101.
- Grauerholz, E., & Pescosolido, B. A. (1989). Gender representation in children's literature: 1900-1984. *Gender & Society*, 3(1), 113-125.
- Hall, C. & Coles, M. (1997). Gendered readings: Helping boys develop as critical readers. *Gender and Education*, 9(1), 61-68, DOI: 10.1080/09540259721457
- Hamilton, M. C., Anderson, D., Broaddus, M., & Young, K. (2006). Gender stereotyping and under-representation of female characters in 200 popular children's picture books: A twenty-first century update. *Sex Roles*, 55(11-12), 757-765.
- Hogrebe, M. C., Nist, S. L., & Newman, I. (1985). Are there gender differences in reading achievement? An investigation using the high school and beyond data. *Journal of Educational Psychology*, 77(6), 716.
- Hornaday, A. (2016, June 2). Thanks, guys: Going to see the all-female 'Ghostbusters' is now a political act. *Washington Post*. Retrieved from <http://www.washingtonpost.com>.
- Hughes-Hassell, S., & Rodge, P. (2007). The leisure reading habits of urban adolescents. *Journal of Adolescent & Adult Literacy*, 51(1), 22-33.
- Huston A. C., Carpenter C. J., Atwater J. B., & Johnson L. M. (1986). Gender, adult structuring of activities, and social behavior in middle childhood. *Child Development*, 57, 1200–1209.
- Jipson, J., & Paley, N. (1991). The selective tradition in teachers' choice of children's literature: Does it exist in the elementary classroom?. *English Education*, 148-159.
- Johnson, D. R. (2012). Transportation into a story increases empathy, prosocial behavior, and perceptual bias toward fearful expressions. *Personality and Individual Differences*, 52(2), 150-155.

- Johnson, D. R., Cushman, G. K., Borden, L. A., & McCune, M. S. (2013). Potentiating empathic growth: Generating imagery while reading fiction increases empathy and prosocial behavior. *Psychology of Aesthetics, Creativity, and the Arts*, 7(3), 306.
- Katz, P. A., & Ksansnak, K. R. (1994). Developmental aspects of gender role flexibility and traditionality in middle childhood and adolescence. *Developmental Psychology*, 30, 272–282.
- Keegan, R., & Andersen, T. (2016, February 8). Oscars 2016: Academy president acknowledges 'elephant in the room' at luncheon. *Los Angeles Times*. Retrieved from <http://www.latimes.com>.
- Kerfoot, D., & Knights, D. (1994). Power, Identity and the Gender Problematic. *Power/Gender: social relations in theory and practice*, 13, 67.
- Kidd, J. (2015). *Representation*. Routledge.
- Kortenhaus, C. M., & Demarest, J. (1993). Gender role stereotyping in children's literature: An update. *Sex Roles*, 28(3-4), 219-232.
- Lietz, P. (2006). A meta-analysis of gender differences in reading achievement at the secondary school level. *Studies in Educational Evaluation*, 32(4), 317-344.
- Levy, G. D., Taylor, M. G., & Gelman, S. A. (1995). Traditional and evaluative aspects of flexibility in gender roles, social conventions, moral rules, and physical laws. *Child Development*, 515-531.
- Luke, A., Cooke, J., & Luke, C. (1986). The selective tradition in action: Gender bias in student teachers' selections of children's literature. *English Education*, 209-218.
- Martin C. L., Ruble D. N., & Szkrybalo J. (2002). Cognitive theories of early gender development. *Psychological Bulletin*, 128, 903–933.
- Martin, C. L., & Ruble, D. (2004). Children's search for gender cues: Cognitive perspectives on gender development. *Current directions in psychological science*, 13(2), 67-70.
- McCabe, J., Fairchild, E., Grauerholz, L., Pescosolido, B. A., & Tope, D. (2011). Gender in Twentieth-Century Children's Books Patterns of Disparity in Titles and Central Characters. *Gender & Society*, 25(2), 197-226.
- McCreary, D. R. (1994). The male role and avoiding femininity. *Sex Roles*, 31(9-10), 517-531.

- McKenna, M. C., Kear, D. J., & Ellsworth, R. A. (1995). Children's attitudes toward reading: A national survey. *Reading Research Quarterly*, 934-956
- Merisuo-Storm, T. (2006). Girls and boys like to read and write different texts. *Scandinavian Journal of Educational Research*, 50(2), 111-125.
- Murray, C. B., & Mandara, J. (2002). Racial identity development in African American children. *Black children: Social, educational, and parental environments*, 73-96.
- Oakhill, J. V., & Petrides, A. (2007). Sex differences in the effects of interest on boys' and girls' reading comprehension. *British Journal of Psychology*, 98(2), 223-235.
- Peralta, E. (2016, January 18). Amid controversy, Scholastic pulls picture book about Washington's slave. *National Public Radio*. Retrieved from npr.org.
- Pho, D. M., Chorn, S., Rustad, M., Kwaymullina, A., & Hartness, J. (2015). *Invisible 2: Personal essays on erepresentation in SF/F*. Jim C. Hines.
- Poarch, R., & Monk-Turner, E. (2001). Gender roles in children's literature: A review of non-award-winning "easy-to-read" books. *Journal of Research in Childhood Education*, 16(1), 70-76.
- Price, Anne. 2006. *Children's catalog*. 19th ed. New York: H.W. Wilson Company.
- R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Rogers, K. (2016, July 19). Leslie Jones, star of 'Ghostbusters,' becomes a target of online trolls. *New York Times*. Retrieved from <http://www.nytimes.com>
- Rubin, D. B. (2011). Causal inference using potential outcomes. *Journal of the American Statistical Association*.
- Ruble, D. N., Taylor, L. J., Cyphers, L., Greulich, F. K., Lurye, L. E., & Shrout, P. E. (2007). The role of gender constancy in early gender development. *Child Development*, 78(4), 1121-1136.
- Särndal, C. E., Swensson, B., & Wretman, J. (2003). *Model assisted survey sampling*. Springer Science & Business Media.
- Schau, C. G., & Scott, K. P. (1984). Impact of gender characteristics of instructional materials: An integration of the research literature. *Journal of Educational Psychology*, 76(2), 183.

- Scaub, M. (2016, April 27). How did 'Space Raptor Butt Invasion' by Chuck Tingle become a Hugo finalist? *Los Angeles Times*. Retrieved from <http://www.latimes.com>.
- Scott, K. P. (1986). Effects of sex-fair reading materials on pupils' attitudes, comprehension, and interest. *American Educational Research Journal*, 23, 105–116.
- Segel, Elizabeth. 1986. "As the twig is bent . . .": Gender and childhood reading. In *Gender and reading: Essays on readers, texts, and contexts*, eds. E. A. Flynn and P. P. Schweickart. Baltimore: The Johns Hopkins University Press.
- Seltzer, S. (2011, February 14). The imminent whitewashing of the *Hunger Games* heroine. *Jezebel*. Retrieved from <http://www.jezebel.com>.
- Signorella, M. L., & Liben, L. S. (1984). Recall and reconstruction of gender-related pictures: Effects of attitude, task difficulty, and age. *Child Development*, 55, 393–405
- Signorella M. L., Bigler R. S., & Liben L. S. (1997). A meta-analysis of children's memories for own-sex and other-sex information. *Journal of Applied Developmental Psychology*, 18, 429–445.
- Spender, U. (1973). *Man made language*. London: Routledge & Kegan Paul.
- Stewart, D. (2012, March 26). Racist Hunger Games fans are very disappointed. *Jezebel*. Retrieved from <http://www.jezebel.com>.
- Summers, E. G., & Lukasevich, A. (1983). Reading preferences of intermediate-grade children in relation to sex, community, and maturation (grade level): A Canadian perspective. *Reading Research Quarterly*, 347-360.
- Taxel, J. (1981). The outsiders of the American revolution: The selective tradition in children's fiction. *Interchange*, 12(2), 206-228.
- Tepper, C. A., & Cassidy, K. W. (1999). Gender differences in emotional language in children's picture books. *Sex Roles*, 40(3-4), 265-280.
- Tin, M. (1965). Comparison of some ratio estimators. *Journal of the American Statistical Association*, 60(309), 294-307.
- Tognoli, J., Pullen, J., & Lieber, J. (1994). The privilege of place: Domestic and work locations of characters in children's books. *Children's Environments*, 272-280.
- Turner-Bowker, D. M. (1996). Gender stereotyped descriptors in children's picture books: Does "curious Jane" exist in the literature? *Sex Roles*, 35(7-8), 461-488.

Vežner, T. (2016, December 8). MOA's black Santa drew raves — and then a backlash online. *Pioneer Press*. Retrieved from <http://www.twincities.com>.

Worthy, J., Moorman, M., & Turner, M. (1999). What Johnny likes to read is hard to find in school. *Reading Research Quarterly*, 34(1), 12-27.