WORKING PAPER · NO. 2021-44

What We Teach About Race and Gender: Representation in Images and Text of Children's Books

Anjali Adukia, Alex Eble, Emileigh Harrison, Hakizumwami Birali Runesha, and Teodora Szasz

APRIL 2023

Becker Friedman Institute
FOR ECONOMICS AT UCHICAGO

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

What we teach about race and gender: Representation in images and text of children's books[*]

Anjali Adukia
University of Chicago and NBER

Alex Eble
Columbia University

Emileigh Harrison
University of Chicago

Hakizumwami Birali Runesha
University of Chicago

Teodora Szasz
University of Chicago

March 2023

Abstract

Books shape how children learn about society and norms, in part through representation of different characters. We introduce new artificial intelligence methods for systematically converting images into data and apply them, along with text analysis methods, to measure the representation of skin color, race, gender, and age in award-winning children's books widely read in homes, classrooms, and libraries over the last century. We find that more characters with darker skin color appear over time, but the most influential books persistently depict characters with lighter skin color, on average, than other books, even after conditioning on race; we also find that children are depicted with lighter skin than adults on average. Relative to their growing share of the U.S. population, Black and Latinx people are underrepresented in these same books, while White males are overrepresented. Over time, females are increasingly present but appear less often in text than in images, suggesting greater symbolic inclusion in pictures than substantive inclusion in stories. We then present analysis of the supply of, and demand for, books with different levels of representation to better understand the economic behavior that may contribute to these patterns. On the demand side, we show that people consume books that center their own identities. On the supply side, we document higher prices for books that center non-dominant social identities and fewer copies of these books in libraries that serve predominantly White communities. Lastly, we show that the types of children's books purchased in a neighborhood are related to local political beliefs.

[*]Contact: adukia@uchicago.edu, eble@tc.columbia.edu, harrisone@uchicago.edu, runesha@uchicago.edu, tszasz@uchicago.edu.
 We include the acknowledgements at the end of the document.

Education teaches children about the world, its people, and their place in it. Much of this happens through the curricular materials society presents to children, particularly in the books they read in school and at home (Giroux, 1981; Apple and Christian-Smith, 1991; Jansen, 1997; Van Kleeck, Stahl and Bauer, 2003; Steele, 2010). These lessons are conveyed, in part, through the inclusion or exclusion of characters of different identities in the images and text of books. Given that the content of books used for education has been shown to shape the later life beliefs of those exposed to them (Fuchs-Schündeln and Masella, 2016; Cantoni et al., 2017; Arold, Woessmann and Zierow, 2022; Arold, 2022), the presence or absence of these characters can contribute to how children see themselves and others, as well as their strengths and possible futures. In light of persistent racial and gender inequality in society (Darity and Mason, 1998; O'Flaherty, 2015; Blau and Kahn, 2017; Quillian et al., 2017; Craemer et al., 2020), and the potential importance of identity and representation in contributing to beliefs, aspirations, academic effort, and outcomes (Dee, 2005; Bian, Leslie and Cimpian, 2017; Gershenson et al., 2018; Porter and Serra, 2020; Riley, 2022), the content of books offers a policy lever for addressing these and related structural inequalities.

In this paper, we analyze representation in the content of children's literature. Specifically, we develop and apply computer science tools from the fields of computer vision and natural language processing to measure the representation of skin color, race, gender, and age in the images and text of influential children's books which are likely to appear in homes, classrooms, and libraries over the past century. These artificial intelligence tools allow for more scalable and systematic measurement than what would be possible using the traditional approach to content analysis, which historically has been done primarily "by hand" using human coders (Bell, 2001; Neuendorf, 2016; Krippendorff, 2018). We use these tools to measure how representation varies by identity, over time, and by type of book. We then present descriptive evidence of economic forces that contribute to these patterns.

Our main data set comprises children's books recognized by awards featured by the Association for Library Service to Children starting in the 1920s. We then divide these award-winning books into two main collections. Our first collection of books receive recognition for their literary or artistic value without explicit intention to highlight an identity group (i.e., the Newbery and Caldecott awards). We call this the "Mainstream" collection of books because of their general usage in mainstream outlets in the United States (U.S.), such as schools and libraries. Using daily book checkout data from a major public library system, we document that after books are recognized by a Mainstream award, they are checked out four times as often as other children's books. Using purchase-level data from over 1.5 million children's book sales, we find that books which were recognized by a Mainstream award sell over five times as many copies per title as other children's books. This corroborates qualitative accounts of how receipt of a Mainstream award establishes a book's membership in the "canon" of children's literature, as well as other accounts of changes in the sales of children's books after receipt of an award (Smith, 2013; Cockcroft, 2018; Koss, Johnson and Martinez, 2018). It also highlights the particular societal influence these books may have and underscores the importance of understanding the messages they transmit. Books in our second collection received recognition for both their literary or artistic value and for how they highlight experiences of specific identity groups. These include awards such as the Coretta Scott King and Rise Feminist awards. We term these the "Diversity" collection. Given their focus, we posit that they provide a potential "upper bound" on representation in children's books in the market.

Our analysis characterizes several dimensions of the representation of skin color, race, gender, and age in the content of these children's books, including how levels of representation have endured, or changed, over time. We find that, over time, these books include more characters with darker skin, but those in the Mainstream collection are significantly more likely to depict lighter-skinned characters than those in the Diversity collection. This pattern remains even when comparing pictured characters with the same predicted race classification. Across all collections, children are more likely than adults to be shown with lighter skin, despite there not being a definitive biological foundation for any systematic difference. Regardless of the reasons behind this difference, our estimates show that lighter-skinned children see themselves represented more often in these books than do darker-skinned children. In addition, we show that Black and Latinx people have been historically underrepresented relative to their share of the U.S. population, corroborating prior work on the representation of race in smaller subsets of these collections of books (Valadez, Sutterby and Donaldson, 2013; Koss, 2015; Koss, Johnson and Martinez, 2018).

We compare the incidence of female-presenting appearances in images to female mentions in text, and we see that females are consistently more likely to be visualized (seen) in images than mentioned (heard) in the text. This suggests there may be symbolic inclusion of females in pictures without their substantive inclusion in the actual stories. Furthermore, females are persistently less likely than males to be represented in the text of books in our sample overall and over time. This finding is consistent across all of the measures we use: pronoun counts, specific gendered terms, gender of famous individuals, and predicted gender of character first names. This generalizes results from prior analysis of the representation of gender in studies focusing on smaller subsets, or a small number of specific features contained in these books (c.f., Weitzman et al. 1972, Davis 1984, Crisp and Hiller 2011). Our analysis of age reveals another surprising result: even though these books are targeted to children, adults are depicted more often than children in both images and text.

Our results build on the rich existing history of manual content analysis. Prior work documents low levels of representation of females and historically minoritized racial groups (c.f., Weitzman et al. 1972; Williams Jr et al. 1987; Hamilton et al. 2006; Koss 2015). These studies often focus on representation solely in prominent places in the images and text, for example, in the images on the cover of the book or in text regarding the main character. We confirm these results in a much larger number of books and in a far greater number of sites within each book than would be possible via manual content analysis, given its time and other cost constraints. These advantages allow us to characterize certain parameters – such as trends in representation over time – more conclusively than prior contributions (c.f., Clark et al. 1999; Crisp and Hiller 2011; Koss, Johnson and Martinez 2018). We also ask several novel questions, for example, characterizing representation in images and text at the intersection of multiple sites of exclusion - skin color, race, gender, and age - and comparing representation between images and text within a given book.

The second part of our paper describes and explores a set of economic forces which may contribute to these patterns of representation, and which can help explain how the messages in these books may propagate through society and across generations. We first discuss theoretical and empirical work characterizing these forces on both the supply- and demand-side, and then present descriptive evidence of their incidence.

On the supply side, prior research on the economics of the media suggests that, due to fixed costs and other market frictions, books centering non-dominant social identities will be under-produced relative to demand for them, and these books will be priced at a higher level than other books (Waldfogel, 2003, 2007). Examining book-level price and purchase data, we find evidence consistent with both phenomena. We also show that there are fewer copies of children's books recognized for highlighting underrepresented identities in libraries that serve predominantly White communities.

On the demand side, we draw from related theoretical work on the economics of identity from Akerlof and Kranton (2000), which suggests that people are more likely to consume books which center identities similar to their own. Using purchase-level data on book consumption linked to consumer demographics and checkout-level data from libraries, we find several patterns consistent with this. Males purchase books with fewer female words and images than females. White purchasers, on average, consume books with characters that have lighter skin color; whereas, Black and Latinx purchasers consume books with characters that have darker skin, on average. In a related analysis tracking trends over time, we document that as the market share of under-represented identities grows, so does their likelihood of being represented in these books.

To understand how local book consumption relates to local consumer beliefs, we link our individual-level book purchase data to the Cooperative Election Study (CCES), a nationally representative, stratified sample survey collecting information about general political attitudes connected to respondent demographics. We show that the type and volume of books purchased in a given neighborhood align with the political viewpoints held by residents of that neighborhood on issues related to race and immigration: in areas where people hold more progressive views, the books purchased contain more diverse representation than do books purchased by people in areas with more conservative views.

In summary, our paper makes three key contributions. First, we develop and hone a series of tools from the field of computer vision to systematically process images into analyzable measures of representation; this includes introducing a novel computational method to measure skin color. Second, we apply these image-to-data tools alongside established natural language processing tools to measure the representation of skin color, race, gender, and age in the images and text contained within a century of influential children's books, and document how this changes over time. Third, we describe economic forces on the supply and demand side that may contribute to these levels of representation, and then present empirical evidence showing how the pressures from these forces may contribute to persistent overrepresentation of historically dominant identities. Using data on local book consumption and local consumer beliefs, we then show that the levels of representation contained in the books people buy are highly correlated with their views on race and immigration. Given that the books used to teach children shape the beliefs these people hold when they are adults (Fuchs-Schündeln and Masella, 2016; Cantoni et al., 2017; Arold, 2022; Arold, Woessmann and Zierow, 2022), the patterns in children's book purchases we document may help explain the persistence and intergenerational transmission of related beliefs (Bian, Leslie and Cimpian, 2017; Dhar, Jain and Jayachandran, 2018; Eble and Hu, 2022).

This paper proceeds as follows. Section I presents background on the importance of representation. Section II describes the books in our data and their influence. Section III discusses prior work on content analysis. Section IV describes the image and text analysis tools. Section V presents the patterns of representation we uncover. Section VI presents descriptive evidence underlying market forces influencing levels of, and trends in representation. Section VII concludes.[1]

[1]The appendix includes: further analysis; a methods appendix containing a discussion of the potential

I Background: The Importance of Representation

Institutional practices, public policies, and cultural representations reflect values that society assigns to specific groups. In a broad range of cultural products, from news media and history books to children's movies, people who do not belong to the culturally dominant group are often absent or portrayed through negative stereotypes (Weitzman et al., 1972; O'Kelly, 1974; Stewig and Knipfel, 1975; Dobrow and Gidney, 1998; Balter, 1999; Witt, 2000; Towbin et al., 2004; Brooks and Hébert, 2006; Martin, 2008; Paceley and Flynn, 2012; Daniels, Layh and Porzelius, 2016). Research from different disciplines suggests that this inequality in representation is a means through which societal inequality in other outcomes can persist. For example, the genderedness of representations in language and folklore across societies are both negatively correlated with gender equity in education, labor force participation, and other social roles (Jakiela and Ozier, 2018; Michalopoulos and Xue, 2021). In addition, debates over the content of what is taught in schools – exemplified by recent attention, controversy, and confusion over the concept of critical race theory – underscore the need to catalog and know what is taught via curricular materials, and what is absent.

One mechanism through which inequality of representation may contribute to inequality in outcomes is through its potential to instill beliefs about who belongs in which societal domains. In particular, experiences of exclusion from certain spaces in society may reduce the "capacity to aspire" in those spaces (Appadurai, 2004), and the absence of identity-specific positive examples of success can lead to a distorted view of the path from present action to future outcomes (Wilson, 2012; Genicot and Ray, 2017; Eble and Hu, 2020). This forms a potential self-reinforcing loop: not seeing such examples may diminish a child's expected return to effort. If that change in expectation reduces actual effort, it may lower performance, thus reinforcing the message behind the (once-erroneous) message. This highlights the importance of addressing inequality in representation within educational content.

Curricular materials are designed and used with the intent to shape children's development into adults and their views of the world around them, and are likely to make important contributions to the formation of children's social preferences (Cappelen et al., 2020; Alan et al., 2021; Dhar, Jain and Jayachandran, 2022). Several studies in economics have shown that exposure to variation in content among textbooks ranging from subjects as diverse as history and religion can lead to variations in later-life beliefs (Fuchs-Schündeln and

benefits and concerns to using AI models, a cost effectiveness analysis, further information on the methods used, and a discussion the relationships between computational content analysis and traditional manual approaches; additional analyses quantifying the increase in library checkout demand for books following receipt of an award; details on award criteria; and qualitative interviews with suppliers of children's books.

Masella, 2016; Cantoni et al., 2017; Arold, Woessmann and Zierow, 2022; Arold, 2022).[2,3] In education research, scholars have shown how children's literature can be used in middle school language arts and social science curricula to shape beliefs about self, community, and civic action (Tyson, 2002; Levstik and Tyson, 2010).[4]

These materials also have the potential to shape how children view *others* of different identities. When children do or do not see others represented, their conscious or unconscious perceptions of their own potential and that of groups with identities different than theirs can be molded in detrimental ways and can erroneously shape subconscious defaults. For example, the representations that children see can shape the beliefs of members of the dominant group about the capacity of members of the underrepresented group to participate in different spheres of society (Hughes, Bigler and Levy, 2007; Marx, Ko and Friedman, 2009; Plant et al., 2009; Alrababah et al., 2021).

Broadening representation to be more inclusive also has been shown to influence the beliefs, actions, and learning of children. In economics alone, changes in representation have been shown to influence these outcomes for females (Beaman et al., 2012; Stout et al., 2011; Porter and Serra, 2020), and, separately, people of underrepresented racial and ethnic identities regardless of gender (Kearney and Levine, 2020; Riley, 2022). While not a panacea, such "subject-object identity match" – e.g., teacher-student identity match, or content-reader identity match – can help improve academic performance for students, and may function via a wide range of potential channels.[5]

II Context: Award-Winning Children's Books

We focus on the content of series of books that are particularly likely to appear in the homes, schools, and libraries of a large proportion of children in the U.S. Specifically, we study the representation contained in the images and text of books recognized by any of 19 awards administered or featured by the Association for Library Service to Children (ALSC),

[2]In psychology, there is mixed evidence whether deliberately manipulated exposure to content shapes child beliefs; see the review in Bigler (1999), as well as the pair of randomized controlled trials reported in Hughes, Bigler and Levy (2007).

[3]Though children are more likely than adults to change their beliefs in response to stimuli (Gopnik et al., 2017), recent evidence from political science shows that a change in the content of the media consumed can change even adults' beliefs (Broockman and Kalla, 2022).

[4]The age of the child consuming the books may influence the way the content is parsed and internalized. This is also likely to vary with the nature of the content.

[5]These include, but are not limited to: by reducing stereotype threat (Steele and Aronson, 1995); by potentially increasing the perceived likelihood of different possible futures for the individual (Wilson, 2012); and by expanding the perceptions and assumptions of those in majority-represented groups who thereby may be less likely to limit access to opportunities (Plant et al., 2009; Alrababah et al., 2021).

a division of the American Library Association (ALA).[6] These began honoring children's books in 1922, and continue to the present.

In this section, we describe these awards and how we separate them into "collections" of awards with distinct sets of goals. We then provide descriptive analyses quantifying changes in book consumption associated with being recognized by these awards.

II.A Collections of Books

In our analyses, we divide books into "collections." These reflect commonalities in goals across the various awards, and allow us to characterize how representation differs between sets of books recognized by awards with different goals. Many of our analyses focus on comparing representation between books in two primary collections: (i) "Mainstream" books considered to be of high literary or artistic value,[7] and (ii) "Diversity" books selected because of how they center experiences of specific underrepresented identity groups in addition to their high literary value.

Mainstream Collection. The Mainstream collection comprises books recognized by either the Newbery or Caldecott awards, the two oldest children's book awards in the U.S. The Newbery Medal, first awarded in 1922, is given to authors of books that are considered to be the "most distinguished contribution to American literature for children." The Caldecott Medal, first awarded in 1938, is given to illustrators of "the most distinguished American picture books for children." These books are explicitly chosen for their literary or artistic quality and not their popular appeal per se. Books receiving these awards are considered to be of general interest to all children and are likely to be incorporated into mainstream outlets for children, such as school libraries and curricula (ALSC, 2007; Koss, Johnson and Martinez, 2018). We provide further evidence of the empirical relationship between recognition and consumption of books in Section II.B. The primary goal for studying these books is to understand the representation contained in a set of books to which a large proportion of children in the U.S. are exposed.

Diversity Collection. The Diversity collection comprises book awards featured by the ALSC that center the experiences of excluded or marginalized identities. These books are also likely to be placed on "diversity lists" during events such as Black History Month or

[6]We selected children's book awards featured on the ALSC website at the time of writing this paper, many of which are administered by different organizations.

[7]We use the term "Mainstream" to refer to these books in order to capture the fact that, as shown by scholars of these books – such as Koss (2015), Koss, Johnson and Martinez (2018), and Koss and Paciga (2020) – these books are the most influential award-winning books for the outcomes we study. We do not intend this to connote any affirmation of centrality or default beyond its description of the historical prominence of these books.

Women's History Month. We study the representation contained in these books for multiple reasons: one, to estimate a potential "upper bound" on representation in children's books in the market; two, to measure the efficacy of these books in highlighting the identity on which they focus; and three, to measure the levels of representation of historically excluded identities beyond the identity on which a given award focuses. We use this last feature to assess the extent to which these books have greater, similar, or less representation of identities which exist at the intersection of multiple sites of exclusion.

This collection includes books recognized by the following awards: American Indian Youth Literature, Américas, Arab American, Asian/Pacific American Award for Literature, Carter G. Woodson, Coretta Scott King, Dolly Gray, Ezra Jack Keats, Middle East, Notable Books for a Global Society, Pura Belpré, Rise Feminist,[8] Schneider Family, Skipping Stones Honor, South Asia, Stonewall, and Tomás Rivera Mexican American awards. The first of these awards was the Coretta Scott King Award, created in 1970 specifically to recognize African American authors and illustrators of books that "demonstrate an appreciation of African American culture"; this award was introduced, in part, because no African American writer had been recognized by a Newbery or Caldecott medal up to that point. Other awards were created more recently, such as the South Asia Book Award, which began in 2012.

We also create smaller collections of these awards that highlight the following specific identities: people of color, African American people, females, people with disabilities, and people who identify as lesbian, gay, bisexual, transgender, and/or queer (LGBTQIA+). We show the list of corpora by collection and their relative sample sizes in Appendix Figure B1.

Each award has a single "medalist" or "winner" of the award. Many awards also have recognition for leading contenders, who are often called "honorees." In our main analysis we refer to the superset of these two groups as those "recognized" by the award. In some analysis in Section II.B, however, we examine trends in consumption separately by medalists and honorees. In Appendix F, we describe the criteria used by each award for recognizing books in greater detail.

We present collection-level summary statistics of the books in our sample in Table 1, which include average representation of skin color, putative race, gender, and age.

II.B Quantifying the Importance of Mainstream Awards

Mainstream awards are considered to be highly influential, with recognition by either the Newbery or Caldecott Awards placing books into the "canon" of children's literature and making them a common feature in homes and libraries (Smith, 2013; Koss and Paciga,

[8]The Rise Feminist Award was formerly known as the Amelia Bloomer Award.

2020). Winners are commonly featured in venues that are part of children's learning experiences, from book fairs and catalogues to school curricula and summer reading lists (Knowles, Knowles and Smith, 1997). Publishers in the industry take cues from the winners for guidance in what to publish, given the large boost in sales that the award stimulates, and children's librarians ensure the books' presence in their inventories once they receive the award (Nilsen, 1971; Weitzman et al., 1972).

We further establish the importance of these awards in children's experiences by estimating the relationship between receipt of these awards and book popularity. Our analyses use data on three measures of book consumption: (1) library checkouts, (2) book purchases, and (3) internet searches. Each measure captures a different – but not mutually exclusive – set of consumer preferences.

Library Checkout Data. We draw from publicly available, book-level, daily checkout data from the Seattle Public Library system spanning the period from 2005 to 2017. Public libraries aim to serve all members of their communities, regardless of socioeconomic status. Library usage is common in the U.S., with approximately half of the population accessing a public library at least once each year (Horrigan, 2015).

Book Purchase Data. We obtain book purchasing data from the Numerator OmniPanel, a large panel data set with information from over 1 billion shopping trips from over 44,000 retailers from 2017-2020 We limit our analyses to purchases of children's books. Each purchase is matched to detailed demographic information on the consumer making the purchase, including their gender, race, and the genders and number of their children. We describe book purchaser characteristics in Appendix Table A1. For example, wealthier people and people with more formal education are more likely to purchase children's books.

Google Trends Data. We use data on the volume of internet searches from Google Trends as a measure of general interest in the book awards found within our sample.[9] We limit our analysis to awards that have topic IDs in the Google Trends data.[10] Data measuring search interest for each topic ID are scaled on a range of 0 to 100 based on a topic's search proportion to total searches in the U.S. over a given time range (e.g., the week of December 12, 2016). We sum weekly search interest across all topic IDs corresponding to awards in a given collection to get aggregate weekly search interest for that collection.

[9]Google Trends filters Google search requests to remove duplicate searches, uncommon searches, and searches with special characters. Google Trends draws from a random sample of internet searches.

[10]Awards with topic IDs include the Amelia Bloomer Project (renamed Rise Feminist), Caldecott Medal, Coretta Scott King Award, Ezra Jack Keats Book Award, John Newbery Medal, Pura Belpré Award, Schneider Family Book Award, and Stonewall Book Award.

We present three event studies that show average daily library checkouts (Figure 1a), average daily purchases (Figure 1b), and average weekly search interest by collection (Figure 1c), centered around the time when awards are announced.[11] In Figures 1a and 1b, we disaggregate the data by books which were recognized by a Mainstream medal or honor in that year, books which were recognized by a Diversity medal or honor in that year, and all other children's books.[12]

First, we see that library checkouts of books selected for Mainstream awards increase substantially after announcement of awards with a larger increase for books that received the medal as opposed to receiving an honorable mention.[13] This persists for at least two years after the award announcement, during which average daily checkouts of books in the Mainstream collection plateau at a rate approximately four times that of the comparator groups. The increase in library checkout rates for books in the Diversity collection after the award announcements are substantially smaller in magnitude and, as expected, we see no change around the award announcement of checkouts for other children's books.[14]

Second, we see a similarly sustained increase in purchases of books belonging to both the Mainstream and Diversity collections after the award announcements, again with a larger increase for Mainstream books. This finding corroborates past analyses of publisher-level data on book sales, which document large gains in sales – of similar or even larger magnitudes – after a book receives an award (Nilsen, 1971; Weitzman et al., 1972; Cockcroft, 2018).

Finally, we find similar patterns in internet search interest: Google search volume for awards belonging to the Mainstream collection is approximately seven times higher than search interest for awards belonging to the Diversity collection, with a spike in search interest immediately following the announcement of the awards.

As a whole, this evidence suggests that Mainstream books have greater influence than other children's books, and children are more likely to be exposed to the messages in these books. This is consistent with and reinforces the findings of previous qualitative analysis of their central role in children's literature and previous quantitative analysis of book sales data.

[11]We describe the empirical specification and data cleaning details in the Data Appendix.

[12]These include books that did not receive one of the awards in our study, but they may have received recognition from a different source.

[13]Most of these awards are presented annually, and many award recipients are announced at the ALA's Midwinter Meeting, which typically occurs near the end of January. To be eligible for these awards, a book must be published between February of the previous year and January of that year.

[14]We discuss and present our analyses of these data in greater detail in Appendix E).

III Prior Work and the Need for Scalable Measurement Tools

The field of content analysis studies the content of books, including the representations contained within them. Historically, content analysis has been conducted primarily by humans reading carefully through text, images, or other media while coding the presence of certain words, themes, or concepts by hand (Neuendorf, 2016; Krippendorff, 2018). Prior work has studied the content of some of these award-winning books, including the representation of gender and, more recently, race. An influential study by Weitzman et al. (1972) examined the gender representation throughout the text of 18 Caldecott award recipients published over a five-year period, documenting that females were less likely than males to be represented in the content of the books; when they were represented, these representations often reinforced traditional gender roles. Many studies since have measured the representation of race, gender, and other identities in various, smaller subsets of these books published in specific time windows (c.f., Kolbe and La Voie 1981; Davis 1984; Williams Jr et al. 1987; Clark, Lennon and Morris 1993; Crisp and Hiller 2011; Koss 2015; Koss, Johnson and Martinez 2018). They find that the books in their samples often underrepresent women and people of color, relative to males and White people, though there is not consensus as to whether these patterns attenuate or persist over time.[15]

The time and other costs it takes to perform manual content analyses constrain the sample size and scope of the analysis that can be performed in a given study. The sample sizes of most studies range from between a few dozen books to – with rare exceptions – at most one or two hundred (Weitzman et al., 1972; Davis, 1984; Crisp and Hiller, 2011; Koss, Johnson and Martinez, 2018). The few studies with a larger scope (500 to several thousand books) focus only on one or a small number of sites of representation – for example, the title of the book, the illustrations on its cover, or the main character instead of the full content of the book (Weitzman et al., 1972; McCabe et al., 2011; Koss, 2015; Koss, Johnson and Martinez, 2018).

By using computational content analysis to measure representation in both the images and the text of these books, and by expanding the set of tools available to do so, we build on and advance the findings and the scope of the existing content analysis literature. Here we summarize a set of key advantages – and thus advances – of a computational approach. We then discuss a set of potential limitations related to our approach and describe our efforts to measure cost-effectiveness and validity.

Improved speed and reduced cost allow the study of more books. First, computational

[15] These differences often coincide with differences in focus or choice of sample collection or time period.

content analysis can be used to systematically analyze features in large bodies of content in a short amount of time. Due to their size, these bodies of content were previously beyond the reach of traditional manual content analysis. Using these tools, we characterize the representation of all detected gendered terms, named characters, and pictured characters detected in over 1,100 books. This is one or two orders of magnitude larger than most prior studies. To conduct further computational analysis of an even larger collections of books would incur minimal additional cost beyond the digitization of the material.

Greater scope for measurement within each book. Computational tools are able to measure more sites of representation in each book. This includes both the ability to analyze all pictured and named characters detected in the book's images and text – as opposed to just the main characters, as is common in much manual content analysis (c.f., Koss 2015; Krippendorff 2018) – and to analyze a wider variety of features of each character. By contrast, resource constraints limit the number of characters and dimensions of representation that can be measured using manual analysis. Studies that use manual content analysis on a larger sample explicitly indicate compensating the cost implications by focusing on a smaller number of prominent features, such as the book's title, the images on its cover, and/or the identities of only the main characters (Kortenhaus and Demarest, 1993; McCabe et al., 2011; Koss, 2015; Koss, Johnson and Martinez, 2018).

Another contribution we make in this direction is to measure the incidence of representation of identities at the intersections of multiple sites of exclusion in a large number of books, and a large number of sites within each book. Different aspects of identity – such as race, gender identity, class, sexual orientation, and disability – do not exist separately from each other, but rather are inextricably linked (Crenshaw, 1989, 1990; Ghavami, Katsiaficas and Rogers, 2016). The notion of "intersectionality" refers to the unique experiences of people whose identities lie at one or multiple intersections of marginalized identities. For example, the experiences of Black women cannot merely be summarized by a description of the experiences of all women and, separately, the experiences of all Black people. It is important to note that intersectionality does not merely refer to an "interaction effect" (e.g., between race and gender), but rather the distinct experiences of individuals whose identities exist at intersections of multiple dimensions of marginalization.

We draw on a central insight from the large body of work on intersectionality in our analysis: when analyzing representation of different dimensions of identity, such as race and gender, it is critical to characterize the power imbalances and their manifestations that lead to greater disadvantage among individuals at the intersection of multiple marginalized identities. We acknowledge that a more developed intersectional analysis requires a widereaching analysis of norms, rules, laws, and history that is beyond the scope of our study. Instead, the starting point for our analysis is that a key site of power – and thus of potential power imbalances – are the messages contained within the material we use to teach children. More specifically, the inclusion or exclusion of identity groups in this content is a fundamental expression of power, as it signals to the reader the spaces that these identities do or do not occupy in society (Crenshaw, 1989, 1990; Davis, 2008; Ghavami, Katsiaficas and Rogers, 2016). Children's books are an important site of the exercise of societal power, given the potential for such content to shape the beliefs, norms, and conceptions of history that the next generation will adopt (Fuchs-Schündeln and Masella, 2016; Cantoni et al., 2017; Arold, Woessmann and Zierow, 2022; Arold, 2022). In this paper, we explore whether there is differential representation of identities at the intersections of multiple sites of marginalization within dimensions of skin color, race, gender, and age.

Greater flexibility and scalability. Separate from scope, our approach has the benefit of yielding greater flexibility and scalability. In a given study, if re-analysis or new analysis is required after the initial coding, the fixed costs of identifying, hiring, and training coders are again incurred. In computational content analysis, the only additional costs are the costs of digitizing material, the computational power necessary to re-run the analysis, and human input to adjust the code. Our approach avoids these and other related costs, allows for greater flexibility in expanding or changing a study's scope mid-stream, either by adding dimensions of analysis within books, or by adding additional content.

Reliability. In manual content analysis, inter-rater reliability is a core concern which increases with scale (Neuendorf, 2016; Krippendorff, 2018).[16] In computer-driven analysis, however, these concerns do not vary with scale, as the traits of the coder are held constant.

Limitations. There are important limitations to our overall analysis. First, while we focus on representation in light of its important role in the processes we describe above, it is only one component of the complex, larger societal processes we are trying to describe. There are myriad structural barriers to racial and gender equality woven throughout the organizations, laws, and customs of our society (Darity and Mason, 1998; O'Flaherty, 2015; Blau and Kahn, 2017; Quillian et al., 2017; Muhammad, 2019; Chetty et al., 2020). Second, there are limitations even within our focus on representation. The key historical advantage of manual content analysis, as opposed to computer-led content analysis, has been its ability to measure more complex and nuanced understanding than those a computer may capture

[16]Once the AI is trained, it conducts its analysis with the same level of replicability, irrespective of scale. In manual content analysis, the cost of maintaining reliability of raters increases as the number of raters increases, as it incurs additional costs of training and supervision to ensure fidelity.

(Rosenberg, Schnurr and Oxman, 1990; Linderman, 2001). We focus our analysis on measuring the *presence* of different identities, a domain for which computer-driven analysis is particularly suitable.[17] In this regard, computational content analysis and manual content analysis represent complementary approaches, and there are important limitations to our application of AI tools to these content analysis tasks.

Cost-effectiveness and validation. In Appendix Section D.A, we describe the cost-effectiveness and other advantages of using AI in our specific context. In that section, we also discuss two important dimensions of our work. First, we describe how we use manual content analysis to validate our computer driven measures of representation. Second, we explain how both manual and computational content analysis reflect human-introduced biases in measurement, and describe how these biases can be minimized.

IV Methods and Data

In this section, we describe the methods we use to create data from the images and text in books.

IV.A Methods: Images as Data

Currently, images are neither widely nor systematically analyzed in social science research despite the richness of information they contain, as alluded to by the maxim "a picture is worth a thousand words." This leaves an important data source "on the table" (i.e., unused), in contrast to the use of text as data, which has seen growing attention from social science in the past 15 years (Gentzkow and Shapiro, 2010; Gentzkow, Shapiro and Taddy, 2019; Kozlowski, Taddy and Evans, 2019). Images may be particularly important in children's books, especially for those who are not yet textually literate (Pressley, 1977; Sadoski and Paivio, 2013), and several studies have shown that better comprehension can result from learning with text and pictures (i.e.,multimedia) compared to learning with text only (Levie and Lentz, 1982; Vekiri, 2002; Fletcher and Tobias, 2005; Eitel et al., 2013).

We introduce, develop, and apply tools for computational analysis of the content of images. Specifically, these tools first identify pictured faces of characters and then classify their skin color, "putative" race,[18] gender, and age. We depict this process in Figure 2a and refer to it as our "Image-to-Data Pipeline."

[17] Understanding patterns in the manner in which characters are represented is also important, and we are pursuing this work in separate projects (c.f. Adukia et al. (2022a), Adukia et al. (2022b).

[18] We define putative race to be the race that society assigns to a person.

IV.A.1 Image Feature Classification: Face Detection

Our first step in converting images to data is to detect the face of each pictured character. Machine-led face detection, however, poses a set of complex problems. First, images in these books consist of both illustrations and photographs. Because the current state-of-the-art face detection models were trained exclusively on photographs, these models are likely to undercount faces in illustrated images. This concern is amplified by the large proportion of illustrations in our data: in a random sample of manually labeled images, we found that over 80 percent were illustrations, as opposed to photographs. Second, these images contain both human and non-human characters. Non-human characters could have human skin colors (e.g., different shades of beige and brown), non-typical skin colors (e.g., blue or green), or monochromatic skin colors (e.g., grayscale or sepia). Third, characters could be shown in different poses, such as facing the viewer, shown in profile, or facing away from the viewer, a challenge for models trained to recognize faces shown from the front.

To address the potential undercounting of characters in illustrations, we trained a custom transfer learning model to detect and classify both illustrated and photographic faces using Google's AutoML Vision (Zoph and Le, 2017).[19,20] We trained our face detection model using a manually labeled data set of 5,403 illustrated faces from our sample, which contains a wide variety of illustrated characters.[21] This process is described in greater depth in Szasz et al. (2022), and we present further detail on it in Methods Appendix D.

IV.A.2 Image Feature Classification: Skin Color

Skin color is an important dimension of how humans categorize each other. Distinct from race, skin color is itself a site of historical and ongoing discrimination with impacts on health and the labor market (Hersch, 2008; Monk Jr, 2015). From a measurement perspective, it is a parameter for which we can use computers to more clearly measure the "ground truth," since the computer directly observes the color of each individual pixel as compared to the categorization of putative race, which varies by observer and cultural context.

[19] Transfer learning is a process which facilitates the use of a pre-trained model as a "shortcut" to learn patterns from data on which it was not originally trained. This mitigates concerns around having a sufficiently large amount of manually labeled data necessary to train deep learning models, particularly in the absence of public data sets using illustrations.

[20] Google is migrating workflows from AutoML to Vertex AI. They have similar functionality, but our models in this paper used AutoML. People who wish to use these approaches in future will use Vertex AI.

[21] We refer to this data set as IllusFace 1.0 (Szasz et al., 2022). We refer to our face detection model as FDAI (face detection using AutoML trained on illustrations). We use two parameters to evaluate the performance of our face detection model: "precision" and "recall." Our face detection model has 93.4 percent precision and 76.8 percent recall in our testing data. In other words, 6.6 percent of the faces we identify may not, in truth, be faces (a false positive), while the model may neglect to identify one in 4 "true" faces (a false negative).

Our skin color classification method involves a three-part process: (1) "segmenting" the skin portion of each face to separate the parts of the face which contain skin from other facial features; (2) extracting the predominant colors in the identified skin and collapsing these colors into a single representative skin color; and (3) constructing measures of skin color. Figure 2a illustrates this process. We discuss each of these steps broadly below and in greater detail within the Methods Appendix.

Skin segmentation. We begin by isolating skin components from non-skin components of each detected face using a deep learning approach called Fully-Connected Convolutional Neural Network Continuous Conditional Random Field (FC-CNN CRF).[22] This process of "skin segmentation" comprises three steps (Jackson, Valstar and Tzimiropoulos, 2016; Zhou, Liu and He, 2017; Beyer, 2018; Lu, 2018). First, we apply a fully-connected convolutional neural network (FC-CNN).[23] This allows us to predict periphery landmarks such as the edges of the facial skin area, eyes, nose, and mouth. Second, we then use these predicted landmarks to extract a convex hull "mask" for the targeted facial region. Third, we refine this mask by applying a continuous conditional random field (CRF) module, which predicts the labels of neighboring pixels (i.e., whether they are predicted to be skin or not skin) to produce a more fine-grained segmentation result.[24] We classify skin color from the resulting face mask.

Representative skin color. We then identify the predominant colors in this face mask (e.g. the segmented skin) by using k-means clustering to group the colors of each pixel into distinct clusters in RGB color space. k-means clustering is a traditional unsupervised machine learning algorithm whose goal is to group data containing similar features into k clusters.[25] For our analysis, we partition all the pixels in the segmented skin into five clusters (i.e., where k takes a value of five), and we drop the pixels in the smallest two clusters as they tend to represent shadows, highlights, or non-skin portions of the detected face. We take the centroid of each of the remaining three largest clusters – which provide the dominant skin colors in the segmented skin – and use a linear mapping to convert these three values from RGB color space into the CIELAB, or L*a*b*, color space.[26] After this conversion, we collapse the dominant skin colors into a single color by taking the weighted average of

[22]Further information about how our skin segmentation approach improves upon traditional approaches can be found in the Methods Appendix D.B.2.

[23]FC-CNN is a type of convolutional neural network (CNN) where the last fully-connected layer is substituted with a convolutional layer that captures locations of the predicted labels.

[24]CRF is a class of statistical modeling using a probabilistic graphical model.

[25]We used the k-means clustering function in the the scikit-learn Python library Sculley (2010).

[26]We convert colors from RGB space to L*a*b* space before averaging because L*a*b* color space – unlike RGB color space – is perceptually linear.

their L*a*b* values, where the weights correspond to the proportion of pixels assigned to the cluster from which each of the top three dominant skin colors came. This weighted average provides our measure of each face's representative skin color.

Skin color classification: Perceptual tint and skin color type. Once we have a representative skin color, we can measure how light or dark the skin color of each face is on a scale of 0-100 (where 0 is the darkest and 100 is the lightest) using the L* value from the representation of each face's representative skin color in L*a*b* color space. This measure reduces the dimensionality of skin color to a single value and provides us with our main skin color measure of interest which we call "perceptual skin tint."[27] A given numerical change in the skin tint value can be interpreted as a similar perceived change in the darkness/lightness of a color. We also divide this continuous measure of skin tint into three terciles (darker, medium, or lighter) for a coarser, but more intuitive, skin color classification.

We also separate the representative skin colors into three types: (1) polychromatic human skin colors (e.g., brown, beige), (2) monochromatic skin colors (e.g., grayscale), and (3) polychromatic non-typical skin colors (e.g., blue, green). We discuss how we separate skin colors into these three types in Methods Appendix Section D.B.3. In Figure 3, we show the representative skin colors of over 44,000 individual faces detected in each collection by the three skin color types present in these images.[28] The x-axis indicates perceptual tint and the y-axis indicates vibrancy of each representative skin color.

IV.A.3 Image Feature Classification: Race, Gender, and Age

In order to classify putative race, gender, and age of detected faces in images, we trained a multi-label classification transfer learning model using Google's AutoML Vision platform. This model was trained on the UTKFace public data set which contains over 20,000 faces manually labeled with race, gender, and age (Zhang and Qi, 2017).[29,30]

[27] A more common term for L* is "perceptual lightness," but to de-center and de-emphasize "lightness" or "brightness" relative to "darkness," we refer to the concept as "perceptual tint," or "skin tint."

[28] We show these for each collection by decade for human skin colors (Appendix Figure B2), monochromatic skin colors (Appendix Figure C1), and non-typical skin colors (Appendix Figure C2). We find that in the earlier decades of the Mainstream collection, there was a greater proportion of monochromatic images, with a general trend over time to have more polychromatic images. In the Diversity collection, and in particular the People of Color collection, there is a consistently high proportion of monochromatic images, perhaps representing the use of historical black-and-white photographs. Note that even though we detect over 54,000 faces in our sample of children's books, we are only able to get a usable skin segmentation for 81 percent of the faces because a CNN-based skin segmentation approach does not work on all illustrated faces.

[29] The labels in the data set include: Gender (female or male), Age (infant (0-3), child (4-11), teenager (12-19), adult (20-64), senior (65+)), Race (Asian (a combination of Asian and Indian), Black, White, and others (e.g., Latinx, Middle Eastern)).

[30] The resulting model has 90.6 percent precision and 88.98 percent recall in our testing data. We provide additional detail in the Methods Appendix.

Our model assigns probabilities that a detected face is of a given race, gender, and age, respectively. Within each dimension, we classify a face with the identity to which the model gives the highest predicted probability.[31] The main limitation of this model is that it was trained on photographs, which means that the predictions will be more accurate for photographs of faces than illustrated faces.[32]

IV.B Methods: Text as Data

In this section, we describe the tools we use to measure representation in the text of books. Researchers have manually analyzed (i.e., by hand) the messages contained in text of printed material for centuries, a process which is highly resource intensive in terms of both labor and time (Neuendorf, 2016; Krippendorff, 2018). Recent work by economists and sociologists showcases how the computational speed and power of (super)computers can be harnessed to conduct computational text analysis, greatly accelerating the speed of work which would have traditionally been done manually (Gentzkow, Kelly and Taddy, 2019; Kozlowski, Taddy and Evans, 2019). We draw from this work and, in particular, a series of natural language processing tools that take bodies of text – e.g., from a book – and extract various features of interest. In Figure 2b, we show our process of extracting text from digitized books and then analyzing it; we refer to this as our "Text-to-Data Pipeline." We describe this process in further detail in Methods Appendix D.C.

Digitizing text. We begin by extracting text from digital scans of the books using optical character recognition (OCR). This process converts text into ASCII which then encodes each character to be recognizable by computers. We derive our textual measures of race, gender, and age by enumerating the features of these text data, specifically various

[31] Previously, many existing artificial intelligence models that classified putative race had a high error rate, both misclassifying the putative race of identified people and, in "one-shot" models that identify existence of people and their putative race simultaneously, misclassifying people as non-human (Fu, He and Hou, 2014; Nagpal et al., 2019; Krishnan, Almadan and Rattani, 2020). Ongoing work attempts to acknowledge and address these disparities (Buolamwini and Gebru, 2018; Mitchell et al., 2019). We acknowledge that race is a human-made construct that exists for political and economic purposes (Roberts, 2011; Logan, 2022) – and so, as a result, any attempt to classify race with either a human or a computer is an imperfect exercise that will yield imperfect results. Our analysis by race looks across collections within race, so any error within a race would be consistent across collections (i.e., racial categories would be classified similarly across the Mainstream and Diversity collections).

When labeling gender, we recognize that binary classifications are imperfect and focus only on the performative aspect of gender presentation, as they are trained based on how humans classify images. Furthermore, because we are classifying character gender based on the character's appearance, our measurements use the same binarized gender classification to assess the perceived presentation of gender, i.e. whether the character is female-presenting or male-presenting, rather than female or male per se. Future work should incorporate the classification of fluid and nonbinary gender identities.

[32] In Szasz et al. (2022), we curate the CBFeatures 1.0 data set, a manually labeled data set of illustrated faces that can be used as training data to more precisely predict the race, gender, and age of faces detected in illustrations in future work.

types of single term counts, the presence of famous people, and the first names of characters.

Text analysis: Token counts (Gender and Age). We generate counts of different "tokens" associated with gender and age.[33] To calculate gender representation in text, we calculate the number of male and female gendered pronouns along with a list of other gendered terms such as queen and husband. To measure representation of age in text, we generate lists of gendered terms associated with children, or "younger," individuals (e.g., girl, nephew) and gendered terms associated with adults, or "older," individuals (e.g., woman, uncle). The vocabulary used for each of these lists is shown in Appendix Section D.C.2.

Text analysis: Named Entity Recognition (Race and Gender). We measure the representation of race and gender among named characters in these stories, be they fictional or historical, using a tool called Named Entity Recognition (NER). NER identifies and segments "named entities," or proper nouns. There are two types of named entities that we identify: (1) famous characters and (2) first names of characters.

Famous individuals. Exposure to salient examples of historical figures or celebrities from marginalized backgrounds can lead to meaningful changes in social attitudes towards people who hold those identities, as well as changes in beliefs about one's self, and improvements in academic performance among children who share those identities (Marx, Ko and Friedman, 2009; Plant et al., 2009; Alrababah et al., 2021). To identify mentions of famous characters, such as Martin Luther King Junior or Amelia Earhart, we match the entities identified by NER that have at least two names (for example, a first and last name) with a pre-existing data set, Pantheon 2.0, that contains data from over 70,000 Wikipedia biographies, which provides information on gender for each famous individual (Yu et al., 2016).[34] We then manually code putative race for each identified person.[35] This generates a data set of 2,697 famous people. We count the number of unique books in which each famous person is mentioned as well as the number of times they are mentioned in each book.

Character first names. We then measure the gender of characters who are identified via NER but not identified as "famous." We extract the first word (name) of each of these named entities and estimate the probability that it is female (or male) using

[33] A token is a maximal sequence of non-delimiting consecutive characters. In our context, a token is an individual word.

[34] The Pantheon 2.0 curators run a classifier over the English text of the Wikipedia biographies to extract demographic information.

[35] Note that coding of putative race is subject to the individual biases and perceptions of each human coder and may be classified with error. We collapse the following identities: East Asian, Middle Eastern, and South Asian into the Asian category; North American Indigenous peoples and South American Indigenous peoples into the Indigenous category; and African American and Black African into the Black category. If an individual was coded as having more than one race, they were then classified as Multiracial.

data on the frequency of names by gender in the U.S. population from the Social Security Administration (SSA).[36] If the predicted probability that a name is female is greater than 50 percent, we classify the name as female. Otherwise, we classify the name as male.[37] For example, in the SSA data, the proportion of people named Cameron who identify as female is 9.16 percent. We therefore assign a probability of 90.84 percent that the name Cameron is male, and classify the name as male.

IV.C Data Collection, Aggregation and Analysis

To analyze representation, we collected and digitized the books recognized by the awards in our sample, using both library and online sources. Our final sample comprises 1,130 books recognized by at least one award.[38] We divide these books into different collections, as described in Section II.A. We then transform digitized page scans into data on the images and text in these books using the methods described in this section.

We report results for the following measures of representation in images and text. For the detected skin color of faces in images, we report the raw perceptual tint and, separately, bin these values into terciles. For race, we measure race of famous figures mentioned in text and predicted race of faces in images. For gender, we measure pronoun counts, gendered term counts, predicted gender of character first names, and gender of famous figures in text; and the predicted gender of faces in images. We also present an aggregate of all words with a gender association, which we refer to as "gendered words."[39] For age, we measure predicted age of faces and the ages associated with gendered terms in text.

To generate our estimates of representation, we first summarize each measure at the book level, and then calculate the average across all books in a given collection, both overall and over time. For example, to estimate the average percent of female faces in a collection,

[36]We predict gender with the *gender* package available in R which uses SSA data (Mullen, 2020). Using this method, we are able to make gender predictions for approximately 60,000 names.

[37]To test how accurate these predictions are, we predicted the gender of each famous person in our data using their first names and compared these predictions to their gender identified using Wikipedia and found that our predictions were 96.35 percent accurate. We do not classify race using first names only. Other recent text analysis has shown that conventional methods for classifying race of names fail to successfully distinguish between Black people and White people (Garg et al., 2018).

[38]Our books include those recognized over the period 1923-2019. We include both books that win the medal of the award, and those books that are honored by the award committee but not ultimate recipients of the award in a given year. Some books are recognized by more than one award. The 19 award corpora comprise 3,447 total books which either won an award or received an honorable mention; we obtained and digitized 1,130 of these books using both library and online resources. In the sample, we have all but 16 Mainstream medalists (3 Newbery winners and 13 Caldecott winners).

[39]This includes predicted gender of character first names, gender of famous characters, gendered pronouns, and specific gendered terms such as queen and husband, to generate a composite measure of gender representation in text.

we first calculate the percent of female faces in each book in the collection and then take the average across books. This ensures that each book contributes equally to our collection-level measures of skin color, race, gender, and age representation, regardless of book length.

We generate these estimates at the book level and then aggregate them to the collection level, both overall and, separately, over time. While different awards commence in different years, we study all books ever recognized by these awards, rather than limiting the analysis to years in which all awards are active. Because the use of many books persists over time, a teacher, librarian, or parent may be at least as likely to select a book considered to be a "classic" that they know (thus an older book) rather than one more recently published.[40]

V Results

In this section, we describe patterns of representation of skin color, race, gender, and age in the images and text of these books across collections and time.[41]

Skin color. We begin by characterizing patterns, across collections and over time, in the skin color of the characters pictured in images. We focus our discussion on characters with human skin colors. Results for characters with monochromatic or non-typical skin colors can be found in Appendix Section C and generally show similar patterns as the characters with human skin colors. Figure 4a shows the distribution of perceptual tint for detected faces in the Mainstream and Diversity collections. These figures show that the faces in the Diversity collection have darker skin tints, on average, than those in the Mainstream collection.[42] A Kolmogorov-Smirnov test rejects the equality of the two distributions ($p < 0.001$); in other words, the distributions of skin colors in pictured characters in the two collections are statistically distinct. Furthermore, the distribution of skin color tint in the Mainstream collection has a much smaller variance than that of the Diversity collection: a test of the null hypothesis that the two variances are equal also rejects equality with $p < 0.001$. This implies that there is a greater diversity of skin color tint shown in the Diversity collection.

We next examine the proportion of character faces in each skin color tercile – darker, medium, or lighter. Our results show that over time, the proportion of characters with skin colors in the darker and medium skin color terciles increases relative to those in the lighter

[40]For example, picture books such as *The Snowy Day* (1962) and novels such as *Charlotte's Web* (1952) were published and recognized by a Mainstream award long before 1970, when the awards in the Diversity collection first began to recognize books. Nonetheless, both of these books are still part of the canon of children's literature and remain frequently used in libraries and classrooms.

[41]A previous version of this paper (available here: https://www.nber.org/papers/w29123) includes some results which were removed in the revision process.

[42]Appendix Figures C3 and C4 demonstrate that this result holds regardless of image color type: monochromatic or non-typical skin colors.

skin color tercile, in both the Mainstream and the Diversity collections (Figure 4b). The distribution of skin color across the three terciles in books in the Mainstream collection from 2010-2019 is similar to that in the Diversity collection from 1970-1979. A related but distinct parameter of interest is the mean value of perceptual skin tint. Unlike our result for the distribution of skin color in faces across terciles, we find that average perceptual tint has changed less over time (Appendix Figure C5).

Figure 4c shows the distributions across these terciles for all seven collections. For both Mainstream and Diversity collections, the medium skin color tercile is the most represented, with almost half of all faces in both collections falling in this tercile. In the Mainstream collection, however, lighter skin is in the second most common tercile of skin color (approximately one third of faces), while in the Diversity collection, darker skin comprises the second most common skin color tercile (approximately 40 percent of faces). This suggests that the Diversity collection is more representative of characters that have darker skin tints. Of the seven collections, the Mainstream collection has the lowest proportion of faces falling in the darker skin color tercile and the Female collection has the greatest proportion.[43]

We then explore how skin color representation varies by race, gender, and age (Figure 5). We see that the Mainstream collection is more likely to show characters *within* a given race as lighter than their counterparts in the Diversity collection (panel A).[44] Given the minoritization of females and those with darker skin color, we test for a difference in representation at the intersection of female gender identity and darker skin tint. We find no evidence of a significant difference between the skin color distributions of faces classified as females and of faces classified as males (panel B).

We then see when children are depicted in images, they are more likely to be shown with lighter skin color than adults, regardless of the collection in which the image appears (panel C).[45,46] We are aware of no definitive biological justification for this systematic difference in the representation of skin colors by age. There are many possible determinants of potential differences. One might expect to see adults depicted with darker skin color, for example, if they have greater exposure to the sun from more outside labor. One might

[43] Appendix Figure C2 shows that the method of classifying "human" vs. "non-typical" polychromatic skin colors may underestimate the number of darker-skinned faces if the browns that are used do not follow the polychromatic $R \geq G \geq B$ rule as described in the Methods Appendix. However, Appendix Figure C4 shows that this does not change the patterns in skin color representation by collection over time.

[44] We see the same result for monochromatic faces in Appendix Figure C6a.

[45] One concern could be that the algorithms are trained to classify faces as being more likely to be a child if the skin color of the detected face is lighter, which then would attenuate the number of children detected.

[46] In Appendix Figure B3, we present the representation of skin color and age by the percentage presence in each of the coarser categories.

also hypothesize that children pictured are products of mixed-race couples which may lead to children having lighter skin, on average, than adults. However, this phenomenon would more likely result in a compression of the skin color distribution rather than a shifting of the distribution. Moreover, interracial relationships were prohibited by "anti-miscegenation" laws in many contexts for a substantial portion of our study period and their incidence remains low. On the other hand, children could be depicted as having darker skin, on average, for a number of other potential reasons. For example, evidence of the breakdown of melanin over the life course suggests that there may be reason to expect the skin tint of adults to be lighter than that of children (Sarna et al., 2003). Nonetheless, the pattern we find of children being represented with lighter skin than adults is consistent across collections. While there are many potential interpretations of this pattern, some include brightness being used to connote innocence (e.g., of childhood), supernatural features (e.g., of angels), or another type of emphasis which separates the character from the rest of the context. Exploration of the reasons behind this phenomenon merits further work beyond the scope of our study.

Putative race. We then explore trends in racial representation of famous individuals over time (Figure 6).[47] Onto this time series we overlay the U.S. population share of different races, by decade, using census data. We see that in the Mainstream collection, relative to their U.S. population shares, Black people and Latinx people have been historically underrepresented. The last three decades, however, have shown increasing parity in representation of Black famous individuals. We see that despite increases in a diversity of representation over time, the average individual included – whether a famous person or a pictured character – is a White individual, regardless of collection.[48]

White adults (and children) are more likely to be pictured than adults (and children) of any other racial category across all collections (Appendix Figure B6). Juvenile ageism, a term coined in Westman (1991), refers to the notion that social systems ignore the interests of children (de la Fuente-Núñez et al., 2021). From an intersectional perspective, this also

[47] Appendix Figure B4 shows a similar version of this graph with non-standard axes to more clearly view changes in groups with small population proportions.

[48] Appendix Figure B5 shows the proportion of famous figures broken down by race overall. We find that, in all collections, the famous figures mentioned are predominantly White. In the Mainstream collection, over 90 percent of famous figures are White. Conventional content analyses of the race of main characters in Caldecott and Newbery award-winning books find qualitatively similar results (Koss, Johnson and Martinez, 2018; Koss and Paciga, 2020). The African American collection is the only collection to have a majority identity other than White represented; in it, Black people are the most represented, comprising 50 percent of the famous people in that collection. In other collections, Black people comprise 7 to 29 percent of famous figures mentioned. Other groups appear far less frequently. Famous people of Asian, Latinx, Indigenous and Multiracial identities account for between 3 and 11 percent of famous people *combined*, a high level of inequality in representation relative to population averages. The U.S. census estimates that only 60 percent of the population is non-Latinx White (2019).

means that children of color, whose identities fall at the intersection of at least two sites of societal marginalization, are least likely to be seen by readers. Figure 6 shows that White individuals have been overrepresented among mentioned famous figures relative to their population share since the 1930s, particularly White males (Figure 7). Both White males and females are predominantly included as pictured characters (Figure 8).[49]

Our results here also show that when children see females in these books, they are seeing mostly White females. This relates to a key prediction from studies of intersectionality: that identities at the intersection of multiple sites of exclusion may face even greater disadvantage than would be predicted by individual, group-specific patterns. Specifically, the message sent by this pattern of representation is that when women inhabit prominent spaces in society (e.g., in the historical and fictional accounts contained in curricular materials), this is primarily limited to White women. However, that same figure reveals the surprising result that, conditional on the person being classified as Asian, Black, or Latinx + Others, the Mainstream collection is more likely than the Diversity collection to represent the person as a woman. The Female collection, on the other hand, is far more likely than the Mainstream collection to represent people classified as Asian, Black, or Latinx + Other as females. This suggests that, on average, books in the Female collection are the most attentive to the power imbalances that come from the intersection of multiple sites of exclusion, at least in terms of including the presence of females of color.

Among famous figures, after White males and females, Black males comprise the next most represented group (5-37 percent of famous people). The representation of Black females (between 2 and 8 percent of famous people, except in the African American collection, where they comprise 13 percent) is consistently less than that of Black males, despite their approximately equal shares in the population. Conditional on the famous person being Black, we see greater representation of females in the Mainstream and Female collections than in the Diversity or African American collections (the representation of Asian and Latinx people is often close to zero for this measure, making comparison difficult). This highlights that even within collections of books curated to highlight a given racial identity, we see less representation of people at the intersection of multiple dimensions of marginalization than of those who occupy only one such dimension.

In Appendix Table A2, we list the five most frequently mentioned famous people

[49] Appendix Figure B7 shows that most pictured characters are classified as being White. Appendix Figure B8 shows that we classify almost half of pictured characters in the Mainstream collection as female-presenting. We map share of faces by predicted race on their respective shares of the U.S. population in Appendix Figure B9. Appendix Figure B10 shows the proportion of characters in images and text by race and gender over time.

overall, including their race and gender. The most uniquely mentioned person in the Mainstream collection is George Washington; in the Diversity collection, it is Martin Luther King Junior. For the Mainstream collection, all five of the most commonly mentioned people are White males. For the Diversity collection, all five are males, three of whom are Black (Martin Luther King Junior, Frederick Douglass, and Langston Hughes) and two of whom are White (Abraham Lincoln, George Washington). In the Female collection, where one might anticipate the presence of more females, the three most uniquely mentioned people are males (John F. Kennedy, Martin Luther King Junior, and Jimmy Carter) and the fourth is a female (Betty Friedan).[50]

Gender. We then explore the representation of gender. We first measure the incidence of words with any gender association, which includes pronouns and other gendered terms, the gender of the famous people mentioned in the text, and the gender classifications for character first names. In Table 1 and Figure 9, we present average book-level proportions of female words out of all gendered words. For all collections except those books specifically recognized for highlighting females, we observe fewer female words than male words. Table 1 shows that the proportion of gendered words that are female in these collections is between 34 and 45 percent, as opposed to 56 percent in the Female collection. Figure 9a shows that this proportion increases gradually over time, but remains below the U.S. population share of females for all collections in every decade, except for the Female collection.

In Figure 9b, we show how these distributions change over time. In both collections, the skewness of the distribution of our measure of book-level gendered words changes over time, becoming less right-skewed in more recent years. In addition the representation contained in the median book has moved closer to equality.

In Figure 9c, we show the distribution of the book-level proportion of female words for each collection. The Mainstream collection is the most male-skewed of all collections, and in all distributions except that of the Female collection, the central tendency is skewed towards more male representation. The Female collection, which we would expect to be more female-centered, appears less female-skewed than the Mainstream collection is male-skewed.

Our results are robust to restricting analysis to each type of gendered word: gendered pronouns, gendered terms, or first names (Appendix Figure B11). This addresses the concern that we could be misattributing changes in gender representation to changes in the historical grammatical convention to use what were then considered "gender-neutral" pronouns (e.g. he,

[50] Appendix Tables A3 and A4 show this for the top five females and top five males, respectively, uniquely mentioned in each collection. Appendix Table A5 shows the most uniquely mentioned famous figure by collection for each decade.

his) and related terms (e.g. waiter, actor); for example, if an author writing in an earlier era wanted to include more female representation, we would see this reflected in the proportion of named female characters but not in the proportion of female pronouns but we do not see this skewed pattern in our results. The robustness of our results to this sample restriction demonstrates that our results are not driven by measurement error stemming from changes over time in this historical convention. Our results are also robust to restricting analysis of gender representation to gender of famous figures. Famous figures transmit more implicit information to a child than generic terms or characters by virtue of their identity in society.[51] In Figure 7, we show that ninety percent of the famous figures in the Mainstream collection mentioned were male, for example, and even the Female collection includes more unique famous males than females. Overall, less than a third of famous figures in the books we study are female (Appendix Figure B11).

Next, we describe the representation of gender in the images of these books.[52] We show the proportion of faces in each collection identified as female in Figure 8 and Appendix Figure B8a. In the majority of the collections, fewer than half of the detected faces are classified as female-presenting. In the Female and Ability collections, respectively, however, our model classifies 71 and 67 percent of the faces as female. Appendix Figure B8b shows that, unlike for text, the incidence of representation of women in images is relatively consistent over time. For example, in the Mainstream collection, female-presenting faces comprise between 39 and 51 percent of all detected character faces over time.[53]

We then compare representation of gender across images and text. In Figure 10, we show a scatterplot of collection-by-decade average proportions of female words on the x-axis and the average proportion of female-presenting faces on the y-axis. It shows that females are more likely to appear in images rather than text, which means that females are more likely to be visualized (seen) than mentioned in the story (heard). One interpretation of this pattern is that authors or illustrators may perfunctorily include additional females in pictures, giving the appearance of equity while not actually having them play an important role in the story. It also highlights that on average, females are represented less than half of the time in both images and text.[54]

[51]This can occur through any of a number of channels, for example via role model effects (Dee, 2005; Porter and Serra, 2020) or via effects on more general social preferences and beliefs (Plant et al., 2009; Alrababah et al., 2021)

[52]This exercise demonstrates the limitations of existing AI approaches. Compared to the state of the art, a human would be better able to more accurately classify individuals who identify as transgender or non-binary.

[53]We show a similar pattern when using a continuous measure of the average probability that a face is classified as being female in Appendix Figure B12.

[54]In Appendix Figure B13, we show these results for females by race in which we see Black and Latinx

Age. Finally, we describe the representation of people by age in the images and text of our books. In Table 1, we show that, across all collections, adults are more likely to be present in both images and text. Three to 19 percent of characters presented in images are classified as children, and 17 to 28 percent of age-specific gendered words refer to children. In Appendix Figure B14a, we show the proportion of pictured character faces by age and gender. Regardless of gender, in both images and text, we show that there are more adults than children depicted in the books in each collection.[55] We also see in Appendix Figure B9c that adults are overrepresented relative to their U.S. population share, meaning that adult depictions are more likely to be privileged in books targeted to children. Children of color are the least likely to be pictured, even in the People of Color or African American collections (Appendix Figure B6).

In Appendix Figure B14b, we show the age classifications of gendered words (e.g., girl vs. woman). Similar to images, we see that older people are more likely to be mentioned than younger people. In most books, the distribution of young people by gender is similar, though in the Female collection, girls are approximately twice as likely to appear as boys. In gendered words specific to adults, however, men appear more often than women.

VI Economic and Social Factors Underlying Representation in Books

In this section, we investigate a series of economic and social factors which may contribute to the patterns of representation of skin color, race, gender, and age in prominent children's books that we document in Section V. Some of these patterns include, for example: that White people and males are overrepresented relative to their population share; that skin pigment in pictured characters has trended darker over time; and that women are more often pictured in images than they are mentioned in text. In this section, we aim to shed light on potential economic forces determining what publishers produce and what consumers consume that, in turn, may contribute to these patterns.

First, we discuss relevant prior theoretical and empirical research related to the economics of the media and, separately, the economics of identity, to conceptually characterize a set of market forces which may influence the patterns of representation within children's books. For clarity, we separate these into demand- and supply-side forces. Second, we generate a series of stylized facts that relate the representational content in the children's books that we study to this series of demand and supply forces suggested by prior literature. Our

females less represented.

[55] One concern may be that the age classification algorithms are primarily trained on adult faces, and therefore may overclassify adults; however, we see consistent ratios of adult presence to children presence in images and in text.

analysis uses individual-level data on book purchases and purchaser demographics, alongside library-branch level data on library acquisitions linked to neighborhood demographic characteristics, in the cross-section and over time. We also estimate the relationship between historical trends – first historical events, followed by changes over time in social mores and then, separately, in market shares of consumers of different identities – and the representation we see in books. Finally, we explore how local political beliefs relate to the consumption of books with different levels of representation.

VI.A Related Literature on Market Forces Driving Supply and Demand

Demand for representation in children's books. A consumer's demand for representation in the images and text of books they purchase may be affected by their identities in various ways. Our analyses describe and explore two main channels for this link from identity to demand.

The first is through demand for shared-identity, or "homophilic" representation (Jackson, 2010). This stems from the idea that people seek out and enjoy psychic utility from associating with – or even seeing – others similar to the self. This consumer preference of "utility from homophily" would lead consumers to be more likely to purchase children's books with characters that match the identities of themselves or their children.

The second is informed by the notion that deviating from social norms is costly; Akerlof and Kranton (2000) call these costs "identity losses." This force can lead to demand for representation that hews closely to the (perceived) status quo. Applied to our setting, this suggests that consumers who have identities that have been historically over-represented in media have been socialized to suffer greater identity losses from consuming content that does not center their (socially dominant) identities than historically under-represented consumers, because consuming such content deviates from the perceived status quo or social norm. For example, males might suffer greater identity losses than females from reading a book with a female main character than females would from reading a book with a male main character. Furthermore, this force of "status-quo bias" in consumption of books would push consumers of all identities to be more likely to consume children's books containing characters with socially dominant identities than those containing characters with other identities. This is reflected in a result from Bernheim (1994) showing that under certain conditions, people will adapt their preferences to match broader societal preferences.

Supply of representation in children's books. Prior work on the economics of the media also points to some key supply-side forces that are likely to contribute to the levels of and trends in representation that we document. This work shows, both theoretically and empirically, that in media markets with startup costs, search costs, and other frictions, supply will cater primarily to the preferences of the majority group rather than proportionally to the individual preferences of various groups of consumers present in the market (Waldfogel, 2003, 2007). Ceteris paribus, these forces would reduce the supply of differentiated products targeted to the demands of various identity-specific subgroups of consumers. Given the various fixed costs faced by the publishing industry (Waldfogel, 2007; Berry and Waldfogel, 2010), publishers of books targeted at the general market – such as those in the Mainstream collection – may choose to publish more books which feature characters whose social identity matches the majority of children in the market. This, of course, would come at the expense of publishing fewer books containing characters of other identities. Such a pattern is in line with phenomena described in Waldfogel (2007), labeled there as the "tyranny of the market."

A corollary of this idea is that, as the market share of a given group changes because of shifting demographics, so should the supply of books catering to that group. This follows Acemoglu and Linn (2004) and DellaVigna and Pollet (2007), who show that market size can be predicted from demographic profiles of birth cohorts, and that this, in turn, shapes profitability and innovation in a wide range of markets, including pharmaceuticals, toy and bicycle manufacturing, and life insurance.

A second supply-side force in such markets is a "pricing-in of representation." This refers to the notion that books which deliberately elevate non-dominant identities may sell fewer copies, leading publishers to increase their prices to cover the fixed costs of production for these books (e.g., author advances, printing start-up costs).[56]

Our analysis puts aside a few key aspects of these markets, such as supply on the extensive margin. We discuss these and other limitations later in this section. We also supplement this with analysis of qualitative data collected from a series of semi-structured interviews with professionals who currently work at or recently worked at libraries, publishing houses, and children's bookstores, and/or who served on book award selection committees. We report these in Appendix G.

VI.B Empirical Analysis of Economic Forces

In this section, we present a series of empirical analyses probing the economic (supply and demand) factors in publishing decisions (and selection) of children's books. We analyze book consumption data from the Numerator OmniPanel linked to book-level representation

[56] This is isomorphic with another possible explanation for higher prices consistent with our summary of prior work on the supply-side forces leading to these patterns: if publishers are less likely to supply books which deliberately elevate non-dominant identities, a given level of demand met with low levels of supply would also lead to higher prices.

levels that we estimate.

We first present analyses of book consumption that document patterns which are consistent with demand-side utility from homophily. We estimate the correlations between book purchaser identity and the average representation in these books in Table 2. In Panel A, we show that purchasers who have a son purchase books with two percent fewer female names as a proportion of all gendered names, and one percent fewer female words as a proportion of all gendered words, as compared to purchasers that have no children. We see a roughly symmetric preference for books with a greater proportion of female names and female gendered words between purchasers who have a daughter and purchasers who have no children. Finally, we see that purchasers with daughters purchase books that, on average, have more similar representation levels of gender in text and images (i.e., books in which females are more equally "seen" and "heard") (column 4). In Table 2, Panel B, we see that males' purchasing patterns exhibit a slight revealed preference for books with more male words and faces. Specifically, compared to female purchasers, males purchase books with 1 to 2 percent fewer female words, names, and faces. We find no evidence of a difference in the ratio of female representation in images and in text across male and female purchasers.

In Table 3, we show the relationship between purchaser race/ethnicity and the representation of skin color and putative race in books purchased. These results are also consistent with the notion of utility from homophily. In column 1, we see that purchasers who identify as Black or as Latinx are more likely to buy books that contain pictured characters with darker skin color, on average, than purchasers who identify as White. In columns 2-5, we show similar results for mentions of famous individuals by putative race. We find positive and statistically significant estimates for Asian, Latinx, and Black consumers purchasing books that contain more mentions of famous people who share their own racial identity. White people, in turn, are more likely than other groups to purchase books with predominantly White famous people.

We also document utility from homophily using inventory data from branches of the Seattle Public Library system. In Table 4, we show that public libraries in communities with a higher proportion of White, non-Hispanic residents contain more books from the Mainstream collection (column 1) and fewer books from our Diversity collection (column 2). We show in columns 3 and 4 that the results are robust to controlling for measures of household income within a community. If we assume that adults are making the majority of purchasing decisions, then the overrepresentation of adults and underrepresentation of children that we estimate (even in these books targeted at children) is also consistent with utility from homophily.[57]

We also estimate whether there is higher demand for Mainstream books, which we show to disproportionately represent males and White people, than for Diversity books. Using consumer panel data on children's book purchases from Numerator, we see in Table 5 that Mainstream books sell more than twice as many copies per title (83) than Diversity books (33). In the Seattle Public Library data, books in the Mainstream collection receive approximately four times as many checkouts per title than do books in the Diversity collection. While only suggestive, this pattern is consistent with the phenomenon of status-quo bias in the following way: the much larger volume of purchases for Mainstream collection books than for Diversity collection books is highly likely to contain purchases by consumers whose racial or ethnic identities are not centered in these books, but who may have demand for the status quo (i.e., the centering of the dominant group). A weakness of this link, of course, is that we cannot quantify the extent of this with the data we have available.. Also consistent with the demand-side force of status-quo bias, a majority of the books purchased in our data have predominantly male-focused content, despite the fact that most of the purchasers in our sample are female (Appendix Table A1). Together, these findings relate to the patterns showing the overrepresentation of White people, and males, in both images and text, documented in Figures 7-9, Appendix Tables A2-A4, and Appendix Figures B5-B8.

On the supply side, we find evidence supporting the notion that suppliers cater primarily to the dominant group (what Waldfogel (2007) describes as tyranny of the market). Specifically, we find that White famous figures are over-represented in the text of Mainstream books relative to the share of White people in the U.S. population (c.f., Figure 6). In the Seattle Public Library inventory data, we see that these libraries stock twice as many copies of books belonging to the Mainstream collection than books belonging to the Diversity collection (Table 5, Panel B).[58] Finally, we show evidence in Table 5, Panel A, that the average price of books in the Diversity collection is 22 percent higher than those in the Mainstream collection, which is consistent with the idea that representation is being priced in by suppliers of books.

VI.C Historical Trends and Representation

We next explore how changes in representation in the Mainstream collection over time may be associated with historical events, trends in societal attitudes towards issues related

[57] Additionally, adults are both the producers of the content and the decision-makers on the award-selection committees. Utility from homophily would predict that their preferences for book content, even in these roles, may reflect their identities as adults.

[58] Number of library copies serve as a measure of supply.

to race and gender, and changes in market share of various identity groups.

We begin by exploring how changes in representation may track salient historical events, such as the Black Lives Matter and #MeToo movements, or the first person of a given identity to inhabit a major societal role, such as the first female Supreme Court justice or Black president. We show the time series of the average skin color of pictured faces (Appendix Figure B15) and the average percentage of gendered words (Appendix Figure B16), with a selected set of relevant salient historical events overlaid upon the graph with vertical black lines. We emphasize that this narrative exercise is descriptive rather than causal and hypothesis-generating rather than providing a confirmatory test of any hypothesized relationship. We observe that these major historical events are often accompanied by a temporary change in representation, similar to estimates of how racial attitudes respond to economic downturns (Jayadev and Johnson, 2017).

We then explore how representation of race and gender tracks social attitudes over time. We use data from the General Social Survey (GSS), a repeated cross-sectional survey collecting attitudes from a nationally representative sample of people in the U.S. several times per decade since 1972 (Smith et al., 2021). We find that attitudes towards Black individuals – as measured by the likelihood that a person "would vote for a qualified Black candidate for president" – have trended more egalitarian, coinciding with a trend towards darker average perceptual tint in the skin color of character faces (Appendix Figure B17a). Similarly, we see a trend in attitudes towards greater gender equality – as measured by people's acceptance of egalitarian gender roles – which coincide with a trend towards more equal inclusion of females and males in the text of books (Appendix Figure B17b).

We can also characterize the correlation between changes in market share and the representation of race and gender in books over time. Following existing studies estimating this type of relationship (Acemoglu and Linn, 2004; DellaVigna and Pollet, 2007), we calculate the market share of various race and gender groups and use this to estimate whether there is a statistically detectable relationship between market share and representation of the group in the books we study. For race, we use the share of racial groups in the US population according to the decadal census. For gender, while the share of females in the census is relatively stable, we can instead use the female labor force participation rate as a measure of market share. We conceive of this as capturing the (relative) consumer power of females relative to males.[59]

We find a positive and significant relationship between the market share of Asian,

[59] A related test for future research would be to correlate market share with prices. Because the price data we use do not extend prior to 2017, this analysis is beyond the scope of our study.

Black, and White people in a given decade and their representation in books from the Mainstream collection published in that decade (Appendix Table A6). We find no evidence of a correlation between market share and representation of Latinx people and their representation in books, but we believe this is primarily an artefact of the very low representation of this group in books we study.[60] Also, census data on Latinx individuals are only available beginning in 1970 and we are only able to predict whether the race of a detected face is "Latinx + Others," both of which lead to noisier estimates. For gender, too, we find a positive and significant association. The female labor force participation rate is strongly related to the proportion of gendered terms contained in books over time. While we find no such correlation with the representation of gender in images, we suspect this is primarily because, throughout our period of study, representation of gender in images is closer to parity than it is in text.[61]

These results help explain the trends in representation in children's books over time that we document in Section V.[62] In this section, we show that these results are correlated with broader changes in overall societal mores. This aligns with findings from sociology on the patterns of changes in racial beliefs over time (Schuman et al., 1997) and the linkages between beliefs – particularly racial beliefs – and behavior (Ajzen et al., 2018). It also corresponds to theoretical predictions of the evolution of social preferences (Bernheim, 1994; Sobel, 2005). Bernheim (1994) predicts that people's preferences will adapt to what they think are social preferences. Similarly, Sobel (2005) predicts that preferences are informed by a desire for reciprocity. In our setting, greater demand for a diverse set of representations could come from awareness of increasing diversity in the U.S. population, and, as we see in the CCES data, (gradually) increasing acceptance of racial equality for Black people.

VI.D Local Beliefs and Book Consumption

We have documented that demand for representation in children's books is related to the identities of the consumer. In this subsection, we provide evidence that demand for representation in children's books is also related to consumer beliefs.

[60]These patterns can also been seen visually in Figure 6, which shows the relationship over time between population share and representation by race and ethnicity in text.

[61]These associations between market share and representation of specific identify groups coupled with projected changes in U.S. demographics from the US Census Bureau suggest that representation of non-Hispanic white will continue to decrease over the next 40 years along with their proportion of the U.S. population. Similarly, we would expect to see a continued increase in the representation of racial groups that are projected to have proportional increases in the U.S. population such as Black and Asian individuals (Vespa, Medina and Armstrong, 2020). We are not aware of any long-term projected change in female labor force participation and so do not anticipate a change in female representation stemming from this linkage.

[62]See Figures 4, 6, and 9, and Appendix Figures B4, B9, and B10.

We analyze cross-sectional variation in consumer beliefs and book consumption, drawing from the Cooperative Election Study (CCES), a nationally representative, stratified sample survey administered by YouGov. The survey collects information about general political attitudes linked with respondent demographic data. We draw from the 2017 CCES data set because it was the earliest survey year for which book purchase data were available. We merge these data with Numerator data on the number of books from the Mainstream and Diversity collections purchased, by zip code, from 2017-2020

In Table 6, we show that a greater number of purchases of books from the Diversity collection is associated with a smaller proportion of individuals who believe that undocumented immigrants should be deported (column 1),[63] a smaller proportion of individuals who believe that federal funds should be withheld from localities that do not follow federal immigration laws (column 2), and a larger proportion of individuals who believe that White people in the U.S. have certain advantages because of the color of their skin (column 3). We see no association between the number of book purchases from the Diversity collection and the percent of people who are angry that racism exists (column 4); this is likely because most respondents (80 percent) answer yes to this question, as opposed to only 37 percent who believe that undocumented immigrants should be deported.

Combined with our analysis of the representations contained in these books, and seen through the lens of other research showing how the content of children's books can shape adult beliefs (Fuchs-Schündeln and Masella, 2016; Cantoni et al., 2017), the evidence we provide here suggests that children's books may be an important factor in the intergenerational transmission of societal values.

VI.E Limitations

In this section, we discuss some limitations of our investigation of the economic forces behind the levels of representation we find.

The first limitation of this investigation is that it is descriptive, rather than causal, and exploratory, rather than confirmatory. We conduct and report a series of descriptive analyses of relationships in the cross-section and over time. We anticipate that the stylized facts we present will serve as hypothesis-generating, instigating further work to characterize these relationships with experimental, quasi-experimental, and structural methods.

A second limitation is that there exist a series of potential contributors to the results analyzed in this section beyond the supply and demand forces explored above. Our analysis attempts to characterize and investigate evidence for forces that influence what consumers

[63]In the CCES, the wording of the question referred to "illegal" immigrants.

choose to purchase. We do not explore factors that may influence what consumers *choose not* to purchase; for example, there is scope for for discrimination against certain identities to drive some of these results. This force could exert itself on the decisions of purchasers, publishers, and awards committees. Its impact would be in addition to – but separate from – the forces we explicitly explore. Another related limitation is a potential market response from publishers to the preferences of different award-granting committees. There is necessarily a limited number of books that can receive major awards. If these major awards increase consumption of books that receive those awards, publishers may actively try to produce books that are more likely to receive these awards, reinforcing whatever patterns of representation that publishers perceive the relevant awards committee to prefer. Because membership on awards committees is confidential, analysis of their preferences beyond what we present here exceeds the reach of our study.

Separately, we observe that the effect of utility from homophily is attenuated for book purchasers who are not White, in comparison to White purchasers. We attribute this, in part, to status-quo bias. We acknowledge, however, that part of this pattern may also arise because of the potentially higher costs that are associated with consuming books that highlight characters with non-dominant identities. These higher costs may come from at least two sources – financial and psychic – which we cannot fully disentangle. The first source may be increased financial cost stemming from there being fewer options available in the larger market centering non-dominant identities, leading to a higher price (i.e., pricing-in diversity). The second source may be from increased psychic costs given that the demand for homophily by members of the dominant group may be amplified by status-quo bias, while this may not be the case for other groups.

Additionally, our empirical analysis of the relationship between content and consumer demographics is limited to the content in award-winning books. In Section II, we document that these awards are strongly correlated with what is purchased and consumed in homes, libraries, and schools. While we might wish to draw from a representative sample of the "universe" of children's books, such a sample is less well-defined and likely has lower per-book influence than our analysis sample. One related challenge is how to appropriately account for the award itself influencing consumer preferences.

Finally, the findings related to skin color can not be further explained in the scope of our economic analysis. We do not have skin color information for individuals in the larger population, so we can not examine the relationship between consumer skin color and revealed preference related to content. These are important phenomena to document nonetheless, given the importance of the role that the messages in these books play in potentially shaping children's development. We leave exploration of their potential causes to future research.

VII Summary and Concluding Remarks

The books we use to educate our children teach them about the world in which they live. The way that people are – or are not – portrayed in these books demonstrates who can inhabit different roles within this world and, in so doing, can shape subconscious defaults.

The content of images is an important but understudied dimension of this and other social processes related to education and belief formation. Per the adage "a picture is worth a thousand words," images in particular convey numerous messages to the reader, and the images contained in the content we use to teach children are likely to be particularly influential in processes of child belief formation and development. Social scientists are leaving data on the table by not systematically measuring the content of these messages implicitly and explicitly sent to the viewer.

In this paper, we make three primary contributions. First, we introduce computer vision methods to convert images into data on skin color, putative race, gender, and age of pictured characters. Second, we apply these image analysis tools – in addition to established natural language processing methods that analyze text – to award-winning children's books to document the representations to which children have been exposed over the last century. This uncovers many sites of inequality of representation in these books, both confirming results found in prior, manual content analysis of smaller sets of these award winning-books, as well as revealing many novel dimensions of inequality in representation in both the images and text of these books. Third, we analyze linkages between economic forces on the demand and supply side described in prior research and the representation levels that we measure. Our analysis reveals a series of stylized facts showing how these economic forces may contribute to the levels of representation we document. This includes evidence that demand for representation in children's books, as demonstrated by local purchasing patterns, is related to consumers' personal and political beliefs. Our results suggest how the demand for representation may be a channel through which beliefs about race and gender could propagate across generations through the messages contained in the books parents purchase for their children.

Our approach has a few key limitations. First, artificial intelligence tools reflect the biases of the human coders that trained the models, in ways distinct from but consistent with traditional content analysis conducted entirely manually. Second, the measures of representation that we use are imperfect. Our measures of gender identity neglect measurement of non-binary and gender-fluid identities. Because race is a multifaceted construct of human categorization that is ill-defined, efforts to measure it are inherently difficult. Third, the algorithms we use do not perfectly detect faces or isolate the skin from faces, generating measurement error. Fourth, our analysis consists of a numerical accounting of different characters through simple representational statistics, i.e., *whether* characters are included. However, this is not a holistic measure of representation. If a character is depicted in a reductive or stereotypical manner, then their representation may send messages which may reinforce existing inequality, despite deliberate efforts to improve equality in numerical representation. An important avenue for future work will be to further develop tools that can measure *how* people are represented and thus capture the messages sent by the manner of their portrayal. Finally, we were able to access and analyze 1,130 of the 3,447 books recognized by these awards. To the extent that book presence or absence in the sources we consulted (library and online) is related to book traits, we may generate a biased estimate of representation in the larger universe of these books. We argue, however, that our ability to access these books is most likely to be positively correlated with consumers' ability to access them, such that our estimates are likely to closely track the levels of representation in the books to which children are actually exposed.

The image-to-data tools we introduce allow for the systematic measurement of characteristics in visual data that were previously beyond the reach of empirical researchers. This contribution is in the spirit of other recent work introducing new sources of data to the economic study of social phenomena, such as text (Gentzkow and Shapiro, 2010; Gentzkow, Shapiro and Taddy, 2019), geospatial imagery (Burchfield et al., 2006; Henderson, Storeygard and Weil, 2012), and traditions of folklore (Michalopoulos and Xue, 2021). Practically, we aim to instigate the use of these tools by scholars in a wide range of fields. This may include, for example, analysis of representation in the historical record, or in other visual media such as television programming (Jensen and Oster, 2009; La Ferrara, Chong and Duryea, 2012; Kearney and Levine, 2019), advertising (Bertrand et al., 2010; Lewis and Rao, 2015), and textbooks (Fuchs-Schündeln and Masella, 2016; Cantoni et al., 2017).

The findings in this study – and the power of the tools we use to generate them – generate hypotheses that can motivate and inform subsequent research on the causes and consequences of representation in children's books. Computational tools allow researchers to systematically measure what content children see in their curricular materials with a greater speed and lower costs than previously possible, while reducing discrepancies across researchers and inaccuracies due to human error. Such measurements, paired with causal inference tools and complementary manual content analysis for measuring features beyond the reach of current computational tools, could be used to advance prior work on the impact of book content on children's beliefs and later life outcomes (Fuchs-Schündeln and Masella, 2016; Cantoni et al., 2017; Arold, Woessmann and Zierow, 2022; Arold, 2022), for example, linking exposure to different levels of representation with formation of beliefs, preferences, and societal outcomes. These same measurements could also be used to better understand the objective functions of different publishers, and how these change over time and in response to societal events.

The "optimal" level of representation is a normative question beyond the scope of this paper, but the actual representation in books is something that can be measured and, given some reasonable set of goals, improved upon. Computational tools will directly contribute to lasting improvement of the practice of education, both by helping guide curriculum choices and by assisting publishers and content creators to prospectively assess representation in the creation of new content. More broadly, they can help inform and contribute to ongoing and future efforts to understand how the representation contained in content contributes to, and can be used to reduce inequality in human development.

Acknowledgements

For helpful feedback, we thank two editors, four anonymous referees, Barbara Atkin, Karen Baicker, Chris Blattman, Anna Brailovsky, Tom Brock, Mary Gen Davies, Steven Durlauf, Lisa Englehardt, James Evans, James Fenske, Alex Frankel, Adam Gamoran, Claudia Goldin, Josh Goodman, Austan Goolsbee, Jon Guryan, Andrew Ho, Rick Hornbeck, Caroline Hoxby, Larry Katz, Dennis LeLoup, Ro'ee Levy, Susanna Loeb, Jens Ludwig, Jonathan Meer, Martha Minow, Sendhil Mullainathan, Shirley Mullin, Derek Neal, Matt Notowidigdo, Steve Raudenbush, Cybele Raver, Heather Sarsons, Fred Stafford, David Uminsky, Miguel Urquiola, Angelino Viceisza, Alessandra Voena, Melanie Xue, and others including many seminar audiences. For financial support, we thank UChicago BFI, UChicago CDAC, NAEd/Spencer, and UChicago Career Advancement. The research reported here was also supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A200478 to the University of Chicago. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. For excellent research assistance, we thank Fabiola Alba-Vivar, Ryan Atkinson, Callista Christ, Marliese Dalton, Anjali Das, Maya Escueta, Saloni Gupta, Shavonna Hinton, Camilo Ibáñez, Juan Miguel Jimenez, Jalnidh Kaur, Zipporah Klain, Jarvis Lam, Erica Lin, Ping-Jung Liu, Simon Mahns, Noah McLean, Karla Monteiro, Jan Morgan, Ifeatu Oliobi, Raj Shukla, Todd Takeno, Bhargavi Thakur, Jeffrey Tharsen, and Celia Zhu. We also thank Ashiyana and Kairav Adukia-Hornbeck for manual coding assistance. For access to important resources, we thank UChicago RCC, Heather McGuire, LaShanda Howard-Curry, Bridget Madden, and Kalli Mathios. Researchers' own analyses calculated (or derived) based in part on data from Market Track, LLC dba Numerator and marketing databases provided through the Numerator Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Numerator data are those of the researcher(s) and do not reflect the views of Numerator. Numerator is not responsible for and had no role in analyzing or preparing the results reported herein.

References

Acemoglu, Daron, and Joshua Linn. 2004. "Market size in innovation: theory and evidence from the pharmaceutical industry." The Quarterly Journal of Economics, 119(3): 1049–1090.

Adukia, Anjali, Callista Christ, Anjali Das, and Ayush Raj. 2022a. "Portrayals of race and gender: Sentiment in 100 years of children's literature." Proceedings of the ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies.

Adukia, Anjali, Patricia Chiril, Callista Christ, Anjali Das, Alex Eble, Emileigh Harrison, and Hakizumwami Birali Runesha. 2022b. "Tales and tropes: Gender roles from word embeddings in a century of children's books." Proceedings of the 28th International Conference on Computational Linguistics (COLING), 3086—-3097.

Ajzen, Icek, Martin Fishbein, Sophie Lohmann, and Dolores Albarracín. 2018. "The influence of attitudes on behavior." The Handbook of Attitudes, 197–255.

Akerlof, George, and Rachel Kranton. 2000. "Economics and identity." The Quarterly Journal of Economics, 115(3): 715–753.

Alan, Sule, Ceren Baysan, Mert Gumren, and Elif Kubilay. 2021. "Building social cohesion in ethnically mixed schools: An intervention on perspective taking." The Quarterly Journal of Economics, 136(4): 2147–2194.

Alrababah, Ala, William Marble, Salma Mousa, and Alexandra Siegel. 2021. "Can exposure to celebrities reduce prejudice? The effect of Mohamed Salah on Islamophobic behaviors and attitudes." 115(4): 1111–1128.

ALSC. 2007. "The Newbery and Caldecott Awards: A guide to the Medal and Honor books." Association for Library Service to Children, American Library Association.

Appadurai, Arjun. 2004. "The capacity to aspire: Culture and the terms of recognition." Culture and Public Action, 59: 62–63.

Apple, Michael, and Linda Christian-Smith. 1991. The Politics of the Textbook. Routledge.

Arold, Benjamin W. 2022. "Evolution vs. creationism in the classroom: The lasting effects of science education." Working Paper.

Arold, Benjamin W, Ludger Woessmann, and Larissa Zierow. 2022. "Religious education in school affects students' lives in the long run." CESifo Forum, 23(03): 40–43.

Balter, Rochelle. 1999. "From stigmatization to patronization: The media's distorted portrayal of physical disability." American Psychological Association.

Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova. 2012. "Female leadership raises aspirations and educational attainment for girls: A policy experiment in India." Science, 335(6068).

Bell, Philip. 2001. "Content analysis of visual images." In The Handbook of Visual Analysis. Chapter 2, 11–34. Thousand Oaks, CA:Sage.

Bernheim, B Douglas. 1994. "A theory of conformity." Journal of Political Economy, 102(5): 841–877.

Berry, Steven, and Joel Waldfogel. 2010. "Product quality and market size." The Journal of Industrial Economics, 58(1): 1–31.

Bertrand, Marianne, Dean Karlan, Sendhil Mullainathan, Eldar Shafir, and Jonathan Zinman. 2010. "What's advertising content worth? Evidence from a consumer credit marketing field experiment." The Quarterly Journal of Economics, 125(1): 263–306.

Beyer, Lucas. 2018. "Github: PyDenseCRF."

Bian, Lin, Sarah-Jane Leslie, and Andrei Cimpian. 2017. "Gender stereotypes about intellectual ability emerge early and influence children's interests." Science, 355(6323): 389–391.

Bigler, Rebecca S. 1999. "The use of multicultural curricula and materials to counter racism in children." Journal of Social Issues, 55(4): 687–705.

Bishop, Rudine Sims. 1990. "Mirrors, windows, and sliding glass doors. Perspectives: Choosing and Using Books for the Classroom, 6 (3)." Perspectives: Choosing and Using Books for the Classroom, 6(3): ix–xi.

Blau, Francine D, and Lawrence M Kahn. 2017. "The gender wage gap: Extent, trends, and explanations." Journal of Economic Literature, 55(3): 789–865.

Boer, Diana, Katja Hanke, and Jia He. 2018. "On detecting systematic measurement error in cross-cultural research: A review and critical reflection on equivalence and invariance tests." Journal of Cross-Cultural Psychology, 49(5): 713–734.

Broockman, David, and Joshua Kalla. 2022. "The manifold effects of partisan media on viewers' beliefs and attitudes: A field experiment with Fox News viewers." OSF Preprints, 1.

Brooks, Dwight E, and Lisa P Hébert. 2006. "Gender, race, and media representation." Handbook of Gender and Communication, 16.

Buolamwini, Joy, and Timnit Gebru. 2018. "Gender shades: Intersectional accuracy disparities in commercial gender classification." Conference on Fairness, Accountability and Transparency, 77–91.

Burchfield, Marcy, Henry G Overman, Diego Puga, and Matthew A Turner. 2006. "Causes of sprawl: A portrait from space." The Quarterly Journal of Economics, 121(2): 587–633.

Cantoni, Davide, Yuyu Chen, David Y Yang, Noam Yuchtman, and Y Jane Zhang. 2017. "Curriculum and ideology." Journal of Political Economy, 125(2): 338–392.

Cappelen, Alexander, John List, Anya Samek, and Bertil Tungodden. 2020. "The effect of early-childhood education on social preferences." Journal of Political Economy, 128(7): 2739–2758.

Chetty, Raj, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter. 2020. "Race and economic opportunity in the United States: An intergenerational perspective." The Quarterly Journal of Economics, 135(2): 711–783.

Clark, Heather D., George A. Wells, Charlotte Huët, Finlay A. McAlister, L. Rachid Salmi, Dean Fergusson, and Andreas Laupacis. 1999. "Assessing the quality of randomized trials: reliability of the Jadad scale." Controlled Clinical Trials, 20(5): 448–452.

Clark, Roger, Rachel Lennon, and Leanna Morris. 1993. "Of Caldecotts and kings: Gendered images in recent American children's books by Black and non-Black illustrators." Gender & Society, 7(2): 227–245.

Cockcroft, Marlaina. 2018. "Caldecott and Newbery Medal wins bring instant boost to book sales." School Library Journal, 64(2).

Craemer, Thomas, Trevor Smith, Brianna Harrison, Trevon Logan, Wesley Bellamy, and William Darity Jr. 2020. "Wealth implications of slavery and racial discrimination for African American descendants of the enslaved." The Review of Black Political Economy, 47(3): 218–254.

Crenshaw, Kimberlé. 1989. "Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics." University of Chicago Legal Forum, 1989(8).

Crenshaw, Kimberlé. 1990. "Mapping the margins: Intersectionality, identity politics, and violence against women of color." Stanford Law Review, 43: 1241.

Crisp, Thomas, and Brittany Hiller. 2011. "Telling tales about gender: A critical analysis of Caldecott Medal-winning picturebooks, 1938-2011." Journal of Children's Literature, 37(2): 18–29.

Daniels, Elizabeth A, Marlee C Layh, and Linda K Porzelius. 2016. "Grooming ten-year-olds with gender stereotypes? A content analysis of preteen and teen girl magazines." Body Image, 19.

Darity, William A, and Patrick L Mason. 1998. "Evidence on discrimination in employment: Codes of color, codes of gender." Journal of Economic Perspectives, 12(2): 63–90.

Das, Abhijit, Antitza Dantcheva, and Francois Bremond. 2018. "Mitigating bias in gender, age and ethnicity classification: a multi-task convolution neural network approach." Proceedings of the European Conference on Computer Vision (ECCV) Workshops.

Davis, Albert J. 1984. "Sex-differentiated behaviors in nonsexist picture books." Sex Roles, 11(1): 1–16.

Davis, Kathy. 2008. "Intersectionality as buzzword: A sociology of science perspective on what makes a feminist theory successful." Feminist Theory, 9(1): 67–85.

Dee, Thomas S. 2005. "A teacher like me: Does race, ethnicity, or gender matter?" American Economic Review, 95(2): 158–165.

de la Fuente-Núñez, Vânia, Ella Cohn-Schwartz, Senjooti Roy, and Liat Ayalon. 2021. "Scoping review on ageism against younger populations." International Journal of Environmental Research and Public Health, 18(8): 3988.

DellaVigna, Stefano, and Joshua M Pollet. 2007. "Demographics and industry returns." American Economic Review, 97(5): 1667–1702.

Dhar, Diva, Tarun Jain, and Seema Jayachandran. 2018. "Intergenerational transmission of gender attitudes: Evidence from India." Journal of Development Studies, 1–21.

Dhar, Diva, Tarun Jain, and Seema Jayachandran. 2022. "Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India." American Economic Review, 112(3): 899–927.

Dobrow, Julia R, and Calvin L Gidney. 1998. "The good, the bad, and the foreign: The use of dialect in children's animated television." The Annals of the American Academy of Political and Social Science, 557(1).

Eble, Alex, and Feng Hu. 2020. "Child beliefs, societal beliefs, and teacher-student identity match." Economics of Education Review, 77: 101994.

Eble, Alex, and Feng Hu. 2022. "Gendered beliefs about mathematics ability transmit across generations through children's peers." Nature Human Behaviour, 868–879.

Eitel, Alexander, Katharina Scheiter, Anne Schüler, Marcus Nyström, and Kenneth Holmqvist. 2013. "How a picture facilitates the process of learning from text: Evidence for scaffolding." Learning and Instruction, 28: 48–63.

Fletcher, JD, and Sigmund Tobias. 2005. "The multimedia principle." The Cambridge Handbook of Multimedia Learning, 117: 133.

Fuchs-Schündeln, Nicola, and Paolo Masella. 2016. "Long-lasting effects of socialist education." Review of Economics and Statistics, 98(3): 428–441.

Fu, Siyao, Haibo He, and Zeng-Guang Hou. 2014. "Learning race from face: A survey." IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(12): 2483–2509.

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. "Word embeddings quantify 100 years of gender and ethnic stereotypes." Proceedings of the National Academy of Sciences, 115(16): E3635–E3644.

Genicot, Garance, and Debraj Ray. 2017. "Aspirations and inequality." Econometrica, 85(2): 489–519.

Gentzkow, Matthew, and Jesse M Shapiro. 2010. "What drives media slant? Evidence from US daily newspapers." Econometrica, 78(1).

Gentzkow, Matthew, Bryan T. Kelly, and Matt Taddy. 2019. "Text as data." Journal of Economic Literature, 57(3): 535–74.

Gentzkow, Matthew, Jesse M Shapiro, and Matt Taddy. 2019. "Measuring group differences in high-dimensional choices: Method and application to Congressional speech." Econometrica, 87(4): 1307–1340.

Gershenson, Seth, Cassandra Hart, Joshua Hyman, Constance Lindsay, and Nicholas W Papageorge. 2018. "The long-run impacts of same-race teachers." National Bureau of Economic Research.

Ghavami, Negin, Dalal Katsiaficas, and Leoandra Onnie Rogers. 2016. "Toward an intersectional approach in developmental science: The role of race, gender, sexual orientation, and immigrant status." In Advances in Child Development and Behavior. Vol. 50, 31–73. Elsevier.

Giroux, Henry A. 1981. Ideology, Culture, and the Process of Schooling. Temple University Press.

Gopnik, Alison, Shaun O'Grady, Christopher G Lucas, Thomas L Griffiths, Adrienne Wente, Sophie Bridgers, Rosie Aboody, Hoki Fung, and Ronald E Dahl. 2017. "Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood." Proceedings of the National Academy of Sciences, 114(30): 7892–7899.

Hamilton, Mykol C, David Anderson, Michelle Broaddus, and Kate Young. 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): 757–765.

Han, Ted, and Amanda Hickman. 2019. "Our search for the best OCR tool, and what we found."

Henderson, J Vernon, Adam Storeygard, and David N Weil. 2012. "Measuring economic growth from outer space." American Economic Review, 102(2): 994–1028.

Hersch, Joni. 2008. "Profiling the new immigrant worker: The effects of skin color and height." Journal of Labor Economics, 26(2): 345–386.

Horrigan, John. 2015. "Chapter 1: Who uses libraries and what they do at their libraries." Pew Research Center. Accessed July 11, 2022.

Hughes, Julie M, Rebecca S Bigler, and Sheri R Levy. 2007. "Consequences of learning about historical racism among European American and African American children." Child Development, 78(6): 1689–1705.

Jackson, Aaron S, Michel Valstar, and Georgios Tzimiropoulos. 2016. "A CNN cascade for landmark guided semantic part segmentation." European Conference on Computer Vision, 143–155.

Jackson, Matthew O. 2010. Social and Economic Networks. Princeton university press.

Jakiela, Pamela, and Owen Ozier. 2018. "Gendered language." World Bank Policy Research Working Paper, , (8464).

Jansen, JD. 1997. "Critical theory and the school curriculum." Metatheories in Educational Theory and Practice.

Jayadev, Arjun, and Robert Johnson. 2017. "Tides and prejudice: racial attitudes during downturns in the United States 1979–2014." The Review of Black Political Economy, 44(3-4): 379–392.

Jensen, Robert, and Emily Oster. 2009. "The power of TV: Cable television and women's status in India." The Quarterly Journal of Economics, 124(3).

Kane, Michael T. 2013. "Validating the interpretations and uses of test scores." Journal of Educational Measurement, 50(1).

Kearney, Melissa S, and Phillip B Levine. 2019. "Early childhood education by television: Lessons from Sesame Street." American Economic Journal: Applied Economics, 11(1): 318–50.

Kearney, Melissa S, and Phillip B Levine. 2020. "Role models, mentors, and media influences." The Future of Children, 30(1): 83–106.

Knowles, Elizabeth, Liz Knowles, and Martha Smith. 1997. The Reading Connection: Bringing Parents, Teachers, and Librarians Together. Libraries Unlimited.

Kolbe, Richard, and Joseph C La Voie. 1981. "Sex-role stereotyping in preschool children's picture books." Social Psychology Quarterly, 369–374.

Kortenhaus, Carole M, and Jack Demarest. 1993. "Gender role stereotyping in children's literature: An update." Sex Roles, 28(3): 219–232.

Koss, Melanie D. 2015. "Diversity in contemporary picturebooks: A content analysis." Journal of Children's Literature, 41(1): 32.

Koss, Melanie D, and Kathleen A Paciga. 2020. "Diversity in Newbery Medal-winning titles: A content analysis." Journal of Language and Literacy Education, 16(2): n2.

Koss, Melanie D, Nancy J Johnson, and Miriam Martinez. 2018. "Mapping the diversity in Caldecott books from 1938 to 2017: The changing topography." Journal of Children's Literature, 44(1): 4–20.

Kozlowski, Austin C, Matt Taddy, and James A Evans. 2019. "The geometry of culture: Analyzing the meanings of class through word embeddings." American Sociological Review, 84(5).

Krippendorff, Klaus. 1980. "Validity in content analysis." In Computerstrategien fÃr die kommunikationsanalyse. 69–112. Frankfurt, Germany:Campus Verlag.

Krippendorff, Klaus. 2018. Content Analysis: An Introduction to its Methodology. Sage publications.

Krishnan, Anoop, Ali Almadan, and Ajita Rattani. 2020. "Understanding fairness of gender classification algorithms across gender-race groups." arXiv.

La Ferrara, Eliana, Alberto Chong, and Suzanne Duryea. 2012. "Soap operas and fertility: Evidence from Brazil." American Economic Journal: Applied Economics, 4(4): 1–31.

Leslie, David. 2020. "Understanding bias in facial recognition technologies: An explainer." The Alan Turing Institute.

Levie, W Howard, and Richard Lentz. 1982. "Effects of text illustrations: A review of research." Ectj, 30(4): 195–232.

Levstik, Linda S, and Cynthia A Tyson. 2010. Handbook of Research in Social Studies Education. Routledge.

Lewis, Randall A, and Justin M Rao. 2015. "The unfavorable economics of measuring the returns to advertising." The Quarterly Journal of Economics, 130(4): 1941–1973.

Linderman, Alf. 2001. "Computer content analysis and manual coding techniques: A comparative analysis." Progress in Communication Sciences, 97–110.

Logan, Trevon D. 2022. "American enslavement and the recovery of Black economic history." Journal of Economic Perspectives, 36(2): 81–98.

Lu, Conny. 2018. "Github: Face segmentation with CNN and CRF."

Lumini, Alessandra, and Loris Nanni. 2020. "Fair comparison of skin detection approaches on publicly available datasets." Expert Systems with Applications, 160: 113677.

Martin, Ardis C. 2008. "Television media as a potential negative factor in the racial identity development of African American youth." Academic Psychiatry.

Marx, David M, Sei Jin Ko, and Ray A Friedman. 2009. "The "Obama effect": How a salient role model reduces race-based performance differences." Journal of Experimental Social Psychology, 45(4): 953–956.

McCabe, Janice, Emily Fairchild, Liz Grauerholz, Bernice A Pescosolido, and Daniel Tope. 2011. "Gender in twentieth-century children's books: Patterns of disparity in titles and central characters." Gender & Society, 25(2): 197–226.

Michalopoulos, Stelios, and Melanie Meng Xue. 2021. "Folklore." The Quarterly Journal of Economics, 1: 54.

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. "Model cards for model reporting." Proceedings of the Conference on Fairness, Accountability, and Transparency, 220–229.

Monk Jr, Ellis P. 2015. "The cost of color: Skin color, discrimination, and health among African-Americans." American Journal of Sociology, 121(2): 396–444.

Muhammad, Khalil Gibran. 2019. The Condemnation of Blackness: Race, Crime, and the Making of Modern Urban America. Cambridge, MA:Harvard University Press.

Mullen, Lincoln. 2020. "gender: Predict gender from names using historical data." R package version 0.5.4.

Nagpal, Shruti, Maneet Singh, Richa Singh, and Mayank Vatsa. 2019. "Deep learning for face recognition: Pride or prejudiced?" arXiv.

Neuendorf, Kimberly A. 2016. The Content Analysis Guidebook. Sage.

Nilsen, Alleen Pace. 1971. "Women in children's literature." College English, 32(8): 918–926.

O'Flaherty, Brendan. 2015. The Economics of Race in the United States. Cambridge, MA:Harvard University Press.

O'Kelly, Charlotte G. 1974. "Sexism in children's television." Journalism Quarterly, 51(4).

Paceley, Megan S, and Karen Flynn. 2012. "Media representations of bullying toward queer youth: Gender, race, and age discrepancies." Journal of LGBT Youth, 9(4).

Plant, E Ashby, Patricia G Devine, William TL Cox, Corey Columb, Saul L Miller, Joanna Goplen, and B Michelle Peruche. 2009. "The Obama effect: Decreasing implicit prejudice and stereotyping." Journal of Experimental Social Psychology, 45(4): 961–964.

Porter, Catherine, and Danila Serra. 2020. "Gender differences in the choice of major: The importance of female role models." American Economic Journal: Applied Economics, 12(3): 226–54.

Pressley, Michael. 1977. "Imagery and children's learning: Putting the picture in developmental perspective." Review of Educational Research, 47(4): 585–622.

Quillian, Lincoln, Devah Pager, Ole Hexel, and Arnfinn H Midtbøen. 2017. "Meta-analysis of field experiments shows no change in racial discrimination in hiring over time." Proceedings of the National Academy of Sciences, 114(41): 10870–10875.

Riley, Emma. 2022. "Role Models in Movies: The Impact of Queen of Katwe on Students' Educational Attainment." The Review of Economics and Statistics, 1–48.

Roberts, Dorothy. 2011. Fatal Invention: How Science, Politics, and Big Business Re-create Race in the Twenty-first Century. New York:New Press/ORIM.

Rosenberg, Stanley D, Paula P Schnurr, and Thomas E Oxman. 1990. "Content analysis: A comparison of manual and computerized systems." Journal of Personality Assessment, 54(1-2): 298–310.

Sadoski, Mark, and Allan Paivio. 2013. Imagery and Text: A Dual Coding Theory of Reading and Writing. Routledge.

Sarna, Tadeusz, Janice M Burke, Witold Korytowski, Małgorzata Różanowska, Christine MB Skumatz, Agnieszka Zaręba, and Mariusz Zaręba. 2003. "Loss of melanin from human RPE with aging: Possible role of melanin photooxidation." Experimental Eye Research, 76(1): 89–98.

Schaffner, Brian, and Stephen Ansolabhere. 2019. "2017 CCES Common Content."

Schuman, Howard, Charlotte Steeh, Lawrence Bobo, Maria Krysan, et al. 1997. Racial Attitudes in America: Trends and Interpretations. Harvard University Press.

Sculley, David. 2010. "Web-scale k-means clustering." Proceedings of the 19th International World Wide Web Conference, 1177–1178.

Smith, Tom W, Michael Davern, Jeremy Freese, and Stephen Morgan. 2021. "General Social Surveys Cross-Sectional Cumulative Data, 1972-2021." Chicago: NORC.

Smith, Vicky. 2013. "The 'Caldecott effect'." Children and Libraries: The Journal of the Association for Library Service to Children, 1(1): 9–13.

Sobel, Joel. 2005. "Interdependent preferences and reciprocity." Journal of Economic Literature, 43(2): 392–436.

Sokolova, Marina, and Guy Lapalme. 2009. "A systematic analysis of performance measures for classification tasks." Information Processing & Management, 45(4): 427–437.

Steele, Claude M. 2010. Whistling Vivaldi: And Other Clues to How Stereotypes Affect Us (Issues of our Time). WW Norton & Company.

Steele, Claude M, and Joshua Aronson. 1995. "Stereotype threat and the intellectual test performance of African Americans." Journal of Personality and Social Psychology, 69(5): 797–811.

Stewig, John Warren, and Mary Lynn Knipfel. 1975. "Sexism in picture books: What progress?" The Elementary School Journal, 76(3).

Stout, Jane G, Nilanjana Dasgupta, Matthew Hunsinger, and Melissa A McManus. 2011. "STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM)." Journal of Personality and Social Psychology, 100(2): 255.

Szasz, Teodora, Emileigh Harrison, Ping-Jung Liu, Ping-Chang Lin, Hakizumwami Birali Runesha, and Anjali Adukia. 2022. "Measuring representation of race, gender, and age in children's books: Face detection and feature classification in illustrated images." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision.

Tafti, Ahmad P, Ahmadreza Baghaie, Mehdi Assefi, Hamid R Arabnia, Zeyun Yu, and Peggy Peissig. 2016. "OCR as a service: An experimental evaluation of Google Docs OCR, Tesseract, ABBYY FineReader, and Transym." International Symposium on Visual Computing, 735–746.

Towbin, Mia Adessa, Shelley A Haddock, Toni Schindler Zimmerman, Lori K Lund, and Litsa Renee Tanner. 2004. "Images of gender, race, age, and sexual orientation in Disney feature-length animated films." Journal of Feminist Family Therapy, 15(4): 19–44.

Tyson, Cynthia A. 2002. "'Get up offa that thing': African American middle school students respond to literature to develop a framework for understanding social action." Theory & Research in Social Education, 30(1): 42–65.

US Census. 2019. "Census QuickFacts." Accessed March 13, 2021.

Valadez, Corinne, Sandra Murillo Sutterby, and Tammy Francis Donaldson. 2013. "Content analysis of Latino award-winning children's literature." In Preparing effective leaders for tomorrow's schools. , ed. Norma Zunker, 77–104. Corpus Christi, TX:Consortium for Educational Development, Evaluation, and Research; Texas AM.

Van Kleeck, Anne, Steven A Stahl, and Eurydice B Bauer. 2003. On Reading Books to Children: Parents and Teachers. Routledge.

Vekiri, Ioanna. 2002. "What is the value of graphical displays in learning?" Educational Psychology Review, 14: 261–312.

Vespa, Jonathan, Lauren Medina, and David Armstrong. 2020. "Demographic turning points for the United States: Population projections for 2020 to 2060." United States Census Bureau Current Population Reports. Accessed February 24, 2023.

Vezhnevets, Vladimir, Vassili Sazonov, and Alla Andreeva. 2003. "A survey on pixel-based skin color detection techniques." Moscow, Russia.

Vijayarani, S, and A Sakila. 2015. "Performance comparison of OCR tools." International Journal of UbiComp (IJU), 6(3): 19–30.

Waldfogel, Joel. 2003. "Preference externalities: an empirical study of who benefits whom in differentiated-product markets." RAND Journal of Economics, 34(3): 557–569.

Waldfogel, Joel. 2007. The Tyranny of the Market. Harvard University Press.

Weitzman, Lenore J, Deborah Eifler, Elizabeth Hokada, and Catherine Ross. 1972. "Sex-role socialization in picture books for preschool children." American Journal of Sociology, 77(6): 1125–1150.

Westman, Jack C. 1991. "Juvenile ageism: Unrecognized prejudice and discrimination against the young." Child Psychiatry and Human Development, 21: 237–256.

Williams Jr, J Allen, JoEtta Vernon, Martha C Williams, and Karen Malecha. 1987. "Sex role socialization in picture books: An update." Social Science Quarterly, 68: 148–156.

Wilson, William J. 2012. The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy. University of Chicago Press.

Witt, Susan D. 2000. "Review of research: The influence of television on children's gender role socialization." Childhood Education, 76(5): 322–324.

Yu, Amy Zhao, Shahar Ronen, Kevin Hu, Tiffany Lu, and César A Hidalgo. 2016. "Pantheon 1.0, a manually verified dataset of globally famous biographies." Scientific Data, 3(1): 1–16.

Zhang, Zhifei, Song Yang, and Hairong Qi. 2017. "Age progression/regression by conditional adversarial autoencoder." IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Zhou, Lei, Zhi Liu, and Xiangjian He. 2017. "Face parsing via a fully-convolutional continuous CRF neural network." CoRR, abs/1708.03736.

Zoph, Barret, and Quoc V Le. 2017. "Neural architecture search with reinforcement learning." arXiv.

VIII Exhibits: Tables and Figures

Table 1. Summary Statistics

	Mainstream (1)	Diversity (2)	People of Color (3)	African American (4)	Ability (5)	Female (6)	LGBTQIA+ (7)
Collection Totals							
Total Number of Books	495	635	577	130	29	14	15
Range of Years in our Sample	1923-2019	1971-2019	1971-2019	1971-2017	2000-2014	2013-2017	2010-2017
Book-Level Averages: Book Attributes							
Number of Pages	139	148	137	147	213	314	268
Number of Words	24,362	26,520	23,816	26,328	35,273	87,411	56,771
Number of Faces	44	59	60	41	30	30	79
Number of Famous People	3	8	7	9	5	40	13
% Faces - Monochromatic Skin Color	58%	47%	47%	52%	45%	55%	45%
Book-Level Averages: Skin Color							
Perceptual Skin Tint of All Faces	55	44	44	41	46	34	47
Book-Level Averages: Putative Race							
% Faces Classified as Asian	6%	16%	16%	11%	6%	9%	4%
% Faces Classified as Black	2%	13%	13%	22%	8%	21%	3%
% Faces Classified as Latinx + Others	4%	3%	3%	3%	4%	1%	5%
% Faces Classified as White	88%	68%	67%	64%	82%	69%	88%
% Famous People Classified as Asian	3%	7%	7%	1%	3%	8%	5%
% Famous People Classified as Black	5%	22%	23%	55%	8%	21%	8%
% Famous People Classified as Indigenous	0%	1%	1%	0%	0%	1%	0%
% Famous People Classified as Latinx	1%	9%	10%	0%	1%	0%	2%
% Famous People Classified as Multiracial	0%	2%	2%	1%	0%	1%	3%
% Famous People Classified as White	92%	59%	57%	44%	87%	68%	82%
Book-Level Averages: Gender							
% Faces Classified as Female	48%	50%	49%	43%	67%	71%	48%
% Female Gendered Words	34%	43%	42%	40%	42%	56%	45%
% Famous People Classified as Female	15%	22%	20%	24%	28%	37%	41%
Book-Level Averages: Age							
% Faces Classified as Children	19%	14%	14%	10%	19%	3%	18%
% Young Gendered Words	26%	20%	20%	21%	17%	21%	32%

Note: In this table, we present summary statistics (described in the row titles) for each collection of books we analyze (named in the column titles). Percentages may not sum to one due to rounding error.

Table 2. Gender Representation in Book Content by Purchaser Identities

	Dependent Variable: Percent of Female			
	Words	Names	Faces	Images vs. Text
	(1)	(2)	(3)	(4)
Panel A: Gender of Purchaser Child				
Purchaser Has a Daughter	0.032***	0.019**	−0.002	−3.399***
	(0.008)	(0.009)	(0.010)	(1.148)
Purchaser Has a Son	−0.012	−0.020**	0.003	1.437
	(0.008)	(0.009)	(0.010)	(1.137)
Constant	0.385***	0.363***	0.415***	4.456***
(Baseline Group: No Children)	(0.003)	(0.003)	(0.004)	(0.400)
Observations	9,658	9,419	6,680	8,049
Adjusted R^2	0.0020	0.0010	−0.0003	0.001
Panel B: Purchaser Gender				
Male	−0.015***	−0.017***	−0.019***	−0.654
	(0.005)	(0.006)	(0.006)	(0.711)
Other	−0.006	−0.038**	0.024	1.386
	(0.016)	(0.019)	(0.021)	(2.378)
Constant	0.388***	0.370***	0.432***	6.645***
(Baseline Group: Female)	(0.002)	(0.002)	(0.002)	(0.238)
Observations	28,645	28,120	18,737	22,213
Adjusted R^2	0.0003	0.0004	0.0004	−0.00003

Note: We regress indicator variables for whether the purchaser has a daughter or son (Panel A) and purchaser gender (Panel B) on four different measures of female representation contained in a purchased book. The dependent variable in the first column is the percent of female words out of all gendered words where gendered words include all gendered names, gendered pronouns, and gendered terms. The dependent variable in the second column is the percent of female names out of all gendered names. The dependent variable in the third column is the percent of female faces out of all faces detected. The dependent variable in the fourth column is the difference between the third and first column dependent variables. We obtain book-level purchasing data from the Numerator OmniPanel which contains data on purchases made from 2017-2020 and merge it with our curated data on representation in award-winning children's books. We subset purchasing data to include purchases of award-winning children's books which we have digitized that contain at least one gendered word, name, or face. *p<0.1; **p<0.05; ***p<0.01

Table 3. Skin Color and Race Representation in Book Content by Purchaser Identities

	Dependent variable:				
	Average Skin Tint	Percent of Famous Mentions by Race			
		Asian	*Black*	*Latinx*	*White*
Purchaser Ethnicity	(1)	(2)	(3)	(4)	(5)
Asian	−0.074	0.005***	−0.005	0.002	−0.004
	(0.709)	(0.002)	(0.007)	(0.002)	(0.008)
Black/African American	−6.467***	−0.001	0.120***	0.005**	−0.125***
	(0.720)	(0.002)	(0.007)	(0.002)	(0.008)
Hispanic/Latino	−3.287***	0.001	0.014**	0.013***	−0.028***
	(0.645)	(0.001)	(0.006)	(0.002)	(0.007)
Other	−2.409**	0.003	0.023**	−0.002	−0.025**
	(1.031)	(0.002)	(0.010)	(0.003)	(0.011)
Constant	59.240***	0.008***	0.078***	0.007***	0.904***
(Baseline Group: White)	(0.190)	(0.0005)	(0.002)	(0.001)	(0.002)
Observations	14,189	18,219	18,219	18,219	18,219
Adjusted R^2	0.0070	0.0004	0.0160	0.0030	0.0150

Note: We regress indicator variables indicating the race or ethnicity of the purchaser on five different dependent variables. The dependent variable in column 1 represents the average skin tint of characters in each book purchased in our sample. The dependent variables in columns 2-5 represent the percentage of famous people of a different race mentioned in the text of each book purchased in our sample. We get book-level purchasing data from the Numerator OmniPanel which contains data on purchases made from 2017-2020 and merge it with our curated data on representation in award-winning children's books. We subset purchasing data to include purchases of award-winning children's books which we have digitized that contain at least one detected face in column 1 and that contain at least one mention of a famous person in columns 2-5. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 4. Number of Mainstream and Diversity Books in Library Collection by Community Characteristics

	Dependent variable: *Number of Award Winning Children's Books by Collection*			
	Mainstream	Diversity	Mainstream	Diversity
	(1)	(2)	(3)	(4)
% of Population White, Non-Hispanic	0.465***	−1.177***	0.324**	−0.770*
	(0.167)	(0.355)	(0.159)	(0.388)
Median Household Income			0.0002	−0.001
			(0.0002)	(0.0004)
% of Population Below Poverty Line			0.238	−0.531
			(0.447)	(0.778)
Number of Children's Books in Library Branch	0.011***	0.021***	0.011***	0.021***
	(0.0004)	(0.001)	(0.0004)	(0.001)
Total Population	0.0005	−0.002**	0.0005	−0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	−1.245	67.706**	−14.690	100.308*
	(13.427)	(30.033)	(27.152)	(53.866)
Observations	53	53	53	53
Adjusted R^2	0.983	0.984	0.982	0.984

Note: Each observation in the data used to make this table corresponds to a community reporting area (CRA). Each community area is manually matched to its closest Seattle Public Library branch. Each Seattle Public Library branch is matched to at least one CRA. Column 1 shows that the number of books which were recognized by a Mainstream award available in the library branch closest to a given CRA is increasing in the proportion of the CRA population that is White, non-Hispanic. Column 2 shows that this relationship is decreasing for books which were recognized by a Diversity award. Columns 3 and 4 show these results are robust to including measures of household income for a given CRA. Population demographics are taken from the American Community Survey, 5-year Series 2013-2017 accessed through Seattle's Data Portal. Seattle Public Library inventory data as reported on October 1st, 2017 also accessed through Seattle's Data Portal. Standard errors are clustered at the library branch level. Variables containing percentages are scaled so that potential values range from $0 - 100$. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 5. Readership by Collection

Panel A: Average Price and Copies Purchased In Numerator OmniPanel

Collection	Number of Copies Sold (1)	Mean Book Price (2)	Number of Unique Titles (3)	Mean Copies Sold Per Title (4)
Mainstream	40,854	$7.66	493	83
Diversity	35,553	$9.34	1,067	33
All Other Children's Books	1,683,406	$7.42	97,866	17
People of Color	26,899	$9.51	880	31
African American	9,081	$9.95	149	61
Female	4,892	$8.68	120	41
Ability	2,834	$8.70	55	52
LGBTQIA+	2,838	$9.07	34	83

Note: In this table, we present summary statistics (described in the column titles) on prices and quantities for purchases of children's books from different collections (named in the row titles) using book purchase level data from the Numerator OmniPanel from 2017-2020

Panel B: Seattle Public Library Inventory and Checkouts

Collection	Number of Checkouts (1)	Mean Checkouts Per Title (2)	Number of Unique Titles (3)	Mean Library Copies Per Title (4)
Mainstream	107,866	823	131	14.0
Diversity	176,828	200	883	6.6
All Other Children's Books	12,918,820	220	58,785	5.6
People of Color	155,217	206	755	6.6
African American	18,197	236	77	8.3
Female	7,240	97	75	6.5
Ability	13,028	296	44	7.5
LGBTQIA+	8,276	251	33	9.3

Note: In this table, we present summary statistics (described in the column titles) for library book checkouts of children's books from different collections (named in the row titles) using data on library book inventory and checkouts from the Seattle Public Library system between 2005-2017.

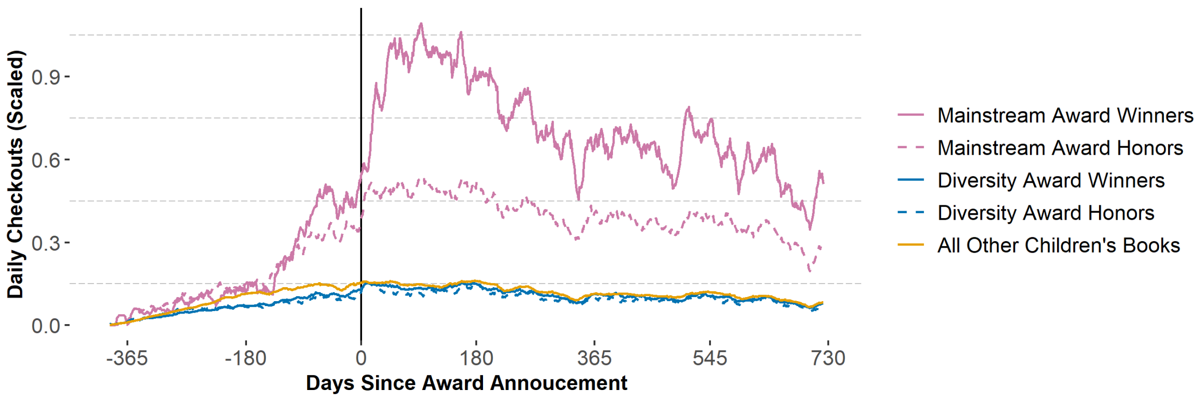
Table 6. Local Beliefs and Children's Book Purchases within Zip Codes

	Dependent variable:			
	% of Respondents who think the U.S. government should		% of Respondents somewhat or strongly agree	
	Identify and deport undocumented immigrants	Withhold federal funds from localities that do not follow federal immigration laws	White people in the U.S. have certain advantages because of the color of their skin	I am angry that racism exists
	(1)	(2)	(3)	(4)
% of Children's Books Purchased that Won a Diversity Award	−0.517***	−0.677***	0.582***	0.117
	(0.107)	(0.107)	(0.109)	(0.087)
% of Children's Books Purchased that Won a Mainstream Award	−0.245**	0.063	0.321***	0.023
	(0.118)	(0.119)	(0.120)	(0.096)
Constant	40.347***	58.045***	52.380***	79.683***
	(0.549)	(0.552)	(0.560)	(0.446)
Observations	9,046	9,046	9,046	9,046
Adjusted R^2	0.003	0.004	0.004	−0.000

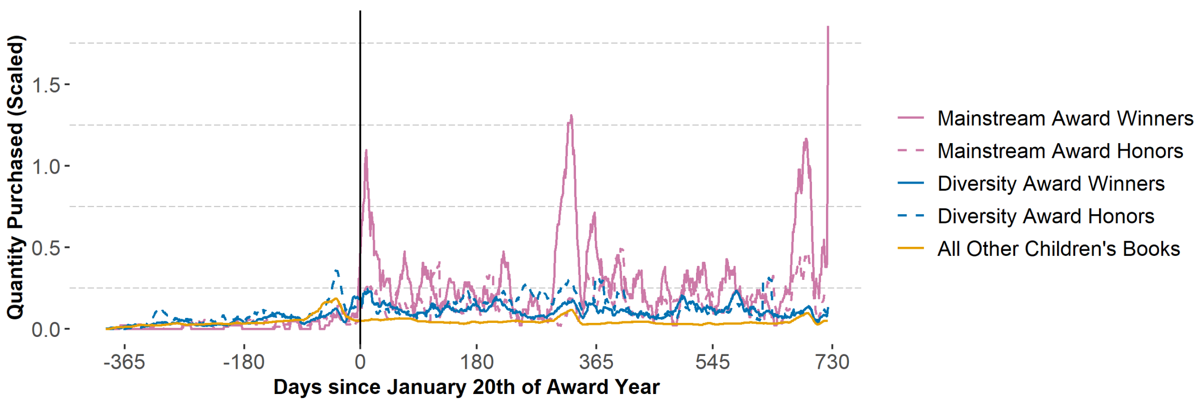
Note: In this table, we regress the percentage of respondents surveyed in a zip code who agree with a statement or policy (described in the column titles) on the percentage of all children's books purchased in that zip code which were recognized by an award in our Mainstream collection and/or Diversity collection. Data on beliefs at the zip code level are drawn from the 2017 Cooperative Election Study Common Content Survey (Schaffner and Ansolabhere, 2019). Data on children's book purchases at the zip code level are drawn from the 2017-2020 Numerator OmniPanel data. Variables containing percentages are scaled so that potential values range from $0 - 100$. *p<0.1; **p<0.05; ***p<0.01

Figure 1. Children's Book Readership Centered Around Award Announcements

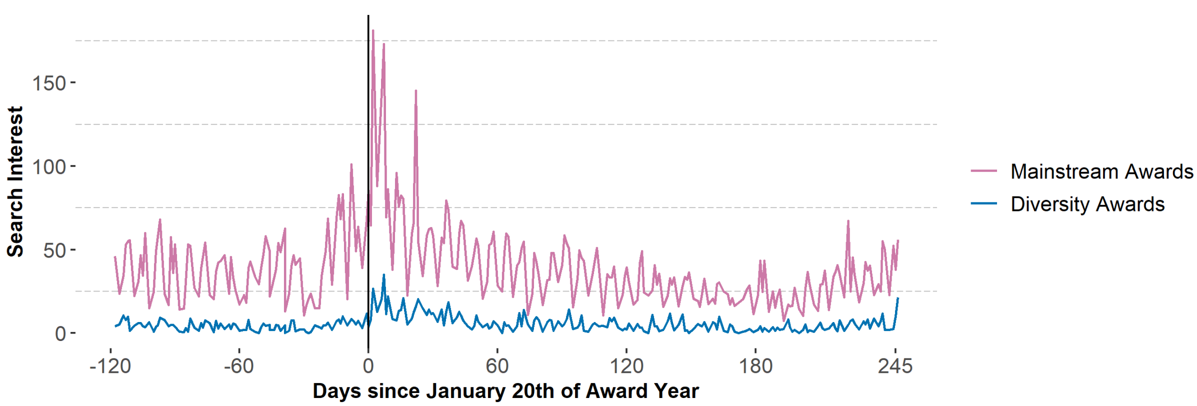
(a) Library Checkouts



(b) Purchases



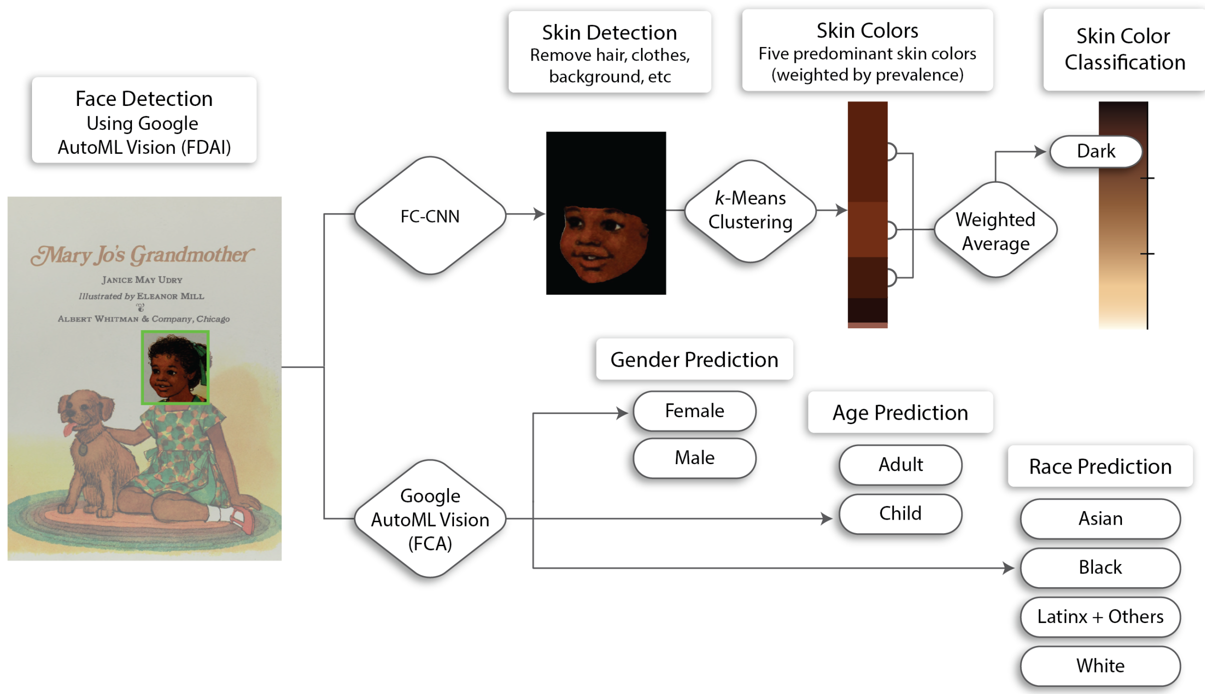
(c) Search Interest



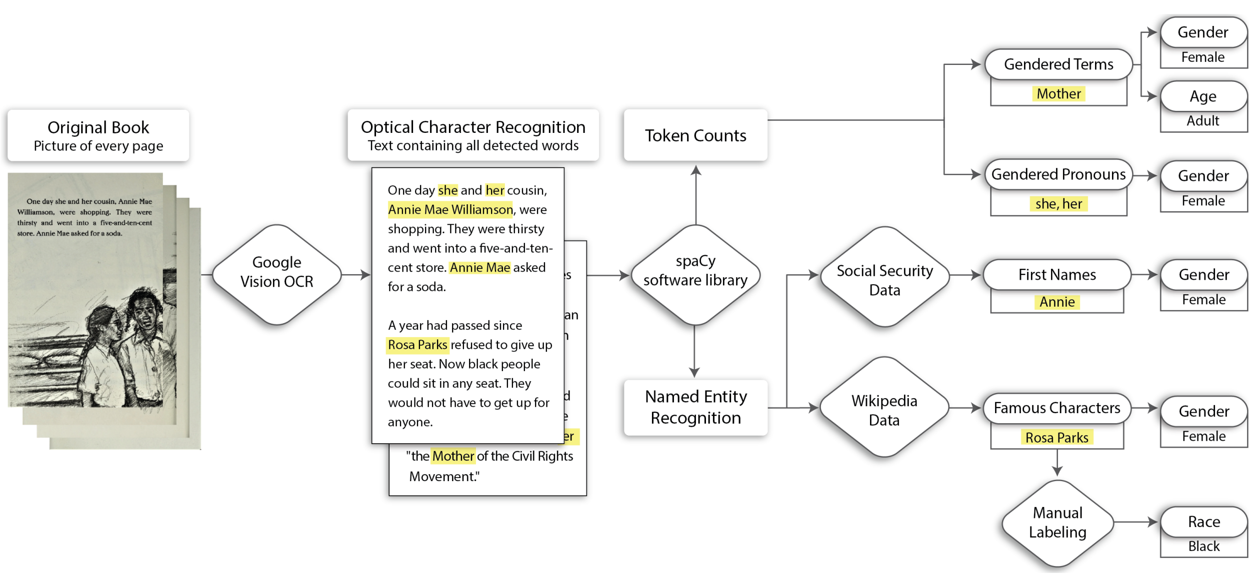
Note: Panel A shows average daily checkouts of children's library books between 2005-2017 from the Seattle Public Library. Panel B shows average daily children's book purchases between 2017-2020 from the Numerator OmniPanel. Both panels are disaggregated by Mainstream books (award winners vs. honorees), Diversity books (award winners vs. honorees), or children's books not recognized by an award in either collection. We scale daily checkouts and purchases by the number of unique titles in each collection and smooth the data with a 14-day moving average. Panel C shows average weekly search interest in the U.S. between 2017-2021 from Google Trends data. We collect search interest for the eight awards with unique topic IDs in Google Trends as described in Section II. Panels are centered around the time of award announcements each year.

Figure 2. Converting Images and Text into Data

(a) Image-to-Data Pipeline

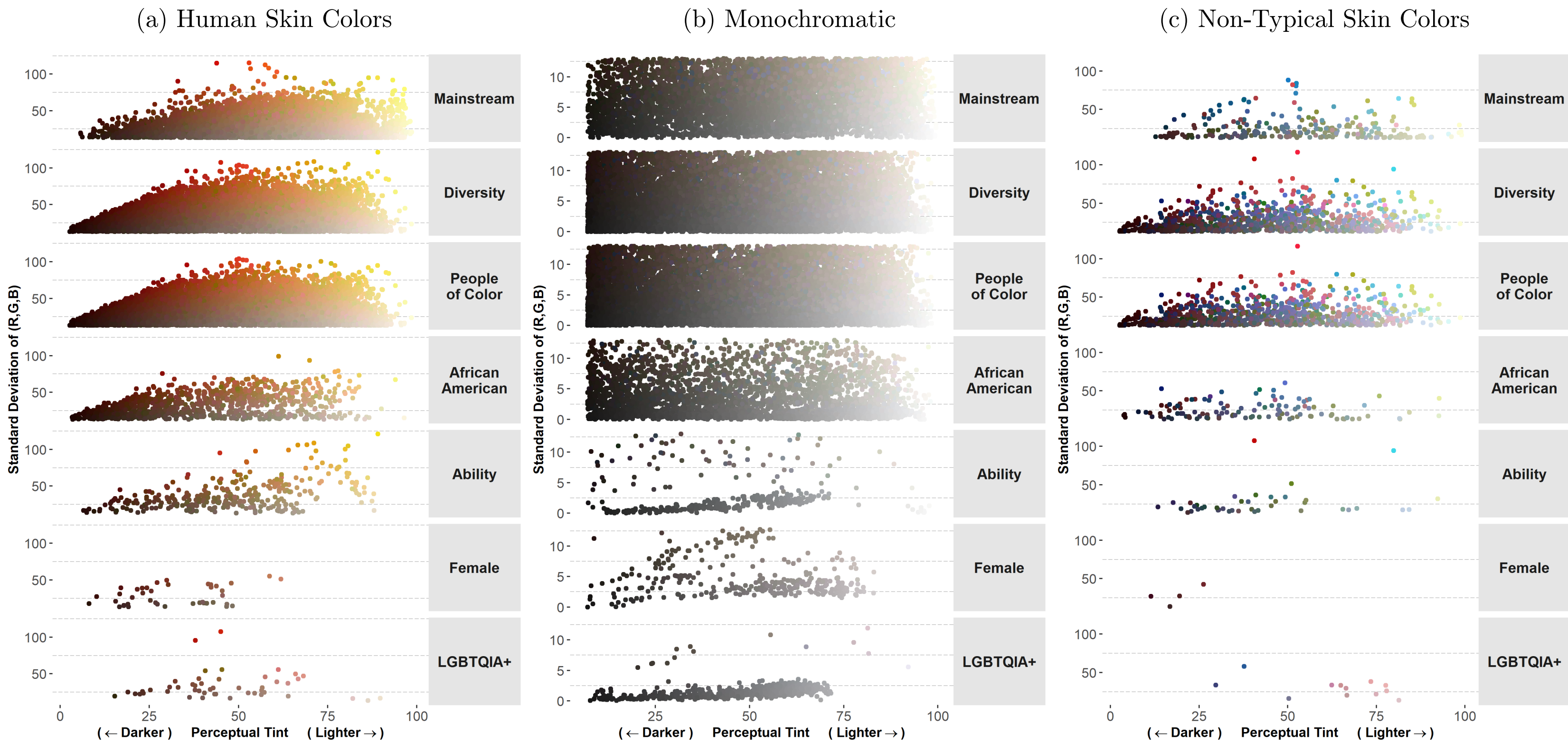


(b) Text-to-Data Pipeline



Note: In this figure, we show how we process scanned book pages into image and text data. In Panel A, we show how we extract image data to construct image measures of skin color, race, gender, and age. In Panel B, we show how we extract and isolate various dimensions of text to construct textual measures of gender, race, and age.

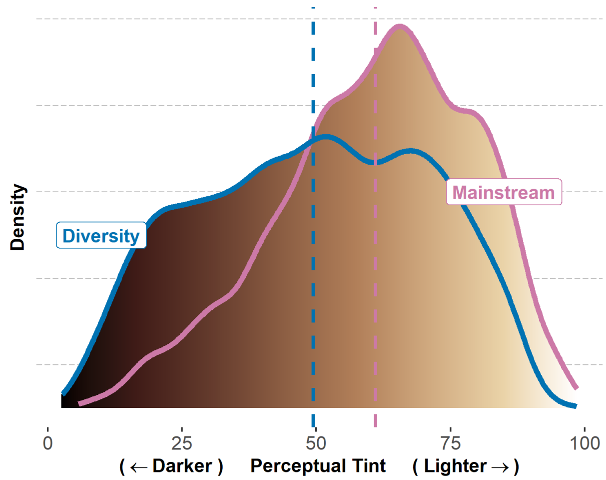
Figure 3. Skin Color Data, by Color Type



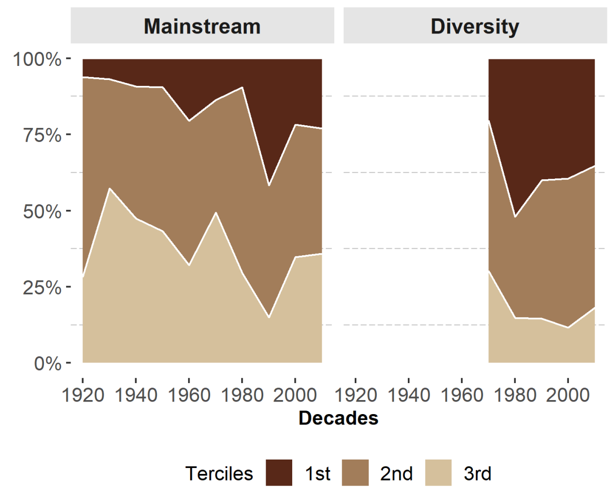
Note: This figure shows the representative skin colors of the individual faces we detect in the images found in the books from each collection. We show these by the three color "types" present in these images: human skin colors (polychromatic skin colors where R $\geq$ G $\geq$ B), monochromatic skin colors (e.g., black and white, sepia), and non-typical polychromatic skin colors (e.g., blue, green). The y-axis indicates the standard deviation of the RGB values of each face. The higher the standard deviation, the more vibrant the color.

Figure 4. Skin Colors in Faces, by Collection: Human Skin Colors

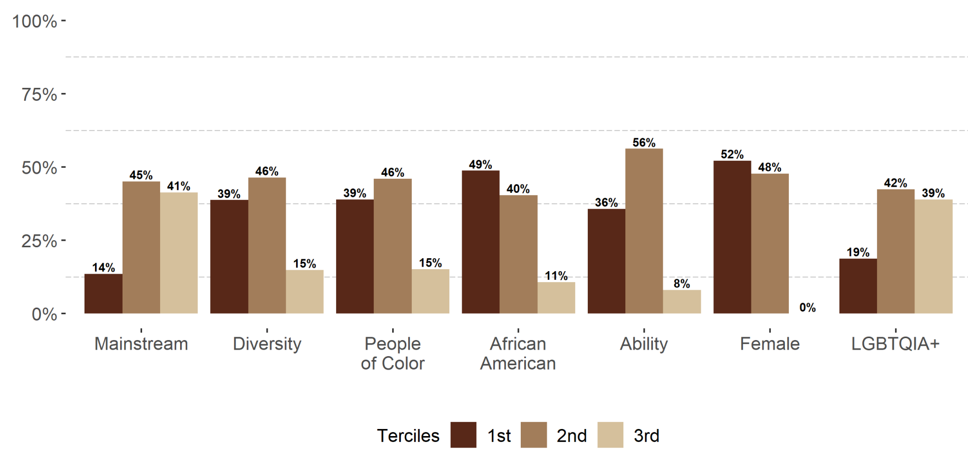
(a) Distribution of Skin Colors



(b) Mean Proportion in Each Tercile, Over Time



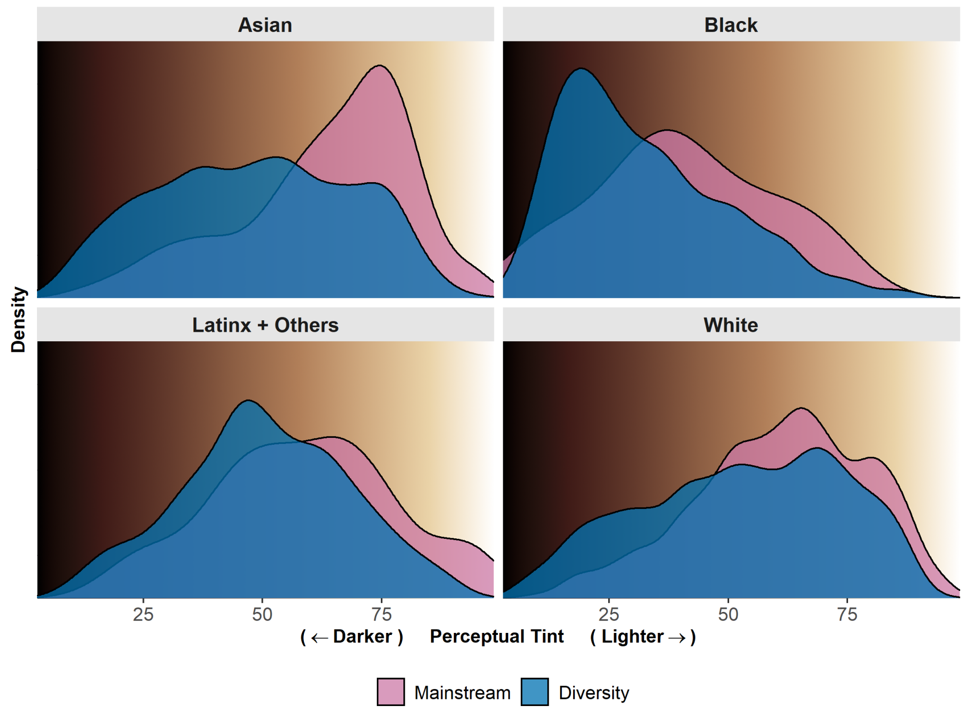
(c) Mean Proportion in Each Tercile, All Collections



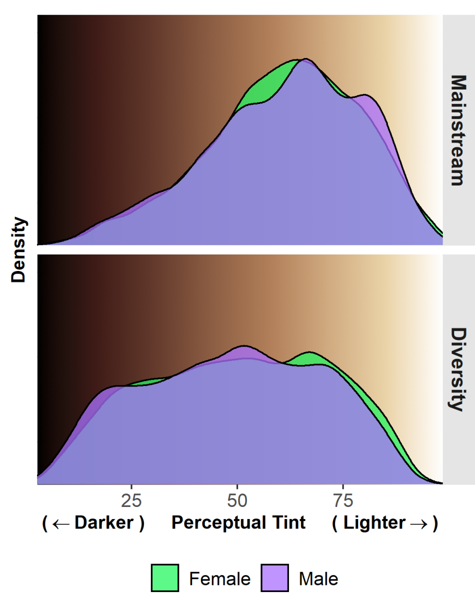
Note: This figure shows our analysis of the representative skin colors of the individual faces we detect in the images found in the books we analyze, focusing on faces considered to be human skin colors (polychromatic skin colors where R $\geq$ G $\geq$ B). Panel A shows the distribution of skin color tint for faces detected in books from the Mainstream and Diversity collections. The mean for each distribution is denoted with a dashed line. In Panel B, we show the average proportion of faces in each tercile, over time, for faces in the Mainstream and Diversity collections. Panel C shows the overall collection-specific average proportion of faces in each skin color tercile for each of the seven collections. Skin classification methods are described in Section IV.A.

Figure 5. Skin Color by Predicted Race, Gender, and Age of Detected Faces

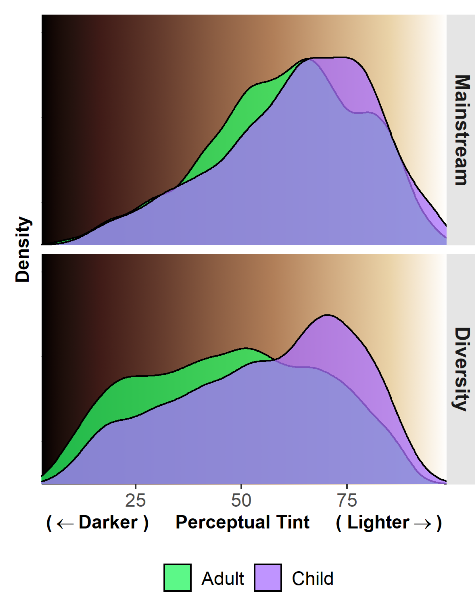
(a) Skin Color Distribution by Race



(b) Skin Color Distribution by Gender

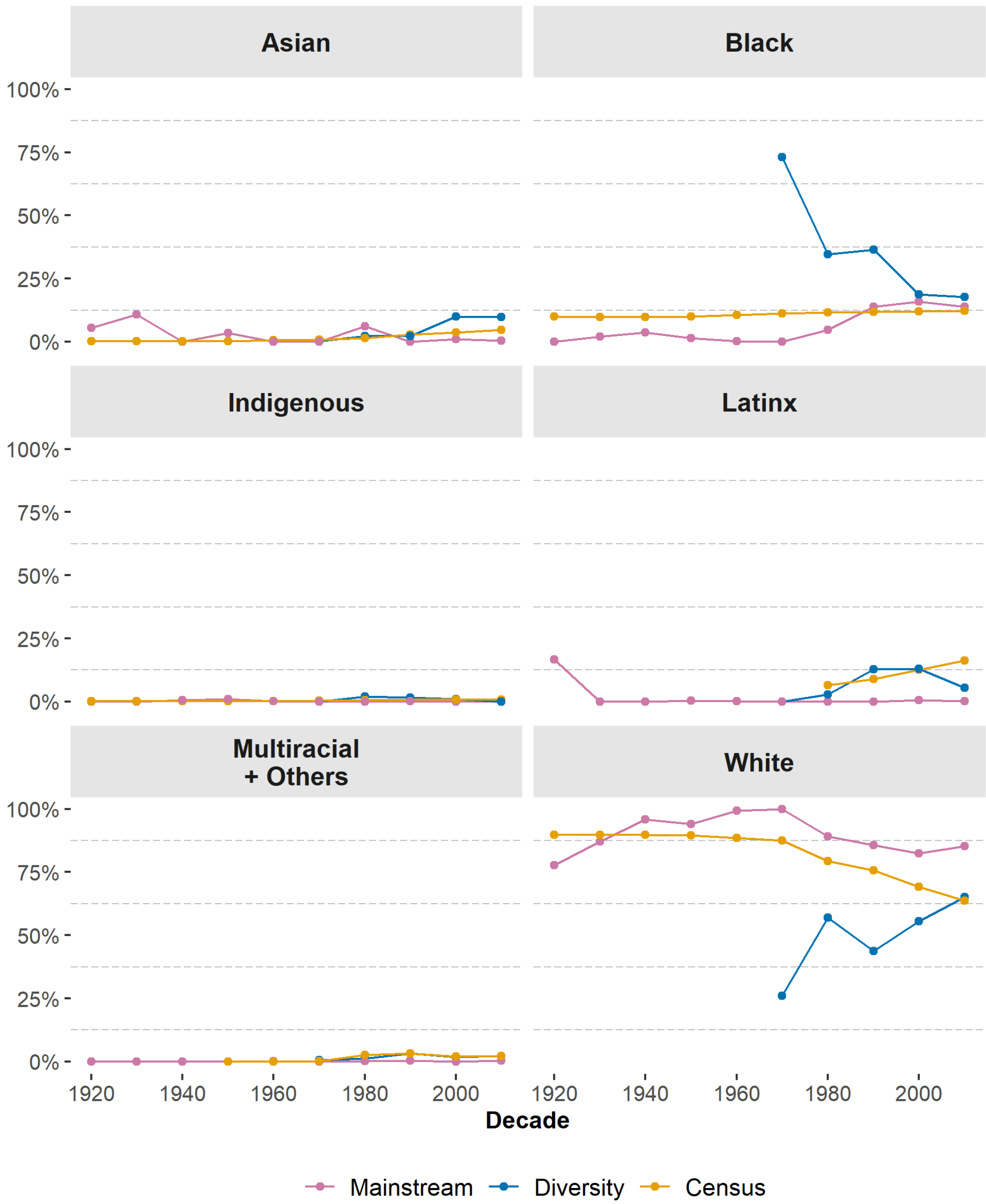


(c) Skin Color Distribution by Age



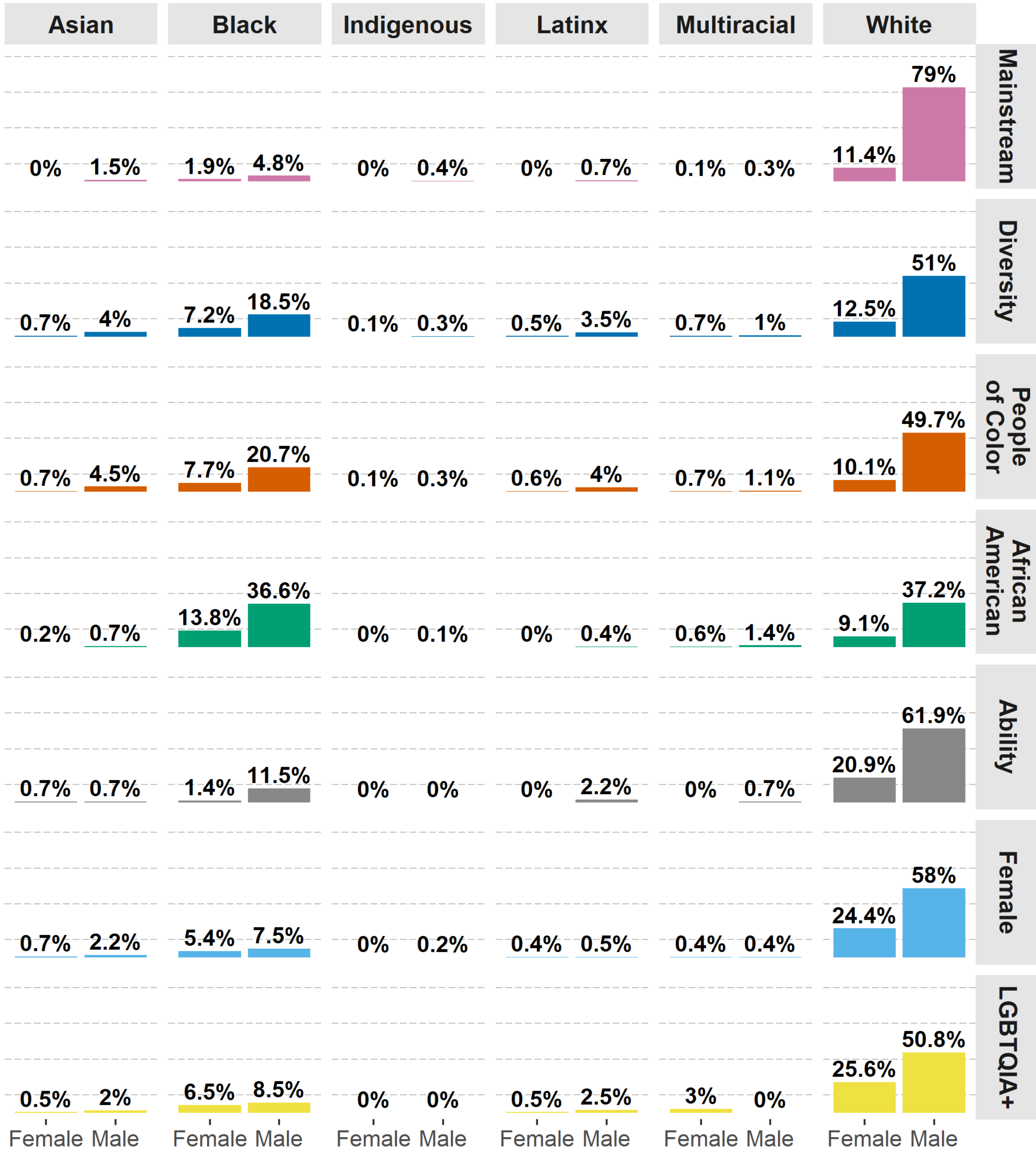
Note: This figure shows the distribution of skin color tint by predicted features of the detected faces in the Mainstream and Diversity collections. Panel A shows differences in the skin tint distributions between collections, conditional on predicted race. Panel B shows differences in the skin tint distributions between faces predicted to be male and faces predicted to be female, conditional on collection. Panel C shows differences in the skin tint distributions between faces predicted to be adults and faces predicted to be children, conditional on collection. Skin tint extracted using methods described in Section IV.A.2. Race, gender, and age were predicted using methods described in Section IV.A.3.

Figure 6. Share of U.S. Population and Famous People in the Text, by Race/Ethnicity



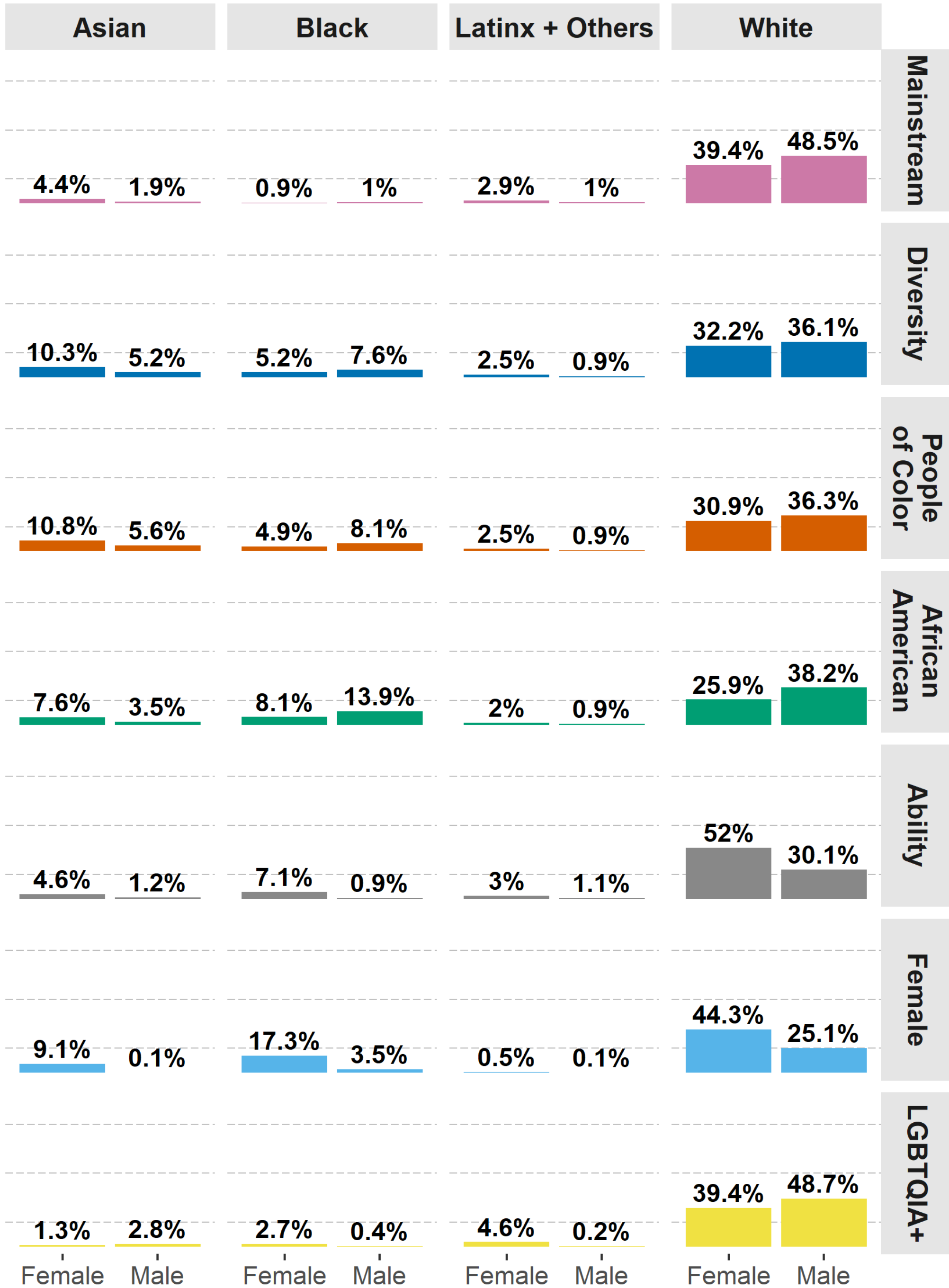
Note: In this figure, we find the percent breakdown of famous people mentioned in a given book by race/ethnicity. For example, if Aretha Franklin was mentioned 3 times in a book and Jimmy Carter is mentioned 2 times, then 60% of the mentions of famous people in that book would be Black. We then show the average percentage breakdown over all books by collection and decade for the Mainstream and Diversity collections. We also show the share of the U.S. population by race/ethnicity for each decade as a comparison. We classify famous people using methods described in Section IV.B. We collapse the following identities: East Asian, Middle Eastern, and South Asian into the Asian category; North American Indigenous peoples and South American Indigenous peoples into the Indigenous category; and African American and Black African into the Black category. If an individual was coded as having more than one race, we classify them as multiracial. See Appendix Figure B4 for a similar version of this graph with non-standard axes to better see changes in groups with small population proportions.

Figure 7. Race and Gender Classifications of Famous Figures in the Text



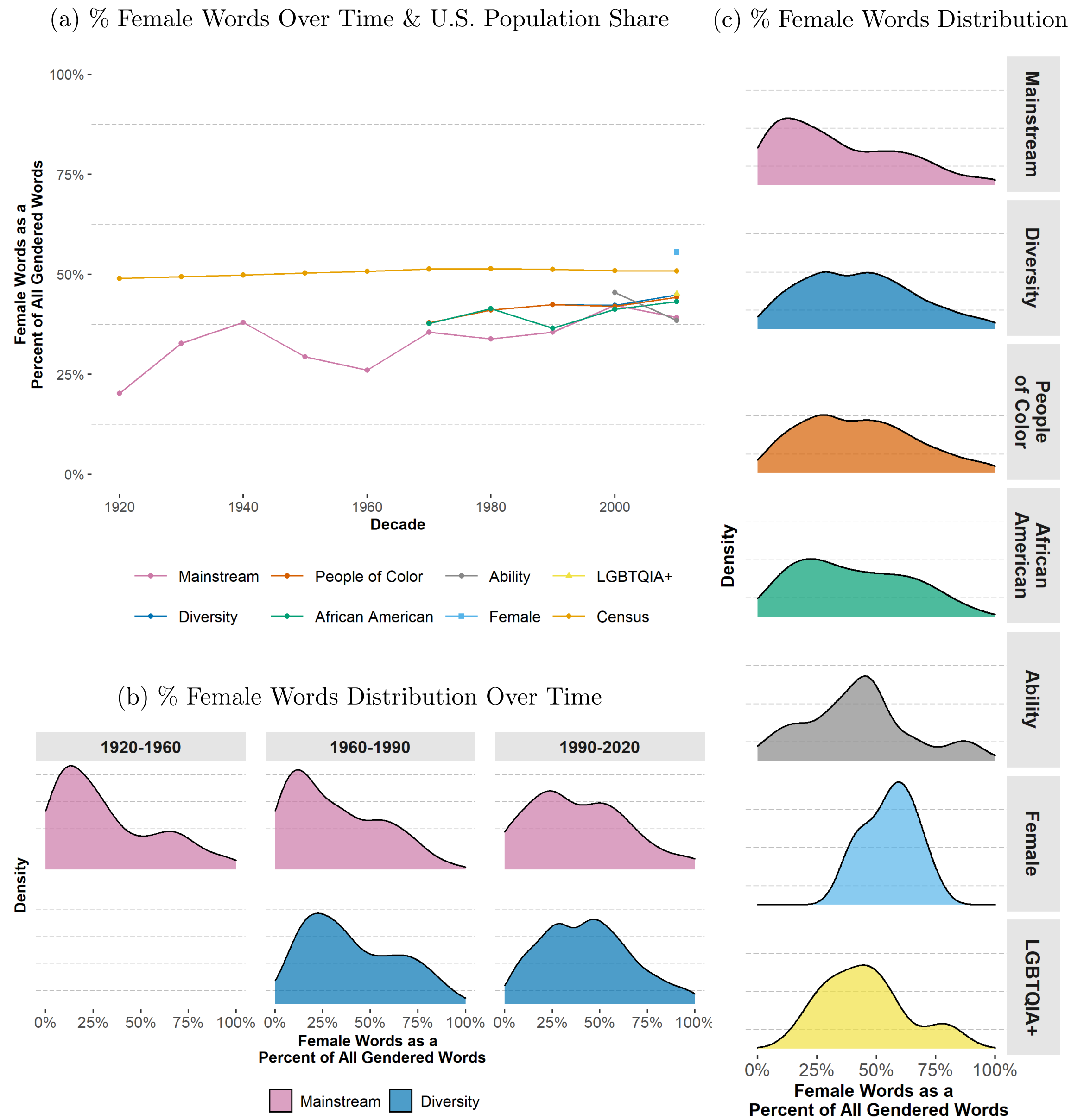
Note: In this figure, we count the number of famous people mentioned at least once in a given book and sum over all books in a collection. We then show the percentage breakdown of these famous people by race and gender. For example, if Aretha Franklin was mentioned at least once in two separate books within the Diversity collection, we would count her twice for that collection. We identify famous individuals and their predicted gender using methods described in Section IV.B. We manually label the race of famous individuals. We collapse the following identities: East Asian, Middle Eastern, and South Asian into the Asian category; North American Indigenous peoples and South American Indigenous peoples into the Indigenous category; and African American and Black African into the Black category. If an individual was coded as having more than one race, we classify them as multiracial. See Appendix Figure B5 for the same figure broken down by race alone.

Figure 8. Race and Gender Predictions of Pictured Characters



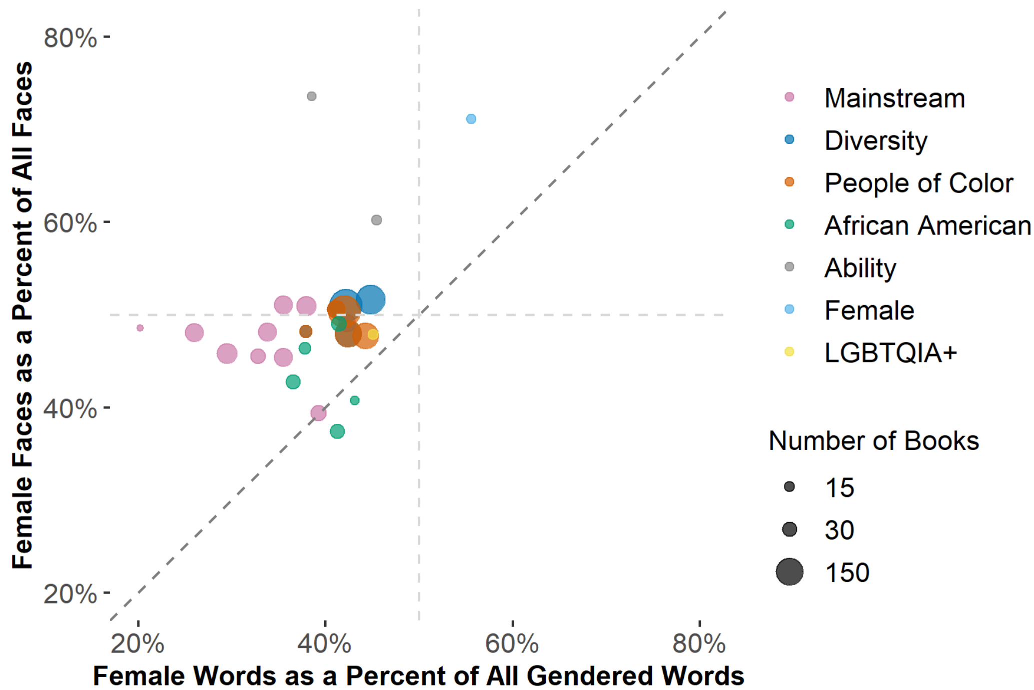
Note: In this figure, we show the proportion of detected faces in all collections by race and gender predictions. Race and gender were classified by our trained AutoML model as described in Section IV.A.3. See Appendix Figure B7 for the same figure broken down by race alone.

Figure 9. Female Words as a Percent of All Gendered Words



Note: In this figure, we show female words as a percentage of all gendered words in three different ways. Panel A shows how this average varies by decade. Panel B shows the distributions over time in the Mainstream and Diversity collections. Panel C shows the distribution over all books in a collection. In this case, gendered words encompass the total number of gendered names, gendered pronouns, and a pre-specified list of other gendered terms (e.g., queen, nephew). We list the pre-specified gendered terms in the Data Appendix.

Figure 10. Female Representation in Images and Text of Children's Books



Note: In this figure, we plot collection-by-decade average percentages of female representation in images (on the y-axis) and female representation in text (on the x-axis). This enables a comparison between the proportion of females represented in the images and the proportion of females represented in the text of the children's books in our sample.

Appendices

A Appendix Tables

Table A1. Summary Statistics for Children's Book Purchasers in Numerator Data

	All Children's Books		*Award-Winning Children's Books*	
Purchaser Demographics	N	Mean	N	Mean
Children				
Has Children	1,547,044	0.73	62,283	0.70
Has Children Ages 0-5	1,188,039	0.23	47,782	0.14
Has Children Ages 6-12	1,188,039	0.46	47,782	0.38
Has Children Ages 13-17	1,188,039	0.23	47,782	0.35
Race/Ethnicity				
Asian	1,506,152	0.06	60,633	0.06
Black/African American	1,506,152	0.04	60,633	0.07
Hispanic/Latino	1,506,152	0.06	60,633	0.08
White/Caucasian	1,506,152	0.81	60,633	0.75
Other Ethnicity	1,506,152	0.03	60,633	0.03
Gender				
Female	1,534,051	0.89	61,714	0.88
Male	1,534,051	0.10	61,714	0.11
Other	1,534,051	0.01	61,714	0.01
Sexuality				
Gay/Lesbian	1,111,247	0.01	41,943	0.02
Straight	1,111,247	0.82	41,943	0.81
Bisexual	1,111,247	0.03	41,943	0.03
Other Sexuality	1,111,247	0.01	41,943	0.01
Prefer Not to Answer	1,111,247	0.13	41,943	0.14
Income				
High Income	1,539,767	0.49	62,031	0.51
Mid Income	1,539,767	0.31	62,031	0.30
Low Income	1,539,767	0.20	62,031	0.19
Education				
Advanced Education	1,548,085	0.25	62,345	0.31
College Education	1,548,085	0.62	62,345	0.58
High School Education	1,548,085	0.12	62,345	0.09
Less than High School	1,548,085	0.02	62,345	0.02

Note: This table shows the sample size and mean of purchaser demographics for children's book purchases in Numerator OmniPanel data from 2017-2020. The first two columns include all children's book purchases. The last two columns include all purchases of a children's book which was recognized by one of the awards in our sample.

Table A2. Top Five Most Mentioned Famous People, by Collection

Collection	Rank	Name	Race	Gender	Mentions	Books
Mainstream	1	George Washington	White	Male	152	32
Mainstream	2	Abraham Lincoln	White	Male	270	25
Mainstream	3	Thomas Jefferson	White	Male	71	15
Mainstream	4	John Adams	White	Male	60	14
Mainstream	5	Benjamin Franklin	White	Male	23	12
Diversity	1	Martin Luther King Junior	Black	Male	282	51
Diversity	2	Abraham Lincoln	White	Male	72	41
Diversity	3	George Washington	White	Male	62	40
Diversity	4	Frederick Douglass	Black	Male	131	30
Diversity	5	Langston Hughes	Black	Male	109	30
People of Color	1	Martin Luther King Junior	Black	Male	263	48
People of Color	2	Abraham Lincoln	White	Male	70	39
People of Color	3	George Washington	White	Male	58	37
People of Color	4	Frederick Douglass	Black	Male	131	30
People of Color	5	Langston Hughes	Black	Male	108	29
African American	1	Langston Hughes	Black	Male	53	17
African American	2	Martin Luther King Junior	Black	Male	130	16
African American	3	Malcolm X	Black	Male	69	12
African American	4	Frederick Douglass	Black	Male	43	12
African American	5	Duke Ellington	Black	Male	25	12
Ability	1	Harold Pinter	White	Male	78	2
Ability	2	Andy Warhol	White	Male	4	2
Ability	3	Marco Polo	White	Male	3	2
Ability	4	Duke Ellington	Black	Male	2	2
Ability	5	Judy Blume	White	Female	2	2
Female	1	John F. Kennedy	White	Male	8	4
Female	2	Martin Luther King Junior	Black	Male	19	3
Female	3	Jimmy Carter	White	Male	15	3
Female	4	Betty Friedan	White	Female	10	3
Female	5	Richard Nixon	White	Male	9	3
LGBTQIA+	1	Alicia Keys	Multiracial	Female	3	3
LGBTQIA+	2	Britney Spears	White	Female	3	3
LGBTQIA+	3	Marilyn Monroe	White	Female	3	3
LGBTQIA+	4	Julia Roberts	White	Female	5	2
LGBTQIA+	5	Alexander Hamilton	White	Male	4	2

Note: This table shows the five most frequently mentioned famous people in each collection, along with their race, their gender, the number of times they were mentioned, and the number of books in which they appeared.

Table A3. Top Five Most Mentioned Famous Females, by Collection

Collection	Rank	Name	Race	Mentions	Books
Mainstream	1	Eleanor Roosevelt	White	30	7
Mainstream	2	Martha Washington	White	9	6
Mainstream	3	Emily Dickinson	White	7	6
Mainstream	4	Shirley Temple	White	12	5
Mainstream	5	Rosa Parks	Black	43	4
Diversity	1	Rosa Parks	Black	157	27
Diversity	2	Harriet Tubman	Black	35	19
Diversity	3	Eleanor Roosevelt	White	42	18
Diversity	4	Coretta Scott King	Black	23	15
Diversity	5	Lena Horne	White	20	14
People of Color	1	Rosa Parks	Black	152	25
People of Color	2	Harriet Tubman	Black	35	19
People of Color	3	Eleanor Roosevelt	White	41	17
People of Color	4	Coretta Scott King	Black	22	14
People of Color	5	Lena Horne	White	20	14
African American	1	Rosa Parks	Black	44	11
African American	2	Coretta Scott King	Black	12	10
African American	3	Zora Neale Hurston	Black	21	9
African American	4	Lena Horne	White	14	9
African American	5	Harriet Tubman	Black	13	9
Ability	1	Judy Blume	White	2	2
Ability	2	Shirley Temple	White	12	1
Ability	3	Anna Lee	White	4	1
Ability	4	Avril Lavigne	White	4	1
Ability	5	Marilyn Vos Savant	White	4	1
Female	1	Betty Friedan	White	10	3
Female	2	Mary Pickford	White	5	3
Female	3	Billie Jean King	White	24	2
Female	4	Katharine Graham	White	14	2
Female	5	Gloria Steinem	White	13	2
LGBTQIA+	1	Alicia Keys	Multiracial	3	3
LGBTQIA+	2	Britney Spears	White	3	3
LGBTQIA+	3	Marilyn Monroe	White	3	3
LGBTQIA+	4	Julia Roberts	White	5	2
LGBTQIA+	5	Patsy Cline	White	3	2

Note: In this table, we show the five most frequently mentioned famous females in each collection, along with their race, the number of times they were mentioned, and the number of books in which they appeared.

Table A4. Top Five Most Mentioned Famous Males, by Collection

Collection	Rank	Name	Race	Mentions	Books
Mainstream	1	George Washington	White	152	32
Mainstream	2	Abraham Lincoln	White	270	25
Mainstream	3	Thomas Jefferson	White	71	15
Mainstream	4	John Adams	White	60	14
Mainstream	5	Benjamin Franklin	White	23	12
Diversity	1	Martin Luther King Junior	Black	282	51
Diversity	2	Abraham Lincoln	White	72	41
Diversity	3	George Washington	White	62	40
Diversity	4	Frederick Douglass	Black	131	30
Diversity	5	Langston Hughes	Black	109	30
People of Color	1	Martin Luther King Junior	Black	263	48
People of Color	2	Abraham Lincoln	White	70	39
People of Color	3	George Washington	White	58	37
People of Color	4	Frederick Douglass	Black	131	30
People of Color	5	Langston Hughes	Black	108	29
African American	1	Langston Hughes	Black	53	17
African American	2	Martin Luther King Junior	Black	130	16
African American	3	Malcolm X	Black	69	12
African American	4	Frederick Douglass	Black	43	12
African American	5	Duke Ellington	Black	25	12
Ability	1	Harold Pinter	White	78	2
Ability	2	Andy Warhol	White	4	2
Ability	3	Marco Polo	White	3	2
Ability	4	Duke Ellington	Black	2	2
Ability	5	Mark Twain	White	2	2
Female	1	John F. Kennedy	White	8	4
Female	2	Martin Luther King Junior	Black	19	3
Female	3	Jimmy Carter	White	15	3
Female	4	Richard Nixon	White	9	3
Female	5	Barack Obama	Black	5	3
LGBTQIA+	1	Alexander Hamilton	White	4	2
LGBTQIA+	2	Adam Lambert	White	3	2
LGBTQIA+	3	Alice Cooper	White	3	2
LGBTQIA+	4	James Dean	White	3	2
LGBTQIA+	5	Michael Jackson	Black	3	2

Note: In this table, we show the five most frequently mentioned famous males in each collection, along with their race, the number of times they were mentioned, and the number of books in which they appeared.

Table A5. Top Mentioned Famous Person, by Collection and Decade

Decade	Mainstream	Diversity	People of Color	African American	Ability	Female	LGBTQ
1920	James Fenimore Cooper *White Male* Charles Darwin *White Male* Mark Twain *White Male*						
1930	Abraham Lincoln *White Male*						
1940	Benjamin Franklin *White Male*						
1950	George Washington *White Male*						
1960	George Washington *White Male*						
1970	Claude Lorrain *White Male* Leonardo da Vinci *White Male*	Frederick Douglass *Black Male*	Frederick Douglass *Black Male*	Frederick Douglass *Black Male*			
1980	George Washington *White Male*	Franklin D. Roosevelt *White Male*	Franklin D. Roosevelt *White Male*	Paul Robeson *Black Male*			
1990	William Shakespeare *White Male*	Martin Luther King Jr. *Black Male*	Martin Luther King Jr. *Black Male*	Martin Luther King Jr. *Black Male*			
2000	Martin Luther King Jr. *Black Male*	George Washington *White Male*	George Washington *White Male*	Langston Hughes *Black Male*	Judy Blume *White Female*		
2010	George Washington *White Male*	Martin Luther King Jr. *Black Male*	Martin Luther King Jr. *Black Male*	Malcolm X *Black Male*	Andy Warhol *White Male*	John F. Kennedy *White Male*	Alicia Keys *Multiracial Female* Marilyn Monroe *White Female* Britney Spears *White Female*

Note: In this table, we show the top most uniquely mentioned famous figure in each collection by decade. When multiple names are listed for a collection within the same decade, it indicates that each of those people were tied for the most mentioned famous person in that collection-by-decade.

Table A6. Correlation between U.S. Demographics and Representation

	Dependent variable: Percent of			
	Faces	Famous People	Female Words	Images vs. Text
	(1)	(2)	(3)	(4)
Panel A: Percent of Labor Force Participation				
Females	-0.08	0.69**	0.45***	-0.36
	(0.15)	(0.29)	(0.15)	(0.22)
Panel B: Percent of Population				
Asian	1.49**	-0.89		
	(0.5)	(0.73)		
Black	1.49***	5.15***		
	(0.25)	(1.36)		
Latinx	-0.53	0.03		
	(0.21)	(0.05)		
White	0.40**	0.33		
	(0.16)	(0.24)		

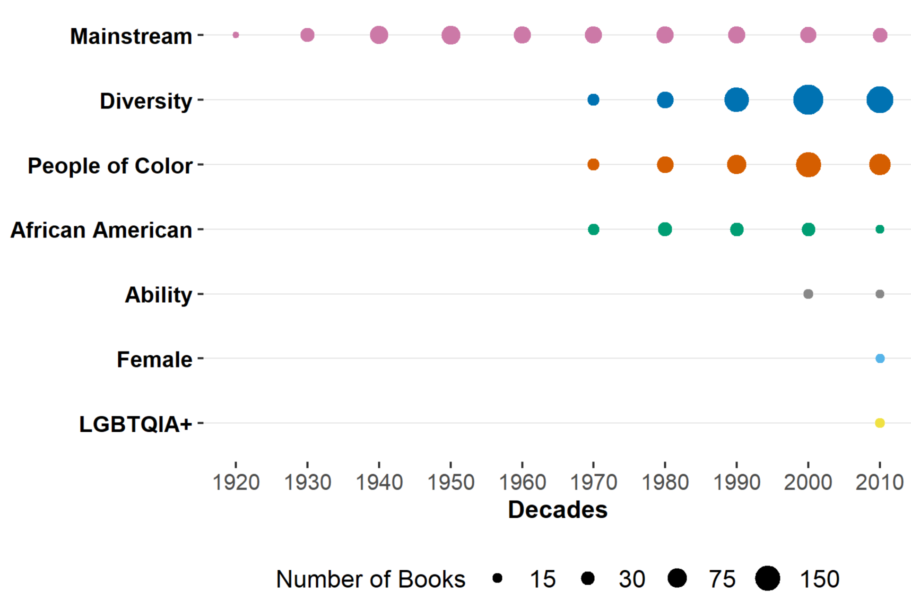
Note: This table estimates the relationship between major demographic parameters (U.S. female labor force participation in Panel A and the racial composition of the U.S. population in Panel B) and representation in the images and text of children's books from our Mainstream collection. We regress a measure of market share or market power – either population share of a given racial group or female labor force participation – for a given race or gender on a measure of their proportional representation in award winning children's books over time. We show each coefficient from these bivariate regressions in this table, with standard errors in parentheses. For example, the first row and column shows the coefficient from a regression of the percentage of female labor force participation on the percentage of female faces in the Mainstream collection over time. Our data on female labor force participation is constructed by taking the yearly average over monthly unadjusted data between 1948-2019 from the U.S. Bureau of Labor Statistics and retrieved from FRED, Federal Reserve Bank of St. Louis. Our data on population breakdown by race is from 1920-2019 U.S. census data. Census information on the proportion of people who are Latinx comes from a response to a question regarding ethnicity and is not mutually exclusive to the other race categories. We construct each race/ethnicity category to be mutually exclusive; for example, we count an individual who identifies as Latinx and White in the Latinx category, not the White category. Census data on ethnicity are only available beginning in 1970.

B Appendix Figures

Figure B1. Books in the Sample

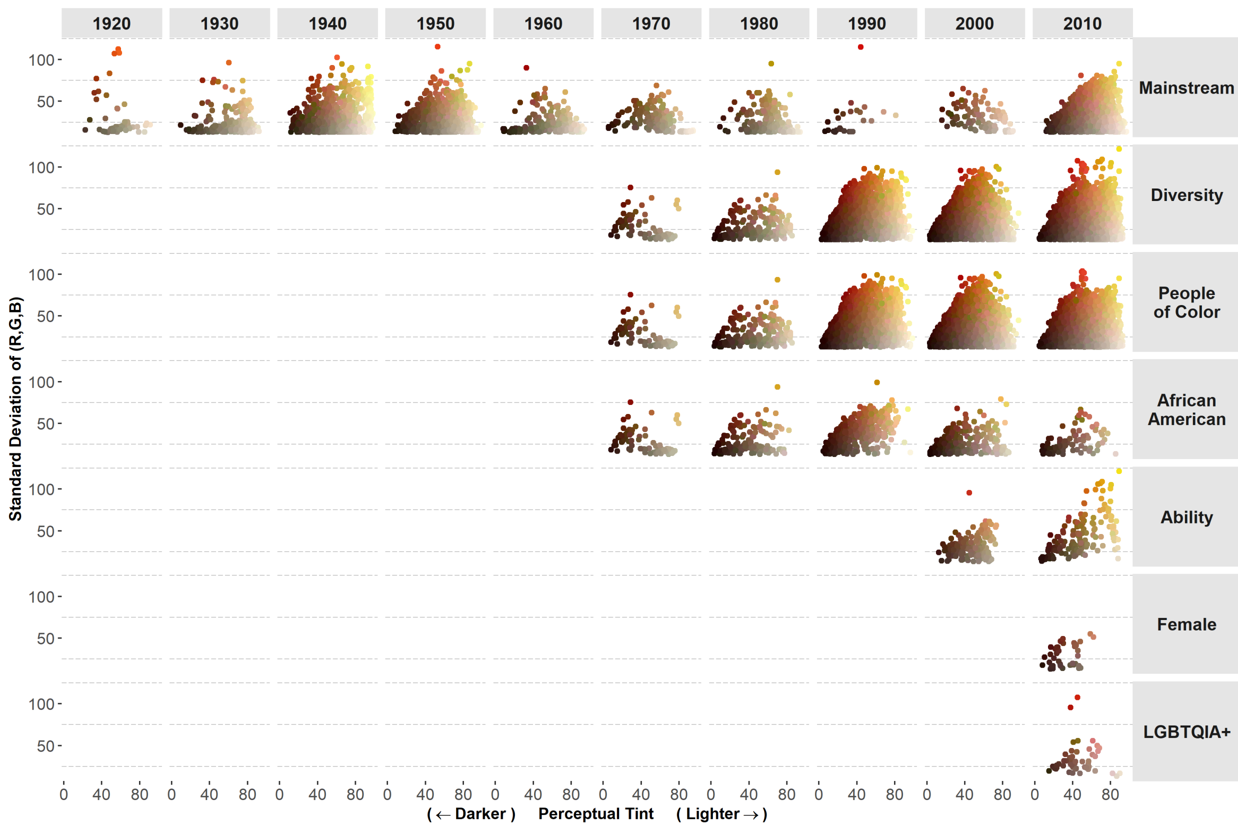
(a) Corpora in Sample, by Collection

(b) Collection Sample Size, by Decade



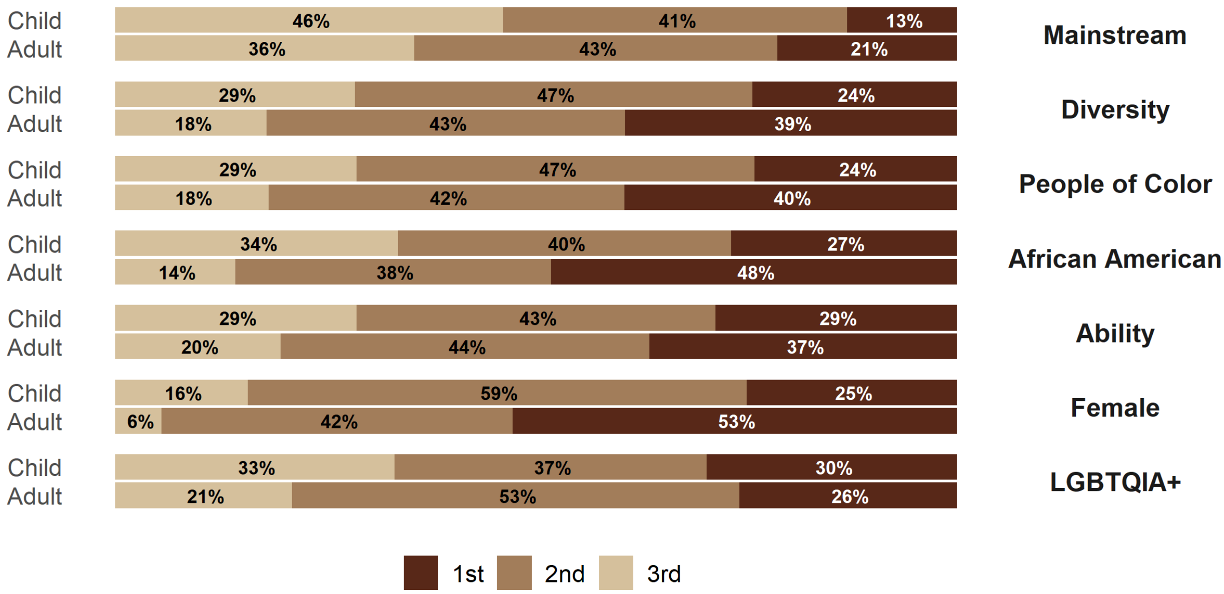
Note: This figure shows the main sources of data we use for our analysis. In Panel A, we list the book awards in our sample, along with the collections into which we group them in our analysis. In Panel B, we show our sample size in each collection, over time.

Figure B2. Skin Color Data Over Time, Human Skin Colors



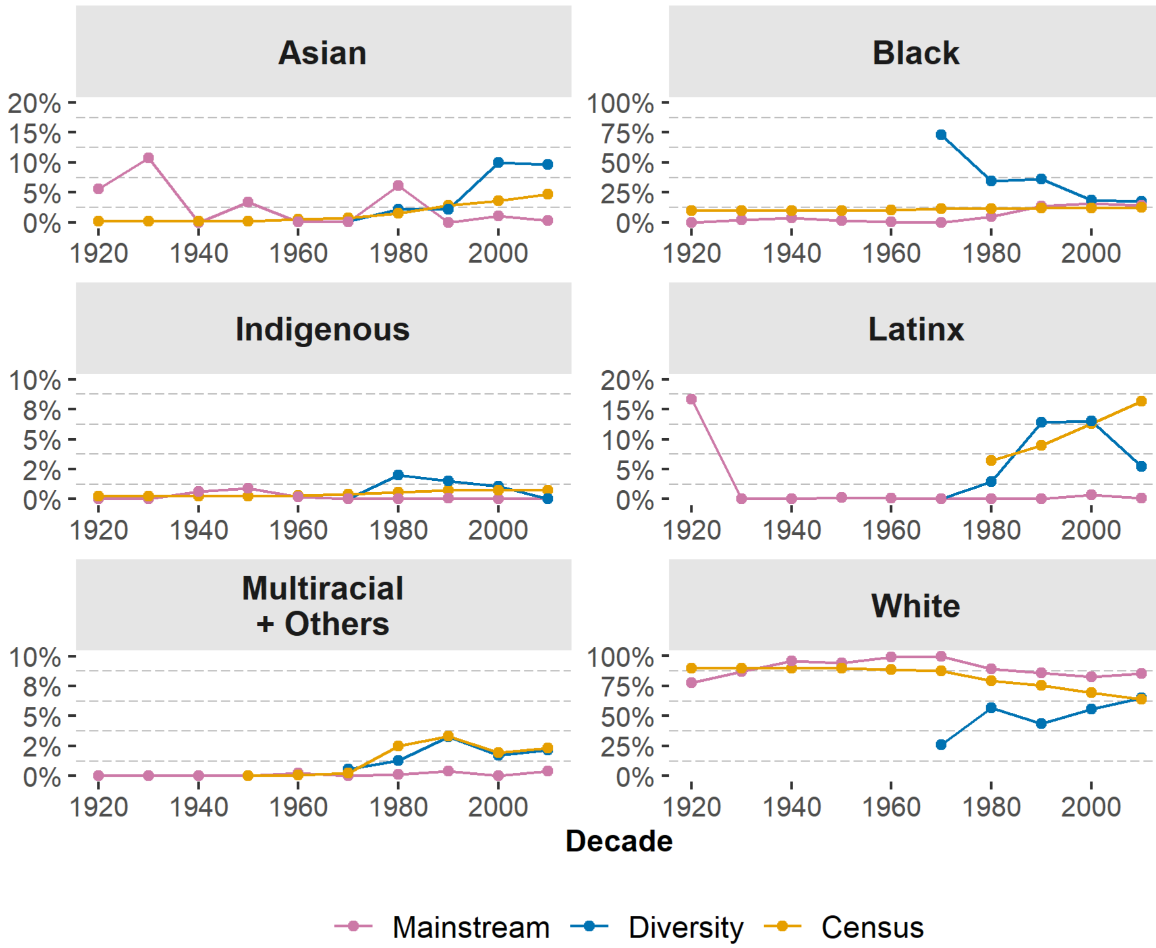
Note: In this figure, we show the representative skin colors for all detected faces with human skin colors (polychromatic skin colors where R ≥ G ≥ B) in each collection-by-decade. As described in Section IV.A, we use our face detection model (FDAI) trained on illustrations to classify faces in images. We determine a face's representative skin color using methods described in Section IV.A.2.

Figure B3. Skin Color Terciles by Age, by Collection



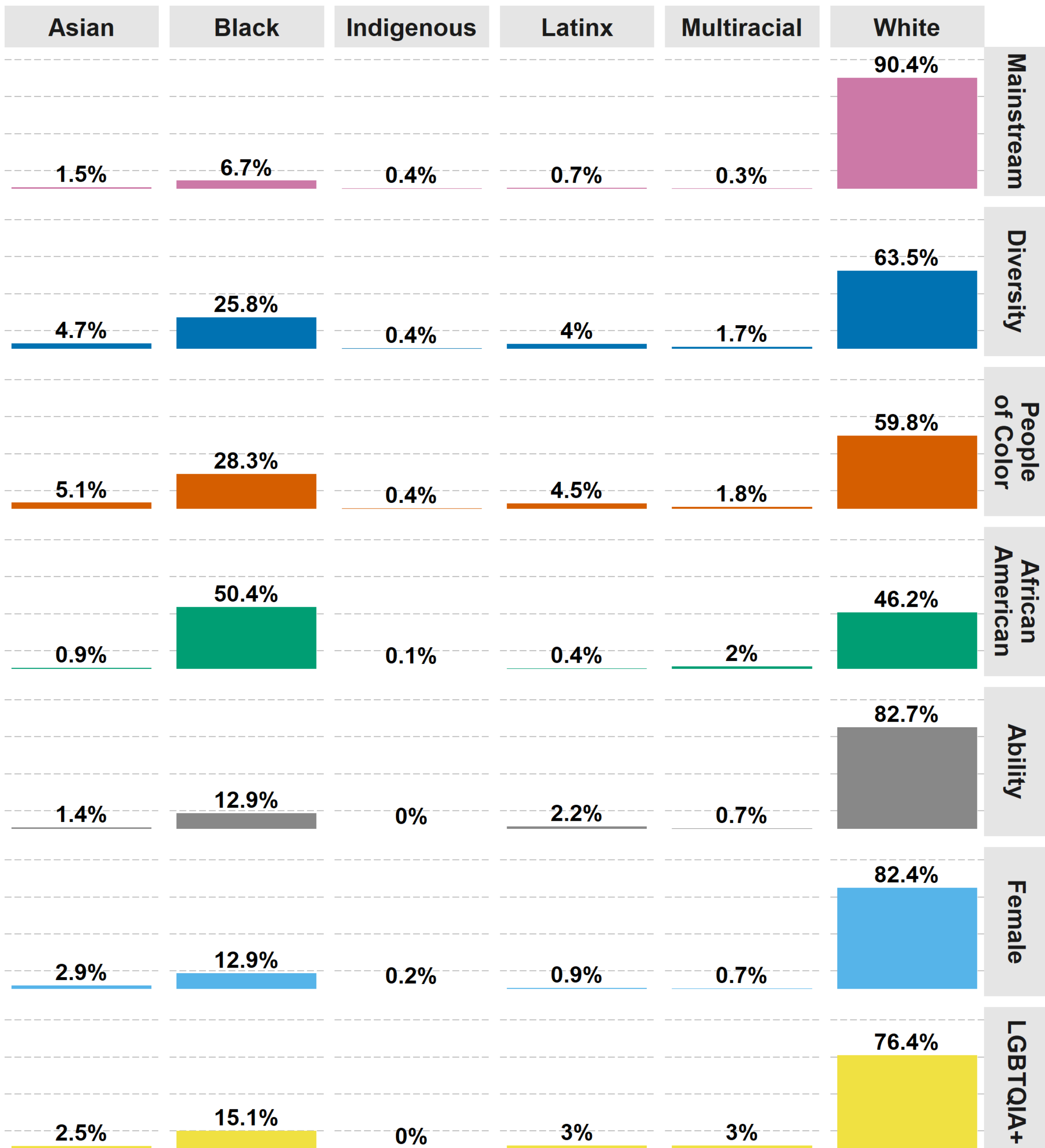
Note: In this figure, we show the proportion of faces in each tercile of the perceptual tint distribution by the classified age (adult vs. child) of the face. We detect faces using our face detection model (FDAI). Within these faces, we classify age using an AutoML algorithm we trained using the UTKFace public data set. Skin tint is determined by the L^* value of a face's representative skin color in L*a*b* space. These figures show the results for images that have human skin colors (defined as polychromatic skin colors where R $\geq$ G $\geq$ B).

Figure B4. Share of U.S. Population and Famous People in the Text, by Race/Ethnicity



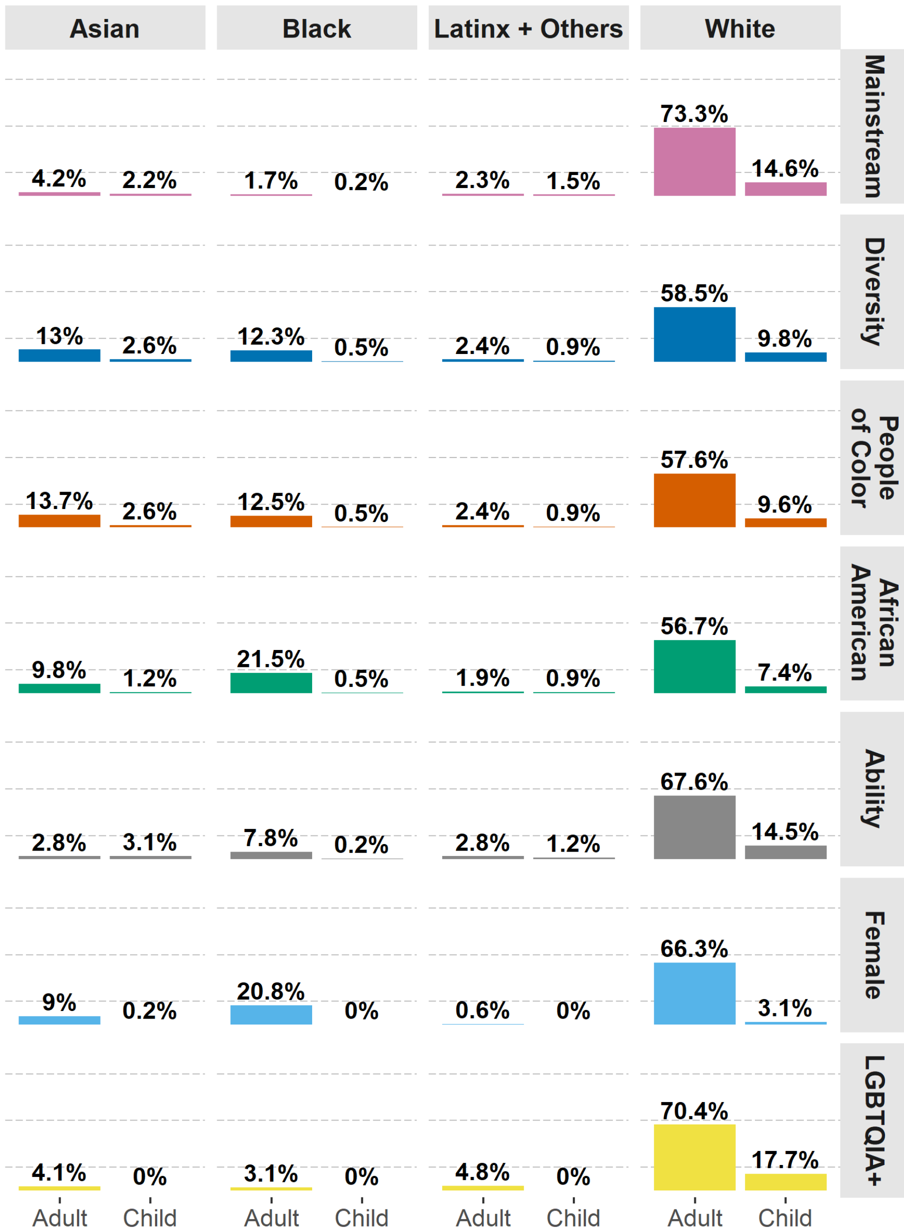
Note: In this figure, we show the percent breakdown of famous people mentioned in a given book by race/ethnicity. For example, if Aretha Franklin was mentioned 3 times in a book and Jimmy Carter is mentioned 2 times, then 60 percent of the mentions of famous people in that book would be Black. We then show the average percentage breakdown over all books by collection and decade for the Mainstream and Diversity collections. We also show the share of the U.S. population by race/ethnicity for each decade as a comparison. We classify famous people using methods described in Section IV.B. We collapse the following identities: East Asian, Middle Eastern, and South Asian into the Asian category; North American Indigenous peoples and South American Indigenous peoples into the Indigenous category; and African American and Black African into the Black category. If an individual was coded as having more than one race, we classify them as multiracial. Note that this is an analog to Figure 6, only with the y-axis collapsed to the maximum level for each race/ethnicity, respectively, to present easier-to-parse patterns for groups with lower levels of representation.

Figure B5. Race Classifications of Famous Figures in the Text



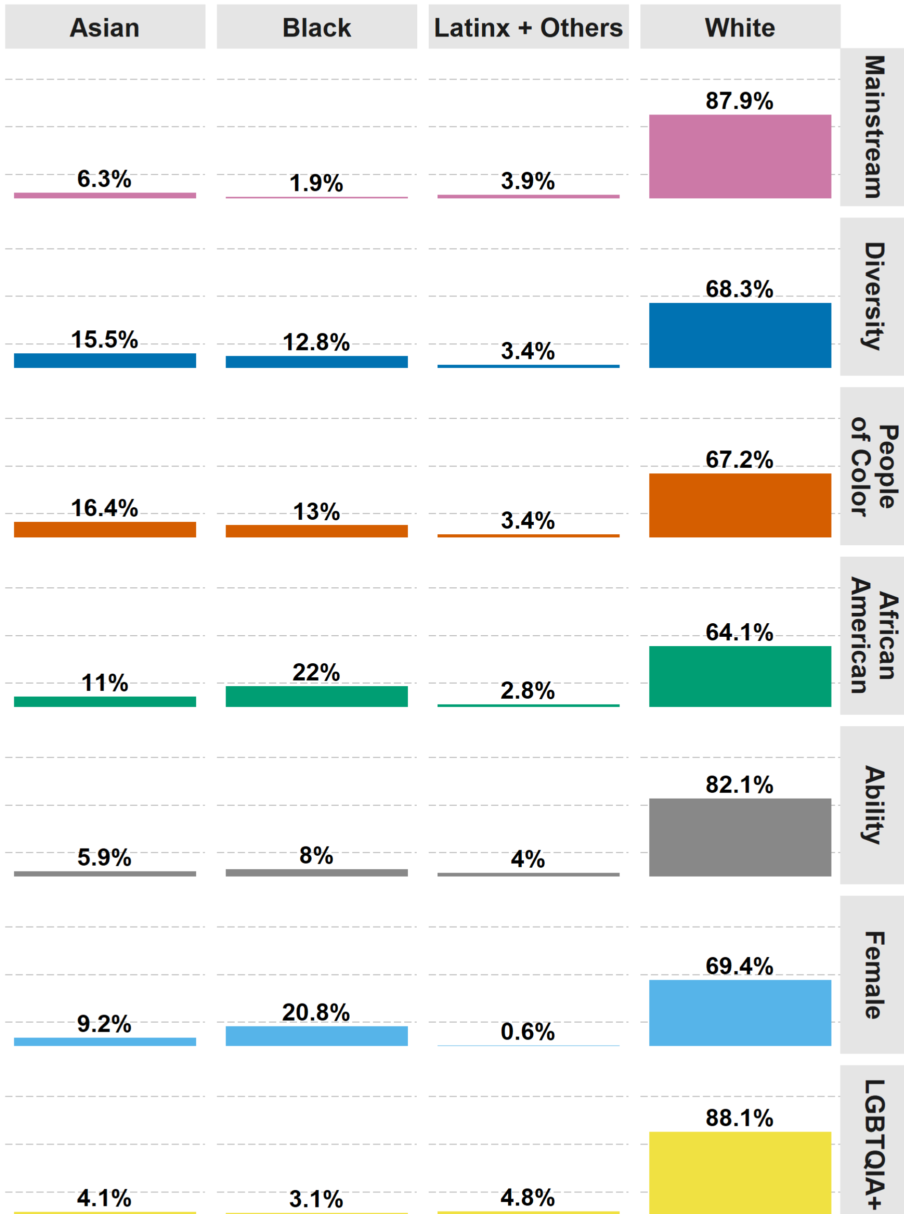
Note: In this figure, we count the number of famous people mentioned at least once in a given book and sum over all books in a collection. We then show the percentage breakdown of these famous people by race. For example, if Aretha Franklin was uniquely mentioned in 3 different books within a collection and Jimmy Carter is uniquely mentioned in 2 books within the same collection, then 60 percent of the unique famous people mentioned in that collection would be Black. We identify famous individuals using methods described in Section IV.B. We collapse the following identities: East Asian, Middle Eastern, and South Asian into the Asian category; North American Indigenous peoples and South American Indigenous peoples into the Indigenous category; and African American and Black African into the Black category. If an individual was coded as having more than one race, we classify them as multiracial.

Figure B6. Race and Age Predictions of Pictured Characters



Note: In this figure, we show the proportion of detected faces in all collections by race and age predictions. Race and age were classified by our trained AutoML model as described in Section IV.A.3. See Appendix Figure B7 for the same figure broken down by race alone.

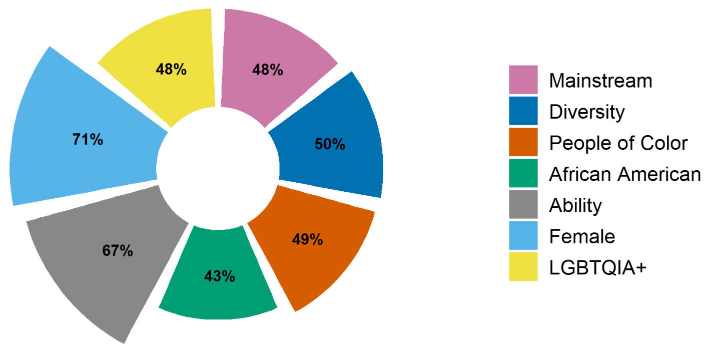
Figure B7. Most Pictured Characters Are Classified as White



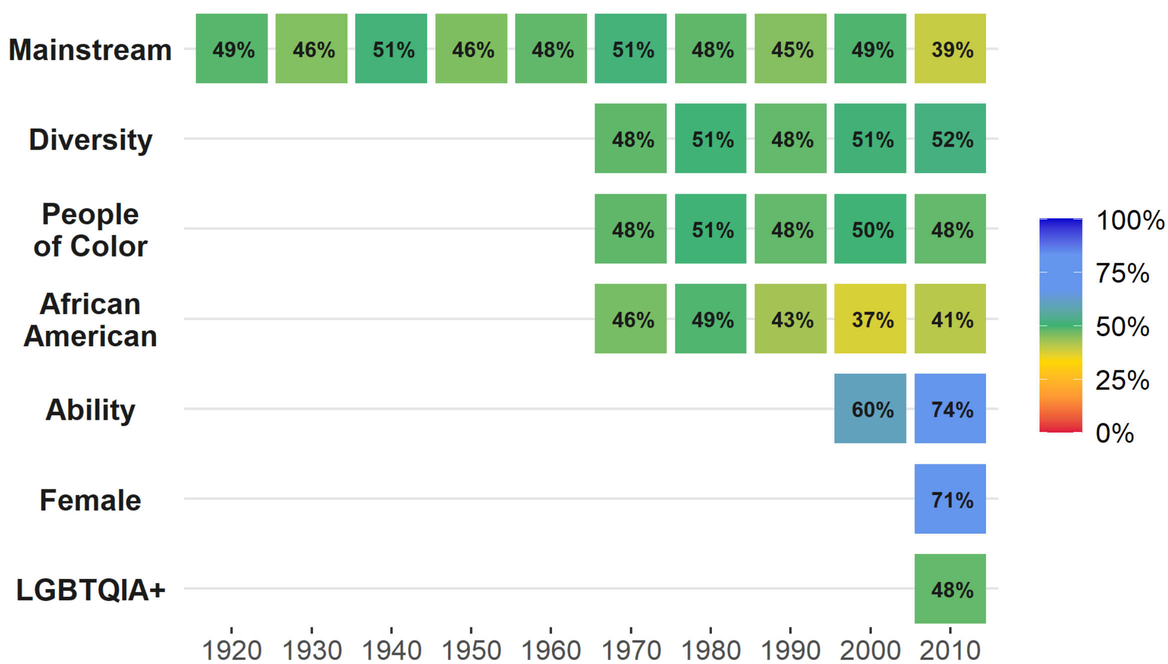
Note: In this figure, we show the proportion of faces in a book which our model labels as a given race averaged over all books in a collection. We detect faces using our face detection model (FDAI) described in Section IV.A.1. Within these faces, we classify race using an AutoML algorithm we trained using the UTKFace public data set.

Figure B8. Proportion of Detected Faces Which Are Female-Presenting

(a) Percent of Female-Presenting Faces Detected, Overall



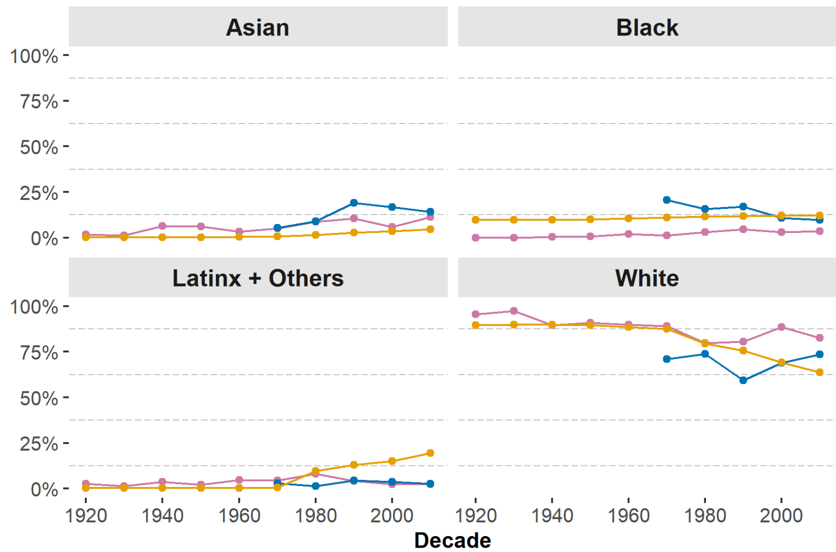
(b) Percent of Female-Presenting Faces Detected, Over Time



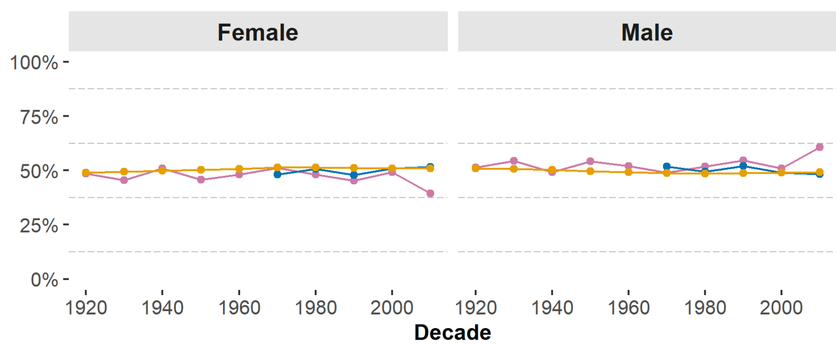
Note: In this figure, we find the proportion of faces in a book which our model labels as female. In Panel A, we show collection-level averages of the proportion of female faces in a given book by averaging over all books in a collection. In Panel B, we show these values over time by averaging the proportion of female faces in a given book by each collection and decade.

Figure B9. Share of U.S. Population and Pictured Characters, by Identity

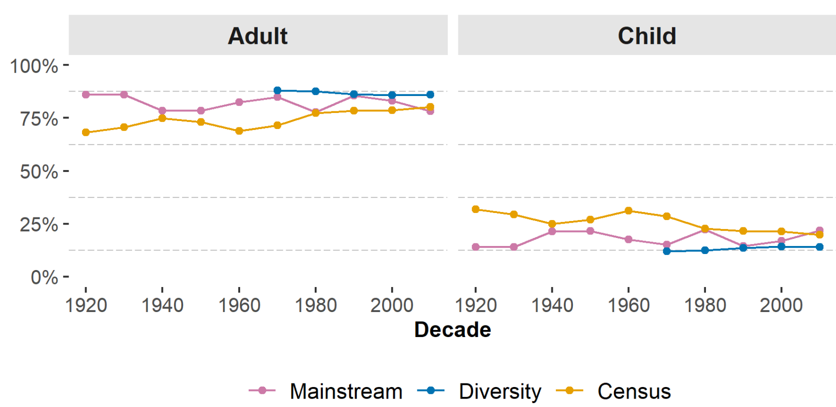
(a) Race/Ethnicity



(b) Gender



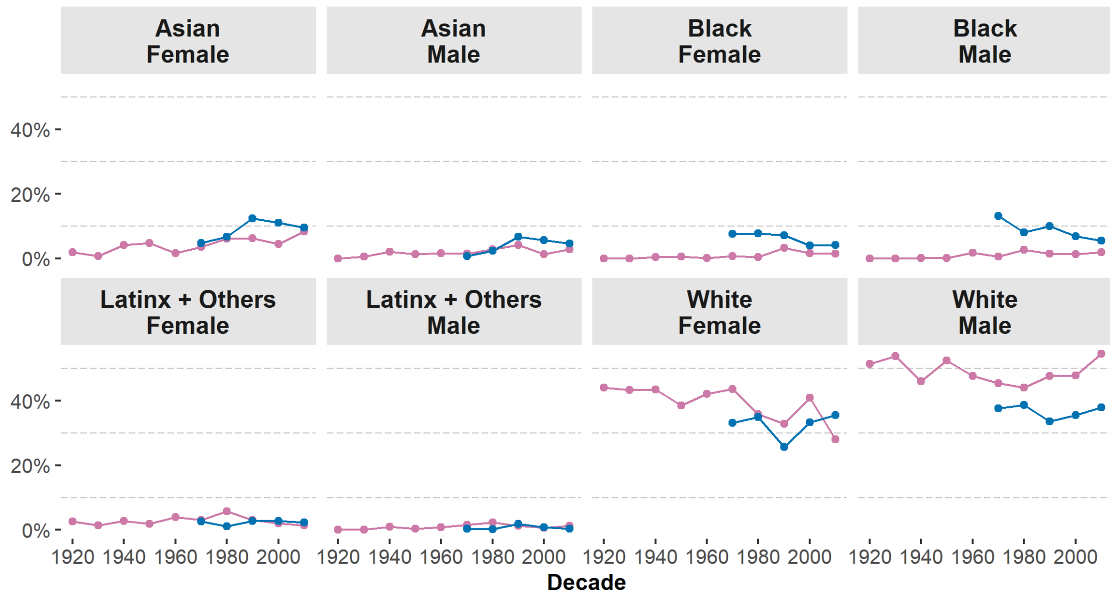
(c) Age



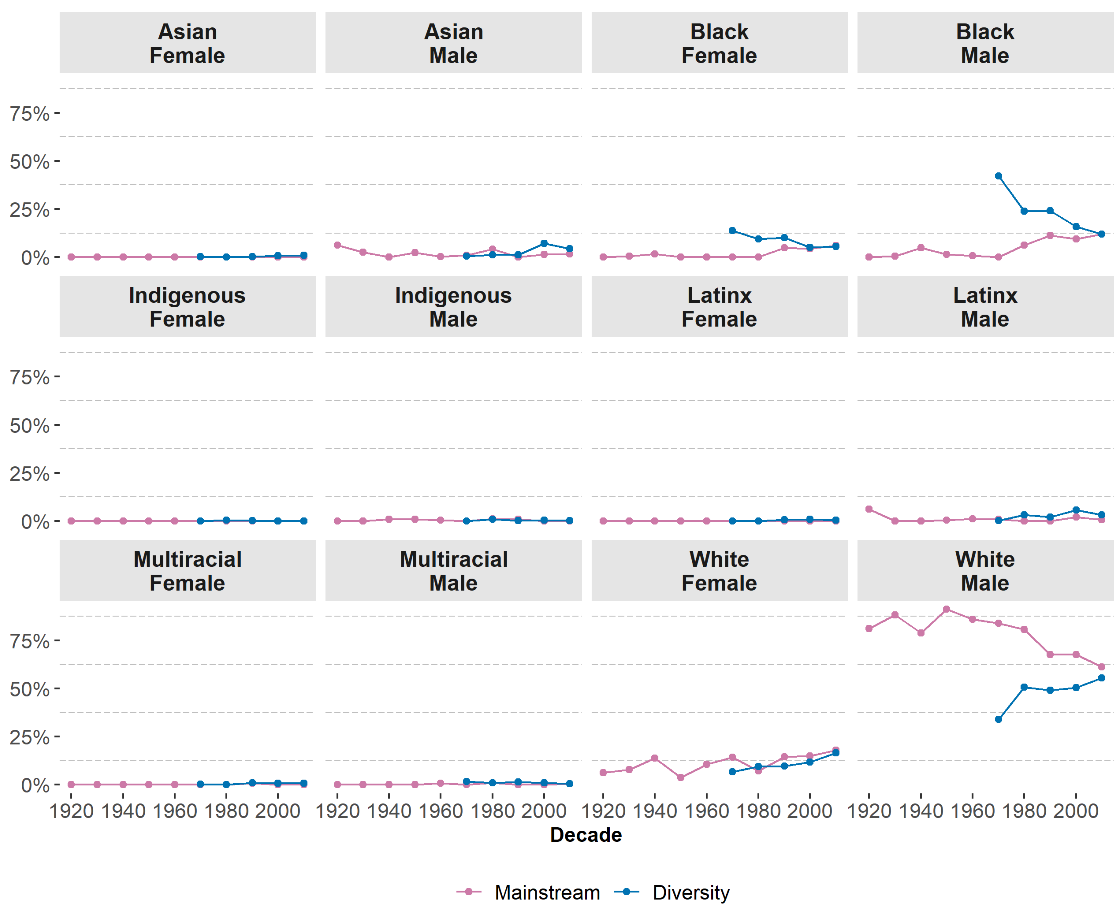
Note: We show the share of the U.S. population of specific identities mapped on the share of the pictured characters classified as a given identity in a given book averaged over all books in collection and decade. In Panel A, we show this by race/ethnicity. Each race/ethnicity category is constructed to be mutually exclusive. In Panel B, we show this by gender. In Panel C, we show this by age group.

Figure B10. Proportion of Characters in Images and Text, by Race and Gender

(a) Detected Faces (Images)

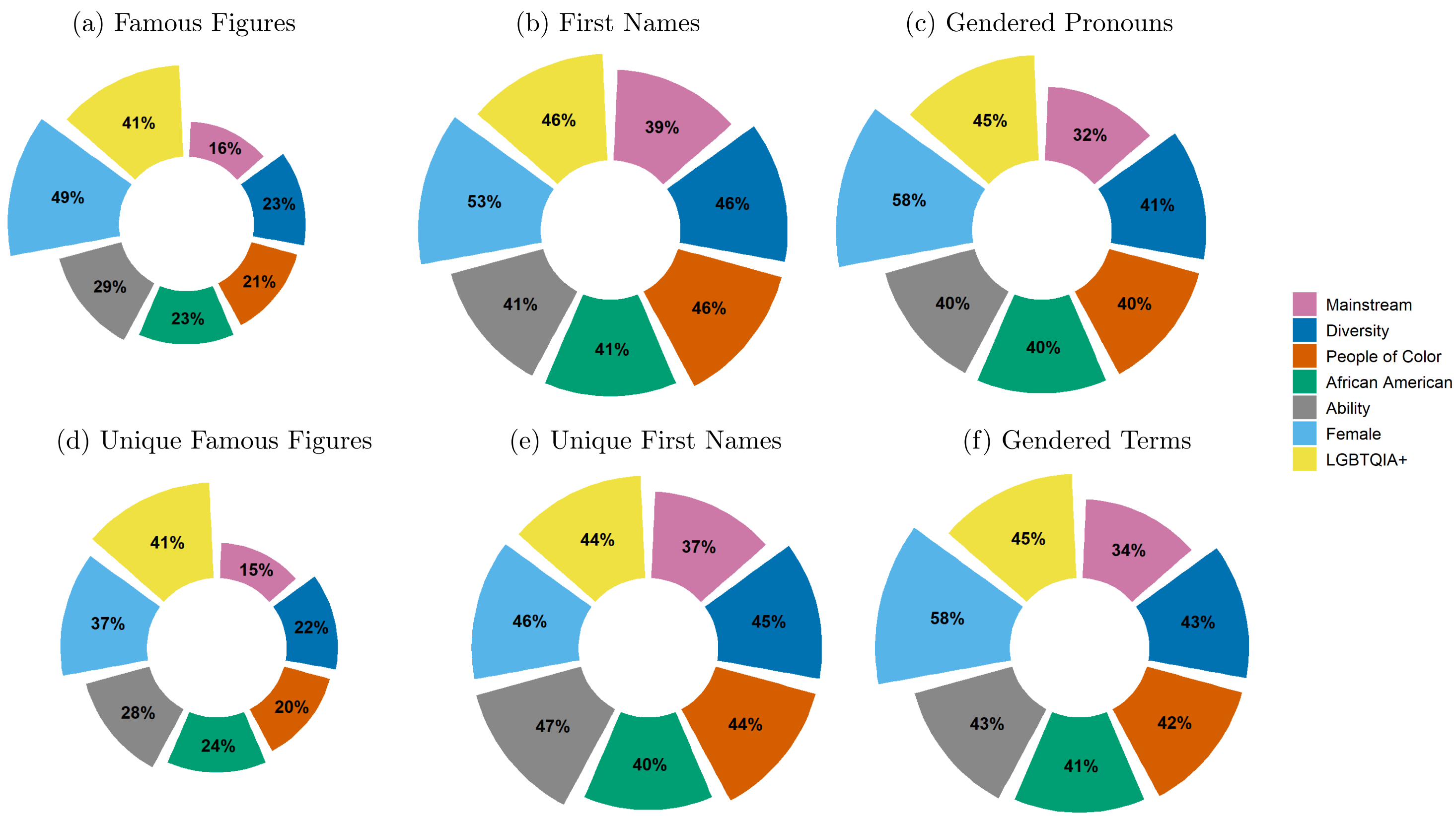


(b) Famous Figures (Text)



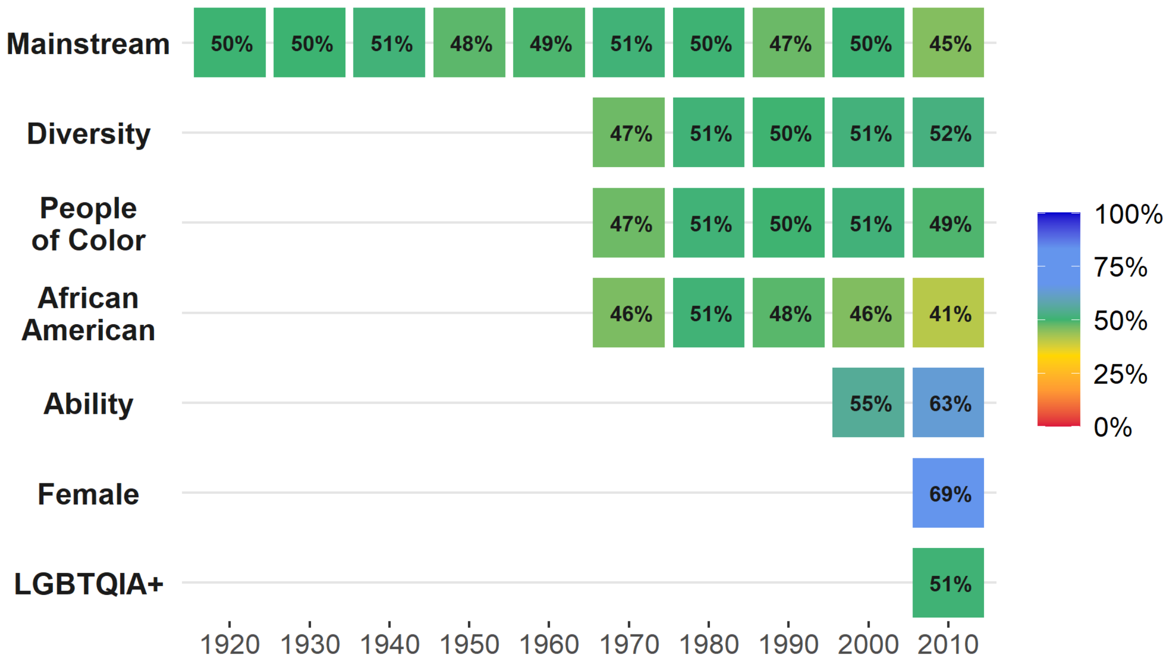
Note: In this figure, we show the share of the characters by race and gender in a given book averaged over all books in a collection and decade. In Panel A, we show this for detected faces in images. In Panel B, we show this for famous figures mentioned in the text.

Figure B11. Female Representation in Text, by Type of Word



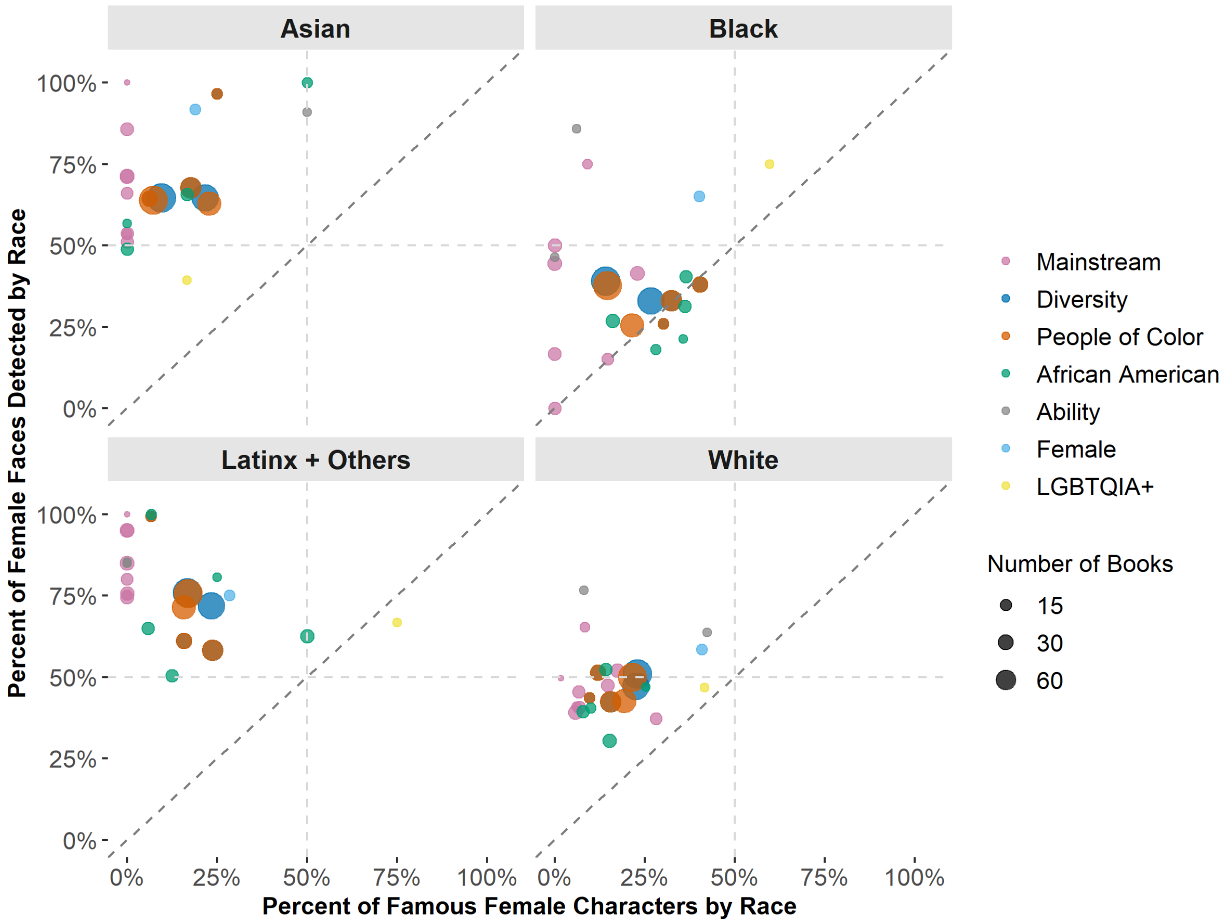
Note: We show the proportion of female representation in the text by collection and type of word. In Panel A, we find the percent breakdown of female famous people mentioned in a given book, averaged over all books in a collection. For example, if Aretha Franklin was mentioned 4 times in a book and Jimmy Carter is mentioned 2 times, then 60 percent of the famous people mentioned in that book would be female. In Panel B, we show the same thing as Panel A, but for mentions of character first names. Panel C shows the percentage of gendered pronouns which are female in a given book, averaged over all books in a collection. In Panel D, we show the percentage breakdown of unique female famous people in a collection. For example, if Aretha Franklin was uniquely mentioned in 3 different books within a collection and Jimmy Carter is uniquely mentioned in 2 books within the same collection, then 60 percent of the unique famous people mentioned in that collection would be female. In Panel E, we show the same thing as Panel D but for unique character first names. Panel F shows the percentage of female terms such as queen or nephew (full list provided in Data Appendix) in a given book, averaged over all books in a collection.

Figure B12. Average Probability a Face is Female, by Decade and Collection



Note: In this figure, we present the average probability that a face was classified as being female in a given collection by decade. We classify gender using an AutoML algorithm trained on the UTKFace public data set.

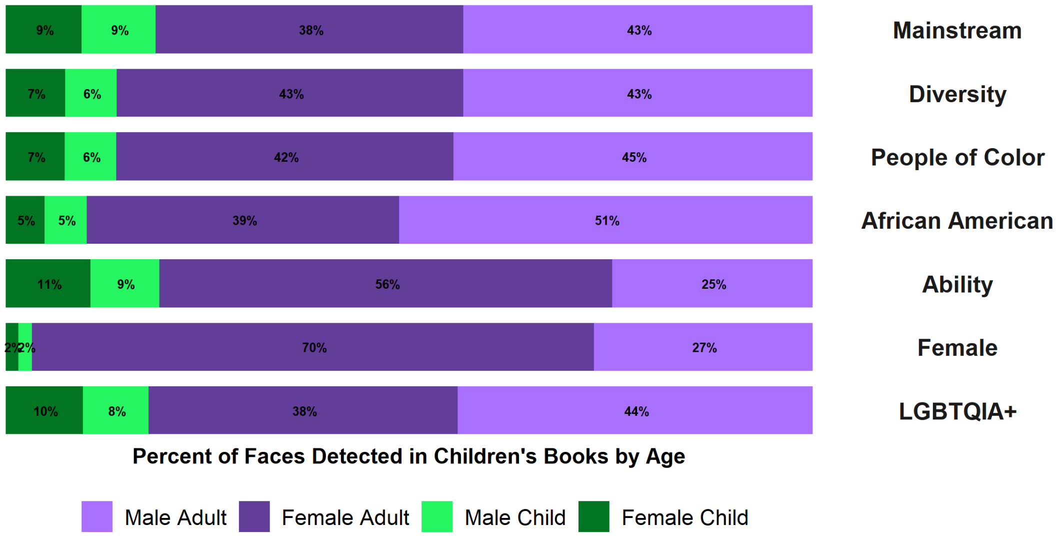
Figure B13. Race and Gender Representation in Images and Text



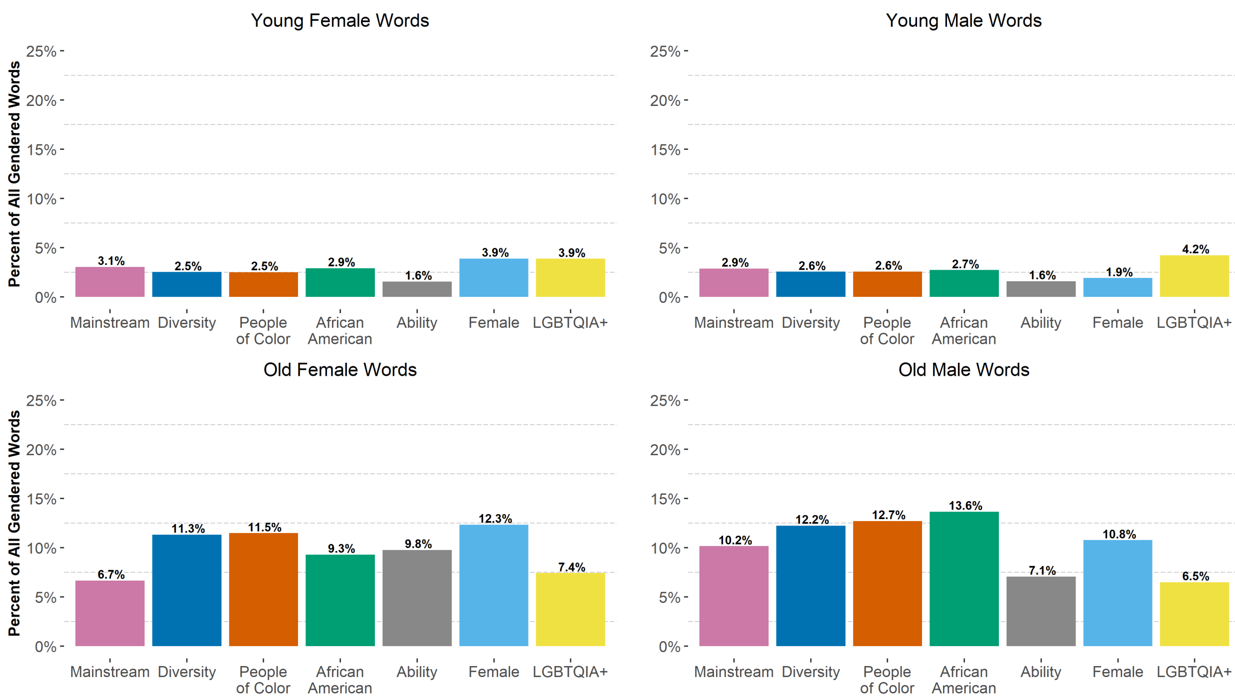
Note: In this figure, we plot female faces by race as a proportion of all faces with a given race classification on the y-axis and famous female characters by race as a proportion of all famous characters with a given race classification on the x-axis.

Figure B14. More Adults than Children in Images and Text

(a) Percent of Faces by Predicted Age Group and Gender

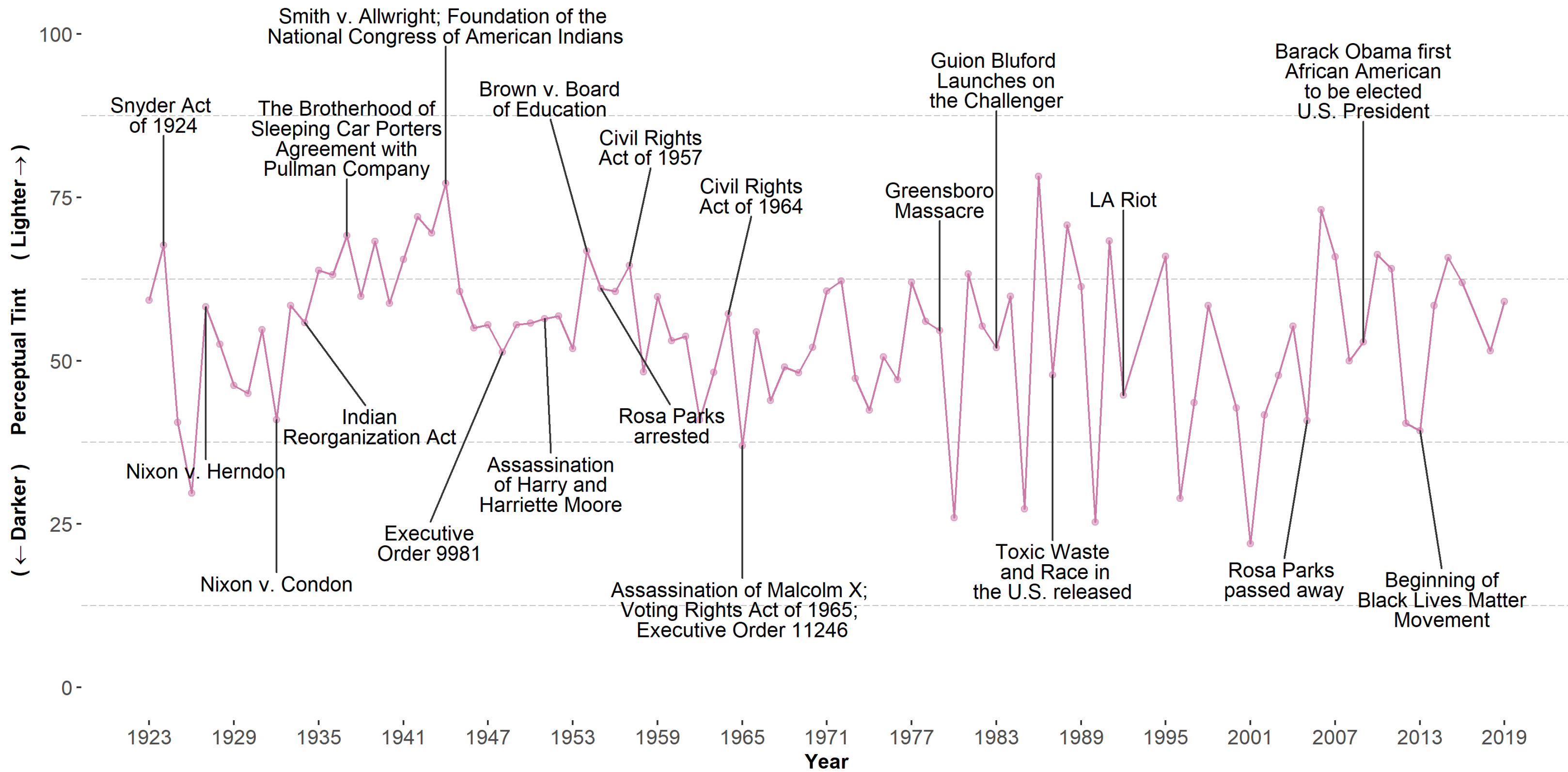


(b) Percent of Gendered Words by Age Group



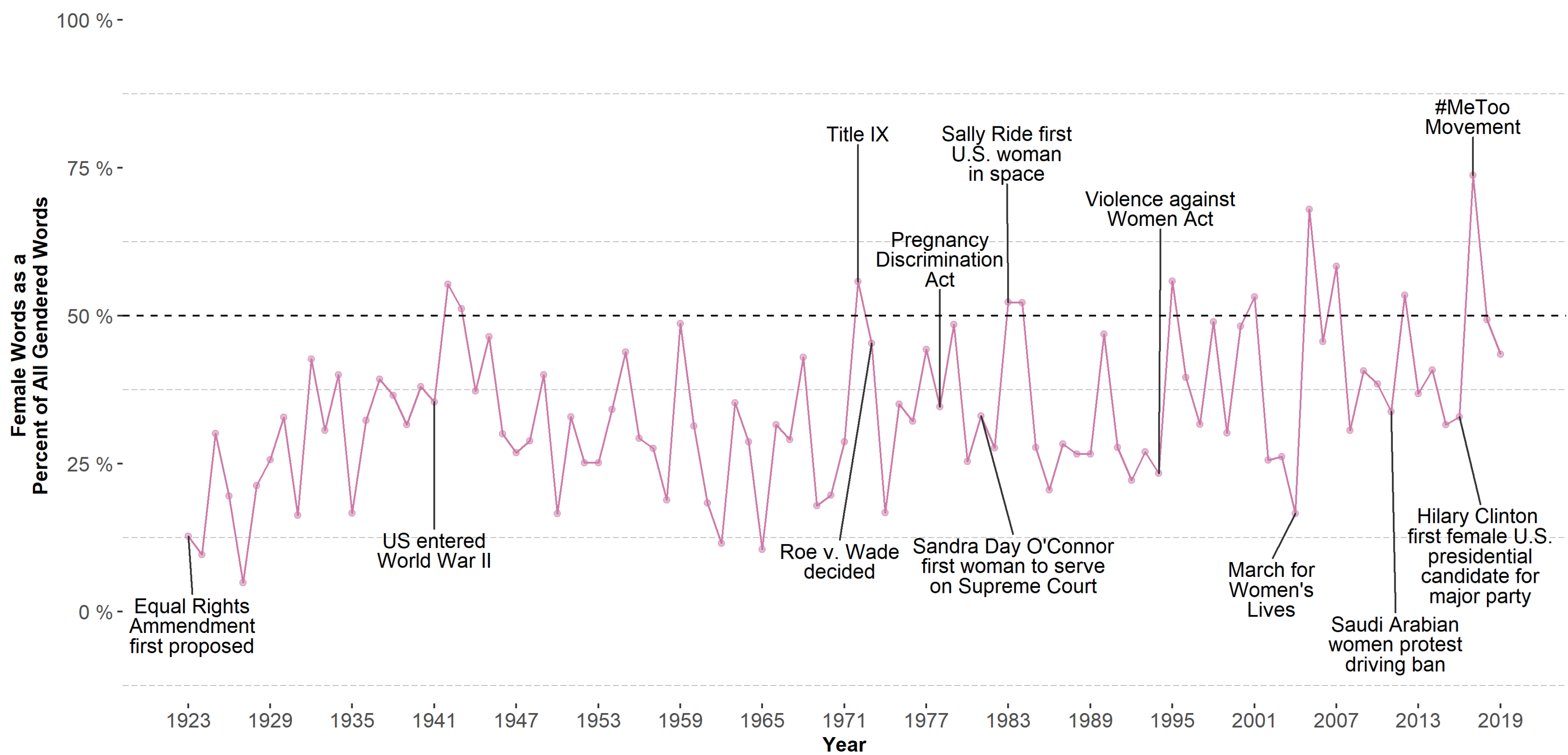
Note: In this figure, we show analysis of the representation of age and gender. In Panel A, we show analysis of predicted age and gender in the faces in images. Specifically, we plot the proportion of identified faces classified in each age (adult vs. child) and gender (female vs. male) category. In Panel B, we show analysis of age and gender in text. Specifically, we plot the proportion of words that refer to specific gender-age combinations (e.g., female adults or male children) as a percent of all gendered words in the book. Gendered words encompass the total number of gendered first names, gender predictions of famous characters, gendered pronouns, and a pre-specified list of other gendered terms (e.g., queen, nephew). We list the pre-specified gendered terms in the Data Appendix.

Figure B15. Mainstream Representation of Skin Color Throughout Historical Events



Note: In this figure we juxtapose measures of representation of skin color of pictured character faces from the Mainstream collection with the timing of salient historical events.

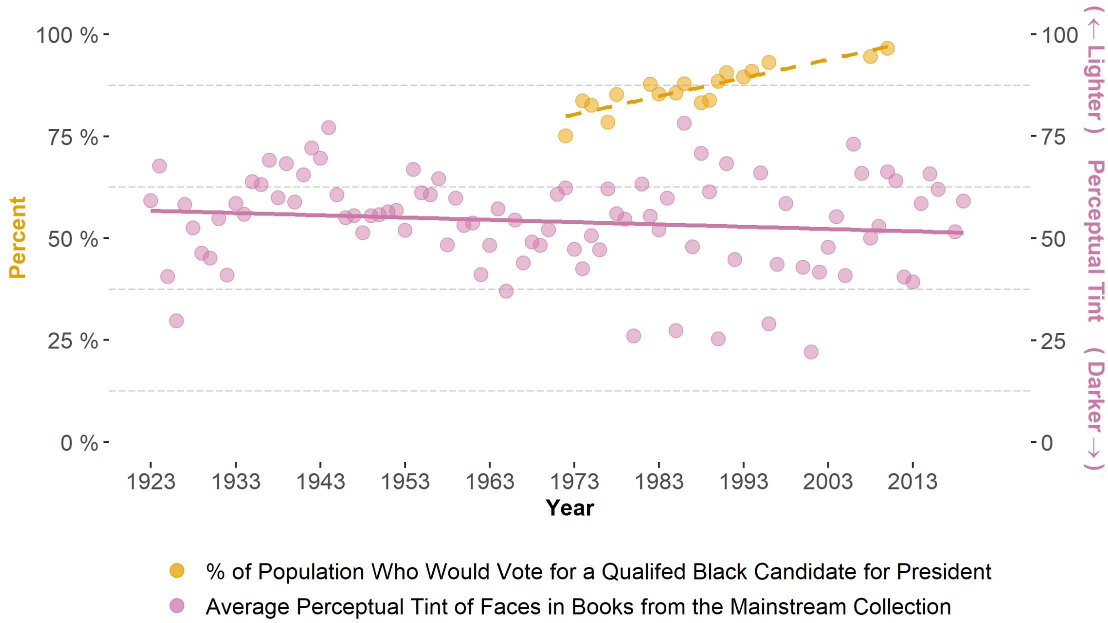
Figure B16. Mainstream Representation of Gender Throughout Historical Events



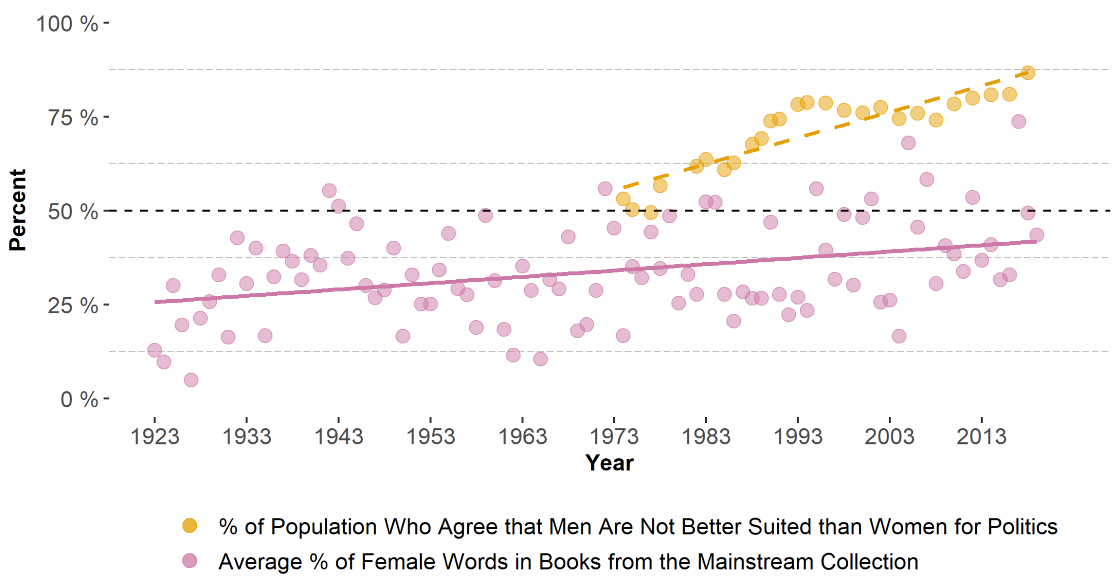
Note: In this figure we juxtapose textual measures of gender representation from the Mainstream collection with the timing of salient historical events.

Figure B17. Mainstream Representation and Social Attitudes Over Time

(a) Race



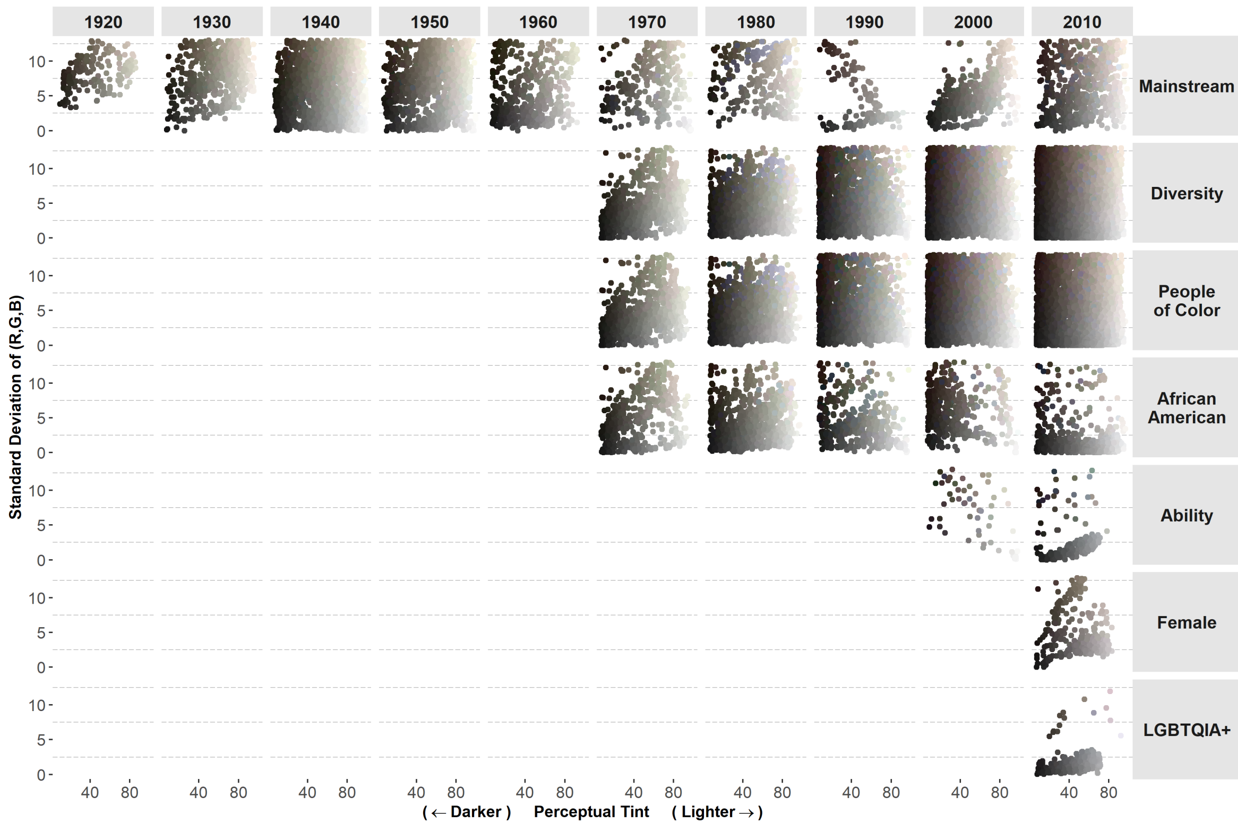
(b) Gender



Note: In this figure we compare trends in social attitudes with yearly representation in the Mainstream collection over time. In Panel A, we show the proportion of respondents who would vote for a qualified Black candidate for president along with the average skin tint of faces found in books within the Mainstream collection by year. In Panel B, we show the proportion of respondents who agree that men are not better suited than women for politics along with the average percent of female words in books within the Mainstream collection by year. Our data on social attitudes comes from the General Social Survey (GSS).

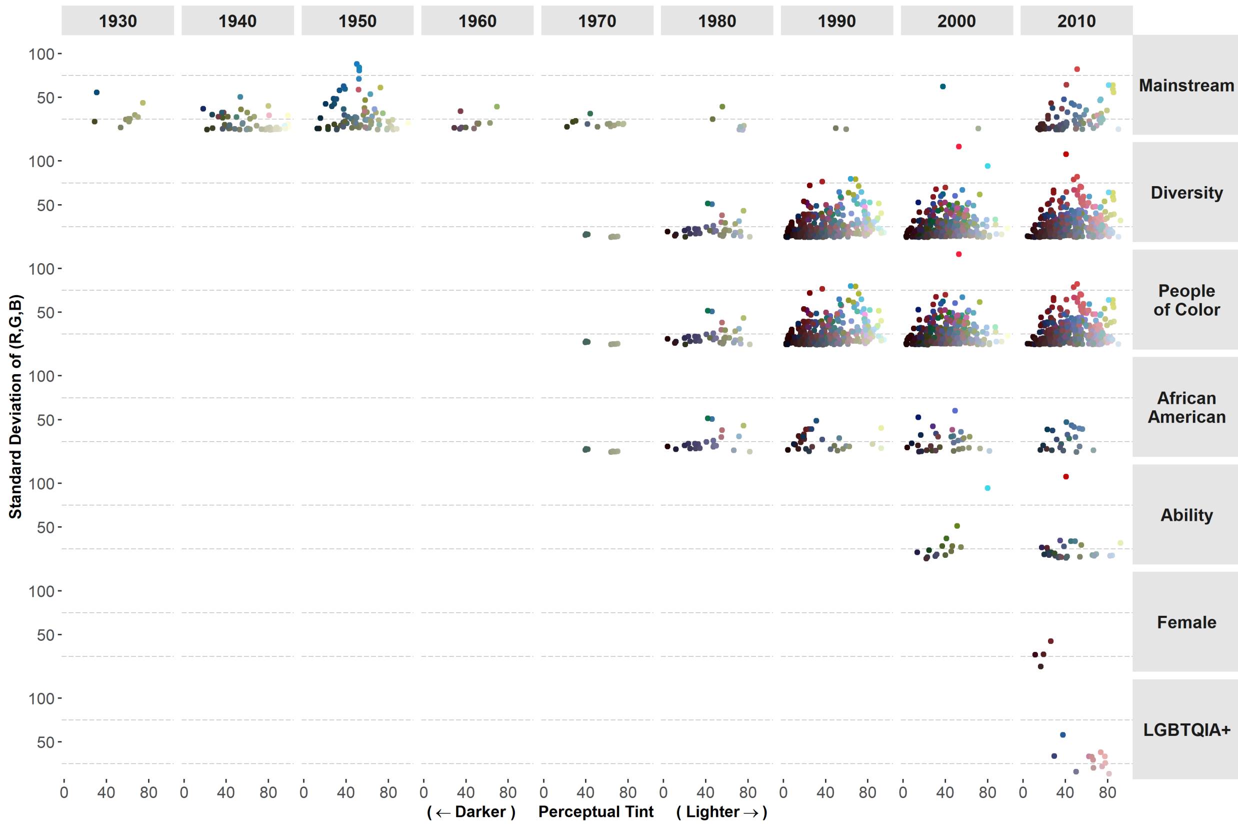
C Non-Typical Skin Color Appendix

Figure C1. Skin Color Data Over Time, Monochromatic Skin Colors



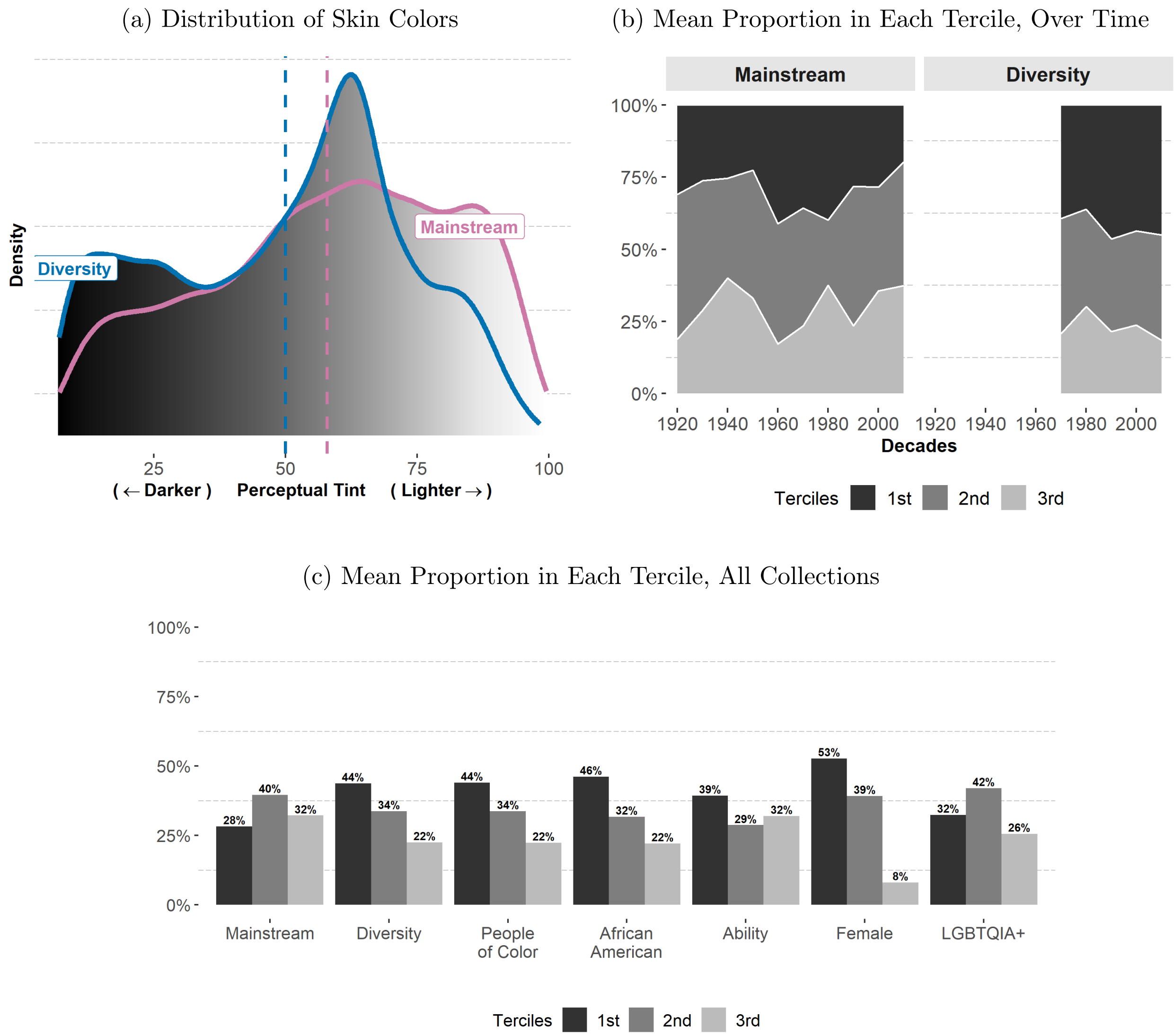
Note: In this figure, we show an analog to Figure B2, here focusing on the representative skin colors for all detected faces with monochromatic skin colors (e.g., black and white) in each collection-by-decade. As described in Section IV.A, we use our face detection model (FDAI) trained on illustrations to classify faces in images. We determine a face's representative skin color using methods described in Section IV.A.2.

Figure C2. Skin Color Data Over Time, Polychromatic Non-Typical Skin Colors



Note: In this figure, we show an analog to Figure B2, here focusing on the representative skin colors for all detected faces with non-typical skin colors (e.g., blue or green) in each collection-by-decade. As described in Section IV.A, we use our face detection model (FDAI) trained on illustrations to classify faces in images. We determine a face's representative skin color using methods described in Section IV.A.2. The data shown in this figure begin in the 1930s, as opposed to in the 1920s as in Figures B2 and C1 found in the Non-Typical Skin Color Appendix, because we detect no faces with polychromatic non-typical skin colors in books from the 1920s.

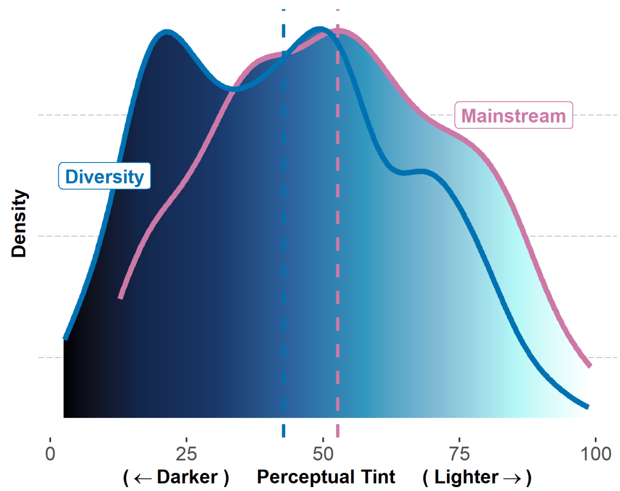
Figure C3. Skin Colors in Faces, by Collection: Monochromatic Skin Colors



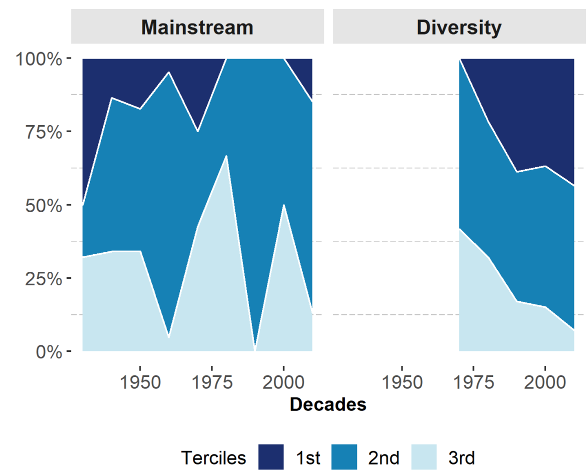
Note: This figure shows our analysis of the representative skin colors of the individual faces we detect in the images found in the books we analyze. This is an analog to Figure 4, only here we focus on monochromatic faces. Panel A shows the distribution of skin color tint for faces detected in books from the Mainstream and Diversity collections. The mean for each distribution is denoted with a dashed line. In Panel B, we show the average proportion of faces in each tercile, over time, for faces in the Mainstream and Diversity collections. Panel C shows the overall collection-specific average proportion of faces in each skin color tercile for each of the seven collections. Skin classification methods are described in Section IV.A.

Figure C4. Skin Colors in Faces, by Collection: Polychromatic Non-Typical Skin Colors

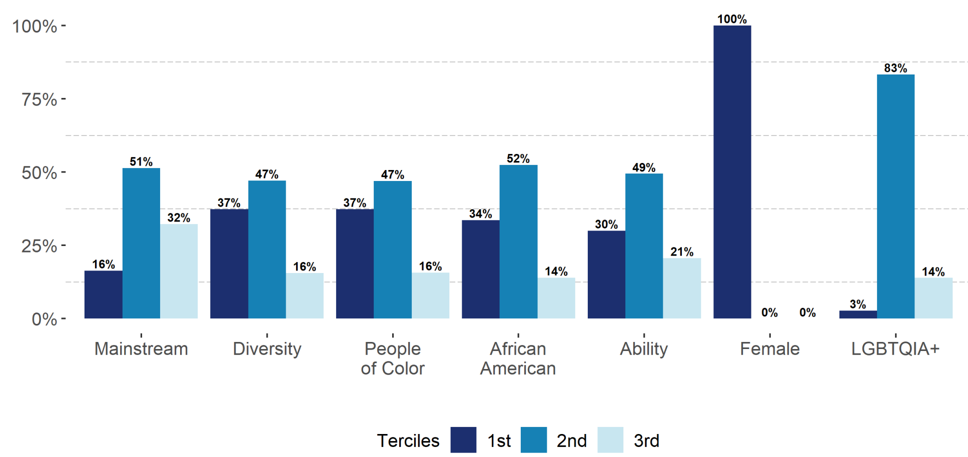
(a) Distribution of Skin Colors



(b) Mean Proportion in Each Tercile, Over Time



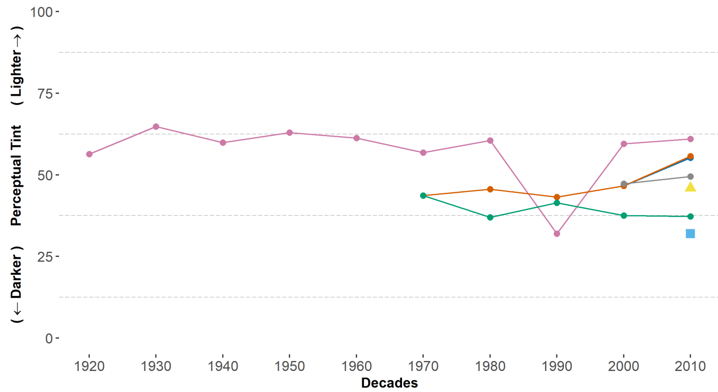
(c) Mean Proportion in Each Tercile, All Collections



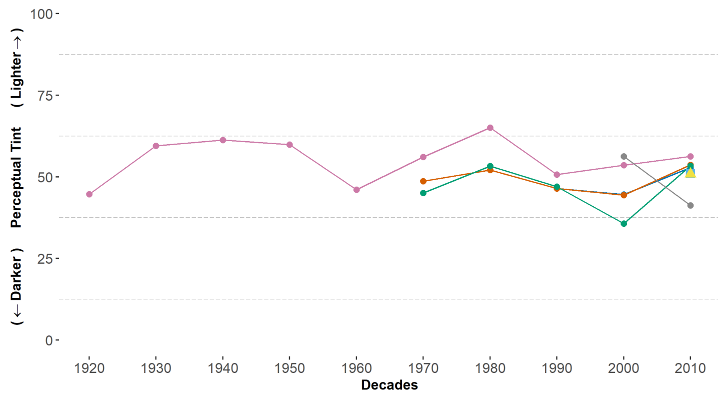
Note: This figure shows our analysis of the representative skin colors of the faces detected in the books we analyze. This is an analog to Figure 4, only here we focus on faces that have non-typical skin colors. Panel A shows the distribution of skin color tint for faces detected in books from the Mainstream and Diversity collections. The mean for each distribution is denoted with a dashed line. In Panels B and C, we show the average proportion of faces in each tercile of the perceptual tint distribution across all books in a collection. In Panel B, we show the average proportion of faces in each tercile, over time, for faces in the Mainstream and Diversity collections. Panel C shows the overall collection-specific average proportion of faces in each skin color tercile for each of the seven collections. Skin classification methods are described in Section IV.A.

Figure C5. Skin Colors over Time, by Collection

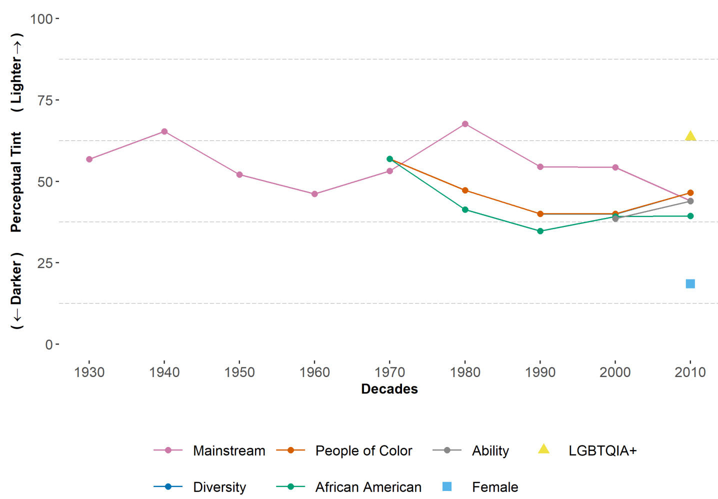
(a) Human Skin Colors



(b) Monochromatic Skin Colors



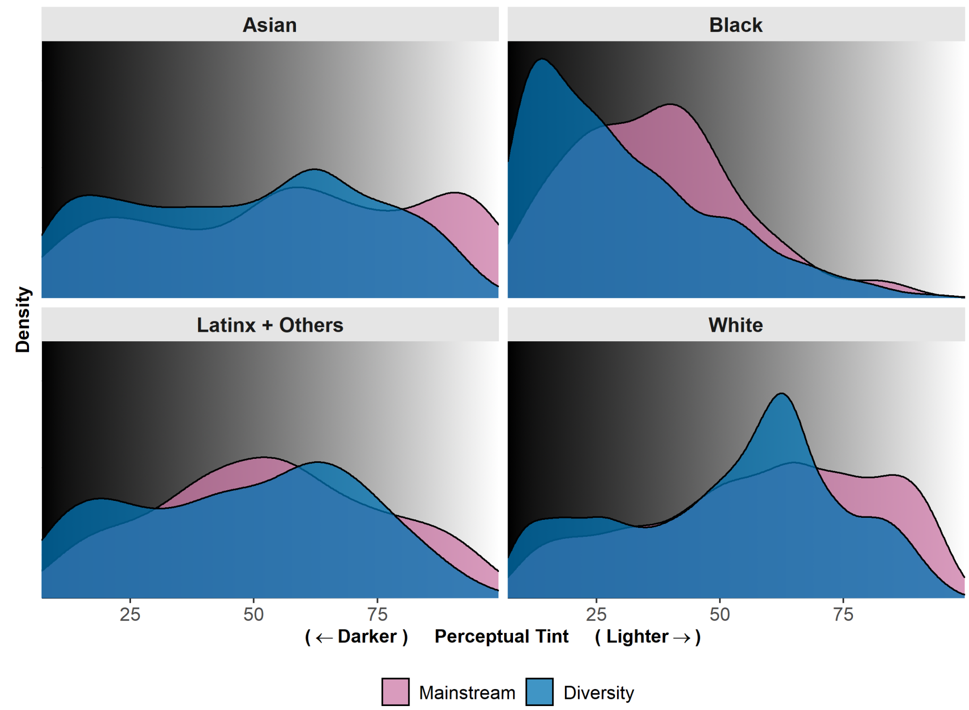
(c) Polychromatic Non-Typical Skin Colors



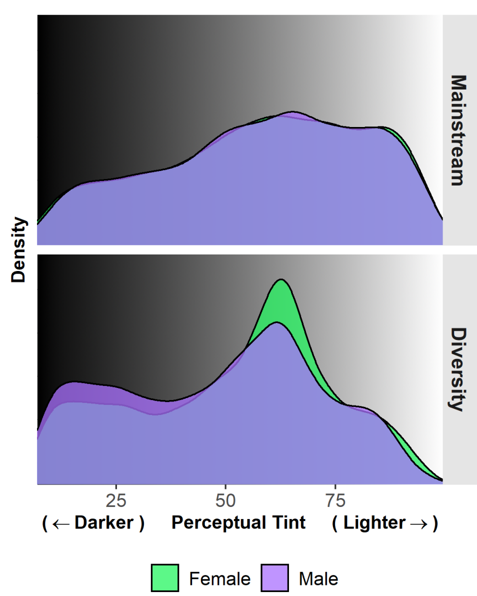
Note: This figure shows the average skin tint over time in our sample of award-winning children's books. We first take the average skin tint for all faces in a given book, then we average across all books in a given year. We separate the faces by skin color type, Panel A shows the average skin tint for all faces with human skin colors, Panels B and C show the same thing as Panel A but for Monochromatic and Polychromatic Non-typical Skin Colors, respectively.

Figure C6. Skin Color by Predicted Race of Pictured Characters: Monochromatic Faces

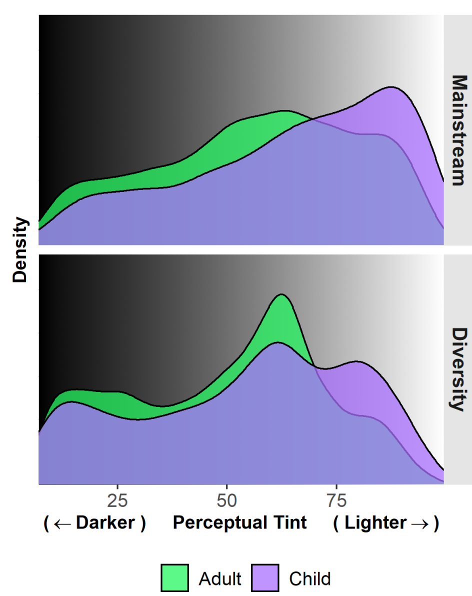
(a) Skin Color Distribution by Race



(b) Skin Color Distribution by Gender



(c) Skin Color Distribution by Age



Note: This figure shows the distribution of skin color tint by predicted race, gender, and age of the detected faces in the Mainstream and Diversity collections. This is an analog to Figure 5, only here focusing on faces depicted in a monochromatic color scheme (e.g., black and white).

D Methods Appendix

D.A Comparing AI with Manual Content Analysis

In this section, we describe some of the biases inherent in manual and AI-run content analysis. We then describe how we used manual content analysis to validate our measures. Finally, we conduct a cost-effectiveness analysis which highlights a key advantage of our approach – far greater reach in terms of the ability to measure representation in an entire book, to respond nimbly to changes in analysis plans, and significantly lower cost.

D.A.1 AI is Only Human

Historically, content analysis to measure representation has been done "by hand" using human coders (Bell, 2001; Neuendorf, 2016; Krippendorff, 2018). Such analysis provides deep understanding but can generally only be done on a small set of content and necessarily reflects human behavior and biases. While artificial intelligence tools also reflect bias in their training data and algorithms, they can be standardized, are more replicable, and can be applied to a much larger sample than manual content analysis permits. An additional advantage of this approach is that, by following a set of clearly-defined steps for implementation which rely primarily on computers, it minimizes variation in results stemming from researcher-specific biases or idiosyncrasies.

Our paper brings a set of artificial intelligence tools to bear on the field of content analysis. These tools are powerful, computer-driven methods. They are designed by humans and, in many cases, trained with initial human input. We use them because they offer a few key advantages. The first is scale: because algorithms are automated, they allow for analysis of a much larger set of content than would be possible using conventional, "by hand" methods. The second is adaptability: we can rapidly change one dimension of measurement and re-run the analysis at low cost. Were we to do this via hand-coding, the cost would increase linearly with each addition or adjustment (see Section D.A); with AI-based analysis, the marginal cost of such additions or adjustments is much lower.

Measuring representation in content via any means will generate some errors in measurement. In traditional content analysis, analysts may misclassify some images or text. If this occurs at random, this can be treated as standard measurement error, which would be captured via estimating inter-rater reliability (Neuendorf, 2016; Krippendorff, 2018). If, however, traits of the analyst systematically influence their coding, then error from misclassification may be non-classical, leading to a bias in expectation (Krippendorff, 1980). This can arise, for example, if an analyst's identity (e.g., one's race and/or gender) causes them to classify content differently than analysts of different identities (Boer, Hanke and He, 2018).

These same biases appear in AI models. Many AI models, including those we use, are trained using a set of data which are first labeled by humans. Furthermore, nearly all models are either fine-tuned, evaluated, or both, based on their performance relative to human classification. As a result, the bias in classical content analysis is "baked into the pie" for computer-driven content analysis (Das, Dantcheva and Bremond, 2018).

Most face detection models are trained using photographs of humans – particularly White humans – which could lead us to undercount people of color and illustrated characters if the model were less able to identify characters on which it was not trained. To address this, we trained our own face detection model using 5,403 illustrated faces from the Caldecott and Newbery corpora (discussed in Section IV.A.1). A similar problem with under-detection of certain types of faces could also appear in the skin segmentation process, as we relied upon a series of convolutional neural networks to identify skin, rather than on human-performed identification of the skin region of faces.

These issues persist when classifying features. In the case of gender, for example, all public data sets with labels for gender that we encountered have a binary structure, limiting classification to "female" or "male," and neglecting to account for gender fluidity or nonbinary identities. Furthermore, intrinsic to these models is the general assumption that we can predict someone's gender identity using an image of their faces (Leslie, 2020). Similar problems beset the task of classifying putative race (Fu, He and Hou, 2014; Nagpal et al., 2019; Krishnan, Almadan and Rattani, 2020). Resolving these problems is an active field of inquiry, and recent scholarship has suggested several promising paths forward for doing so (Buolamwini and Gebru, 2018; Mitchell et al., 2019).

While AI is a product of and therefore reflects human biases, human biases are also intrinsic to traditional "by-hand" content analysis. Manual coding necessarily can only reflect the biases of the individual coders. We observed that the identities of the manual labelers on our team led to non-classical error, particularly in the classification of race of the pictured characters in images. We therefore use multiple measures for each identity to try to understand the extent of this potential measurement error. For example, in addition to the manually coded putative race of famous figures, we examine also examine skin color of detected characters.

While we primarily use AI tools to study representation, we end this section by emphasizing that AI and manual coding provide complementary understanding of content. The tools we use are meant to rapidly estimate how a human might categorize these phenomena. They are motivated by human perception and, ultimately, their performance is also evaluated based on how accurately they can determine how a human might perceive the representations in images and text. Our use of these tools depends on human input at each stage, from the conception of tools and the labelling of training data, to the evaluation of the tools' accuracy and the way that we interpret their results. We see our efforts adding the strengths of recent advances in computational science to content analysis as a natural extension of the rich history of human-driven analysis in this field.

D.A.2 Validation

Drawing from validation theory, we conducted traditional manual content analysis to validate our measures (Kane, 2013; Neuendorf, 2016). To do so, we hand-coded representations in 30 short stories and poems for children written and illustrated by a variety of authors and illustrators from a third grade reading textbook published in 1987. This helped us to evaluate the plausibility of our measures and also identify messages our tools failed to detect, clarifying limitations of computer-led content analysis. Regardless of whether we use manual coding or computer vision, the broad patterns are the same. Over 50 percent of the characters/detected faces and gendered words are male and the skin colors depicted are skewed away from darker-skinned individuals.

D.A.3 Cost-Effectiveness

It took approximately 40 hours to code the entire book (400 pages at an average of 6 minutes per page).[64] While the length of time needed to code "by hand" varies with the grade level of the books in our sample, we estimate that it would have taken us over 16,000 hours to hand-code the 162,872 pages in our sample of children's books. At an hourly wage of between $15 and $20, we estimate this work would have cost between $244,000 to $326,000.

D.B Images as Data

D.B.1 Image Feature Classification: Face Detection Methods

To train our face detection model, we split our manually labeled data set into training (80 percent of the data), validation (10 percent of the data, used for hyper-parameter tuning), and testing (10 percent of the data, used for evaluating the model).[65]

[64]Hand-coding of pages entails documenting a wide variety of features in image and, separately, text, which is a time- and detail-intensive process. Our estimate of six minutes per page represents a lower bound on the time needed to perform the type of analysis we conducted. In this case, for example, the manual coders did not count every token that could be related to gender, nationality, and color.

[65]The validation data are used for hyper-parameter tuning to optimize the model architecture. Hyper-parameter tuning involves "searching" for the optimal values of the hyper-parameters. Examples of hyper-parameters include learning rate, number of epochs (number of times the model goes through the whole data set), and different activation functions of the model that can be tuned to improve the accuracy of the model. FDAI is using Google Cloud infrastructure and functions to test different hyperparameter configurations and chooses the set of hyperparameters that maximize the model's accuracy.

The manually labeled test data are kept separate from the training and hyper-parameter tuning algorithms.[66] The models compare results from the algorithms to the manual labels in the test data to evaluate the accuracy of the algorithms.

We use two specific parameters that are commonly used to evaluate the performance of this class of model: "precision" and "recall."[67] Precision is the proportion of items which are correctly assigned a label out of all items that *are assigned* that label. For example, precision for detected faces is the number of actual faces out of all regions in an image that our model classifies as a face (that might not always be a face). Recall, on the other hand, tells us the percentage of items that are correctly assigned a label out of all items that *should be assigned* that label. In the case of recall for faces, recall is the proportion of the number of correctly detected faces out of the actual number of faces in the book.[68] Formally:

$$precision = \frac{true\ positives}{true\ positives + false\ positives}$$

$$recall = \frac{true\ positives}{true\ positives + false\ negatives}$$

The higher the precision, the fewer false positives the model produces. In other words, precision tells us from all the test examples that were predicted with a certain label, which ones are truly of that label? On the other hand, the higher the recall, the fewer false negatives the model produces. In other words, recall tells us, from all the test examples that should have had the label assigned, how many were actually assigned the label (Sokolova and Lapalme, 2009).

Our face detection model has 93.4 percent precision and 76.8 percent recall.

D.B.2 Image Feature Classification: Skin Segmentation Methods

Traditional skin segmentation methods assign a skin or non-skin label for every pixel of the cropped face image in which skin features are extracted. These labels are assigned using traditional image processing methods such as thresholding, level tracing, or watershed. These methods, however, face a number of challenges such as the need to take into account skin color (in)consistency across variations in illumination, acquisition types, ethnicity, ge-

[66] The manually labeled data for the face detection model came from data labeled by our research team. The manually labeled data for the feature classification model came from the UTKFace data set.

[67] AutoML has its own functions to calculate the precision and recall of the model. For our purposes, we use the precision and recall that were calculated on the test data. In other words, the model is run on the test data, and then the results generated by the trained model are compared to the results from the manually labeled test data.

[68] Sometimes "recall" is also referred to as "sensitivity."

ometric transformations, and partial occlusions (Lumini and Nanni, 2020). Our FC-CNN CRF method – by combining three different types of networks (an unary network, a pairwise network, and a continuous CRF network) – takes into account the local and global dependencies between the pixels, and considers the location of the pixels when assigning the skin label, preserving the region integrity. The CRF model parses the face image into semantic regions (e.g, eyes, eyebrows, and mouth) for further processing. This is integrated with an unary network for generating the feature map. The pairwise network is then used to learn the pixel-wise similarity based on neighbor pixels. Thus segmentation accuracy is greatly improved compared to traditional pixel-wise methods which do not take into account semantic regions, boundaries, and the correlations between neighbor pixels.

D.B.3 Image Feature Classification: Classifiying Skin Color Types

We classify the representative skin color for each detected face into one of three categories of skin color type: (1) monochromatic skin colors (e.g., grayscale, sepia), (2) polychromatic human skin colors (e.g., brown, beige), and (3) polychromatic non-typical skin colors (e.g., blue, green).

Monochromatic Classification. In the RGB color space, the closer the R, G, and B values are to each other, the less vibrant the color. For this reason, we classify a face as monochromatic if the standard deviation between the R, G, and B values associated with the weighted average of the face's top k skin colors is less than a threshold T. Thus, a given face i is classified as monochromatic using the following equation:

$$Monochromatic_i = \mathbb{1}\left[\sqrt{\frac{(R_i-\mu_i)^2+(G_i-\mu_i)^2+(B_i-\mu_i)^2}{3}} <= T\right] \tag{D1}$$

Where μ_i is equal to the average of the R, G, B values of face i.

Our process of choosing a threshold T proceeded as follows. First, we manually labeled a random sample of 2,836 detected faces (stratified by collection) as either monochromatic or polychromatic. We then calculated the mean squared error between the manual label and our predicted labels using the equation above for every integer value of T between zero and 100. We calculated the average of these mean squared errors using 1,000 bootstrapped samples. The threshold that minimized the mean squared error on average is given by $T = 13$; this provides a classification of images as being monochromatic or not that is 82.9 percent accurate, on average.

Polychromatic Classification. Once we have identified the monochromatic faces, we then separate the remaining faces into two polychromatic color types using the R, G, and

B values associated with the weighted average of a face's top k skin colors: (1) human skin colors and (2) polychromatic non-typical skin colors. This allows us to distinguish between humans and non-human characters who may have colorful skin tints (e.g., aliens, monsters, or characters found in Dr. Seuss books). Specifically, we classify the skin color of the face as a typical human skin color if R $\geq$ G $\geq$ B.[69] Otherwise, it is classified as a polychromatic non-typical skin color.

$$Human_i = [1 - Monochromatic_i] \times \mathbb{1}\,[R \geq G \geq B] \tag{D2}$$

$$NonTypical_i = [1 - Monochromatic_i] \times [1 - Human_i] \tag{D3}$$

We find this method of classifying the skin color of a face as human or non-typical to be 82.1 percent accurate using our set of 2,836 manually labeled faces.

To classify the darkness or lightness of pictured skin colors, we use the perceptual tint, or L* value, associated with the average of the k colors in L*a*b* space. This value ranges from zero to 100 where a value of zero represents the color black and a value of 100 represents the color white, and there is a range of colors in between.

D.B.4 Image Feature Classification: Race, Gender, and Age

To train our feature classification model we use a publicly available labeled data set called UTKFace which is a large-scale face data set consisting of over 20,000 face images with age, gender, and ethnicity labels. The images cover large variation in pose, facial expression, illumination, occlusion, and resolution and cover a large age range of individuals (from 0 - 116 years old) (Zhang and Qi, 2017). We split this data set into three parts: training (80 percent of the data), validation (10 percent), and testing (10 percent). The resulting model has 90.6 percent precision and 88.98 percent recall in our testing data.

Race Classification (Images). The model assigns the probability that a detected face is of a given race category: Asian, Black, Latinx + Others, or White. The race labels in the original model were defined in the UTKFace data set and include: Asian, Black, Indian, Others (where "Others" includes Latinx and Middle Eastern) and White. We combine Asian and Indian predictions into a broader Asian category. Each identified face is assigned the

[69] The boundaries of skin color regions in RGB space from an established pixel-based method of skin classification are defined as $R > 95$ and $G > 40$ and $B > 20$ and $\max\{R, G, B\} - \min\{R, G, B\} > 15$ and $|R - G| > 15$ and $R > G$ and $R > B$ (Vezhnevets, Sazonov and Andreeva, 2003). However, these rules for defining skin color regions are only focused on classifying skin color from photographs. We expand this region in RGB space to account for illustrated skin colors (such as pure white and yellow).

race category which the model gives the highest predicted probability to.[70,71]

Gender Classification (Images). For each face detected, we predict the probability that the face is female- (or male-) presenting. We label a face as female if the predicted probability that the face is female-presenting is greater than 50 percent; otherwise, we label the face as male.

We recognize that these classifications are imperfect and focus only on the performative aspect of gender presentation, as they are trained based on how humans classify images. Future work should incorporate the classification of fluid and nonbinary gender identities.

Age Classification (Images). The model assigns the probability that a detected face is of a given age category (infant, child, teenager, adult, senior). We aggregate these categories into two bins: child and adult. We collapse the probabilities for infant and child into a single "child" bin and those for teenager, adult, and senior into a single "adult" bin. A face is classified as that of a child if the probability assigned to the age categories comprising the aggregated child bin is greater than 50 percent, and as that of an adult otherwise.

D.C Text as Data

D.C.1 Digitizing Text

To extract text from digital scans of books, we use the Google Vision Optical Character Recognition (GVOCR). We input the raw files into GVOCR, which identifies and separates the text in each file from the images (e.g., illustrations and photographs). It then applies its own OCR software to the text sections of the scans, converting the text into ASCII which then encodes each character to be recognized by the computer. This generates the text data we analyze.[72]

[70]Previously, many existing artificial intelligence models that classified putative race had a high error rate, both misclassifying the putative race of identified people and, in "one-shot" models that identify existence of people and their putative race simultaneously, misclassifying people as non-human (Fu, He and Hou, 2014; Nagpal et al., 2019; Krishnan, Almadan and Rattani, 2020). Much work has been done to acknowledge and address these disparities (Buolamwini and Gebru, 2018; Mitchell et al., 2019).

[71]Classifying race is an imperfect exercise that will yield imperfect algorithms with imperfect categories. Our analysis by race looks across collections within race, so any error within a race would be consistent across collections (i.e., Both the Mainstream and Diversity collections would classify people of the same race similarly.)

[72]There are other commonly used OCR interfaces. However, over the past five years, researchers have consistently identified Google Cloud Vision OCR as the best technology for converting images to text. In one study, Tafti et al. (2016) compare the accuracy of Google Docs (now Google Vision), Tesseract, ABBYY FineReader, and Transym OCR methods for over 1,000 images and 15 image categories, and found that Google Vision generally outperformed other methods. In particular, Google Vision's accuracy with digital images was 4 percent better than any other method. Additionally, the standard deviation of accuracy for Google Vision was quite low, suggesting that the quality of OCR does not drastically change from one image to the next. A test of OCR tools by programmers compared the performance of seven different OCR tools

We clean these raw text data to remove erroneous characters and other noise generated by the OCR process, increasing the precision of our measurement of features in the text. The cleaning process removes numerical digits and line breaks but maintains capitalization, punctuation, and special characters. It also standardizes the various permutations of famous names (e.g., all variations of reference to Dr. Martin Luther King Jr. become "Martin Luther King Junior").

D.C.2 Vocab Lists Used in Token Counts

The vocab lists containing all the words we use in our token counts are listed below. These lists may not be comprehensive.

Gendered Terms. The gendered terms we enumerate are as follows.

Female. abuela, abuelita, actress, aunt, auntie, aunties, aunts, aunty, czarina, damsel, damsels, daughter, daughters, emperess, emperesses, empress, empresses, fairies, fairy, female, females, girl, girls, grandma, grandmas, grandmom, grandmother, grandmothers, her, hers, herself, housekeeper, housekeepers, ladies, lady, ma'am, madame, mademoiselle, mademoiselles, maid, maiden, maidens, maids, mama, mamas, mermaid, mermaids, miss, mlle, mme, mom, mommies, mommy, moms, mother, mothers, mrs, ms, nana, nanas, princess, princesses, queen, queens, she, sissie, sissy, sister, sisters, stepmother, stepmothers, titi, tsarevna, tsarina, tsaritsa, tzaritza, waitress, wife, witch, witches, wives, woman, women

Plural Female. aunties, aunts, damsels, daughters, emperesses, empresses, fairies, females, girls, grandmas, grandmothers, housekeepers, ladies, mademoiselles, maidens, maids, mamas, mermaids, mommies, moms, mothers, nanas, queens, sisters, stepmothers, witches, wives, women

Singular Female. abuela, abuelita, aunt, auntie, aunty, czarina, damsel, daughter, emperess, empress, fairy, female, girl, grandma, grandmom, grandmother, her, hers, herself, housekeeper, lady, maam, madame, mademoiselle, maid, maiden, mama, mermaid, miss, mlle, mme, mom, mommy, mother, mrs, ms, nana, princess, queen, she, sissie, sissy, sister, stepmother, titi, tsarevna, tsarina, tsaritsa, tzaritza, wife, witch, woman

Young Female. damsel, damsels, daughter, daughters, fairies, fairy, girl, girls, mademoiselle, mademoiselles, maiden, maidens, miss, princess, princesses, tsarevna

(Han and Hickman, 2019). This analysis also found Google Vision to be superior, specifically when extracting results from low resolution images. In another study that focused on comparing results from multiple image formats (including .jpg, .png, and .tif), Vijayarani and Sakila (2015) found that Google surpassed all other OCR tools. We also tested OCR using ABBYY FineReader and Google Tesseract. Our comparison of their performance relative to manual coding also showed GVOCR performed the best.

Old Female. abuela, abuelita, aunt, auntie, Auntie, aunts, aunty, czarina, emperess, emperesses, empress, empresses, grandma, grandmas, grandmom, grandmother, grandmothers, housekeeper, housekeepers, maam, madame, mama, mamas, mlle, mme, mom, mommies, mommy, moms, mother, mothers, mrs, nana, nanas, queen, queens, stepmother, stepmothers, titi, tsarina, tsaritsa, tzaritza, wife, witch, witches, wives, woman, women

Male. abuelito, abuelo, actor, boy, boys, bro, brother, brothers, butler, butlers, chap, chaps, czar, dad, daddies, daddy, dads, einstein, emperor, emperors, father, fathers, fellow, fellows, gentleman, gentlemen, granddad, granddads, grandfather, grandfathers, grandpa, grandpas, he, him, himself, his, hisself, husband, husbands, king, kings, knight, lad, lads, lord, lords, male, males, man, master, masters, men, merman, mermen, mr, paige, paiges, papa, papas, prince, princes, sir, sirs, son, sons, squire, squires, stepfather, stepfathers, tio, tsar, uncle, uncles, waiter, wizard, wizards

Plural Male. boys, brothers, butlers, chaps, daddies, dads, emperors, fathers, fellows, gentlemen, granddads, grandfathers, grandpas, husbands, kings, knights, lads, lords, males, masters, men, mermen, paiges, papas, princes, sirs, sons, squires, stepfathers, uncles, wizards

Singular Male. abuelito, abuelo, boy, bro, brother, butler, chap, czar, dad, daddy, emperor, father, fellow, gentleman, granddad, grandfather, grandpa, he, him, himself, his, hisself, husband, king, knight, lad, lord, male, man, master, merman, mr, paige, papa, prince, sir, son, stepfather, tio, tsar, uncle, wizard

Young Male. boy, boys, lad, lads, prince, princes, son, sons

Old Male. abuelito, abuelo, butler, butlers, czar, dad, daddies, daddy, dads, emperor, emperors, father, fathers, gentleman, gentlemen, granddad, granddads, grandfather, grandfathers, grandpa, grandpas, husband, husbands, king, kings, lord, lords, man, men, mr, papa, papas, sir, sirs, stepfather, stepfathers, tio, tsar, uncle, uncles, wizard, wizards

E Seattle Public Library Checkouts Data

To study the impact of being recognized by the children's book awards we examine, we analyze data from the Seattle Public Library system on all checkouts from the library between April 2005 and September 2017.[73] Awards are given near the end of January each year to books published in that year or the year before. We analyze checkout data for the award-winning books in our data, alongside all books belonging to the children's and junior book collections published in the year prior to the award, covering award years 2005 to 2017.

We collapse these to a data set measuring the number of collection-by-day checkouts, scaled by the number of books in the collection to generate a measure of the average number of checkouts per book, per day, in each of the three collections. We limit checkout data for each book to approximately one calendar year before the award was given and the two following calendar years.

To generate Figure 1, we re-center the checkout date according to its distance from the date in which the award is given for books published in that year. For example, books published in 2011 would be eligible for an award in 2012. Checkouts from before January 20th, 2012 (The first date of the ALA Midwinter Meeting in 2012) would be given negative values – for example, checkouts on January 10th, 2012, would be –10 days from January 20th, 2012. Checkouts after that date have positive values. Figure 1 shows the results of applying a 14-day moving average to each series of average collection-specific number of checkouts per day (divided by the number of books in that collection to account for the fact that the number of books per collection varies across the Mainstream, Diversity, and all other children's books) over the window of days to award spanning [–400 days, 730 days].

We quantify the post-award increase using a simple event study design. While not causal per se, this allows us to estimate more precisely how much more likely books in each collection are to be checked out after receipt of an award or honor, relative to the rest of the sample. To do so, we use the following equation:

$$checkouts_{cd} = \beta_1 Post + \beta_2 Post * Mainstream + \beta_3 Post * Diversity + \eta_c + \varepsilon_{cd}$$

The dependent variable is the average number of checkouts, per book, in collection c on day d. We regress this on the following variables: whether the day is after January 20th ($Post$) (a noisy estimate of the date when the awards are announced each year); a set of fixed effects for each collection; and an interaction of the $Post$ variable with the $Mainstream$ and

[73] These data are publicly available at https://data.seattle.gov/Community/Checkouts-by-Title/tmmm-ytt6; site accessed on April 15, 2021.

Diversity collection variables.

Table E1. Estimates of the Increase in Daily Checkouts After Receipt of Mainstream and Diversity Awards

	Dependent variable:
Parameter	Estimate
Non-Recognized Children's Books in Library Fixed Effect	0.089***
	(0.004)
Diversity Collection Fixed Effect	0.064***
	(0.004)
Mainstream Collection Fixed Effect	0.165***
	(0.004)
Post	0.032***
	(0.005)
Post × Diversity	0.012*
	(0.008)
Post × Mainstream	0.277***
	(0.008)
Observations	3,387
Adjusted R^2	0.757

Notes: These parameters were generated using the equation given in this subsection of the Data Appendix estimated using data from the Seattle Public Library on daily checkouts. *p<0.1; **p<0.05; ***p<0.01

We present our results in Table E1. This shows that after winning or being honored by an award, Mainstream books are approximately four times as likely as non-recognized children's books in the library to be checked out on any given day. We derive this from calculating the ratio of the post-award checkout rate for the Mainstream collection to that of the non-recognized books. For the Mainstream collection, this is the sum of the $Mainstream$ fixed effect, the coefficient on the "post-award" variable (*Post*), and the coefficient on the interaction term between *Post* and the $Mainstream$ collection, which sums to approximately 0.474. The post-award checkout rate for non-recognized children's books in the library is the sum of the non-recognized children's books in the library fixed effect and the coefficient on *Post*, which sums to approximately 0.121.

An alternate interpretation is that after winning the award, the Mainstream collection books are approximately 2.9 times more likely to be checked out than they were before. This is derived by dividing the sum of coefficients on *Post*, the interaction of *Mainstream* and *Post*, and the *Mainstream* fixed effect, by the *Mainstream* fixed effect. We note that these should be interpreted as suggestive estimates; we define "pre-" and "post-" award using January 20th, an estimate of when news of the award announcements is likely to reach readers, parents, and librarians. Its precise date varies from year to year.

For the Diversity awards, we see a slight change in checkout behavior after January 20th. This can be seen in our estimate of the interaction term between *Diversity* and *Post*, which is statistically significant, but small in magnitude - especially when compared to the coefficient on the interaction term between *Mainstream* and *Post*. Seen through the lens of the calculations above, after receiving an award, Diversity collection books are more than 11 percent *less* likely to be checked out than non-winners; this can be derived analogously, comparing the post-award checkout rate for the Diversity collection – the sum of the *Diversity* fixed effect, the coefficient on *Post*, and the coefficient on the interaction term between *Post* and the *Diversity* collection, which sums to approximately 0.108. The post-award checkout rate for non-winners is the sum of the *Non-winners* fixed effect and the coefficient on *Post*, which is approximately 0.121. Prior to receipt of the award, they were approximately 28 percent less likely to be checked out.

In Table E2, we present an alternative specification where we estimate a similar equation, only with separate parameters for award winners and honorees. This shows broadly similar results, with one exception: winning a mainstream award yields a premium that is 2.5 times as large as merely being an honoree. This is similar to the visual patterns we see in Figure 1 and, more specifically, the distinct post-award increases in checkouts we observe for winners and awardees, respectively.

Table E2. Estimates of the Increase in Daily Checkouts After Receipt of Mainstream and Diversity Awards, Disaggregated by Winners and Honorees

	Dependent variable:
Parameter	Estimate
Non-Recognized Children's Books in Library Fixed Effect	0.089***
	(0.006)
Diversity Honoree Fixed Effect	0.063***
	(0.006)
Diversity Winner Fixed Effect	0.065***
	(0.006)
Mainstream Honoree Fixed Effect	0.165***
	(0.006)
Mainstream Winner Fixed Effect	0.183***
	(0.006)
Post	0.032***
	(0.008)
Post × Diversity Honoree	0.006
	(0.011)
Post × Diversity Winner	0.014
	(0.011)
Post × Mainstream Honoree	0.205***
	(0.011)
Post × Mainstream Winner	0.499***
	(0.011)
Observations	5,645
Adjusted R^2	0.743

Notes: This table is similar to Table E1, except that it separates award premia by whether books were named honorees for a given award, or recipients of the award itself. *p<0.1; **p<0.05; ***p<0.01

F Award Criteria

In this section we give the criteria for award selection for the Newbery and Caldecott awards and provides links to the criteria for the other awards.

F.A Caldecott Medal Criteria

Terms and criteria are listed below.[74] Note that the numbering and itemization follows the formatting as presented on the website and were not altered for consistency.

F.A.1 Terms

The Medal shall be awarded annually to the artist of the most distinguished American picture book for children published by an American publisher in the United States in English during the preceding year. There are no limitations as to the character of the picture book except that the illustrations be original work. Honor books may be named. These shall be books that are also truly distinguished.

The award is restricted to artists who are citizens or residents of the United States. Books published in a U.S. territory or U.S. commonwealth are eligible.

The committee in its deliberations is to consider only books eligible for the award, as specified in the terms.

F.A.2 Definitions

A "picture book for children" as distinguished from other books with illustrations, is one that essentially provides the child with a visual experience. A picture book has a collective unity of story-line, theme, or concept, developed through the series of pictures of which the book is comprised.

A "picture book for children" is one for which children are an intended potential audience. The book displays respect for children's understandings, abilities, and appreciations. Children are defined as persons of ages up to and including fourteen and picture books for this entire age range are to be considered.

"Distinguished" is defined as:

- Marked by eminence and distinction; noted for significant achievement.
- Marked by excellence in quality.
- Marked by conspicuous excellence or eminence.

[74] Terms and criteria downloaded exactly from https://www.ala.org/alsc/awardsgrants/bookmedia/caldecott on July 14, 2022.

- Individually distinct.
- The artist is the illustrator or co-illustrators. The artist may be awarded the medal posthumously.

The term "original work" may have several meanings. For purposes of these awards, it is defined as follows: "Original work" means that the illustrations were created by this artist and no one else. Further, "original work" means that the illustrations are presented here for the first time and have not been previously published elsewhere in this or any other form. Illustrations reprinted or compiled from other sources are not eligible.

"American picture book in the United States" means that books first published in previous years in other countries are not eligible. Books published simultaneously in the U.S. and another country may be eligible. Books published in a U.S. territory or U.S. commonwealth are eligible.

"In English" means that the committee considers only books written and published in English. This requirement DOES NOT limit the use of words or phrases in another language where appropriate in context.

"Published... in the preceding year" means that the book has a publication date in that year, was available for purchase in that year, and has a copyright date no later than that year. A book might have a copyright date prior to the year under consideration but, for various reasons, was not published until the year under consideration. If a book is published prior to its year of copyright as stated in the book, it shall be considered in its year of copyright as stated in the book. The intent of the definition is that every book be eligible for consideration, but that no book be considered in more than one year.

"Resident" specifies that author has established and maintains a residence in the United States, U.S. territory, or U.S. commonwealth as distinct from being a casual or occasional visitor.

The term, "only the books eligible for the award," specifies that the committee is not to consider the entire body of the work by an artist or whether the artist was previously recognized by the award. The committee's decision is to be made following deliberation about books of the specified calendar year.

F.A.3 Criteria

In identifying a "distinguished American picture book for children," defined as illustration, committee members need to consider:

- Excellence of execution in the artistic technique employed;
- Excellence of pictorial interpretation of story, theme, or concept;
- Appropriateness of style of illustration to the story, theme or concept;
- Delineation of plot, theme, characters, setting, mood or information through the pictures;
- Excellence of presentation in recognition of a child audience.

The only limitation to graphic form is that the form must be one which may be used in a picture book. The book must be a self-contained entity, not dependent on other media (i.e., sound, film or computer program) for its enjoyment.

Each book is to be considered as a picture book. The committee is to make its decision primarily on the illustration, but other components of a book are to be considered especially when they make a book less effective as a children's picture book. Such other components might include the written text, the overall design of the book, etc.

Note: The committee should keep in mind that the award is for distinguished illustrations in a picture book and for excellence of pictorial presentation for children. The award is not for didactic intent or for popularity.

[Adopted by the ALSC board, January 1978. Revised, Midwinter 1987. Revised, Annual 2008.]

F.B Newbery Medal Criteria

Terms and criteria are listed below.[75] Note that the numbering and itemization follows the formatting as presented on the website and were not altered for consistency.

F.B.1 Terms

1. The Medal shall be awarded annually to the author of the most distinguished contribution to American literature for children published by an American publisher in the United States in English during the preceding year. There are no limitations as to the character of the book considered except that it be original work. Honor books may be named. These shall be books that are also truly distinguished.
2. The Award is restricted to authors who are citizens or residents of the United States.

[75] Terms and criteria downloaded exactly from https://www.ala.org/alsc/awardsgrants/bookmedia/newbery on July 14, 2022.

3. The committee in its deliberations is to consider only the books eligible for the award, as specified in the terms.

F.B.2 Definitions

1. "Contribution to American literature" indicates the text of a book. It also implies that the committee shall consider all forms of writing—fiction, non-fiction, and poetry. Reprints, compilations and abridgements are not eligible.

2. A "contribution to American literature for children" shall be a book for which children are an intended potential audience. The book displays respect for children's understandings, abilities, and appreciations. Children are defined as persons of ages up to and including fourteen, and books for this entire age range are to be considered.

3. "Distinguished" is defined as:

   - Marked by eminence and distinction; noted for significant achievement.
   - Marked by excellence in quality.
   - Marked by conspicuous excellence or eminence.
   - Individually distinct.

4. "Author" may include co-authors. The author(s) may be awarded the medal posthumously.

5. The term "original work" may have several meanings. For purposes of these awards, it is defined as follows:

   - "Original work" means that the text was created by this writer and no one else. It may include original retellings of traditional literature, provided the words are the author's own.
   - Further, "original work" means that the text is presented here for the first time and has not been previously published elsewhere in this or any other form. Text reprinted or compiled from other sources are not eligible. Abridgements are not eligible.

6. "In English" means that the committee considers only books written and published in English. This requirement DOES NOT limit the use of words or phrases in another language where appropriate in context.

7. "American literature published in the United States" means that books first published in previous years in other countries are not eligible. Books published simultaneously

in the U.S. and another country may be eligible. Books published in a U.S. territory, or U.S. commonwealth are eligible.

8. "Published... in the preceding year" means that the book has a publication date in that year, was available for purchase in that year, and has a copyright date no later than that year. A book might have a copyright date prior to the year under consideration but, for various reasons, was not published until the year under consideration. If a book is published prior to its year of copyright as stated in the book, it shall be considered in its year of copyright as stated in the book. The intent of the definition is that every book be eligible for consideration, but that no book be considered in more than one year.

9. "Resident" specifies that the author has established and maintains a residence in the United States, U.S. territory, or U.S. commonwealth as distinct from being a casual or occasional visitor.

10. The term, "only the books eligible for the award," specifies that the committee is not to consider the entire body of the work by an author or whether the author was previously recognized by the award. The committee's decision is to be made following deliberation about the books of the specified calendar year.

F.B.3 Criteria

1. In identifying "distinguished contribution to American literature," defined as text, in a book for children,

   (a) Committee members need to consider the following:

      - Interpretation of the theme or concept
      - Presentation of information including accuracy, clarity, and organization
      - Development of a plot
      - Delineation of characters
      - Delineation of a setting
      - Appropriateness of style.

      Note: Because the literary qualities to be considered will vary depending on content, the committee need not expect to find excellence in each of the named elements. The book should, however, have distinguished qualities in all of the elements pertinent to it.

    (b) Committee members must consider excellence of presentation for a child audience.

2. Each book is to be considered as a contribution to American literature. The committee is to make its decision primarily on the text. Other components of a book, such as illustrations, overall design of the book, etc., may be considered when they make the book less effective.

3. The book must be a self-contained entity, not dependent on other media (i.e., sound or film equipment) for its enjoyment.

   Note: The committee should keep in mind that the award is for literary quality and quality presentation for children. The award is not for didactic content or popularity.

Adopted by the ALSC Board, January 1978. Revised, Midwinter 1987. Revised, Annual 2008.

F.C Award Information for Diversity Collection

In this section, we provide the website describing each award and its selection criteria, accessed on July 15, 2022. Selection criteria vary by award. At a high level, they share two main goals. One is to recognize excellence in the content of the book. This goal, and the text of the various award criteria given in the links below, tracks closely with the main goals of the Caldecott and Newbery awards. The second goal is to recognize books who portray, recognize, or elevate a specific identity group, for example, people with disabilities or Hispanic Americans. These goals vary widely by award, as each award focuses on a specific identity.

- American Indian Youth Literature Award
  Site: ailanet.org/activities/american-indian-youth-literature-award

- Américas Award
  Site: claspprograms.org/pages/detail/65/About-the-Award

- Name: Arab American Book Award
  Site: arabamericanmuseum.org/book-awards/

- Asian/Pacific American Award for Literature
  Site: apalaweb.org/awards/literature-awards/literature-award-guidelines/

- Carter G. Woodson Book Awards
  Site: woodsonawards.weebly.com/

- Coretta Scott King Book Award

Site: ala.org/rt/emiert/cskbookawards/slction

- Dolly Gray Children's Literature Award
  Site: dollygrayaward.com/
- Ezra Jack Keats Award
  Site: degrummond.org/ezra-jack-keats-book-award-guidelin
- Middle East Book Award
  Site: meoc.us/book-awards.html
- Notable Books for a Global Society
  Site: clrsig.org/nbgs.html
- Pura Belpré Award
  Site: ala.org/alsc/awardsgrants/bookmedia/belpre
- Rise: A Feminist Book Project
  Site: risefeministbooks.wordpress.com/criteria/
- Schneider Family Book Award
  Site: ala.org/awardsgrants/awards/1/apply
- Skipping Stones Youth Honor Awards
  Site: skippingstones.org/wp/youth-honors-award/
- South Asia Book Award
  Site: southasiabookaward.wisc.edu/submission-guidelines/
- Stonewall Book Awards
  Site: ala.org/awardsgrants/awards/177/apply
- Tomás Rivera Mexican American Awards
  Site: education.txstate.edu/ci/riverabookaward/about.html

G Perspectives of Suppliers of Children's Books

We complement our quantitative analysis of the supply and demand pressures on publishers' choice of books with qualitative analysis of data from semi-structured, one-on-one interviews of professionals who currently work at or recently worked at libraries, publishing houses, and children's bookstores, and/or who served on award selection committees. Our interviews began with a prompt that asked a series of questions, first about the processes they used to identify and select books, and then about their perception and understanding of the forces that shape the content of these books.

A few key themes arose from these conversations. The first theme is that many booksellers, publishers, and librarians wish to procure and promote books that highlight people from historically marginalized groups, particularly Black and Latina/o/x people. A common goal across librarians and booksellers was the desire to show children both potential versions of themselves, as well as potential versions of the world they grow up to inhabit. One professional who had served as both a librarian and a bookseller asserted that, when presenting books to children, librarians and booksellers alike wish "to provide each child with both a mirror and a window." This paraphrases the description in Bishop (1990), which argues that the books we give to children should serve as mirrors, windows, and sliding glass doors - in other words, the books should show children visions of themselves, windows onto the reality they inhabit, and doors through which they can step to see imaginary futures they might inhabit, respectively.

The second theme is that, until recently, this desire to present children with both a mirror and a window was very difficult to meet. Several interviewees asserted that this difficulty arose from mainstream publishers not offering sufficient amounts of this content. This corresponds to the economic forces we study in Section VI, wherein books with greater representation of non-dominant societal groups will be under-supplied by the market. One interviewee – the owner of a decades-old children's bookstore in a medium-sized midwestern city – lamented that before the mid-2010's, requests to publishers for books representing people of color was met with the quip: "we don't sell those books because those books don't sell." In response, motivated booksellers such as this professional sought out smaller publishers specializing in such content, such as Lee and Low, a publishing house founded in 1991 to address this shortcoming.[76]

To better understand the process through which books were selected for these awards,

[76] The #WeNeedDiverseBooks movement (diversebooks.org), started in roughly 2012, has also agitated and organized for more equitable representation in books. A relevant resource created to meet this need is the Diverse Book Finder, available at diversebookfinder.org.

we also conducted semi-structured interviews with people involved in the selection committees. Committee members are selected by either election or appointment by the head of ALSC to serve for a one-year term. Committee members review books published in that year, vetting them based on a set of criteria specific to each award. At the end of the term, the committee convenes to discuss candidates and select honorees. Two key themes arose in these discussions: first, the criteria for selection are stable over time, despite the other secular changes in this period.[77] Second, the composition of the award committees generally comprise a circulating group of librarians, booksellers, and educators that refreshes every year.[78] As a result, these awards – particularly those in the Mainstream collection – are likely to reflect the equilibrium of supply from the publishing industry and demand from the annually rotating group of educators and booksellers selected to be on the committees, rather than the idiosyncratic tastes of a few individuals.

[77] We give these criteria for the Mainstream collection awards, and link to those in the Diversity collection, in Appendix F.

[78] According to ALSC bylaws for the Mainstream awards, individuals who served on a committee in one year were ineligible to serve on it in following several years.