ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 21-462

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

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 race, gender, and age in award-winning children's books from the past century. We find that more characters with darker skin color appear over time, but the most influential books persistently depict a greater proportion of light-skinned characters than other books, even after conditioning on race; we also find that children are depicted with lighter skin than adults. 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 report empirical evidence for predictions about the supply of and demand for representation that would generate 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.

VERSION: November 2022

Suggested citation: Adukia, Anjali, Alex Eble, Emileigh Harrison, Hakizumwami Birali Runesha, and Teodora Szasz. (2022). What we teach about race and gender: Representation in images and text of children's books. (EdWorkingPaper: 21-462). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/ska1-an37

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

July 21, 2022

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 race, gender, and age in award-winning children's books from the past century. We find that more characters with darker skin color appear over time, but the most influential books persistently depict a greater proportion of light-skinned characters than other books, even after conditioning on race; we also find that children are depicted with lighter skin than adults. Relative to their

*Contact: team@miielab.com. For helpful feedback, we thank Milena Almagro, Barbara Atkin, Karen Baicker, Anna Brailovsky, Tom Brock, Mary Gen Davies, Steven Durlauf, Alice Eagly, Lisa Englehardt, Allyson Ettinger, James Evans, James Fenske, Alex Frankel, Adam Gamoran, Claudia Goldin, Josh Goodman, Austan Goolsbee, Jon Guryan, Andrew Ho, Rick Hornbeck, Caroline Hoxby, Larry Katz, Ro'ee Levy, Susanna Loeb, Jens Ludwig, Jonathan Meer, Martha Minow, Sendhil Mullainathan, Shirley Mullin, Derek Neal, Anna Neumann, Matt Notowidigdo, Aaron Pallas, Steve Raudenbush, Cybele Raver, Heather Sarsons, Eric Schneider, Fred Stafford, Chenhao Tan, David Uminsky, Miguel Urquiola, Angelino Viceisza, Alessandra Voena, Amy Stuart Wells, Melanie Xue and others including seminar participants at AEFP, Bocconi, Brown, CGD, Discrimination and Disparities, EPC, ETH Zurich, Harvard, NAEd/Spencer, NBER, NYU, SREE, W.T. Grant Fdn., UChicago, UVA, and UW-Madison. 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, Mariel Bedoya-Guevara, Callista Christ, Marliese Dalton, Anjali Das, Maya Escueta, Saloni Gupta, Amara Haider, Shavonna Hinton, Camilo Ibáñez, Juan Miguel Jimenez, Jalnidh Kaur, Zipporah Klain, Jarvis Lam, Erica Lin, Ping-Chang Lin, Ping-Jung Liu, Simon Mahns, Noah McLean, Karla Monteiro, Jan Morgan, Ifeatu Oliobi, Sarita Oré, Raj Shukla, Todd Takeno, Bhargavi Thakur, Jeffrey Tharsen, Qurat ul ain, 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.

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 report empirical evidence for predictions about the supply of and demand for representation that would generate 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.

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 et al., 2022; Arold, 2022), the presence or absence of these characters contributes 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; Riley, 2017; Bian, Leslie and Cimpian, 2017; Gershenson et al., 2018; Porter and Serra, 2020), these representations offer means through which society can either address, perpetuate, or entrench structural inequalities.

In this paper, we develop and apply computer science tools from the fields of computer vision and natural language processing to measure the representation of racial constructs, gender identity, and age in the images and text of influential children's books over the last century. These 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 then explore the economic forces that may contribute to this representation.

Our main data set comprises influential books targeted to children and likely to appear in homes, classrooms, and libraries over the past century. Specifically, we analyze books that have won awards from the Association for Library Service to Children, a division of the American Library Association, starting in 1922. We divide the award-winning corpora into two primary collections: (i) "Mainstream" books considered to be of high literary value but written without explicit intention to highlight an identity group (i.e., the Newbery and Caldecott awards) and (ii) "Diversity" books selected because they highlight experiences of specific identity groups (e.g., the Coretta Scott King and Rise Feminist awards).

We show that receipt of the awards in the Mainstream collection increases consumption of these books by both library borrowers and book purchasers. Using daily book checkout data from a major public library system, we find that books which received a Mainstream award are checked out four times as often on average, relative to other children's books. Using purchase-level data from the Numerator OmniPanel data set on over 1.5 million children's book purchases, we find that books which received a Mainstream award sell over 2.5 times as many unique copies on average as books which received a Diversity award and approximately five times as many unique copies on average 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 (Smith, 2013; Koss, Johnson and Martinez, 2018) as well as other recorded increases in the sales of children's books after receipt of an award (Cockcroft, 2018). It also highlights the particular societal influence these books may have, and underscores the importance of understanding the messages they may transmit.

Our analysis characterizes several dimensions of the representation of race, gender, and age in these books. It also shows how these levels have endured, or changed, over time. We find that these award-winning children's books include more characters with dark skin over time, but those in the Mainstream collection are more likely to depict lighter-skinned characters than those in the Diversity collection, 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 this systematic difference in skin colors across ages in society.[1] Regardless of the reasons why these differences exist, our estimates show that lighter-skinned children see themselves represented more often in these books than do darker-skinned children. In addition, relative to their share of the U.S. population, we see that Black and Latinx people have been historically underrepresented.

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, except in the collection of books specifically selected to highlight the experiences of females. This suggests there may be symbolic inclusion of females in pictures without their substantive inclusion in the actual story. Furthermore, despite being half of the U.S. population, and despite substantial changes in female societal roles over time, females are persistently less likely than males to be represented in the text of books in our sample. This finding is consistent across all of the measures we use: pronoun counts, specific gendered tokens,[2] gender of famous characters, character first names, and

[1]Differences in skin color between children and adults could present in many possible configurations: adults could be darker than children (perhaps due to greater exposure to the sun due to outside labor or due to children of mixed race couples having a more compressed distribution of skin tone values than that of their parents), children could be darker than adults (given evidence of the breakdown of melanin over the life course (Sarna et al., 2003)), or the skin tone of adults and children could be similar, on average.

[2]A "token" refers to a single word such as "queen" or "nephew." We explain this further in Section IV.

geographic origin. Another surprising result is that, even though these books are targeted to children, adults are depicted more than children both in images and text.

We find that White males from North America and Europe comprise the majority of famous figures in the text in all collections. While the Diversity collection portrays a greater proportion of famous figures born outside of the United States or Europe than the Mainstream collection, those portrayed still heavily skew male. The combined representation from all racial groups other than Black or White (e.g., Asian, Latinx) comprises zero to eight percent of all famous people, on average, per collection.

We next discuss economic forces which may contribute to these patterns, and provide suggestive evidence of the ways the messages in these books may propagate through society across generations. On the supply side, we draw a set of predictions from prior theoretical and empirical work on the economics of the media (Waldfogel, 2003, 2007). These suggest 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. Examining data on book sale volume and book prices from over 1.5 million book purchases in the U.S., we find evidence consistent with both phenomena. We also show that there are fewer copies of these books in libraries that serve predominantly White communities. On the demand side, we draw a set of similar predictions from prior theoretical work on the economics of identity from (Akerlof and Kranton, 2000), which suggest that people are more likely to consume books which center identities similar to their own. Using rich data on the consumption of these books from daily library checkouts and consumer-level book purchases matched to demographic information at the neighborhood level, we find evidence consistent with these patterns.

Finally, we estimate how local book consumption relates to local consumer beliefs by linking our individual-level book purchase data to the Cooperative Election Study (CCES), a nationally representative, stratified sample survey of 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, with more purchases of books with more (less) diverse representation among people with more progressive (conservative) viewpoints. Together, our analyses suggest that these economic forces contribute to the patterns of representation that we document.

In summary, our paper makes four key contributions. First, we develop and hone a series of tools from the field of computer vision for scalably and systematically processing images into analyzable data on the skin color, race, gender, and age of detected characters. This process involves three primary components: (1) training the computer to detect faces, (2) classifying skin color, and (3) predicting the race, gender, and age of the faces. Second, we apply these image-to-data tools alongside established natural language processing tools to measure how race, gender, and age have been represented to children in the images and text within almost a century of influential children's books, and document how this representation has changed over time. Third, we characterize economic drivers that may contribute to these levels of representation and present empirical evidence of our key economic predictions on both the demand and supply side, showing how the pressures from distinct economic forces may contribute the persistent overrepresentation of historically dominant identities that we find. Finally, using data on local book consumption and local consumer beliefs, we show that the levels of representation contained in the books people buy are highly correlated with their views on race and immigration. Prior research on how 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) suggests that these patterns of children's book purchases may help explain the persistence and intergenerational transmission of related beliefs.

This paper proceeds as follows. We present background information in Section I. Section II describes the books in our data. Section III discusses our image analysis tools. Section IV discusses our text analysis tools. Section V synthesizes our final measures. Section VI presents our main results, showing our estimates of inequality and inclusion of race and gender in the images and text of different book collections, and how these estimates change or persist over time. Section VII discusses potential economic factors underlying demand for, and supply of, representation in children's books, in addition to examining the relationship between book purchases and local consumer beliefs. Section VIII discusses the potential benefits and concerns to using AI models. Section IX concludes. The appendix includes further analysis, additional information on methods, cost-effectiveness of automated content analysis relative to traditional manual approaches, and additional analysis quantifying the increase in library checkout demand for books following receipt of an award.

I The Importance of Representation and the Challenge of Measurement

In this section, we briefly discuss research on the representation of race and gender and discuss empirical challenges involved in measuring these representations.

I.A Why Should We Care? The Importance of Equality in Representation

Our institutional practices, public policies, and cultural representations reflect the value that society assigns to specific groups. Inequality in representation, therefore, constitutes an explicit statement of inequality in value. If our records of history, culture, and society are disproportionately associated with whiteness and maleness, then the human potential of females, males of color, and non-binary individuals is devalued relative to the privileged group. Furthermore, this can have meaningful implications for societal structure. Across societies, the genderedness of representations in language and folklore, respectively, are strongly negatively correlated with gender equity in education, labor force participation, and other societal roles (Jakiela and Ozier, 2018; Michalopoulos and Xue, 2021). In a broad range of cultural products, from news media and history books to children's books, people who do not belong to the culturally dominant group are typically absent or portrayed through negative stereotypes (O'Kelly, 1974; Stewig and Knipfel, 1975; Dobrow and Gidney, 1998; Balter, 1999; Witt, 2000; Brooks and Hébert, 2006; Martin, 2008; Paceley and Flynn, 2012; Daniels, Layh and Porzelius, 2016). Prior empirical work has documented that the content of books has the potential to shape children's future beliefs and later life outcomes (Fuchs-Schündeln and Masella, 2016; Cantoni et al., 2017; Arold et al., 2022; Arold, 2022).

While myriad structural barriers to racial and gender equality are woven throughout the organizations, laws, and customs of our society (Darity and Mason, 1998; O'Flaherty, 2015; Blau and Kahn, 2017; Muhammad, 2019; Chetty et al., 2020), inequality of representation is an important potential contributor to inequality in outcomes because of its potential to instill the belief that members of the underrepresented group are inherently deficient.[3] Research from different disciplines supports the notion that representation gaps may be linked to socioeconomic inequality. For example, the experience of cultural subjugation may reduce the "capacity to aspire" (Appadurai, 2004). 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 these examples may diminish a child's expected return to effort. If that change in expectation reduces actual effort, it is likely to lower performance, thus reinforcing the message behind the (once-erroneous) message. This pathway underscores the importance of addressing inequality in representation within educational materials.

[3] Several studies in economics have shown that exposure to variations in content among books used to teach children, ranging from subjects as diverse as history and religion, can lead to variations in later life beliefs (Fuchs-Schündeln and Masella, 2016; Cantoni et al., 2017; Arold et al., 2022; Arold, 2022). 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). Furthermore, 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). 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).

Inequality in representation in the context of schools is particularly pernicious because educational materials are explicitly intended to shape students' views of the world around them, and schools make important contributions to the formation of children's social preferences (Cappelen et al., 2020). Importantly, these materials are designed to help children learn about the world, and the messages they contain also have the potential to shape how children view *others* of different identities. Specifically, levels and manners of representation can also 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). When children do or do not see others represented, their conscious or unconscious perceptions of their own potential and that of unrepresented groups can be molded in detrimental ways and can erroneously shape subconscious defaults.

Empirical evidence suggests that the reverse also may be true: improving representation may improve outcomes. Closing the representation gap by revealing previously invisible opportunities may influence beliefs, actions, and educational outcomes for females and, separately, people of underrepresented racial and ethnic identities of all genders (Dee, 2004; Stout et al., 2011; Beaman et al., 2012; Riley, 2017). While not a panacea, such "subject-object identity match" (e.g., teacher-student identity match, or content-reader identity match) can help reduce academic performance gaps among multiple marginalized groups via a wide range of potential channels.[4]

I.B The Importance of Accounting for Intersectionality

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.[5]

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

[4] 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.

[5] It is important to note that intersectionality does not merely refer to an "interaction effect" (e.g., between race and gender), but rather the specific intersections of multiple dimensions of marginalization.

lead to greater disadvantage among individuals at the intersection of multiple marginalized identities. Taking such an intersectional lens is, of course, an ambitious task that requires a wide-reaching analysis of norms, rules, laws, and history. The starting point for our analysis is that a key site of power – and thus of potential power imbalances – is 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 et al., 2022; Arold, 2022).

In this paper, our approach to account for intersectionality is to estimate whether or not there are such power imbalances in the books we study at the intersections of multiple dimensions of marginalization. We pursue this goal in the definition of our sample of which books to analyze and in the way we conduct our analysis.

We study a series of awards that specifically recognize books for how they highlight the representation of excluded or marginalized identities. We measure the levels of representation within them, and contrast these with levels of representation in a highly-regarded set of books with no such aim. With this contrast, we can measure whether, within books deliberately trying to elevate the representation of members of marginalized groups, there remains a power imbalance within the representation of these groups, with greater power held by members of the group who do not exist within one or more additional marginalized groups. This power imbalance can manifest in two ways. First, it can manifest in absolute terms – for example, as compared to population shares. Second, it can manifest in relative terms, that is, as compared to the balance of representation between individuals at one versus multiple sites of marginalization in the books attempting to remedy such power imbalances, as compared to in books where no such effort is explicitly taken.

Our analysis aims to estimate the extent of the following problem: inattention to intersectionality can lead to the omission of the experiences of groups with multiple identities that have been historically and/or contemporaneously excluded from analysis. An effort by publishers to diversify by gender, for example, is likely to overrepresent the experiences of White women relative to women of color, given the relative abundance of White women in popular media. Even those who select content with an eye towards increasing representation of particular groups are themselves often products of a system that reflects the structural racism, sexism, and other drivers of systemic inequality. Thus, even deliberate efforts to address inequality in one dimension of representation may inadvertently perpetuate other inequalities in other dimensions of representation, thereby contributing to the underrepresentation – or exclusion – of people whose identities fall at the intersection of multiple sites of such historical exclusion or subordination (Davis, 2008).

It is important to acknowledge that this type of analysis barely scratches the surface of what we might want to study in a richer interrogation of intersectionality in these texts. Crucially, in future work we aim to measure how these individuals are being portrayed, to characterize whether similar power imbalances arise in the roles in which people who have multiple marginalized identities are and are not portrayed, and to understand the likely causes of this. These goals notwithstanding, our analysis here provides an initial step in understanding these important phenomena.

I.C The Need for Better Measurement Tools

Addressing these system-level issues requires a systematic method for assessing the representations contained in the content used to instruct children. Many individual educators and curriculum developers have worked to address this representation gap by, for instance, expanding their curricula to include individual books that elevate the presence of an identity group. The incidence, levels, and impacts of such efforts, however, are likely to vary dramatically across teachers and schools, and the sheer quantity of content that they have to review or create is too large for any individual to manually track and assess. As a result, educators, administrators, and policymakers currently lack feasible ways to systematically identify such inclusive materials.

The process of identifying such books has historically been done through manual content analysis, which is 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). Because this manual process is time-consuming and therefore costly, resource constraints have limited the scope of such work.

In this paper, we demonstrate how tools from computer vision and natural language processing can be used to systematically analyze features of large bodies of content. We expand and develop tools for image analysis, pairing them with tools from text analysis used in recent work by Caliskan, Bryson and Narayanan (2017), Garg et al. (2018), and Kozlowski, Taddy and Evans (2019). These tools can facilitate broader and more cost-effective measurements of racial constructs, gender identity, and age in images and text in a larger set of content than could be analyzed by any one individual or institution. In Appendix Section D.A we describe the cost-effectiveness and other advantages of using AI in our specific context and describe how we use manual content analysis to validate our computer driven measures of representation.

There are challenges to this type of numeric measurement of representation, however. For example, racial constructs are multi-faceted and often ill-defined. To address this challenge, we measure different facets of the broad concept of race in various ways: skin color, putative race (that is, assigned by society), and birthplace.[6] It is important to focus on these racial constructs, because each of these concepts has been used in systems that perpetrate oppression and inequality by asserting a system of intrinsic hierarchy. In systems of explicit and implicit racism, European facial features are privileged over non-European features, such as those seen as African, Asian, or Indigenous peoples (MacMaster, 2001). In colorism, lighter skin tones are similarly either more desired or more associated with desirable traits relative to darker skin tones (Banks, 1999; Hunter, 2007; Burton et al., 2010; Ghavami, Katsiaficas and Rogers, 2016; Keith and Monroe, 2016; Dixon and Telles, 2017). Separately, current methods are only able to measure binary gender identities, mis-classifying non-binary and gender fluid identities. While the methods we use have this shortcoming, addressing this challenge is an important avenue for future work.

Furthermore, even numeric characterization of the representation of race and gender can be difficult and, if executed improperly, a tool for the perpetuation of bias. Because AI tools are designed by humans, they contain human biases (Das, Dantcheva and Bremond, 2018), and, if used improperly, their use can even perpetuate inequality (Fu, He and Hou, 2014; Nagpal et al., 2019; Krishnan, Almadan and Rattani, 2020). New scholarship shows, however, that careful attention to identifying and addressing these biases allows scholars and practitioners to overcome them while preserving the advantages of this type of computer measurement (Buolamwini and Gebru, 2018; Mitchell et al., 2019). We include a larger discussion surrounding the limitations of AI in Section VIII.

Many approaches to mitigating inequality focus on trying to "fix" the individual (often people from marginalized backgrounds) but such approaches neglect to address the systemic forces that lead to inequalities. Focusing instead on the curricula and other content to which children are automatically exposed provides an opportunity to help "fix the institution," or the part of the larger system (Recalde and Vesterlund, 2020).

[6]A wide range of research studies highlight the importance of both place of birth (Jencks and Mayer, 1990; Brooks-Gunn et al., 1993; Cutler and Glaeser, 1997; Leventhal and Brooks-Gunn, 2000; Sampson, Morenoff and Gannon-Rowley, 2002; Chetty, Hendren and Katz, 2016) and the color of one's skin (Banks, 1999; Hunter, 2007; Burton et al., 2010; Ghavami, Katsiaficas and Rogers, 2016; Keith and Monroe, 2016; Dixon and Telles, 2017) in determining one's chances of economic and social mobility.

II Data

We use our tools to analyze levels of representation in collections of high-profile, award-winning children's books that are consumed widely across households, libraries, and schools in the U.S. We focus on award-winning children's books because they are deliberately chosen and curated by librarians, teachers, and school administrators, and are often selected because they transmit clear narratives about appropriate conduct, an account of important historical moments, or other, often identity-specific messages. As a result, they present a prime opportunity to understand and improve the institutions of education.

II.A Award-Winning Children's Books

We study the representation contained in the images and text of books that received awards administered or featured by the Association for Library Service to Children (ALSC), a division of the American Library Association (ALA). Our sample comprises 1,130 books, and each book in this sample is associated with at least one of 19 different awards.[7]

We divide these award-winning corpora into two primary "collections" of books, which we call the "Mainstream" and "Diversity" collections. Figure 1a presents the full list of awards in our sample and the collection(s) into which we categorized them. Figure 1b and Table 1 show the sample size of each collection.

Mainstream Collection. The Mainstream collection comprises books that have received either 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 quality and not their popular appeal per se. Books receiving these awards are considered to be of general interest to all children and are quickly incorporated into mainstream outlets for children, such as school libraries and curricula (ALSC, 2007; Koss, Johnson and Martinez, 2018). The covers of these books are typically marked by a conspicuous picture of the award. The primary goal for studying these books is to understand the representation of race, gender, and age in a set of books to which a large proportion of children in the U.S. are exposed.

Diversity Collection. The Diversity collection comprises books featured by ALSC that purposefully highlight the representation of excluded or marginalized identities.[8] These

[7] 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.

[8] 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.

books are also likely to be placed on "diversity lists" during events such as Black History Month or Women's History Month. Goals of studying these books include: one, to estimate a potential "upper bound" of representation 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 identities beyond the identity on which a given award focuses, particularly to assess the extent to which they account for intersectional experiences.

This collection includes books that have received 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,[9] 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 highlight African American writers, partly because no such writer had received either the Newbery or Caldecott awards as of that point. Other awards were created more recently, such as the South Asia Book Award in 2012.

We also create smaller collections of these awards that highlight the following specific identity areas: people of color, African American people, females, people with disabilities, and people who identify as lesbian, gay, bisexual, transgender, and/or queer (LGBTQIA+).

While different awards begin in different years, we do not limit the analysis to years in which all awards have books in the sample. The use of books persists over time, and it may be just as likely, if not more likely, for someone to select a book considered to be a "classic" (typically an older book) rather than to select a book more recently published. For example, picture books such as *The Snowy Day* (1962) and novels such as *Charlotte's Web* (1952) came into the collection before 1970, when the first Diversity collection book entered the sample, yet they remain an important part of the canon of children's literature and remain frequently used in libraries and classrooms.

We present summary statistics of the books in our sample, by collection, in Table 1. This includes information such as the number of years each award within a given collection has existed, as well as aggregate information about each collection, including the average length of the books (number of pages, number of words contained) and descriptive statistics of general measures of racial constructs, gender, and age.

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

II.B Why Focus on Award-winning Books?

An award from the ALSC, particularly one of the Mainstream awards, places books into the "canon" of children's literature and makes them a common feature in homes and libraries (Smith, 2013; Koss and Paciga, 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).

We show empirical evidence of the relationship between receipt of these awards and book popularity using data on three measures of book consumption: (1) library checkouts, (2) book purchases, and (3) internet searches. Each of these measures captures different – but not mutually exclusive – sets of consumer preferences.

Library Checkout Data. We draw from publicly available, book-level, daily checkout data from the Seattle Public Library system from 2005 to 2017. Public libraries aim to serve all members of their communities, regardless of socioeconomic status. For example, the mission of the Seattle Public Library "is to provide free and easy access to a vast array of knowledge, ideas, and information by supporting lifelong learning and a love of reading, so that everyone in our community is empowered, informed, and enriched."[10] Library usage is extremely common in the U.S. population - recent survey data estimates that roughly half of Americans have accessed a public library in the past 12 months.[11]

Book Purchase Data. We also draw data from the Numerator OmniPanel, a large panel data set. These data include information from over 1 billion shopping trips from over 44,000 retailers from between 2017 and 2020. Each purchase is matched to detailed demographic information on consumers, including the purchaser's demographics and number of children. We subset these data to only include purchases of children's books. The majority of the books in this panel were purchased on Amazon (88%), with Walmart (3%) and Target (3%) as the next most popular retailers. Wealthier people and people with more formal education are more likely to purchase children's books. We describe the demographic characteristics of these book purchasers in Appendix Table A5.

Google Trends Data. We use internet search data from Google Trends as a measure of general interest in the book awards found within our sample.[12] We limit our analysis to the following eight awards that have unique topic IDs in the Google Trends data: the Amelia

[10] Source: https://www.seattle.gov/documents/Departments/FinanceDepartment/0102Adopted/spl.pdf, accessed July 12, 2022.

[11] Source: https://www.pewresearch.org/internet/2015/09/15/who-uses-libraries-and-what-they-do-at-their-libraries, accessed July 11, 2022.

[12] 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.

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. Using these topic IDs, we can measure weekly search interest across the U.S. for each children's book award. 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 2016-12-04). 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.

We present three event studies that show average daily checkouts (Figure 2a), average daily purchases (Figure 2b), and average weekly search interest by collection (Figure 2c), centered around the time when awards are announced.[13] For Figures 2a and 2b, we disaggregate the data by three collections: (1) books recognized by Mainstream awards in that year; (2) books recognized by Diversity awards in that year; and (3) all other children's books.[14]

We see that library checkouts of books selected for Mainstream awards increase substantially after announcement of awards with a larger increase for books that won the award as opposed to being honored.[15] This persists for at least two years after the award announcement, during which average daily checkouts of the Mainstream collection plateau at a rate approximately four times that of the comparator groups. Library checkout rates of Diversity books do not increase similarly to those of Mainstream books.

In our data on book purchases, we see a sustained increase in purchases for books belonging to both the Mainstream and Diversity collections after the award announcements with a larger increase for Mainstream books. This finding is reflected in the analysis of publishers' book sales data by Cockcroft (2018), which documents large gains in sales – of similar or even larger magnitudes – after a book receives an award.

Finally, we find similar results for 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 particular spike in search interest following the announcement of the Mainstream awards.

This evidence suggests that Mainstream books have greater influence than other chil-

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

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

[15]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.

dren's books and children are more likely to be exposed to the messages in these books, consistent with previous qualitative assessments of their central role in children's literature.

III Images as Data

Images are not currently 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, and appears in stark contrast to the use of text as data, which has seen substantial attention from social science in the past decade (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 children who are not yet textually literate.

We introduce, develop, and apply tools for the automated analysis of the content of images. Specifically, these tools (1) identify pictured faces of characters and (2) classify their skin color, putative race, gender, and age. We depict this process in Figure 3a and refer to it as our "Image-to-Data Pipeline."

III.A Image Feature Classification: Face Detection

Our first step in converting images to data is to detect the face of each pictured character. However, machine-led face detection poses a set of complex problems. First, images in these books consist of both illustrations and photographs. This is particularly notable because the state-of-the-art face detection models were trained exclusively on photographs, leading these models to undercount faces in illustrations.[16] Second, these images contain both human and non-human characters. These 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., greyscale or sepia). Third, characters could be shown in different poses, such as facing the viewer, shown in profile, or facing away from the viewer.

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).[17] Our face detection model uses a manually-labeled data set of 5,403 illustrated faces from our sample that contain a wide variety of illustrated characters.[18] This process is described in depth in Szasz et al. (2022).

[16] 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 illustrated, as opposed to photographic.

[17] 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.

[18] We refer to our face detection model as FDAI (face detection using AutoML trained on illustrations). We refer to this data set as IllusFace 1.0 (Szasz et al., 2022).

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 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).[19] We describe these in further detail in the Methods Appendix.

III.B Image Feature Classification: Skin Color

Skin color is an important and distinct dimension of how humans categorize each other. Starting from the youngest ages, skin color is a salient feature of a character that a viewer is likely to process. Distinct from putative race, skin color is itself a site of historical and ongoing discrimination with clear impacts on health and in 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 may vary by observer and cultural context.

Our skin color classification method involves a three-step process: (1) "segmenting" the skin on the face (isolating the parts of the face which contain skin from other facial features), (2) extracting the dominant colors in the identified skin and collapsing them into a single representative skin color, and (3) constructing measures of skin color. Figure 3a illustrates this process.

Skin segmentation: Fully-Convolutional Conditional Random Field. We begin by isolating skin components from non-skin components of each character's face using a deep learning approach called Fully-Connected Convolutional Neural Network Continuous Conditional Random Field (FC-CNN CRF).[20] 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, which is a type of convolutional neural network (CNN) where the last fully-connected layer is substituted with a convolutional layer that can capture locations of the predicted labels. 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

[19]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.

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

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.[21] The resulting mask provides the skin that we then use to classify skin color.

Representative skin color: *k*-means clustering. 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.[22] 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.[23] After this conversion, we collapse the dominant skin colors into a single color by taking the weighted average of 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. The value of L* from this measure of each face's representative skin color in L*a*b* space provides our main skin color measure of interest: "perceptual tint."[24] This measure of perceptual tint reduces the dimensionality of skin color to a single value and interpret a given numerical change in the color values as a similar perceived change in color. We also separate the representative skin colors into three categories of skin color type: (1) polychromatic human skin colors (e.g., brown, beige), (2) monochromatic skin colors (e.g., greyscale), and (3) polychromatic non-typical skin colors (e.g., blue, green). See Methods Appendix Section D.B.3 for details on how we separate skin colors into these three types. In Figure 4, 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.[25] The x-axis indicates perceptual tint of each representative

[21]Conditional random field (CRF) is a class of statistical modeling using a probabilistic graphical model.

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

[23]It is important to convert colors from RGB space to L*a*b* space before averaging since L*a*b* color space – unlike RGB color space – is perceptually linear.

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

[25]We show these for each collection by decade for human skin colors (Appendix Figure B1), 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

skin color and the y-axis indicates vibrancy of each representative skin color.

III.C Image Feature Classification: Race, Gender, and Age

In order to classify putative race, gender, and age of detected faces in images, we train a multi-label classification model using Google's AutoML Vision platform. Due to the large amount of manually labeled data necessary to train deep learning models and because of the absence of public data sets using illustrations, we apply a transfer learning model trained using the UTKFace public data set to predict race, gender, and age for character faces. The UTKFace data set contains over 20,000 photographs of faces with manual labels of race, gender, and age (Zhang and Qi, 2017).[26]

Our model assigns probabilities that a detected face is of a given race, gender, or age, respectively. We classify a face with the identities to which the model gives the highest predicted probabilities.[27] We split the 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 other words, 9.4 percent of the images assigned a given race, gender, or age label will, in truth, not possess that trait (a false positive), while 11 percent of the images not assigned the label for that trait would, in truth, possess it (a false negative). The main limitation of this model is that it was trained on photographs, while the majority of the faces in our children's books are illustrated.[28]

---

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 faces.

[26] 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)).

[27] 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). 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., identities 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. Future work should incorporate the classification of fluid and nonbinary gender identities.

[28] In a random sample, 84.2 percent of detected faces were illustrated. 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.

IV Text as Data

In this section, we describe the tools we use to measure representation in the text of books. Social scientists 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 automated 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 3b, 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) which converts text into ASCII which then encodes each character to be recognized by the computer. We derive our textual measures of race, gender, and age by enumerating the attributes of features of these text data, which include token (single word) counts, the presence of famous people, and the first names of characters.

Text Analysis: Token Counts. We begin by generating counts of different "tokens" associated with race, gender, and age.[29]

Gender (Token Counts). To calculate gender representation in token counts, we calculate the number of tokens with a gender association. For example, female gendered tokens consist of pronouns and other gendered terms such as she, her, queen, aunt, and girl. Similar examples for male gendered tokens include he, husband, prince, and son.[30]

Age(-by-Gender) (Token Counts). To measure representation of age, we generate a list of "younger" (e.g., princess, boy) and "older" gendered tokens (e.g., queen, man).

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

[30] We also show how gender representation varies on three additional dimensions: one, whether the gendered identity is represented by individuals (singular) or groups (plural); two, whether the character is placed as the subject or object of a sentence; and three, by the age of the gendered word as described in the next paragraph. To analyze singular and plural representation separately, we separate gendered tokens into those referring to singular cases (e.g., daughter) and plural cases (e.g., daughters). To analyze whether the character is the subject or object of a sentence, we generate counts of the number of gendered pronouns that are capitalized versus lowercase, under the assumption that an individual who is the subject of a sentence is in a position of more active importance than the same character when used as the object and thus occupying a more passive role.

Nationality and Color (Token Counts). We measure race through two token proxies: (1) we calculate the proportion of all words that refer to nationalities (e.g., Mexican, Canadian), and (2) we calculate the proportion of all color tokens (e.g., black, white, blue).

Text Analysis: Named Entity Recognition. 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 figures. Exposure to salient examples of historical figures or celebrities from marginalized backgrounds can lead to meaningful change 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, including information on gender and birthplace (Yu et al., 2016).[31] 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.

Gender and Birthplace (Famous figures). We match the Pantheon 2.0 demographic data to each famous figure identified from the NER in our data.

Race (Famous figures). We manually code putative race for each identified person.[32]

Character first names. We then measure the gender of characters who are named but not identified as "famous" using the remaining entities identified from NER.

Gender (Character first names). To identify the names of characters who are not "famous," we extract the first name of each remaining named entity and estimate the probability that the character is female (or male) using data on the frequency of names by gender in the U.S. population from the Social Security Administration (SSA). For example, in the

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

[32] 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.

SSA data, the proportion of people named "Cameron" who identify as female is 9.16 percent. Therefore, if a character is named "Cameron," we assign a probability of 9.16 percent that the character is female. If the predicted probability that a character is female is greater than 50 percent, we classify that name as female. Otherwise, we classify the name as male.[33]

IV.A Text Analysis: All Gendered Words

We aggregate all words with a gender association (including predicted gender of character first names, gender of famous characters, and gendered tokens such as titles, pronouns, or specific gendered terms such as queen and husband) to generate a composite measure of gender representation in text. We refer to this aggregate measure as "gendered words."

V Measures of Representation: Race, Gender, Age

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 find the average probability that a face in a book belonging to the Mainstream collection is female-presenting, we first find the average probability that a face is female-presenting over all the faces in each book in the collection and then take the average across books. This approach ensures that our measures of race, gender, and age representation in each book are equally weighted within a collection. In other words, books with more faces do not receive more weight in the collection averages than books with fewer faces. We summarize our measures of representation in Table 2 and below.

Race Representation. We measure racial constructs through: (1) skin color classification of character faces, (2) race classification of character faces, (3) manually coded race of famous figures, (4) birthplace of famous characters, and (5) counts of words relating to nationalities and, separately, color word token counts.

Gender Representation. We measure representation of gender identity through: (1) gendered pronoun counts, (2) gendered token counts, (3) gender classifications of famous characters, (4) predicted gender of characters based on their first name, and (5) predicted gender of character faces. We construct an aggregate measure of gendered words by combining (1) – (4): gendered pronoun counts, gendered token counts, gender classifications of famous people mentioned in the text, and predicted gender of characters based on their first name. We refer to these aggregate counts as gendered words, or "words" for simplicity.

[33] 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).

Age Representation. We measure representation of age through: (1) age-by-gender word counts (such as father vs. son) and (2) predicted age of detected character faces.

Comparator Data. We draw from U.S. census data to explore how trends in representation track the population share of people by race, gender, and age.[34]

VI Results

In this section, we present our results characterizing the representation of race, gender, and age in the images and text of the books across collections and time.

VI.A Representation of the Construct of Race

We first present our measures of representation related to the social construct of race.

Skin color of faces. We begin by reporting the representation of skin color of characters pictured in images. In Figure 5 we show patterns in human skin color representation across collections overall and over time. Figure 5a 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 tones, on average, than those in the Mainstream collection.[35] A Kolmogorov-Smirnov test rejects the equality of the two distributions ($p < 0.001$) which suggests that the skin color distributions between the two collections are statistically distinct. Furthermore, the distribution of skin color tint in the Mainstream collection also has a much smaller variance than that of the Diversity collection (a test of the null hypothesis that the two variances are equal rejects equality with $p < 0.001$). This implies that there is a greater variety of skin color tint shown in the Diversity collection.

We then examine the proportion of character faces in each skin color tercile: darker, medium, or lighter. Figure 5b shows that, over time, the proportion of characters who have skin colors in the darker and medium skin color terciles is increasing relative to those in the lighter skin color tercile, both for the Mainstream and Diversity collections. Figure 5c 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

[34]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. Similarly, census data on the number of people who identify as "Multiracial" or "Other" are not available for all years in our sample.

[35]Appendix Figures C3 and C4 demonstrate that this result holds regardless of image color type.

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 tones. 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.[36]

Race of pictured character faces. We then examine the predicted race of characters in images. Figure 6 shows that the Mainstream collection is likely to show characters *within* a given race as lighter than their counterparts in the Diversity collection.[37] Figure 7 shows that pictured character faces are overwhelmingly classified as being White males or females.[38] This suggests that while children may see women in these books, they are seeing mostly White women which, may skew their perceptions of who belongs in a given space; specifically, that when women inhabit particular spaces in society, this is limited to White women, further excluding women of color. However, that same figure reveals the surprising result that, conditional on the person being classified as 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 the Female collection does a better job in addressing the power imbalances that come from the intersection of multiple sites of marginalization, at least in terms of including the presence of people from these groups.

Race of famous figures. We show the proportion of famous figures by race and gender in each collection in Figure 8.[39] 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.[40] 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

[36]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. However, Appendix Figure C4 shows that this does not change the patterns in skin color representation by collection over time.

[37]We see the same result for monochromatic faces in Appendix Figure C5.

[38]Appendix Figure B2 shows that most pictured characters are classified as being White. Appendix Figure B3 shows a substantial portion of pictured characters predicted to be female-presenting. We map share of faces by predicted race on their respective shares of the U.S. population in Appendix Figure B4. Appendix Figure B5 shows the proportion of characters in images and text by race and gender over time.

[39]Appendix Figure B6 shows the proportion of famous figures broken down by race alone.

[40]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).

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.[41]

We then explore how these trends in racial representation of famous people track the U.S. population share of different races using census data in Figure 9.[42] In the Mainstream collection, White people – particularly White males – have been overrepresented, whereas Black people and Latinx people have been historically underrepresented, relative to their U.S. population shares.

We also examine how representation of race of famous figures varies by gender. Figure 8 shows that the majority of famous characters in all collections are White males. The next most represented groups are White females (9-26 percent of famous people) and Black males (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 (as detailed above, 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 A1, we list the five most frequently mentioned famous people 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, two of whom are White (Abraham Lincoln, George Washington) and three of whom are Black (Martin Luther King Junior, Frederick Douglass, and Langston Hughes). Even in the Female collection, 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).[43]

[41] The U.S. census estimates that only 60 percent of the population is non-Latinx White (2019).

[42] Appendix Figure B7 shows a similar version of this graph with non-standard axes to more clearly view changes in groups with small population proportions.

[43] Appendix Tables A2 and A3 show this for the top five females and top five males, respectively, uniquely mentioned in each collection. Appendix Table A4 shows the most uniquely mentioned famous figure by collection for each decade.

Birthplace of famous figures. We next examine the representation of famous figures by their places of origin. Learning about real people from different parts of the world can expand a child's understanding of experiences beyond their own, and is another dimension of how humans categorize each other. We show the spatial distribution of birthplaces of famous figures mentioned in the Mainstream and Diversity collections in Figure 10, which presents a map of the world with a dot for each birthplace. This captures the representation of national and subnational identities presented to children. We see that Mainstream books primarily feature famous figures from Europe and the eastern portion of the United States; whereas Diversity books feature famous figures from across the world (Table 1).

However, when we explore how the birthplace of famous people presented in these books varies by gender, we see that males have more geographic heterogeneous representation in terms of birthplaces than females across both the Mainstream and Diversity collections (Appendix Figure B8). Females that are represented are far more likely to be from North America (primarily the United States) and Europe than males, who, particularly in the Diversity collection, come from many more parts of the world. This finding is consistent with a key insight from the study of intersectionality. These collections of books are the result of deliberate efforts to address one dimension of marginalization. Nonetheless, the (relative) exclusion of people at the intersection of multiple dimensions of marginalization – in this case, women born outside of the U.S. and Europe – persists.

Words related to nationality and color. We next look at the construct of race in text by examining the proportion of words related to nationalities (e.g., Kenyan, Canadian) and colors (e.g., black, white, blue) in Appendix Figure B9. These measures, while more straightforward than our other analyses, serve as a barometer for our other measures of race and help illustrate what a simpler approach to content analysis might have yielded. The collection of books that recognize African American experiences is much more likely to mention the words black and white. We then look at mentions of non-race colors such as red and blue as a falsification exercise because colors are more likely to appear in children's content. They are a negligible proportion of words overall and this is consistent over time.

VI.B Representation of Gender Identity

In this section, we show results for our measures of gender representation.

Gender in text. We first report the patterns for an aggregated measure of the textual representation of gender, which includes all counts of gendered tokens, the gender of the famous people mentioned in the text, and the gender classifications for character first names. In Figure 11, we present estimates of the book-level proportions of gendered words and characters which are female. The main pattern we observe is that, for all collections except those books specifically recognized for highlighting females, fewer female words are present than male words. Figure 11a shows that the proportion of female words in these collections is between 34 and 45 percent, as opposed to 56 percent in the Female collection. Figure 11b shows that this proportion is gradually increasing 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 11c, we show the distribution of the book level proportion of female words for each collection. We observe that the Mainstream collection is the most male-skewed of all the collections. The patterns show that in all collections except the Female collection, the central tendency of each distribution is skewed towards more male representation. However, we see that the Female collection – which we would expect to be more female-centered – is less female-skewed than the Mainstream collection is male-skewed.

In Appendix Figure B10, we show how these same distributions of the proportion of female words change over time. We find that books published in recent decades have a roughly unimodal distribution of gendered words centered around parity. In prior decades, the distributions are bimodal, with a more "hollowed out" distribution wherein books are more likely to contain a clear majority of either male- or female gendered words. This is the case for books in both the Mainstream and Diversity collections.

One dimension on which the representation of gender might vary is by type of gendered word. In particular, until recently, grammar rules dictated that male pronouns would be used as "gender-neutral" pronouns, which would then lead us to overstate the male representation in these books. However, the pattern holds when the analysis is restricted to each type of gendered word: pronouns, specific gendered tokens such as "girl," gender of character first names, or gender of famous people mentioned (Appendix Figure B11).

These patterns of discrepancy in the representation of gender in text are consistent across other measures of gender representation, such as whether characters are represented as individuals or groups of females vs. males (Appendix Figure B12) or if characters are represented as the subject (as opposed to the object) of a sentence (Appendix Figure B13).

A related but distinct parameter is the unique number of female and male famous figures mentioned in these books. The specific people who are named in a book transmit more implicit information to a child than generic tokens. By naming these individuals, they take on a greater significance to children. This can influence both child aspiration, as in the role model effects studied in Dee (2005) and Porter and Serra (2020), as well as social preferences and beliefs more generally (Plant et al., 2009; Alrababah et al., 2021). We show our collection-specific estimates of this parameter in Figure 8. On this dimension, inequality in representation of gender is much more severe. In the Mainstream collection, 90 percent of the famous figures uniquely mentioned across all books were male, for example. Even the Female collection is not uniquely more representative of women than of men. Appendix Figure B11 shows that not even one-third of the historical figures mentioned across all books were female. However, when famous females were present, they were mentioned more often. Furthermore, two collections (Diversity and LGBTQIA+) contain similar average proportions of characters who are female as contained in the Female collection.

Gender of pictured characters. Next, we describe the representation of gender in the images of these books. We show the proportion of faces in each collection identified as female in Figure 7 and Appendix Figure B3a. 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 B3b 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.[44,45]

Gender in images and text. We then compare representation of gender across images and text. In Figure 12, 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. In other words, females are more likely to be visualized (seen) than mentioned in the story (heard). This suggests 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.[46]

VI.C Representation of Age

Finally, we briefly discuss the representation of people by age in the images and text of our books. In Table 1, we report that adults are more likely to be present in both images

[44]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 B14.

[45]In Appendix Figure B15, we examine the representation of skin color by gender by showing the perceptual tint of faces, separated by their detected gender. Given some of our other findings, we might have expected to see differences by skin color among females and males pictured in images; in practice, however, we find no evidence of a significant difference between faces classified as females and faces classified as males in terms of the frequency of different skin tones represented.

[46]In Appendix Figure B16, we show these results for females by race in which we see Black and Latinx females less represented.

and text, with 3 to 19 percent of the characters being presented as children in the images and 17 to 32 percent of age-specific gendered words across all collections. In Appendix Figure B17a, 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.[47] We also see in Appendix Figure B4c that adults are overrepresented relative to their U.S. population share. This raises a question as to why adult experiences or depictions are privileged in books targeted to children.

In Appendix Figure B17b, 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. This also shows that, 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. For words specific to gendered adults, however, men are often more likely to appear. This discrepancy is largest in the Mainstream and African American collections, where adult men are approximately 60 percent of adult gendered mentions.

We also study how the representation of skin color varies by age. Figures 13a and 13b present plots of the distribution and percentage of skin color tints. These reveal that when children are depicted in images, they are more likely to be shown with lighter skin tone than adults, regardless of the collection in which the image appears.[48] 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 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. 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 (Sarna et al., 2003) suggests that there may be reason to expect the skin tone of adults to be lighter than that of children. Nonetheless, the pattern we find of children being represented with lighter skin than adults is consistent across

[47] 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.

[48] 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.

collections. While there are many potential interpretations of this pattern, a particularly concerning one is that brightness may be 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.

VII Economic and Social Factors Underlying Representation in Books

Our analysis of the representation in these prominent, award-winning children's books raises a natural question: what determines what publishers produce and what consumers consume? In this section, we draw upon research related to the supply of and demand for media and the economics of identity to help shed light on how the market might generate these patterns in the representation of race and gender within children's books. We first provide a conceptual framework of market forces, both supply and demand, to fix ideas about how these forces contribute to the levels of representation likely to appear in the market for children's books. This yields a set of empirical predictions about purchasing and publishing decisions, which we take to individual-level data on book purchases and demographics, alongside library-branch level data on library acquisitions and neighborhood demographic characteristics.

VII.A Conceptual Framework

Demand for representation in children's books. Existing research suggests that demand for representation in the images and text of books a consumer purchases is affected by the person's various identities in the following two ways.

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.

The second is through demand for "status quo" representation. This is informed by the theory that deviating from social norms is costly; Akerlof and Kranton (2000) call these costs "identity losses." In our case, this means 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 social norm or "status quo." 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. This is also consistent with a prediction from Bernheim (1994), specifically, that people may adapt their preferences to match their beliefs about what societal preferences are.

This framework yields two main predictions that we can take to data.

Prediction 1: Utility from homophily. Consumers will be more likely to purchase children's books with characters that match the identities of themselves or their children.

Prediction 2: Status quo bias. Consumers of all identities will be more likely to consume children's books containing characters with socially dominant identities than those containing characters with other identities.

Supply of representation in children's books. In a frictionless market with no startup costs, we would expect the supply of books to rise to meet the demands of consumers. However, given startup costs, search costs, and other market frictions, supply in many markets instead caters primarily to the preferences of the majority group (Waldfogel, 2003, 2007). When suppliers (e.g., publishers, librarians) cater to the majority, they necessarily limit the number of differentiated products available to consumers. This yields the following predictions for the supply of books:

Prediction 3: Tyranny of the market. Given the 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 – will choose to publish books which feature characters whose social identity matches the majority of children in the market, and fewer books containing characters who identify as a racial minority.

Prediction 4: Pricing-in representation. 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).

This framework abstracts from a few key aspects of these markets, such as supply on the extensive margin. We discuss the limitations of our framework in Section VII.C.

VII.B Evidence for these Predictions

We explore evidence for these predictions with book purchase data from the Numerator OmniPanel and data on the number of copies of each book in branch-level inventory from the Seattle Public Library, both of which contain data on consumer demographics. We match these to the book-level data on representation.

We first present results from our demand-side predictions, starting with Prediction 1, which we refer to as utility from homophily. Using consumer panel data from Numerator merged with our data on representation in award-winning children's books, we estimate the correlation between purchaser identity and the average representation in these books. In Table 3, Panel A, we show that purchasers who have a son purchase books with two percent fewer female names out of all gendered names and one percent fewer female words out of all gendered words on average, 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. In Table 3, Panel B, we see that male purchasers are more likely to purchase books with 1 to 2 percent fewer female words, names, and faces compared to female purchasers.

In Table 4, column 1, we show estimates for correlated consumption of skin tint in pictured faces in books. This analysis reveals additional evidence consistent with utility from homophily. Purchasers who identify as Black or as Latinx are more likely to buy books that contain darker faces, on average, than purchasers who identify as White. In Table 4, columns 2-5, we show similar results for mentions of famous individuals by putative race. We find positive and significant estimates indicating that Asian, Latinx, and Black purchasers each buy books that contain more mentions of famous people who share their own racial identity than all other groups. White people, in turn, are more likely than other groups to purchase books with predominantly White famous people.

We show further evidence for utility from homophily using inventory data from the Seattle Public Library system. In Table 5, 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.

We explore Prediction 2, which we refer to as status quo bias, by using two different data sources to show that there is higher demand for Mainstream books (which disproportionately represent males and White people, as shown earlier) than for Diversity books.

Using consumer panel data on children's book purchases from Numerator, we show in Table 6 that there is less demand for books from the Diversity collection than books from the Mainstream collection as measured by the average number of copies sold per book title. Between 2017 and 2020, we observed an average of 33 copies sold per book title in the Diversity collection and 83 copies sold per book title in the Mainstream collection. Furthermore, using Seattle Public Library data, we observe approximately four times as many checkouts per title for books in the Mainstream collection than we see for books in the Diversity collection.[49]

On the supply side, our empirical analysis of Prediction 3, which we refer to as the

[49]Library checkouts are a measure of demand.

tyranny of the market, references results presented earlier in our study. In our analysis in Section VI, we see that White famous figures are overrepresented in the text of Mainstream books relative to the share of White people in the U.S. population (c.f., Figure 9).

Using Seattle Public Library inventory data, we find that the libraries contain twice as many copies on average for books belonging to the Mainstream collection than books belonging to the Diversity collection (Table 6, Panel B).[50]

We show evidence in support of Prediction 4, which we refer to as pricing-in representation, in Table 6, Panel A. We see that books in the Diversity collection are 22 percent more expensive than those in the Mainstream collection.

VII.B.1 Perspectives of Suppliers of Children's Books

We complement this 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 into 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 main predictions from our conceptual framework, wherein books

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

with greater representation of non-dominant societal groups will be under-supplied by the market.

Further emphasizing this correspondence is the following data from the owner of a decades-old children's bookstore in a medium-sized midwestern city. This person lamented that until the mid-2010's, their requests to publishers for books representing people of color yielded the following response: "we don't sell those books because those books don't sell."

In response to this inadequate supply from major publishers, motivated booksellers such as this professional sought to establish connections with smaller publishers which specialized in such content. For example, this person highlighted work by the publishing house Lee and Low, which was founded in 1991 to address this shortcoming and has grown to be a well-known vendor of such content. Other booksellers or librarians without the bandwidth to make this extra effort, however, were left with few options to choose from should they have the desire to offer a diverse range of representations in the books they sold or lent.[51]

Finally, we conducted a series of semi-structured interviews with people involved in the committees responsible for these awards. The aim of these interviews was to better understand the process through which books were selected. Generally speaking, committee members are selected by election (for the Caldecott, roughly one half of committee members were voted upon) or appointment by the head of ALSC. The individuals on the committee then review books published in a given year, select a roster of potential nominees who meet a set of established criteria for receipt of award, and then at the end of the award period, meet to discuss their choices and those of others, and together come to an agreement. Two key themes came from these discussions: the first is that the criteria for selection are stable over time, despite the other secular changes in this period.[52] The second is that the composition of the award committees generally comprise a circulating group of librarians, booksellers, and educators. Specifically, it was a combination of those selected by the annually elected president and those elected by members of the ALSC. Furthermore, 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. In short, these awards, particularly those in the Mainstream, 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.

[51] 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.

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

VII.C Limitations to the Conceptual Framework

There is a series of other phenomena which may potentially contribute to the results analyzed in this section but which are beyond the reach of our current framework. The first 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.

We observe that the effect of utility from homophily is attenuated for book purchasers who are not White, in comparison to White purchasers. We can attribute this, in part, to status quo bias (Prediction 2). We acknowledge, however, that part of this pattern may also come from a simple case of the law of demand, since it is more costly to consume books that highlight characters with non-dominant identities.

Furthermore, these higher costs 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 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.

VII.D Historical Trends and Representation

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

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 B18) and the average percentage of gendered words (Appendix Figure B19), with a curated set of relevant salient historical events overlaid upon the graph with vertical black lines. This narrative exercise is, by its nature, hypothesis-generating rather than providing a confirmatory test of any hypothesized causal relationship. We observe that each of these major historical events is sometimes accompanied by a temporary change in representation. This is similar to the estimates from Jayadev and Johnson (2017) of how racial attitudes respond to economic downturns.

We then explore how representation of race and gender tracks with social attitudes using 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).[53] We see that attitudes towards Black individuals – as measured by the likelihood that a person "would vote for a qualified Black candidate for president" in the GSS – have trended more egalitarian, at the same time as the average perceptual tint of character faces has become darker (Appendix Figure B20a). Similarly, we see a positive trend of attitudes towards greater gender equality – as measured by people's acceptance of egalitarian gender roles in the GSS. We see a similar trend towards more equal inclusion of females and males in the text of books (Appendix Figure B20b).

These figures suggest that the trends in representation in children's books 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 can be 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.

VII.E Local Beliefs and Book Consumption

We showed evidence in Section VII.A that demand for representation in these children's books is related to the identities of the consumer. In this section, we provide evidence that demand for representation in children's books is also related to consumer beliefs.

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

[53] The GSS is a project of the independent research organization NORC at the University of Chicago, with principal funding from the National Science Foundation.

because it was the earliest survey year for which book purchase data were available. We merge these data with the number of books from the Mainstream and Diversity collections purchased, by zip code, between 2017 and 2020 using Numerator data.

In Table 7, 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),[54] 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 (a separate question).

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.

VIII 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"

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

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 III.A). 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 two other constructs of race – birthplace of famous figures and 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.

IX 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 can shape subconscious defaults. Historical and persistent inequality, both by race and gender and in other dimensions, can be either affirmed or challenged by what we teach children about the world. While many educators, librarians, parents, and school administrators wish to eliminate materials that have overt racial and gender bias and use content that promotes positive messages about all people, such efforts are driven by (only some) individuals, rather than by the systems necessary to ensure all students benefit from these changes. Per the adage "a picture is worth a thousand words," images in particular convey numerous messages. Social scientists are leaving data on the table by not systematically measuring the content of these messages implicitly and explicitly being sent to children through these visual depictions.

In this paper, we make four 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. Third, we characterize the economic forces that contribute to the levels of representation documented by our automated methods. Finally, we show that demand for representation in children's books, as demonstrated by local purchasing patterns, is related to consumers' identities and political beliefs.

We show suggestive evidence from public library checkout data that children may be four times as likely to be exposed to books from the Mainstream collection relative to other children's books. This illustrates the outsized influence that Newbery and Caldecott awardees may have and highlights the importance of understanding what messages children may be encountering in these books.

Our image analysis tools show that books selected to highlight people of color or females increasingly depict characters with darker skin tones over time. Books in the Mainstream collection, however, primarily depict characters with lighter skin tones compared to books in the other collections. Moreover, we see that children consistently have been more likely than adults to be depicted with light skin. Regardless of the reason, these findings show that lighter-skinned children see themselves represented more often in these books than darker-skinned children.

We compare the patterns we find in images to those we find in text. We see that females are more likely to be represented in images than in text over time, consistent with the maxim that women should "be seen but not heard." This suggests there may be symbolic inclusion in pictures without substantive inclusion in the actual story. Across all measures in our study, males, especially White males, are persistently more likely to be represented; this overrepresentation relative to their share in the U.S. population is surprising, particularly given substantial changes in female societal participation over time.

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 inadvertently reinforce existing inequality in 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.

These image-to-data tools allow 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).

We also hope that our findings, and the power of the tools we use to generate them, will motivate and inform subsequent research on the causes and consequences of representation in children's books. Our 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, 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 et al., 2022; Arold, 2022), for example, linking exposure to different levels of representation with formation of beliefs, preferences, and societal outcomes. These same measurements also could be used to better understand the objective functions of different publishers, and how these change over time and in response to societal events.

This also demonstrates how our tools can be used by another key set of stakeholders: the practitioners, policymakers, and parents looking for information to guide their choice of which books or other curricular materials to include in their classrooms, libraries, and homes. 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. To achieve any progress toward such goals, practitioners and publishers require mechanisms to systematically measure and compare the amount and type of representation in the content they consider for inclusion in curriculum or even for prospective consideration for publication. Our tools provide a starting point for this work.

To explore the economic forces that underlie these patterns, we draw a series of predictions from prior theoretical and empirical work on the supply of and demand for media. We take these to purchase-level data on over 1.5 million children's book purchases alongside data on book checkouts from a major public library system. We show that aggregate book sale volume and book prices reflect our predictions for supply-side behavior, and individual purchase patterns and the volume of checkouts of books with different levels of representation reflect our predictions for demand-side behavior. We posit that these forces contribute to the persistent overrepresentation of historically dominant identities that we reveal in our second contribution.

To understand how book consumption relates to local consumer beliefs, we map book purchase data at the local level to surveys of political beliefs. We document patterns that suggest that the demand for representation may contribute to the propagation of beliefs about race and gender across generations, through the messages contained in books.

Inequality in representation, particularly in the materials we use to teach children, is a systemic problem which requires a systemic solution. Our 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. Separately, these tools can help catalyze a wide range of scholarship to systematically use printed content – images, as well as text – as primary source data. This work could, for example, describe patterns of representation in other bodies of content and, subsequently, explore how variation in representation shapes human beliefs, behavior, and outcomes. Finally, these methods can be applied to the study of other text and visual media, from print-based and online news to television and film. By providing research that expands our understanding about the diversity in content, we can help to contribute to work that aims to overcome the structural inequality that pervades society and our daily lives.

References

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.

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, Larissa Zierow, et al. 2022. "Religious education in school affects students' lives in the long run." Vol. 23, 40–43, ifo Institute-Leibniz Institute for Economic Research at the University of ....

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

Banks, Taunya Lovell. 1999. "Colorism: A darker shade of pale." UCLA L. Rev., 47: 1705.

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.

Brooks-Gunn, Jeanne, Greg J Duncan, Pamela Kato Klebanov, and Naomi Sealand. 1993. "Do neighborhoods influence child and adolescent development?" American Journal of Sociology, 99(2): 353–395.

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.

Burton, Linda M, Eduardo Bonilla-Silva, Victor Ray, Rose Buckelew, and Elizabeth Hordge Freeman. 2010. "Critical race theories, colorism, and the decade's research on families of color." Journal of Marriage and Family, 72(3): 440–459.

Caliskan, Aylin, Joanna J Bryson, and Arvind Narayanan. 2017. "Semantics derived automatically from language corpora contain human-like biases." Science, 356(6334).

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, and Lawrence F Katz. 2016. "The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment." American Economic Review, 106(4): 855–902.

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.

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.

Cutler, David M, and Edward L Glaeser. 1997. "Are ghettos good or bad?" The Quarterly Journal of Economics, 112(3): 827–872.

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. 2004. "Teachers, race, and student achievement in a randomized experiment." Review of Economics and Statistics, 86(1): 195–210.

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

Dixon, Angela R, and Edward E Telles. 2017. "Skin color and colorism: Global research, concepts, and measurement." Annual Review of Sociology, 43: 405–424.

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.

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.

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.

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.

Hunter, Margaret. 2007. "The persistent problem of colorism: Skin tone, status, and inequality." Sociology Compass, 1(1): 237–254.

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.

Jencks, Christopher, and Susan E Mayer. 1990. "The social consequences of growing up in a poor neighborhood." Inner-city Poverty in the United States, 111: 186.

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.

Keith, Verna M, and Carla R Monroe. 2016. "Histories of colorism and implications for education." Theory Into Practice, 55(1): 4–10.

Knowles, Elizabeth, Liz Knowles, and Martha Smith. 1997. The reading connection: Bringing parents, teachers, and librarians together. Libraries Unlimited.

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.

Leventhal, Tama, and Jeanne Brooks-Gunn. 2000. "The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes." Psychological Bulletin, 126(2): 309.

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.

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.

MacMaster, Neil. 2001. Racism in Europe: 1870-2000. New York, NY:Palgrave.

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, with a new preface. 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.

Nenkova, Ani, and Kathleen McKeown. 2012. "A survey of text summarization techniques." In Mining text data. 43–76. Springer.

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

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).

Ouyang, Jessica, and Kathy McKeown. 2014. "Towards Automatic Detection of Narrative Structure." 4624–4631.

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.

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.

Recalde, Maria, and Lise Vesterlund. 2020. "Gender Differences in Negotiation and Policy for Improvement." National Bureau of Economic Research Working Paper.

Riley, Emma. 2017. "Increasing students' aspirations: The impact of Queen of Katwe on students' educational attainment." In CSAE Working Paper WPS/2017-13.

Sampson, Robert J, Jeffrey D Morenoff, and Thomas Gannon-Rowley. 2002. "Assessing "neighborhood effects": Social processes and new directions in research." Annual Review of Sociology, 28(1): 443–478.

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.

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.

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

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.

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.

X Exhibits: Tables and Figures

Table 1. Summary Statistics

	Mainstream	Diversity	People of Color	African American	Ability	Female	LGBTQIA+
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: Racial Constructs							
Perceptual Skin Tint of All Faces	55	44	44	41	46	34	47
% 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 Born in Africa	0%	2%	2%	2%	0%	16%	1%
% Famous People Born in Americas	56%	69%	69%	86%	66%	53%	70%
% Famous People Born in Asia	5%	6%	7%	2%	3%	8%	5%
% Famous People Born in Europe	39%	22%	21%	11%	30%	23%	24%
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%
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. Measures of Representation

Measure	Image	Text
Race	• Skin color • Predicted race of face	• Race of famous figures • Birthplace of famous figures • Color token counts • Nationality token counts
Gender	• Predicted gender of face • Probability of gender of face	• Pronoun counts • Gendered token counts • Gender of famous figures • Predicted gender of first names
Age	• Predicted age of face	• Age-by-gender token counts

Note: In this table, we list the different variables we use to measure race, gender, and age in the faces in the images and, separately, the text in children's books.

Table 3. Gender Representation in Book Content by Purchaser Identities

	Dependent variable: Percent of Female		
	Words (1)	Names (2)	Faces (3)
Panel A: Gender of Purchaser's Children			
Purchaser Has a Son	−0.012	−0.020**	0.003
	(0.008)	(0.009)	(0.010)
Purchaser Has a Daughter	0.032***	0.019**	−0.002
	(0.008)	(0.009)	(0.010)
Constant	0.385***	0.363***	0.415***
(Baseline Group: No Children)	(0.003)	(0.003)	(0.004)
Observations	9,658	9,419	6,680
Adjusted R^2	0.0020	0.0010	−0.0003
Panel B: Gender of Purchaser			
Male	−0.015***	−0.017***	−0.019***
	(0.005)	(0.006)	(0.006)
Other	−0.006	−0.038**	0.024
	(0.016)	(0.019)	(0.021)
Constant	0.388***	0.370***	0.432***
(Baseline Group: Female)	(0.002)	(0.002)	(0.002)
Observations	28,645	28,120	18,737
Adjusted R^2	0.0003	0.0004	0.0004

Note: In this table, we regress indicator variables for whether the purchaser has a son or daughter (Panel A) and purchaser gender (Panel B) on three 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 in addition to other gendered words such as daughter or uncle. 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. We get book level purchasing data from the Numerator OmniPanel which contains data on purchases made between 2017 and 2020 and merge it with our curated data on representation in award-winning children's books. We necessarily subset purchasing data to include purchases of award-winning children's books which we have digitized that contain at least one gendered word/name/face. *p<0.1; **p<0.05; ***p<0.01

Table 4. Race and Skin Color Representation in Book Content by Purchaser Identities

	Dependent variable:				
	Average	Percent of Famous Mentions by Race			
	Skin Tint	*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: In this table 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 between 2017 and 2020 and merge it with our curated data on representation in award-winning children's books. We necessarily 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 5. 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 (CSA). Each community area is manually matched to its closest Seattle Public Library branch. Each Seattle Public Library branch is matched to at least one CSA. Column 1 shows that the number of books which won a Mainstream award available in the library branch closest to a given CSA is increasing in the proportion of the CSA population that is White, non-Hispanic. Column 2 shows that this relationship is decreasing for books which won a Diversity award. Columns 3 and 4 show these results are robust to including measures of household income for a given CSA. 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 6. Readership by Collection

Panel A: Average Price and Copies Purchased In Numerator OmniPanel

Collection	Number of Copies Sold	Mean Book Price	Number of Unique Titles	Mean Copies Sold Per Title
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 between 2017 and 2020.

Panel B: Seattle Public Library Inventory and Checkouts

Collection	Number of Checkouts	Mean Checkouts Per Title	Number of Unique Titles	Mean Library Copies Per Title
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 7. Local Beliefs and Children's Book Purchases within Zip Codes

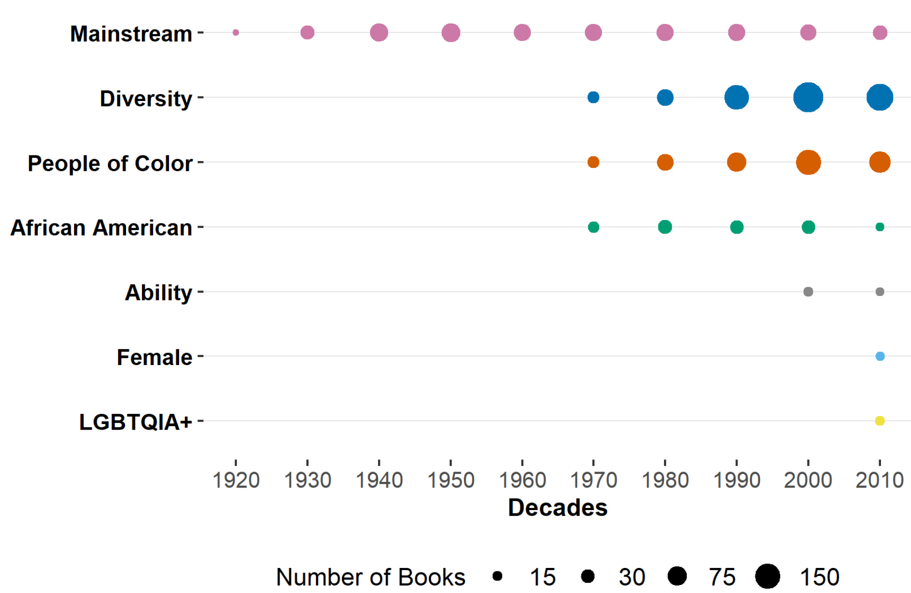
	Dependent variable:			
	% of Respondents who think the U.S. government should		% of Respondents who somewhat or strongly agree	
	Identify and deport undocumented immigrants	Withhold federal funds from localities that do not follow federal immigration laws	White people in the U.S. have certain advantages 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 won 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 and 2020 Numerator OmniPanel data set. Variables containing percentages are scaled so that potential values range from $0 - 100$. *p<0.1; **p<0.05; ***p<0.01

Figure 1. Books in the Sample

(a) Corpora in Sample, by Collection

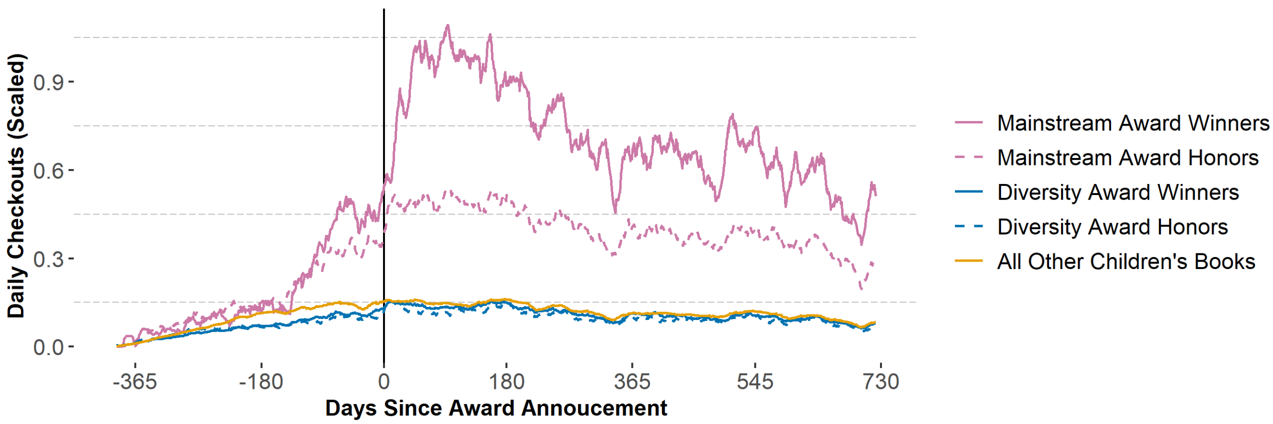
(b) Collection Sample Size, by Decade



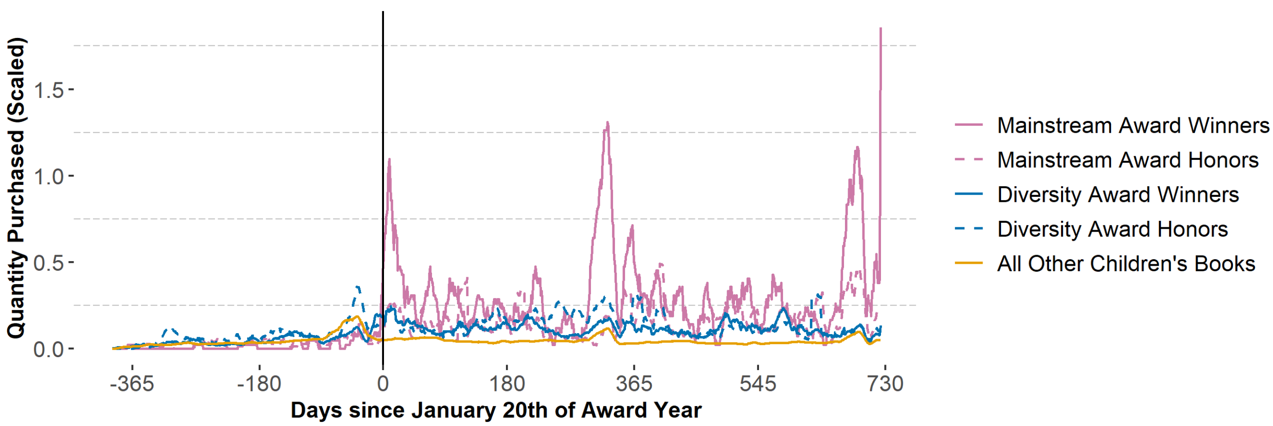
Note: This figure shows the main sources of data we use for our analysis. In Panel A, we list the book awards in our sample, along with the collections into which we group them in our analysis. In Panel B, we show our sample size in each collection, over time.

Figure 2. Children's Book Readership Centered Around Award Announcements

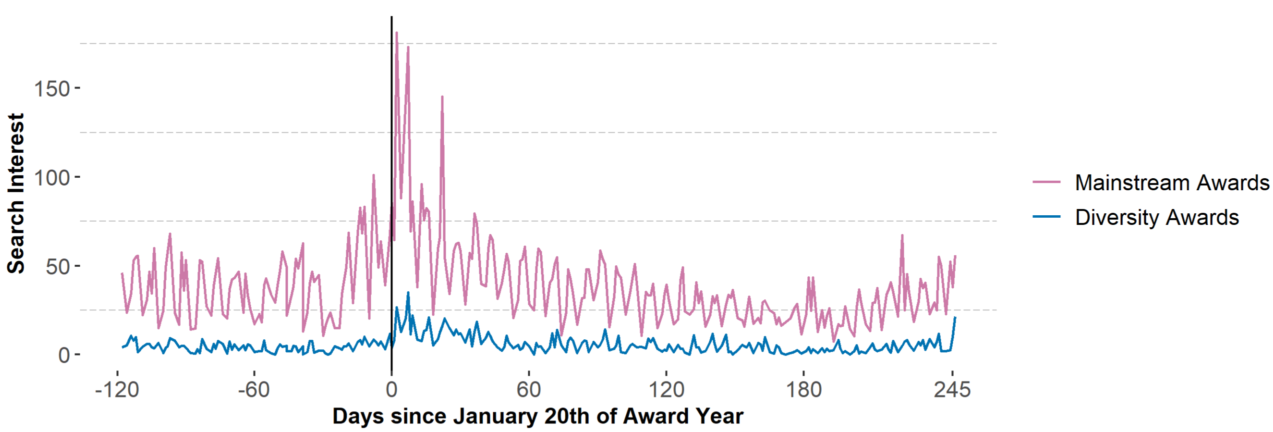
(a) Library Checkouts



(b) Purchases



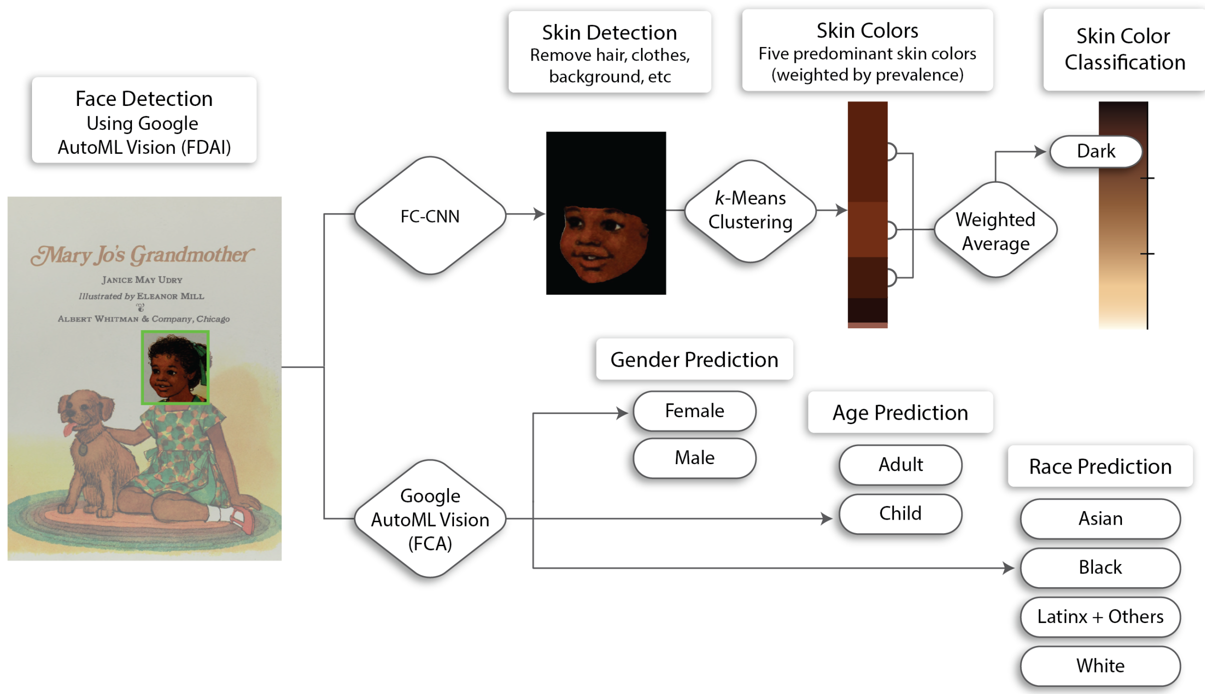
(c) Search Interest



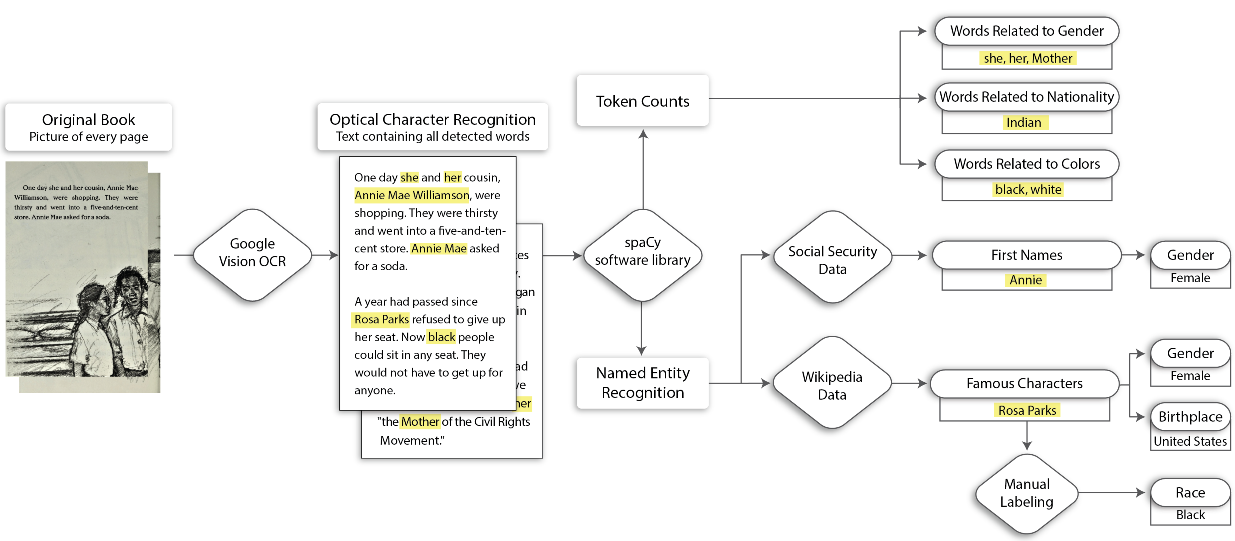
Note: Panel A shows average daily checkouts of children's library books between 2005-2017 from the Seattle Public Library. Panel B shows average daily children's book purchases between 2017 and 2020 from the Numerator OmniPanel. Both panels are disaggregated by whether the book was recognized by a Mainstream award, a Diversity award, or a children's book 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 using a 14-day moving average. Panel C shows average weekly search interest in the U.S. between 2017 and 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.B. All panels are centered around the time of award announcements each year.

Figure 3. Converting Images and Text into Data

(a) Image-to-Data Pipeline



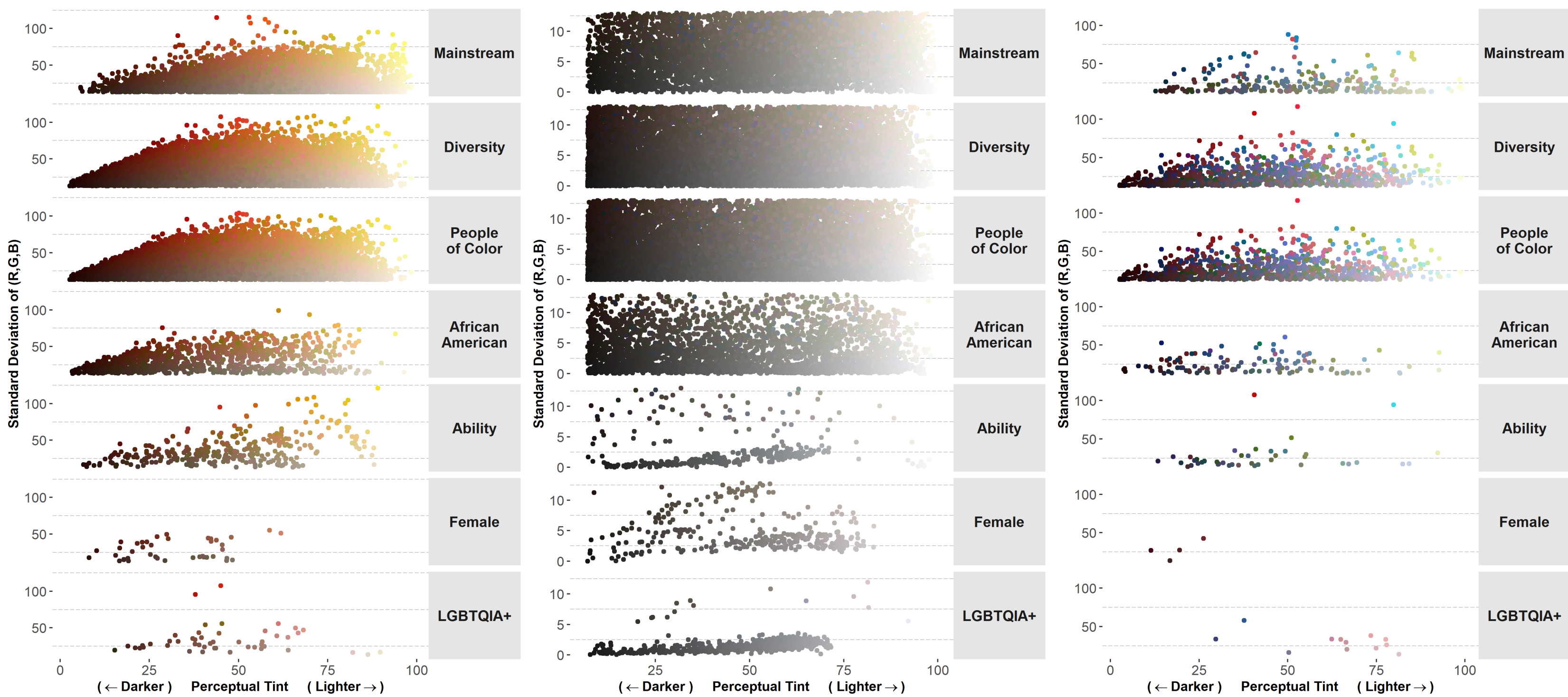
(b) Text-to-Data Pipeline



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 and classify skin color, race, gender, and age. In Panel B, we show how we extract and isolate various dimensions of text, such as names of famous people or words related to gender.

Figure 4. Skin Color Data, by Color Type

(a) Human Skin Colors (b) Monochromatic (c) Non-Typical Skin Colors

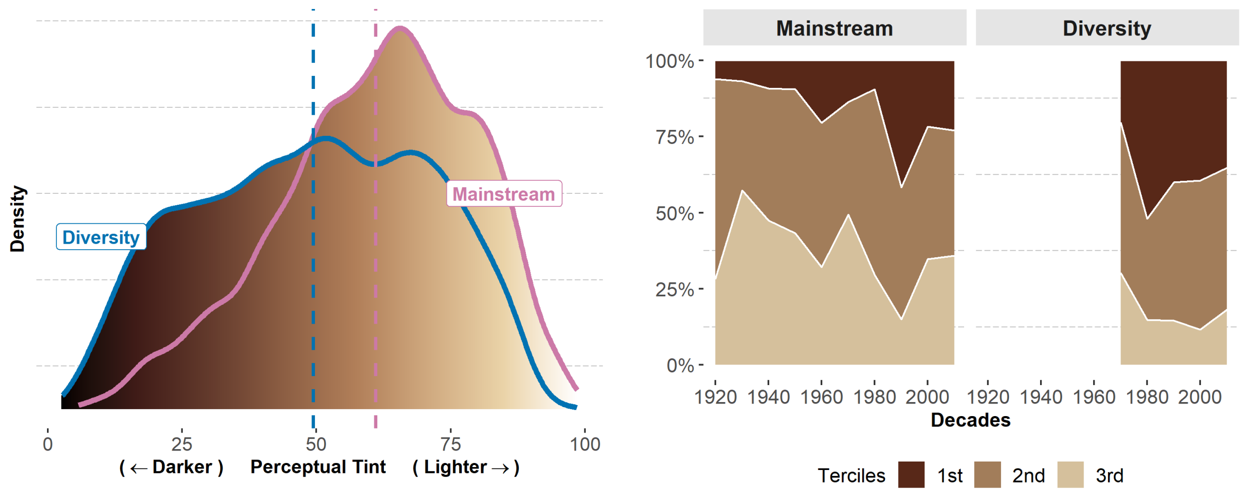


Note: This figure shows the representative skin colors of the individual faces we detect in the images found in the books from each collection. We show these by the three color "types" present in these images: human skin colors (polychromatic skin colors where R ≥ G ≥ B), monochromatic skin colors (e.g., 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.

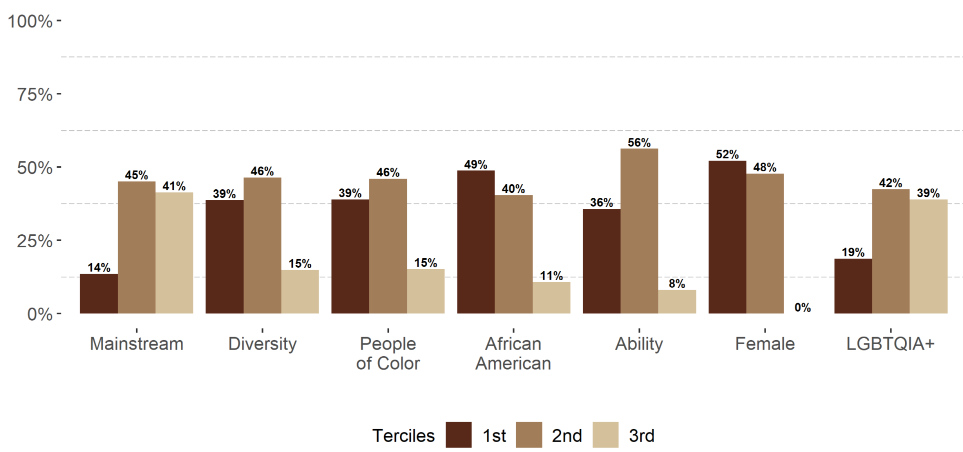
Figure 5. Skin Colors in Faces, by Collection: Human Skin Colors

(a) Distribution of Skin Colors

(b) Mean Proportion in Each Tercile, Over Time

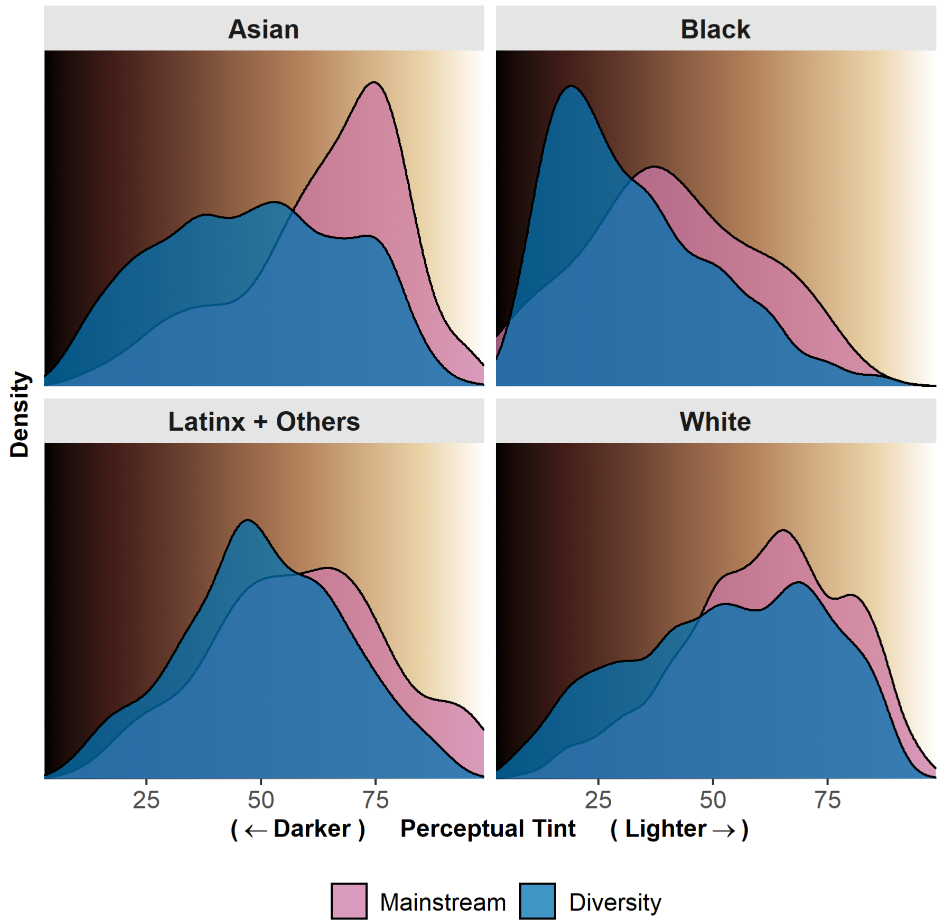


(c) Mean Proportion in Each Tercile, All Collections



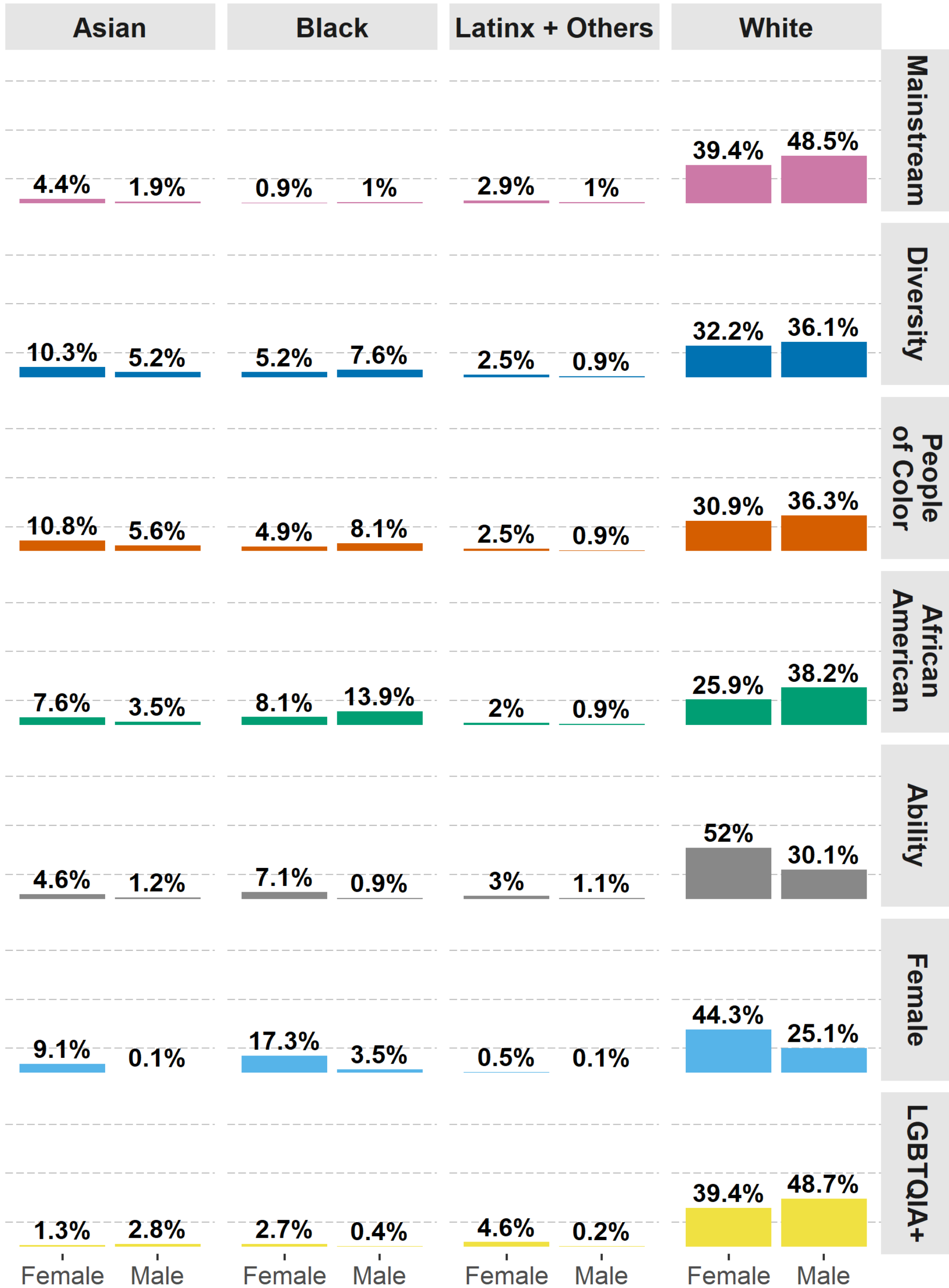
Note: This figure shows our analysis of the representative skin colors of the individual faces we detect in the images found in the books we analyze, focusing on faces considered to be human skin colors (polychromatic skin colors where R $\geq$ G $\geq$ B). Panel A shows the distribution of skin color 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 III.

Figure 6. Skin Color by Predicted Race of Pictured Characters



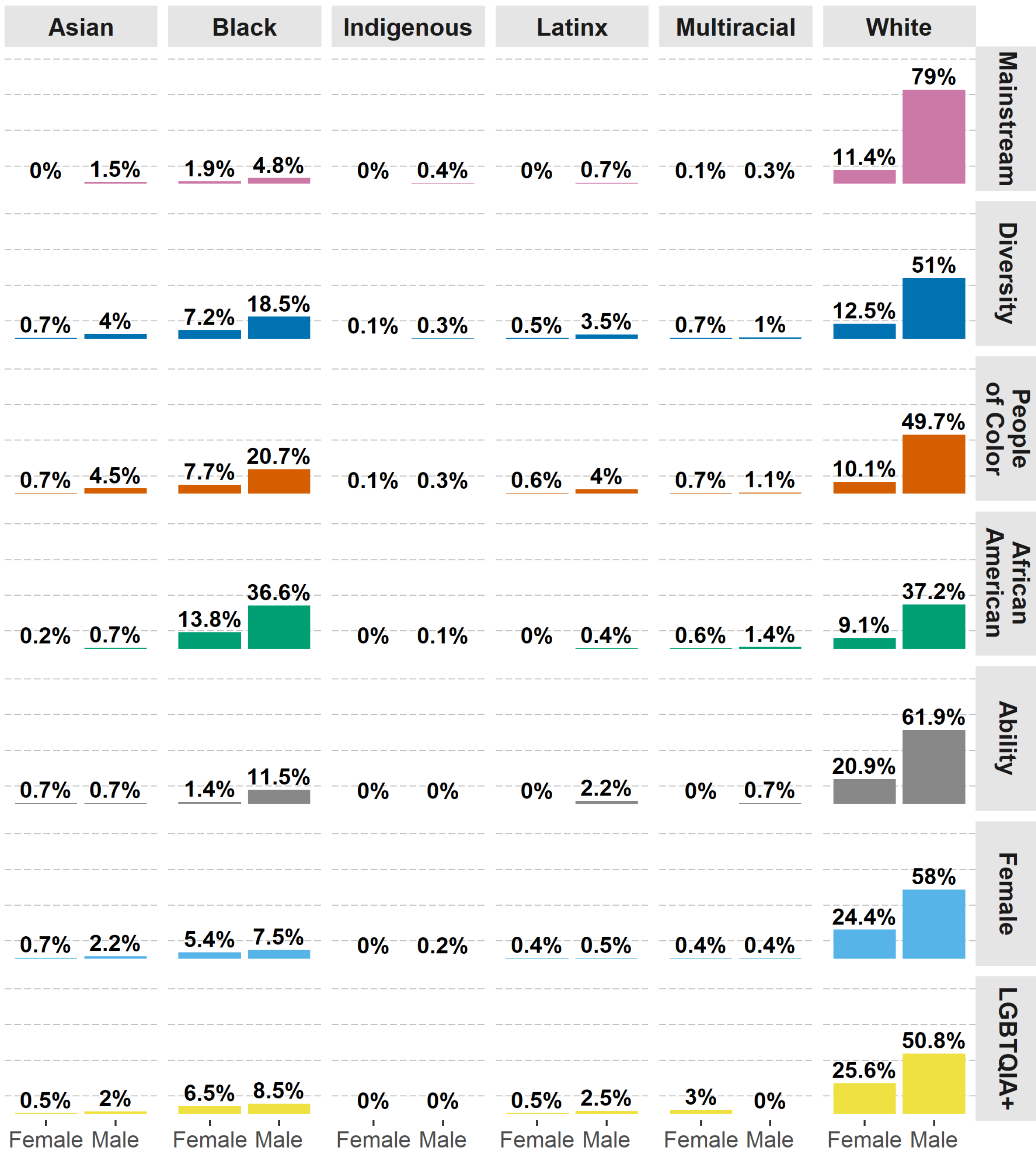
Note: This figure shows the distribution of skin color tint by predicted race of the detected faces in the Mainstream and Diversity collections. Skin tint is determined by the L* value of a face's representative skin color in L*a*b* space. We extract a face's representative skin color using methods described in Section III.B. Race was classified by our trained AutoML model as described in Section III.C.

Figure 7. Race and Gender Predictions of Pictured Characters



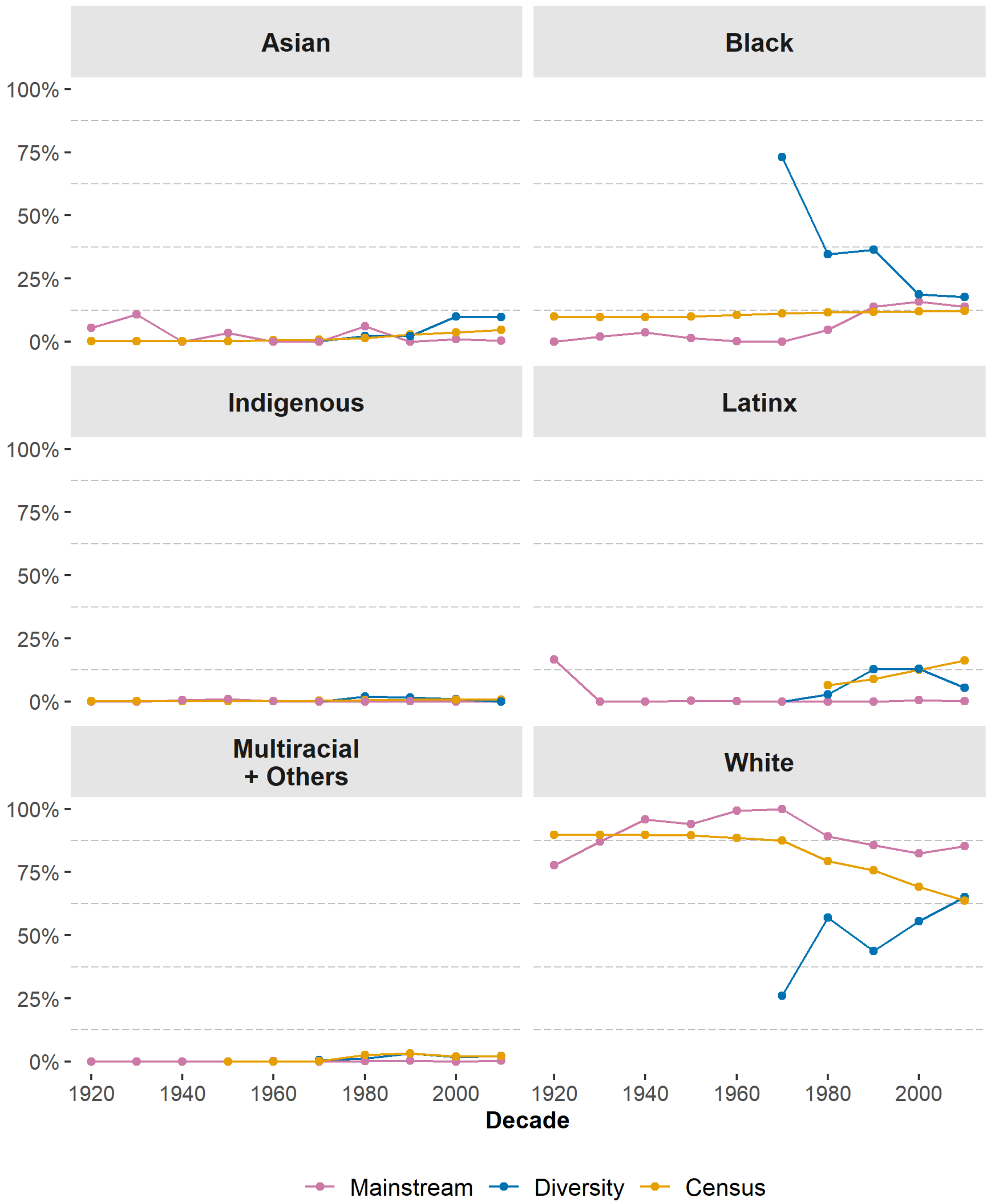
Note: In this figure, we show the proportion of detected faces in all collections by race and gender predictions. Race and gender were classified by our trained AutoML model as described in Section III.C. See Appendix Figure B2 for the same figure broken down by race alone and not both race and gender.

Figure 8. Race and Gender Classifications of Famous Figures in the Text



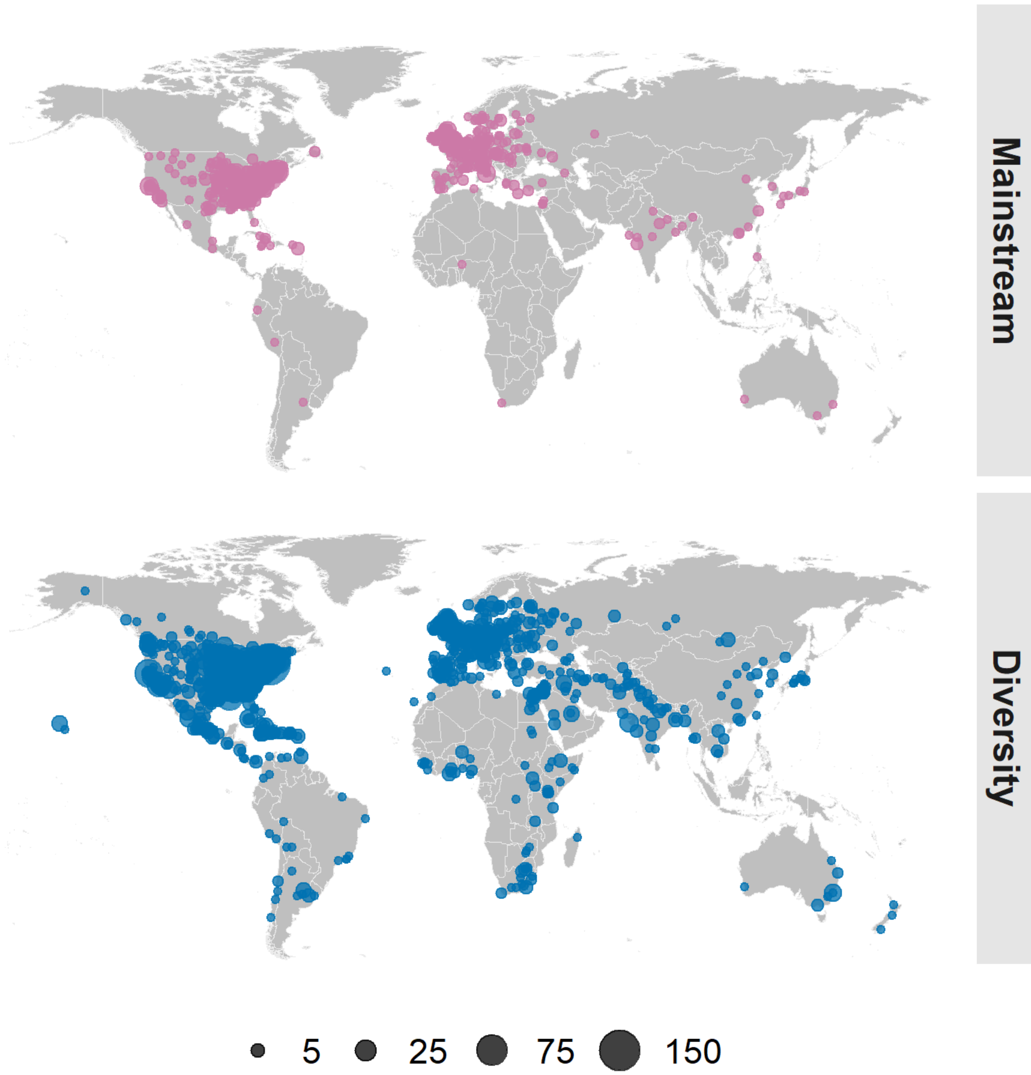
Note: In this figure, we count the number of famous people mentioned at least once in a given book and sum over all books in a collection. We then show the percentage breakdown of these famous people by race and gender. For example, if Aretha Franklin was mentioned at least once in two separate books 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 D.C.3. 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 B6 for the same figure broken down by race alone and not both race and gender.

Figure 9. Share of U.S. Population and Famous People in the Text, by Race/Ethnicity



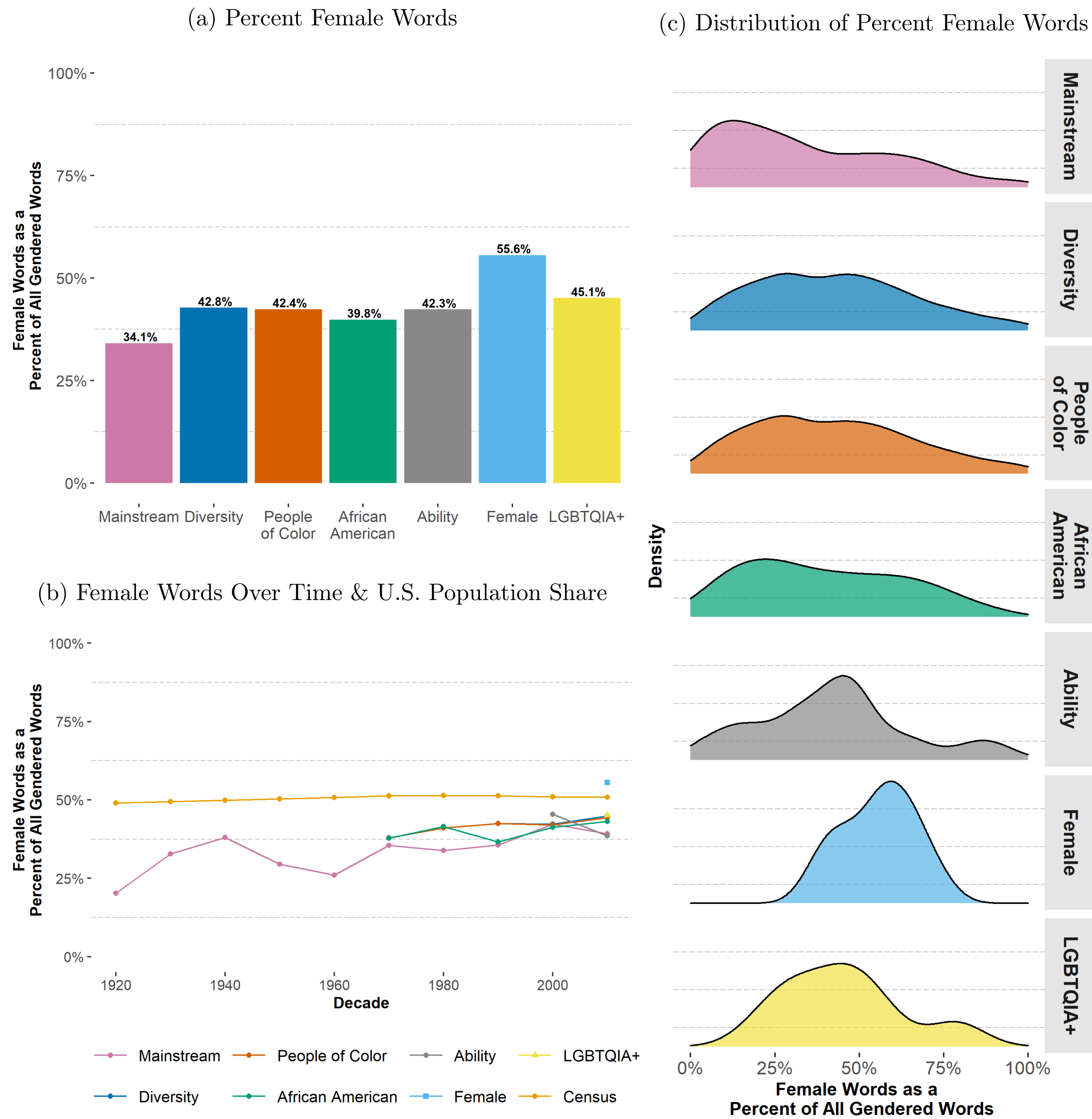
Note: In this figure, we find the percent breakdown of famous people mentioned in a given book by race/ethnicity. For example, if Aretha Franklin was mentioned 3 times in a book and Jimmy Carter is mentioned 2 times, then 60% of the mentions of famous people in that book would be Black. We then show 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 D.C.3. 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 B7 for a similar version of this graph with non-standard axes to better see changes in groups with small population proportions.

Figure 10. Birthplace of Famous Figures



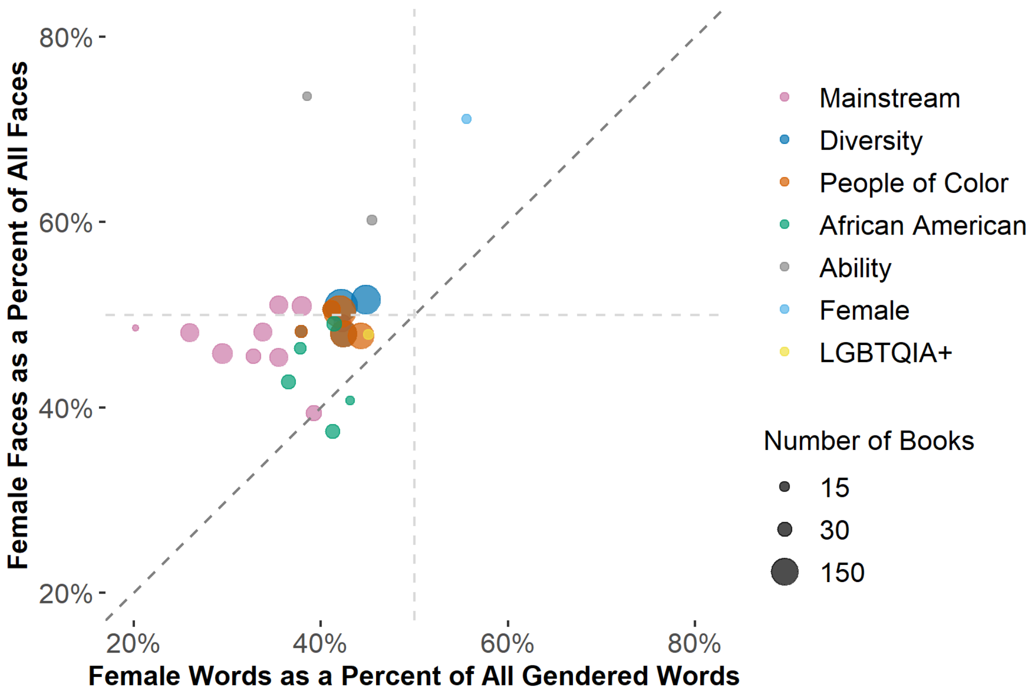
Note: In this figure, we show two maps in which we plot the distribution of the place of birth of the famous people in our books: one for the Mainstream collection and one for the Diversity collection. We identify birthplace using a model trained on text from Wikipedia biographies collected by Pantheon (Yu et al., 2016). If the city/town they were born in was unavailable, we use birth country. Size of dots correspond to the number of famous characters born in a given location that are mentioned at least once in a given book and then aggregated across all books 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 Black. We show this broken down by gender in Appendix Figure B8.

Figure 11. 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 the average percent by collection. Panel B shows how this average varies by decade. Panel C shows the distribution over all books in a collection. In this case, 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 tokens (e.g., queen, nephew). We list the pre-specified gendered tokens in the Data Appendix.

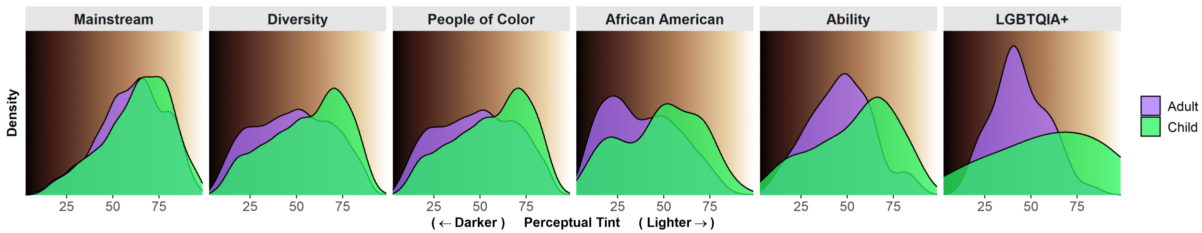
Figure 12. Percent Female Faces vs. Words: Women Should be Seen More Than Heard?



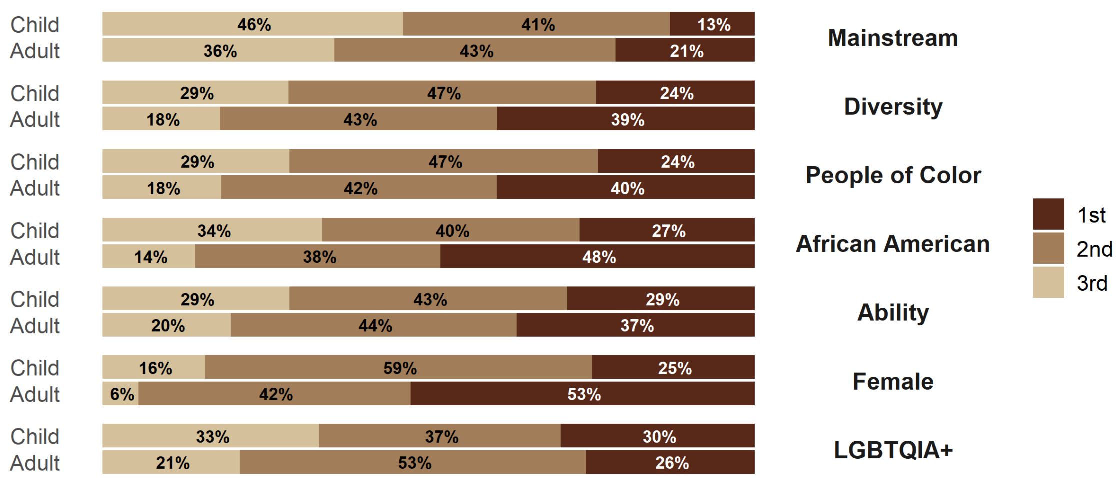
Note: In this figure, we plot collection-by-decade averages of female representation in images (on the y-axis) and female representation in text (on the x-axis). This enables a comparison between the proportion of females represented in the images vs. the text of the children's books in our sample.

Figure 13. Children Are Consistently Depicted with Lighter Skin than Adults

(a) Skin Color Distribution by Age



(b) Proportion of Faces in Skin Color Terciles, by Age



Note: In this figure, we show analysis of how the representation of skin color varies with the predicted age of the person being represented. In Panel A, we show the distribution of skin tint values of representative skin color of detected faces in the Mainstream and Diversity collections by the classified age (adult vs. child) of the face. In Panel B, 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) described in Section III.A. 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. We extract a face's representative skin color using methods described in Section III.B. These figures show the results for images that have identified human skin colors (polychromatic skin colors where R $\geq$ G $\geq$ B). There are not enough detected faces which are classified as children in the Female collection to infer a skin tint distribution, so we omit the Female collection from Panel A.

Appendices

A Appendix Tables

Table A1. 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 A2. 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 A3. 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 A4. 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.

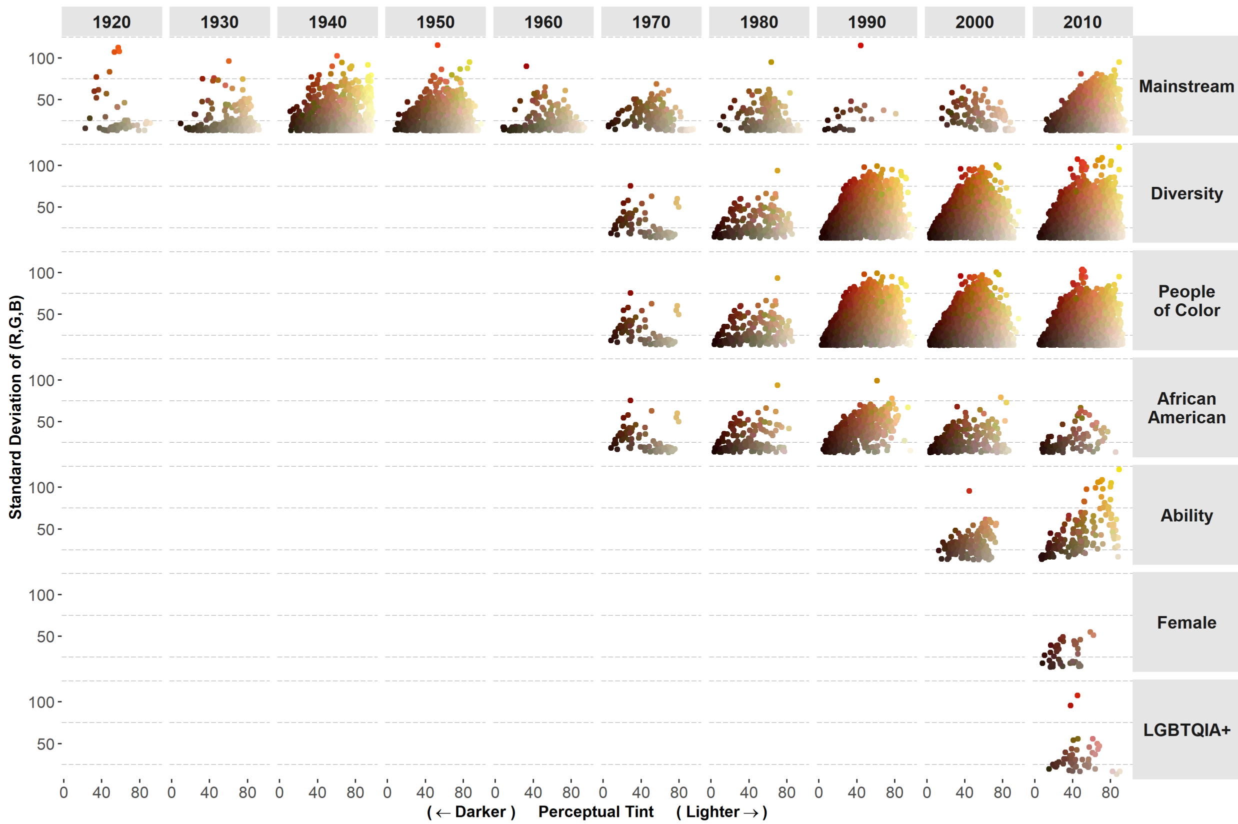
Table A5. Summary Statistics for Children's Book Purchases in Numerator Data

	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
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 describes purchaser demographics for shopping trips in Numerator OmniPanel data between 2017 and 2020. The first two columns show the sample size and mean of purchaser demographic variables over all shopping trips that include the purchase of a children's book. The last two columns show the sample size and mean of purchaser demographic variables over all shopping trips that include children's books which won one of the awards in our sample.

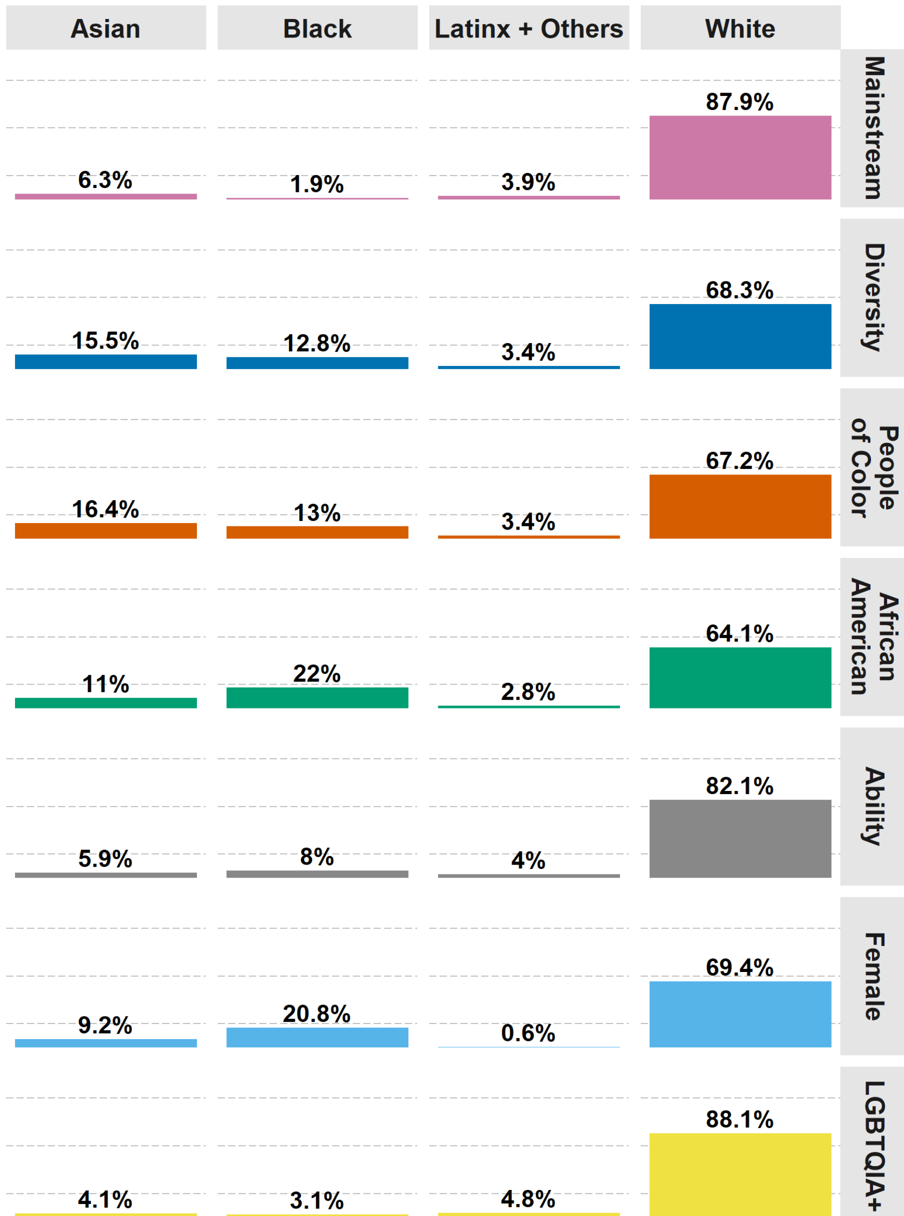
B Appendix Figures

Figure B1. 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 $\geq$ G $\geq$ B) in each collection-by-decade. As described in Section III, 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 III.B.

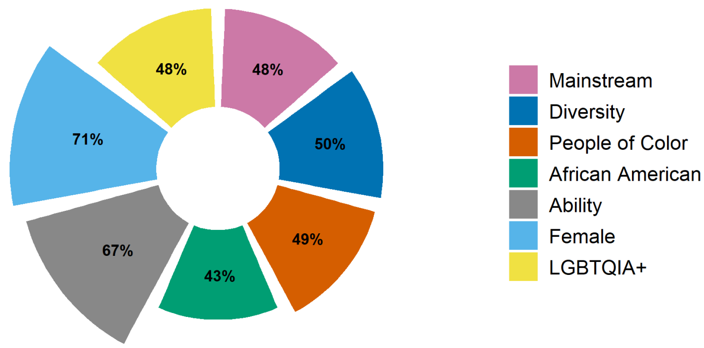
Figure B2. Most Pictured Characters Are Classified as White



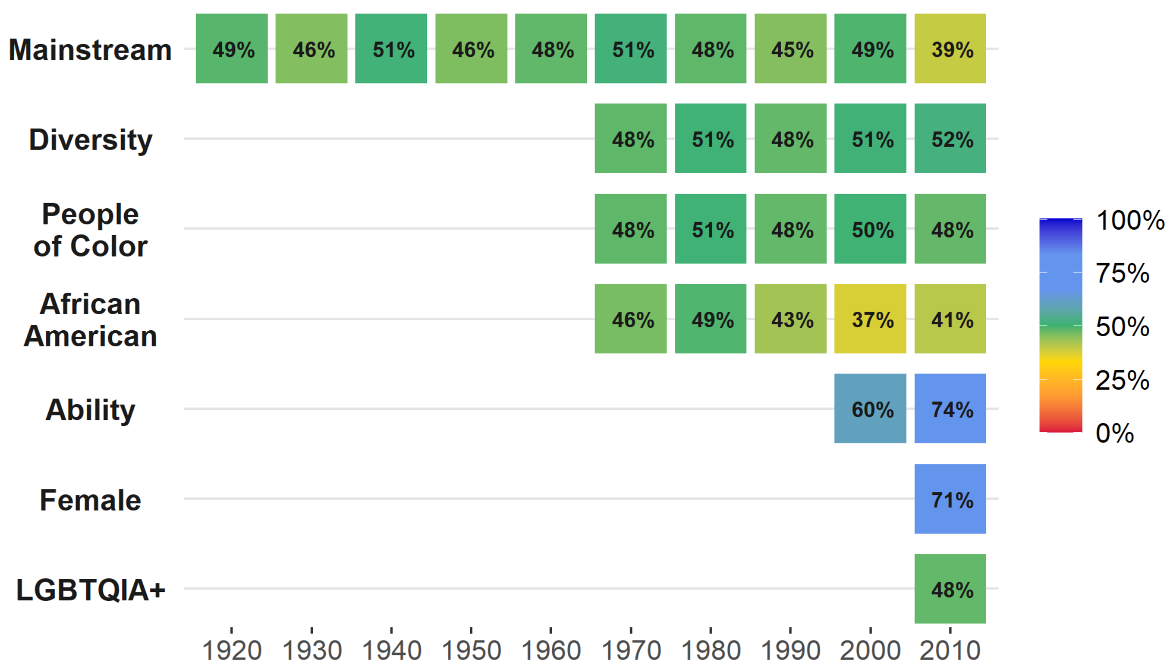
Note: In this figure, we show the proportion of faces in a book which our model labels as a given race averaged over all books in a collection. We detect faces using our face detection model (FDAI) described in Section III.A. Within these faces, we classify age and gender using an AutoML algorithm we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

Figure B3. Proportion of Detected Faces Which Are Female-Presenting

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



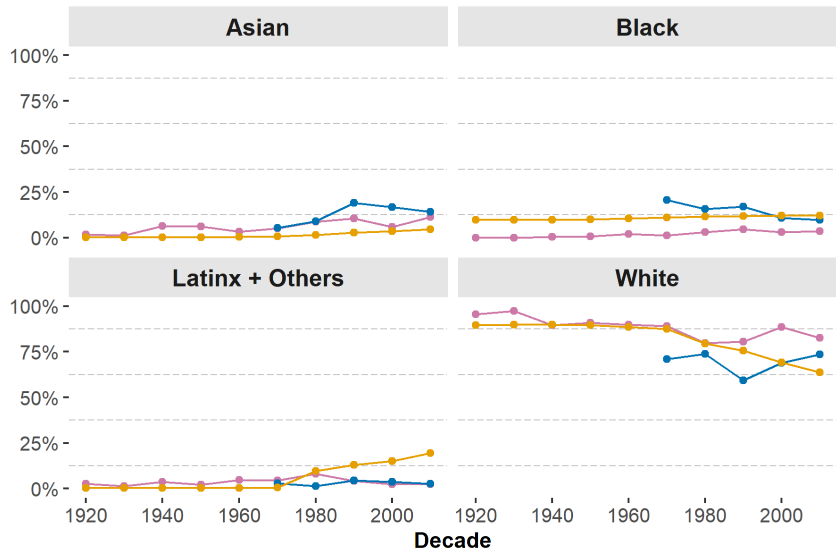
(b) Percent of Female-Presenting Faces Detected, Over Time



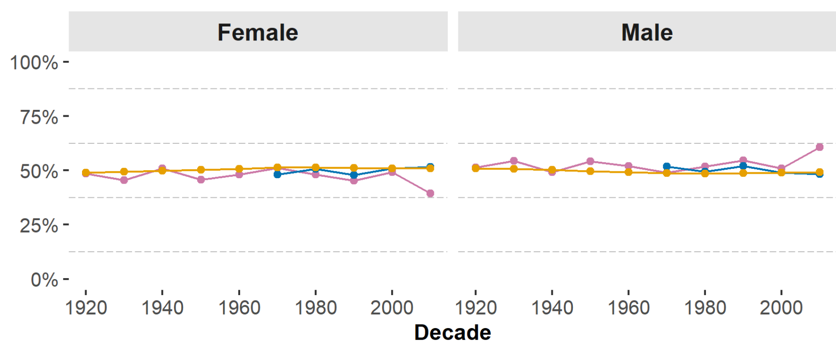
Note: In this figure, we find the proportion of faces in a book which our model labels as female. In Panel A, we show collection level averages of the proportion of female faces in a given book by averaging over all books in a collection. In Panel B, we show these values over time by averaging the proportion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

Figure B4. Share of U.S. Population and Pictured Characters, by Identity

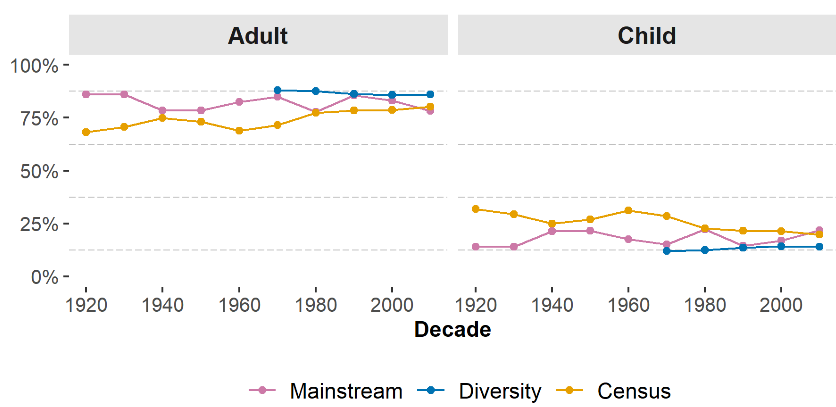
(a) Race/Ethnicity



(b) Gender



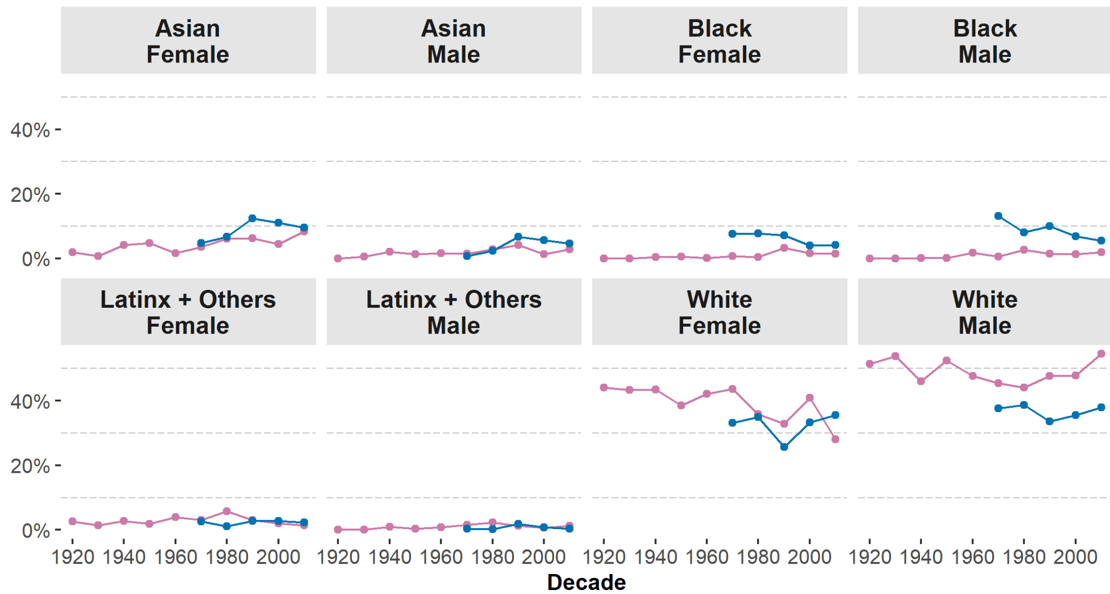
(c) Age



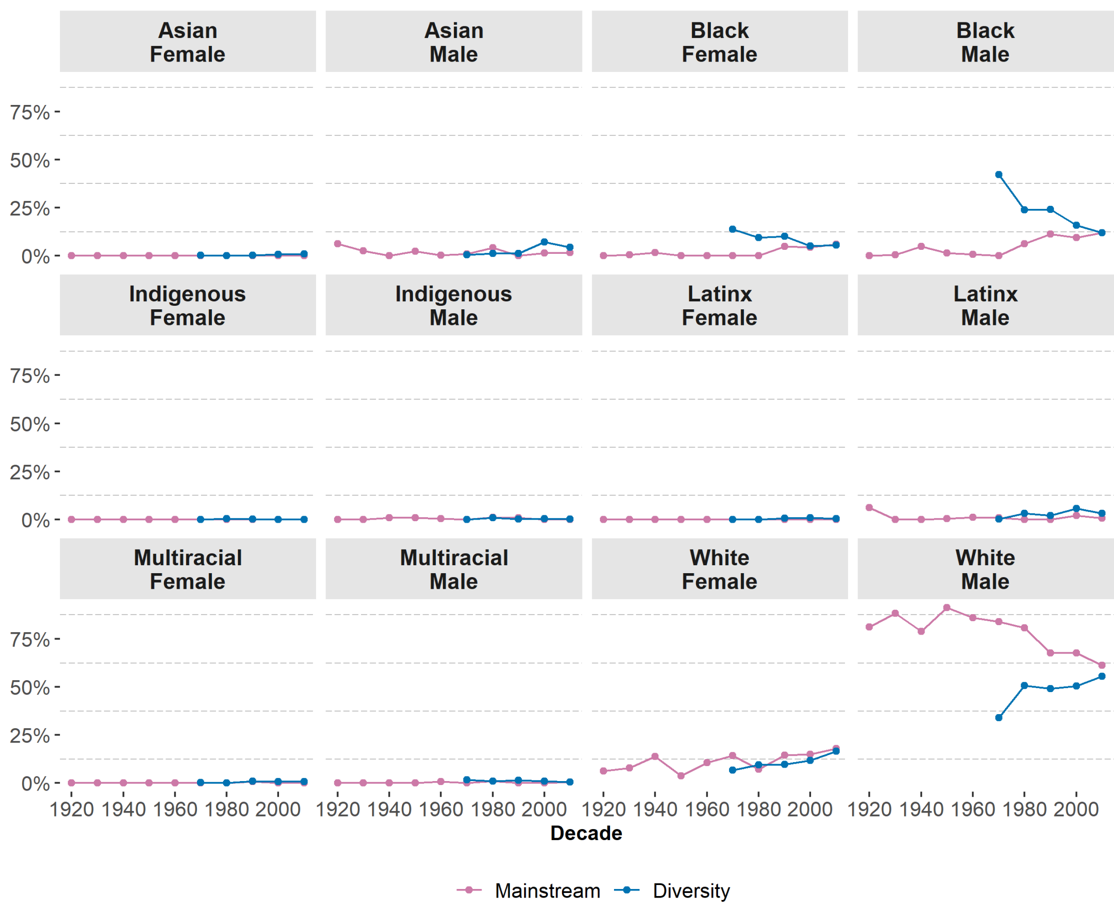
Note: We show the share of the U.S. population of specific identities mapped on the share of the pictured characters classified as a given identity in a given book averaged over all books in collection and decade. In Panel A, we show this by race/ethnicity. Each race/ethnicity category is constructed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

Figure B5. Proportion of Characters in Images and Text, by Race and Gender

(a) Detected Faces (Images)

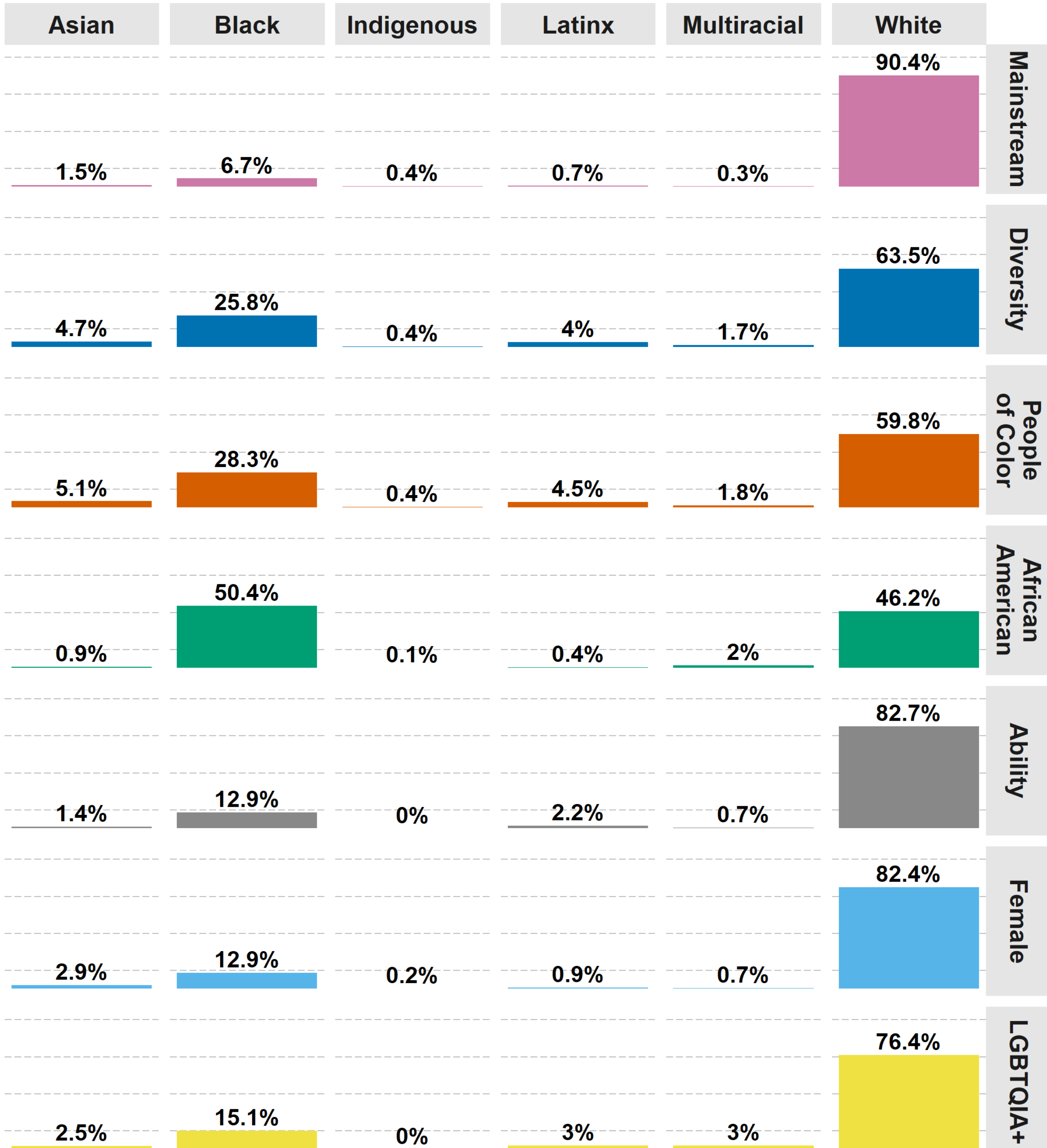


(b) Famous Figures (Text)



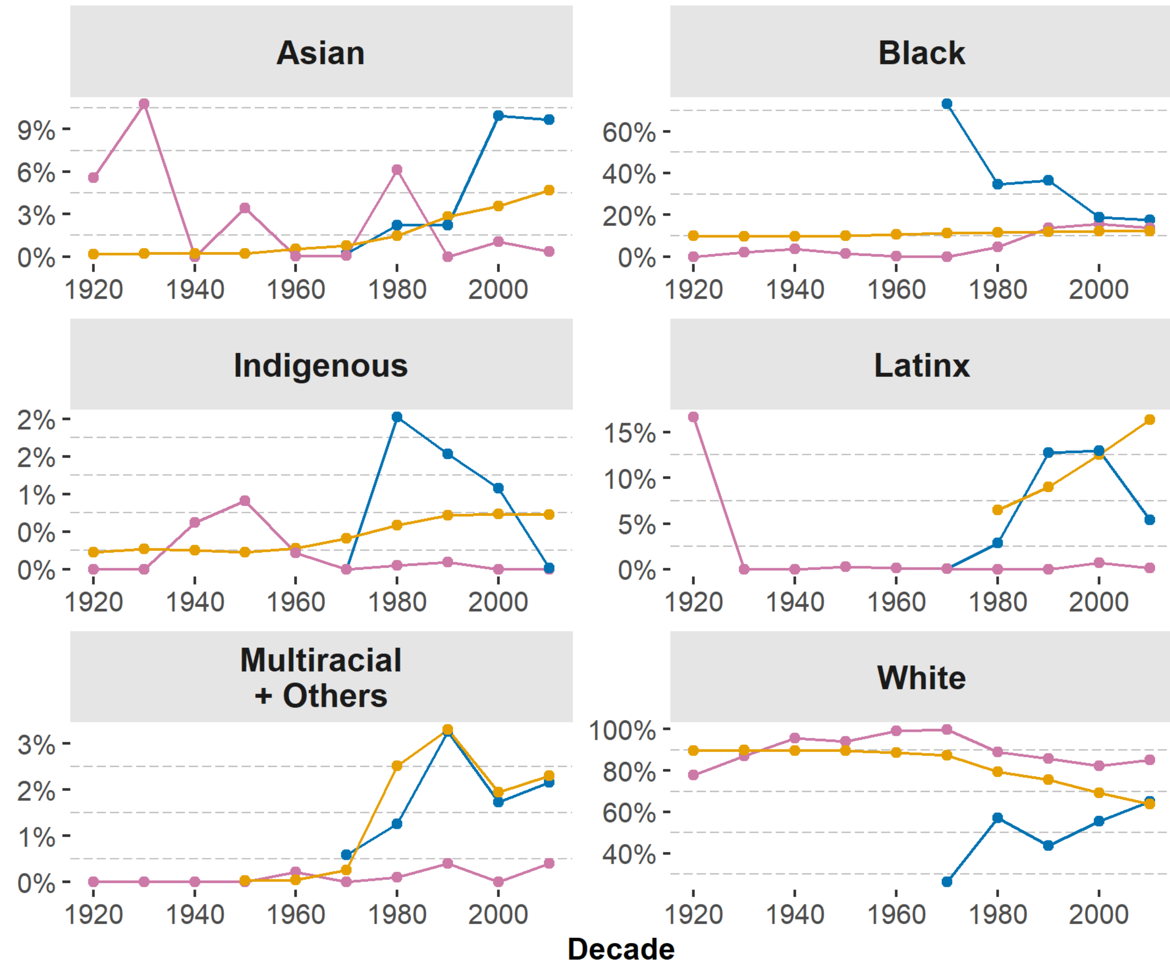
Note: In this figure, we show the share of the characters by race and gender in a given book averaged over all books in a collection and decade. In Panel A, we show this for detected faces in images. In Panel B, we show this for famous figures mentioned in the text.

Figure B6. Race Classifications of Famous Figures in the Text



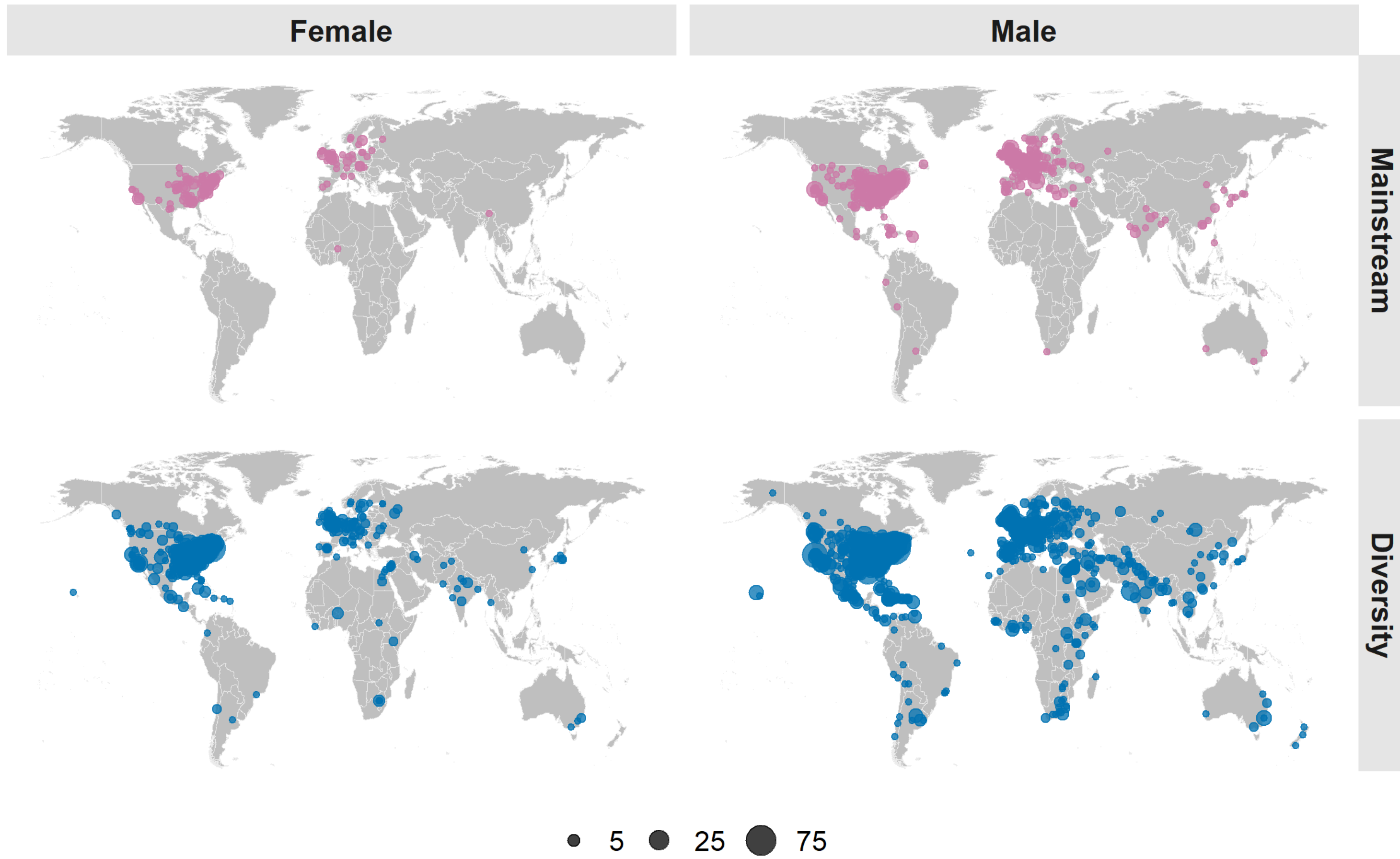
Note: In this figure, we count the number of famous people mentioned at least once in a given book and sum over all books in a collection. We then show the percentage breakdown of these famous people by race. For example, if Aretha Franklin was uniquely mentioned in 3 different books within a collection 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 D.C.3. 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 B7. 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 D.C.3. 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 9, 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.

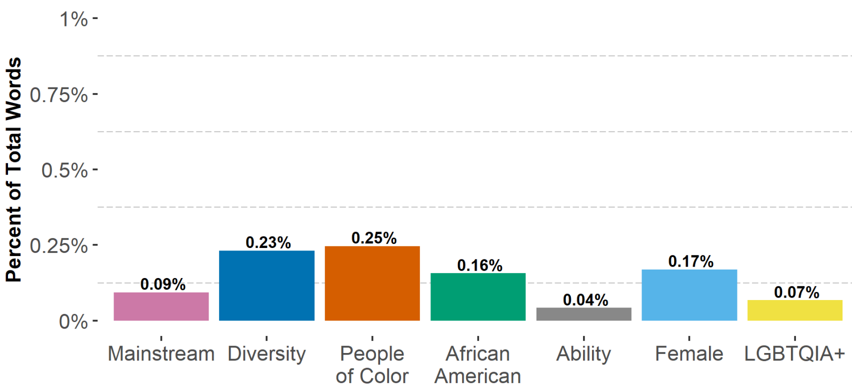
Figure B8. Birthplace of Famous Figures, by Gender



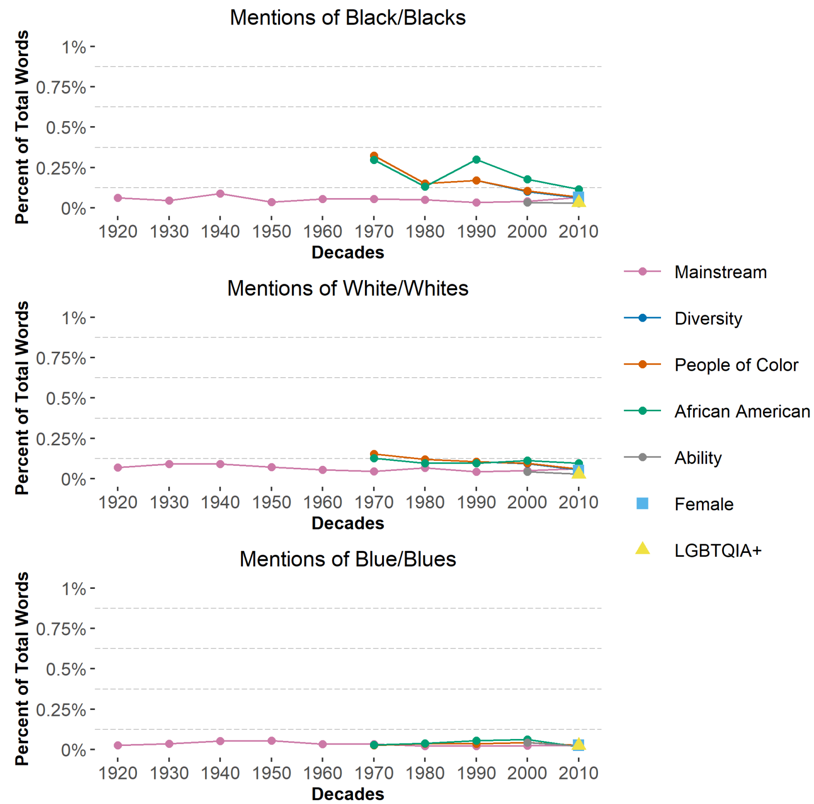
Note: In this figure, we show collection-specific measures of the birthplace of famous figures, separately for females and males. We identify famous individuals as well as their gender and birthplace using methods described in Section D.C.3. If the city/town they were born in was unavailable, we use birth country. Size of dots correspond to the number of famous characters born in a given location that are mentioned at least once in a given book and then aggregated across all books 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 Black. Note that this is an analog to Figure 10, only here with the maps shown separately for female and male famous figures.

Figure B9. Token-Based Proxies for Race: Nationality and Color

(a) Words Related to Nationality

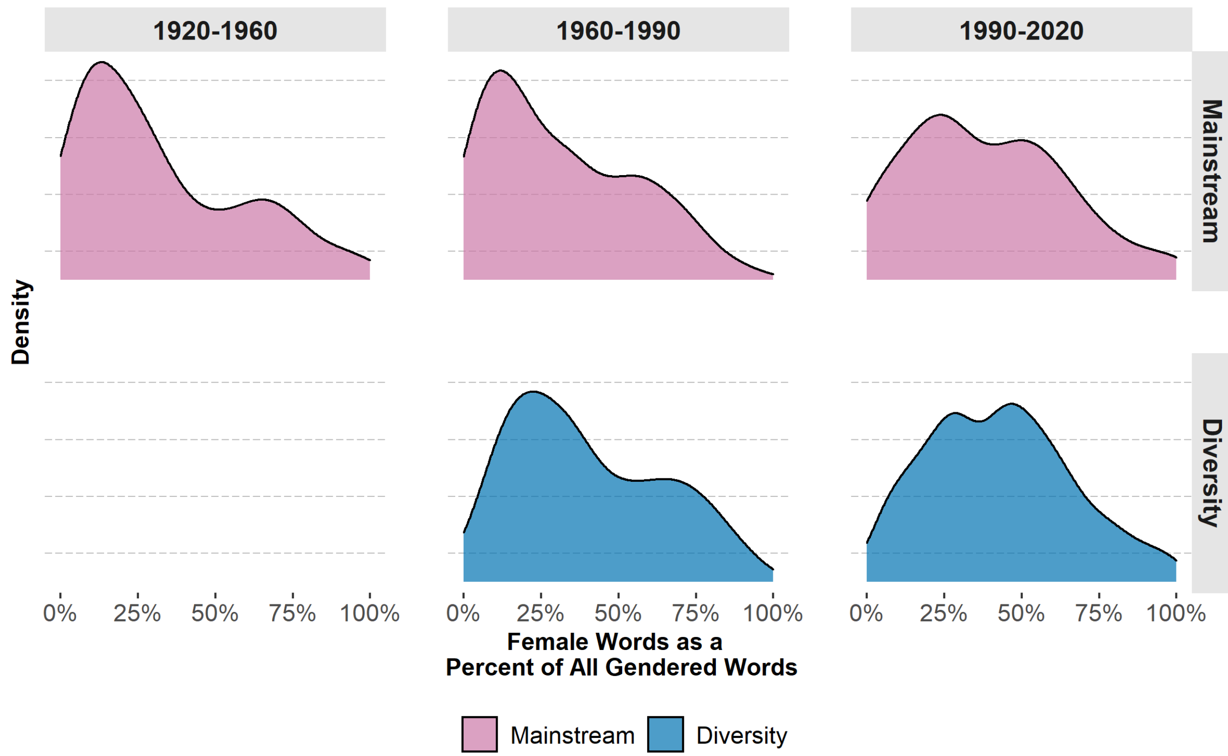


(b) Color Words



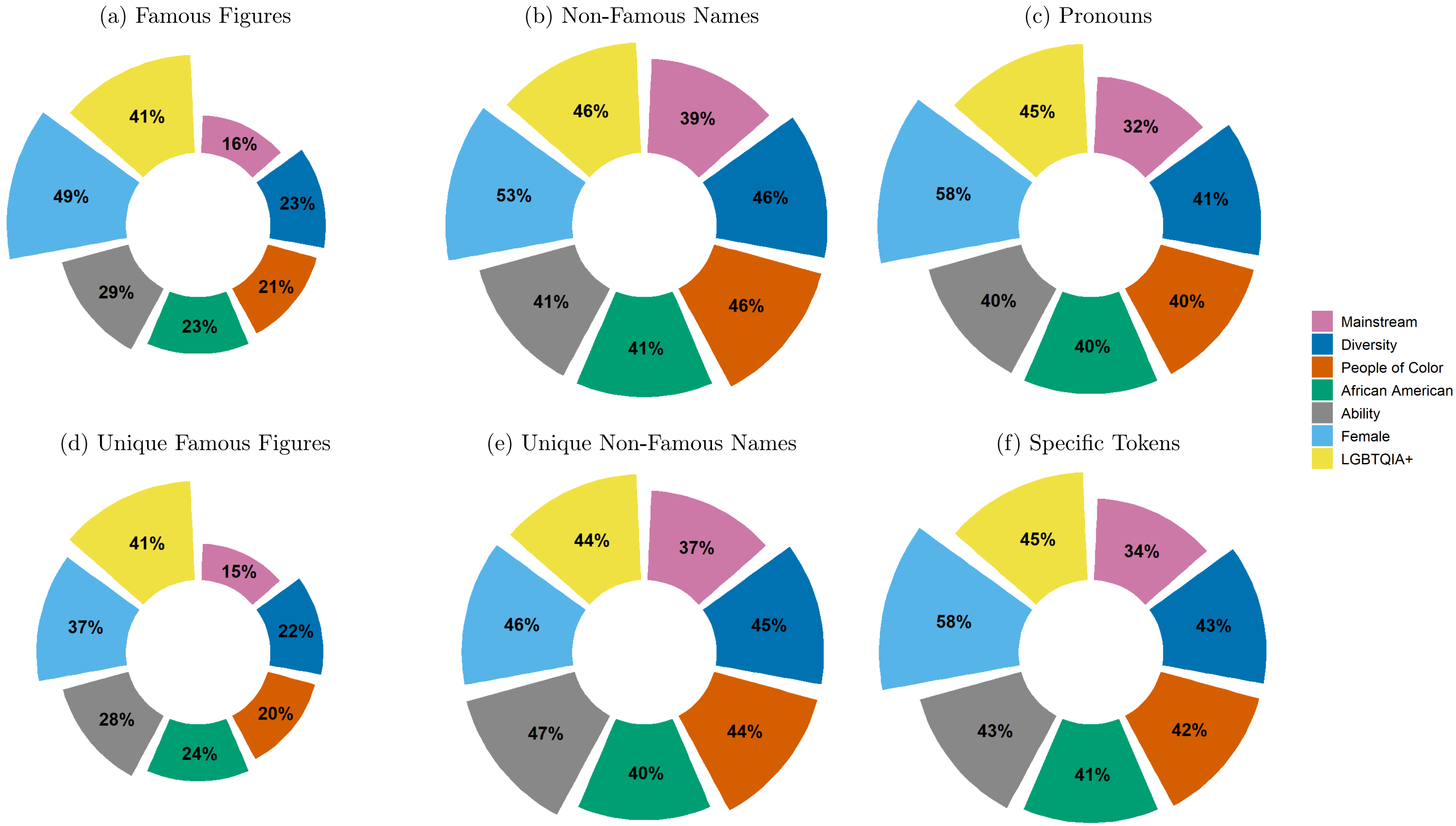
Note: In this figure, we show two measures of the representation of race in text: words related to nationalities and words related to color. In Panel A, we show collection-specific averages of the proportion of words in a book that relate to nationalities. In Panel B, we show collection-by-time averages of mentions of three color words: black, white, and blue – as a proportion of all words in our data. We generated the estimates using a pre-specified list of words (also known as "tokens," as described in Section D.C.1). We provide this list in the Data Appendix.

Figure B10. Distribution of Female Words as a Percent of All Gendered Words, Over Time



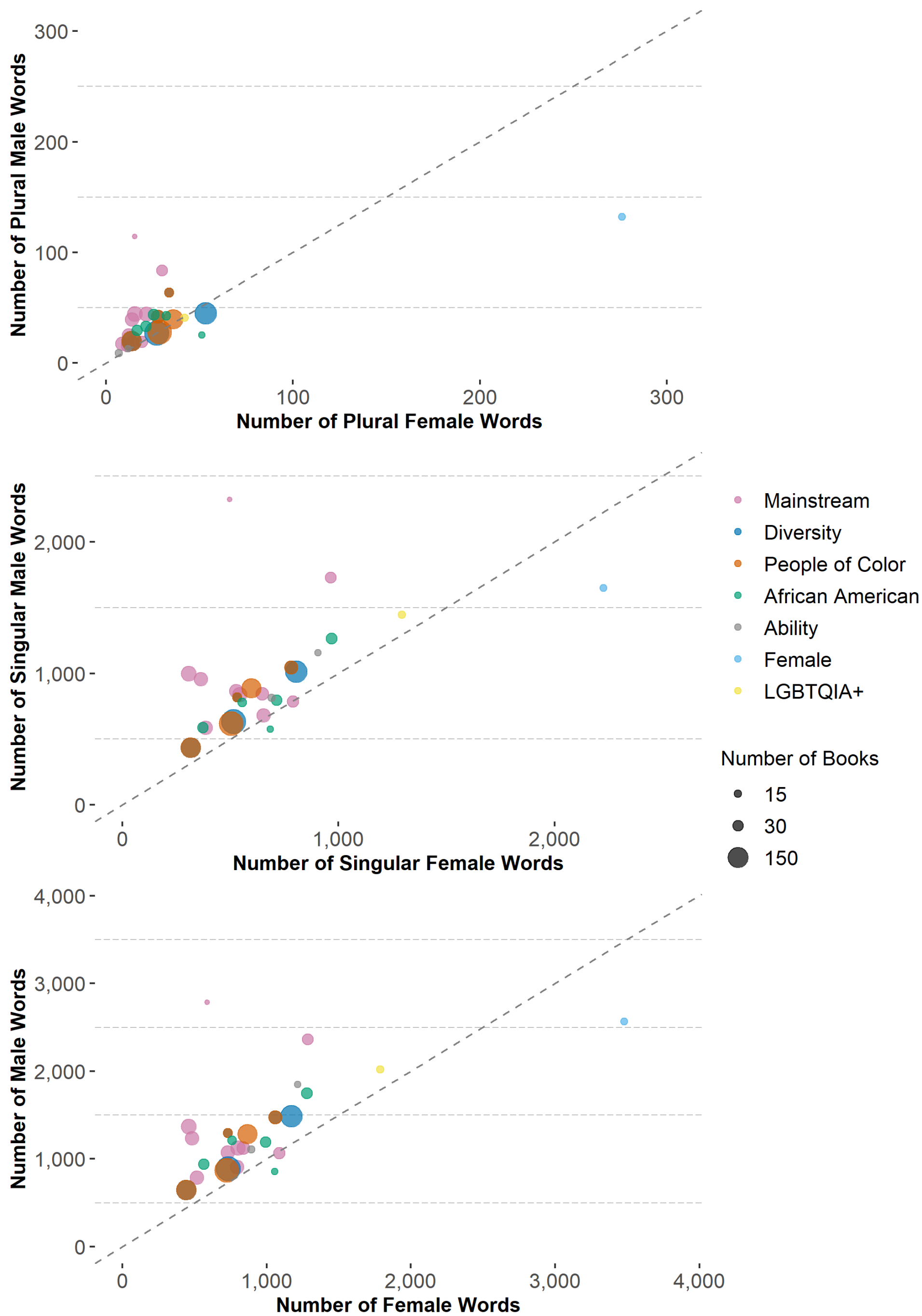
Note: In this figure, we show the distribution of female words as a percentage of all gendered words over time in the Mainstream and Diversity collections. In this case, gendered words encompass the total number of gendered first names, gender predictions of famous characters, gendered pronouns, and a pre-specified list of other gendered tokens (e.g., queen, nephew). We list the pre-specified gendered tokens in the Data Appendix.

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



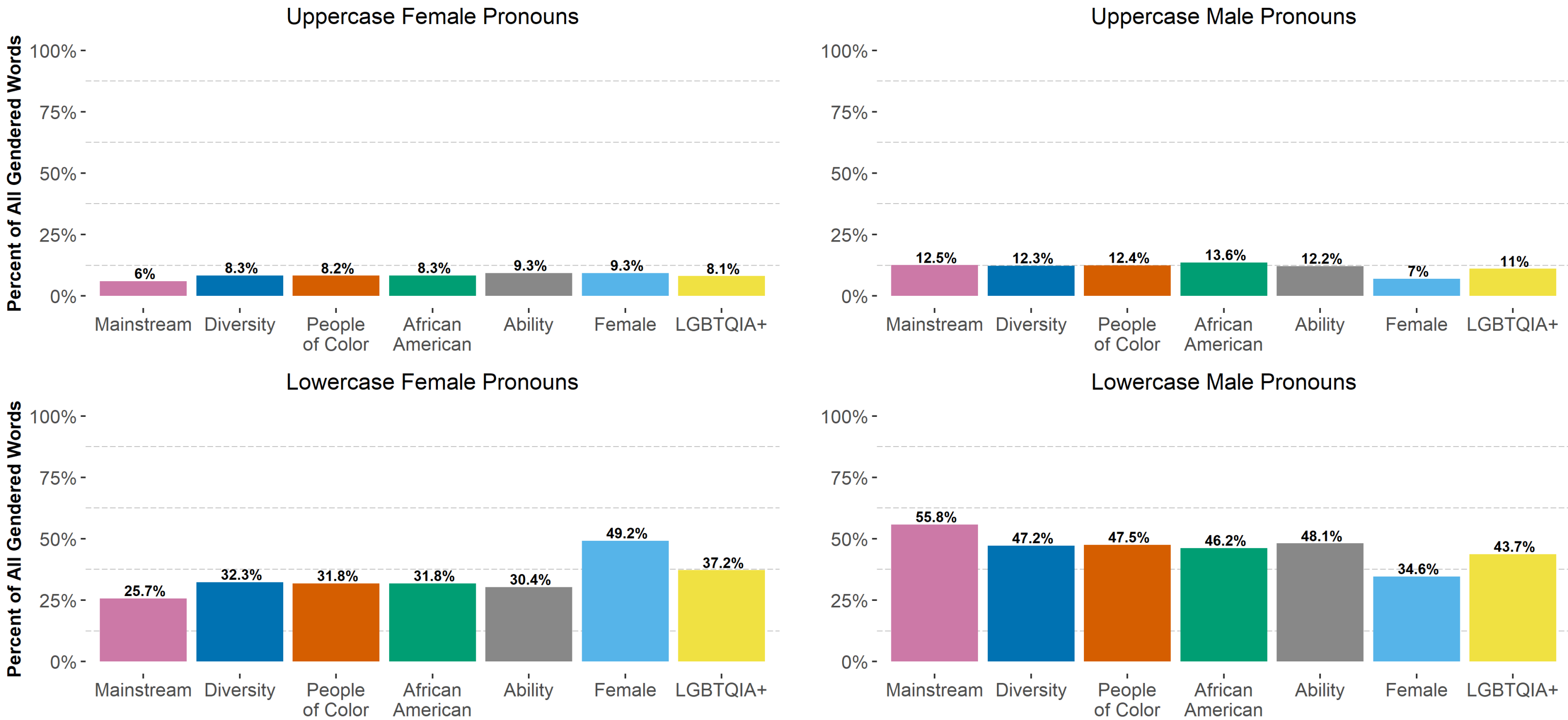
Note: In this figure, we show the proportion of female representation in the text by collection and type of word. In Panel A we find the percent breakdown of female famous people mentioned in a given book, averaged over all books in a collection. For example, if Aretha Franklin was mentioned 4 times 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 non-famous 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 non-famous names. Panel F, shows the percentage of female words from a pre-specified list of gendered words (tokens) such as queen or nephew (full list provided in Data Appendix) in a given book, averaged over all books in a collection.

Figure B12. Gender Representation, by Quantity of Individuals



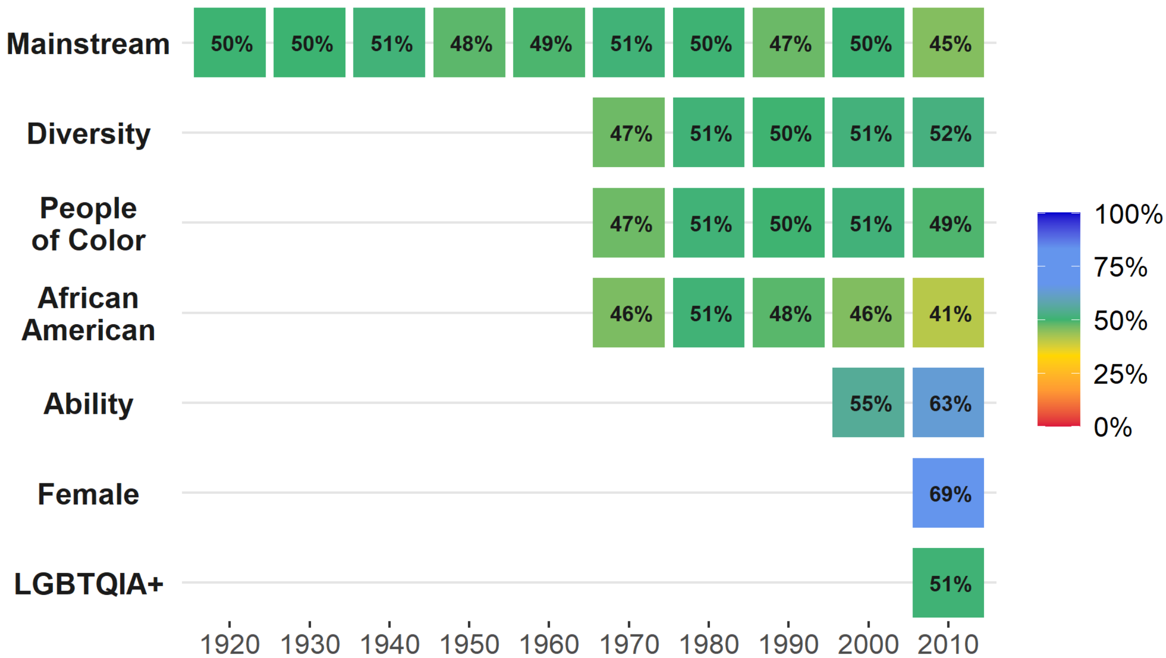
Note: In this figure, we show how gender representation in text varies by whether it is referring to an individual or a group of people; in other words, whether the representation of gender varies by presence of singular (individuals) or plural (groups of people) gendered words. We show collection-specific averages by decade. In the top plot, we show the number of plural male words vs. the number of plural female words; in the bottom plot, we show the number of singular male words vs. the number of singular female words. These male and female words were drawn from a pre-specified list of other gendered tokens (e.g., queen, nephew). We list the pre-specified gendered tokens in the Data Appendix. We also show the total number of male words and the total number of female words for reference.

Figure B13. Proportion of Females and Males Serving as Subjects and Objects of Sentences



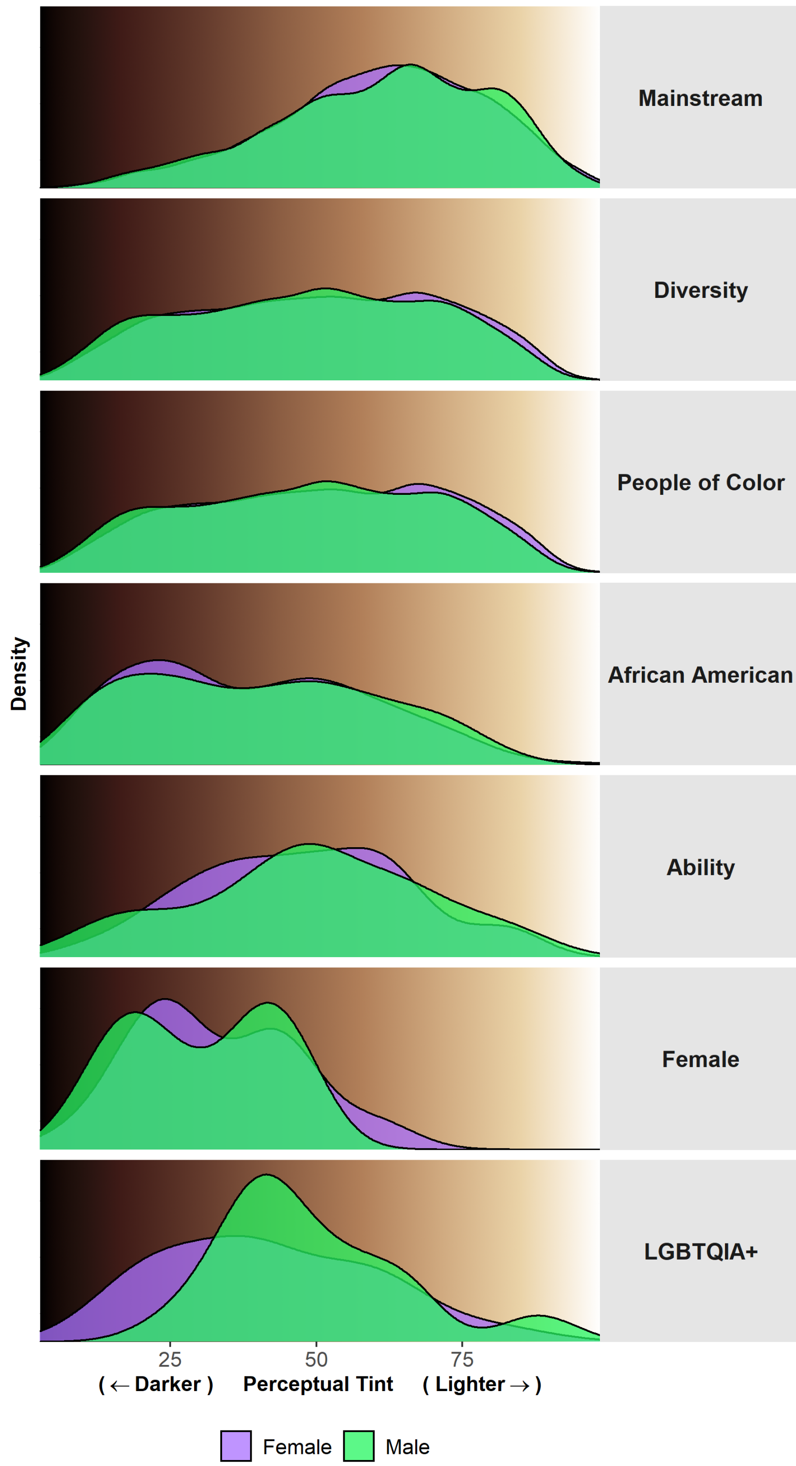
Note: In this figure, we plot the representation of gender by its location in sentences. The top two plots show the average proportion of all gendered pronouns in a book that are uppercase, and the bottom two plots show those that are lowercase. The left plots show the female-related pronouns, and the right plots show the male-related pronouns. We present these separately because an uppercase pronoun is more likely than a lowercase pronoun to be the subject, as opposed to the object, of the sentence in which it appears.

Figure B14. Average Probability a Face is Female, by Decade and Collection



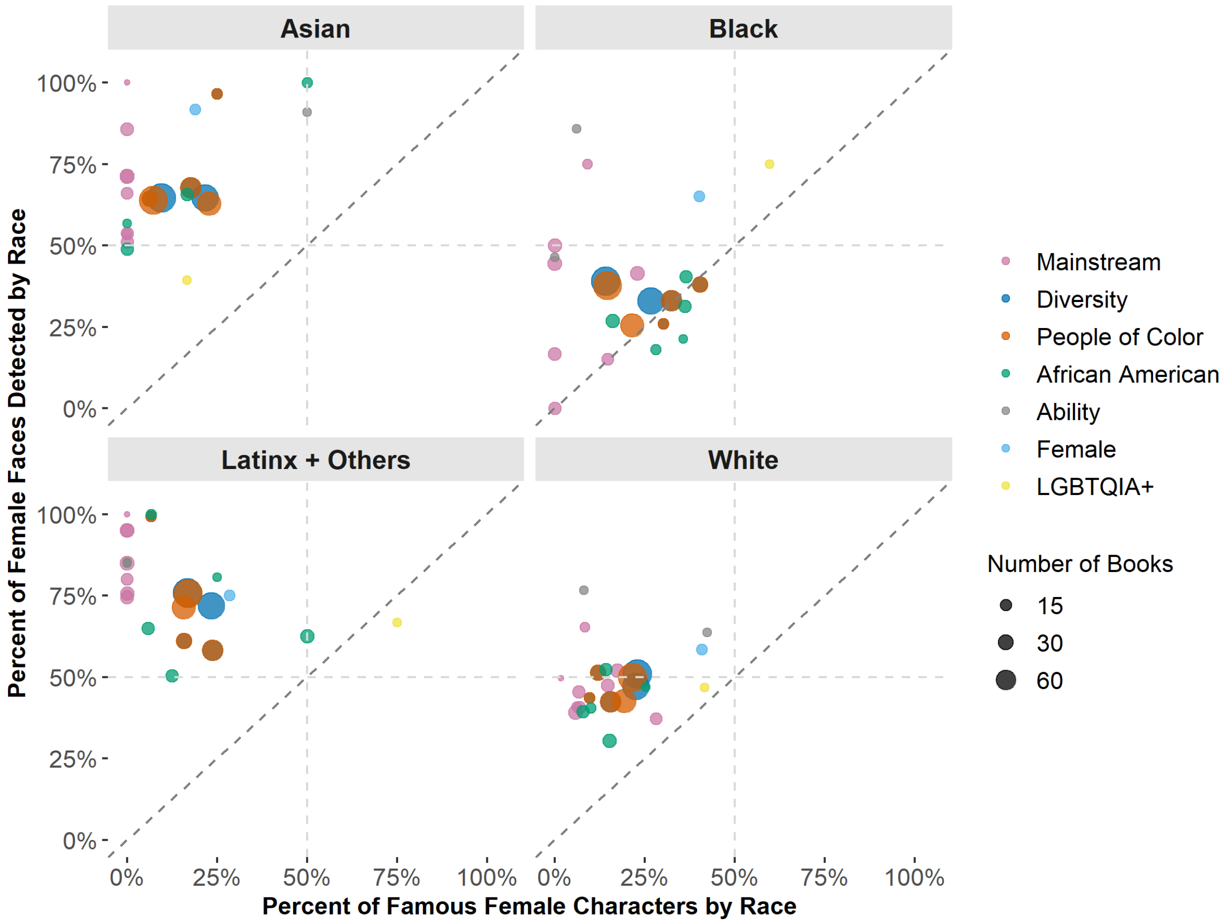
Note: In this figure, we present the average probability that a face was classified as being female in a given collection by decade. We classify gender using an AutoML algorithm trained on the UTKFace public data set.

Figure B15. Distribution of Skin Color, by Gender and Collection



Note: In this figure, we show the distribution of skin tint by gender. We detect faces using our face detection model (FDAI) described in Section III.A. Within these faces, we classify gender 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. We extract a face's representative skin color using methods described in Section III.B. These figures show the results for images that have identified human skin colors (polychromatic skin colors where R $\geq$ G $\geq$ B).

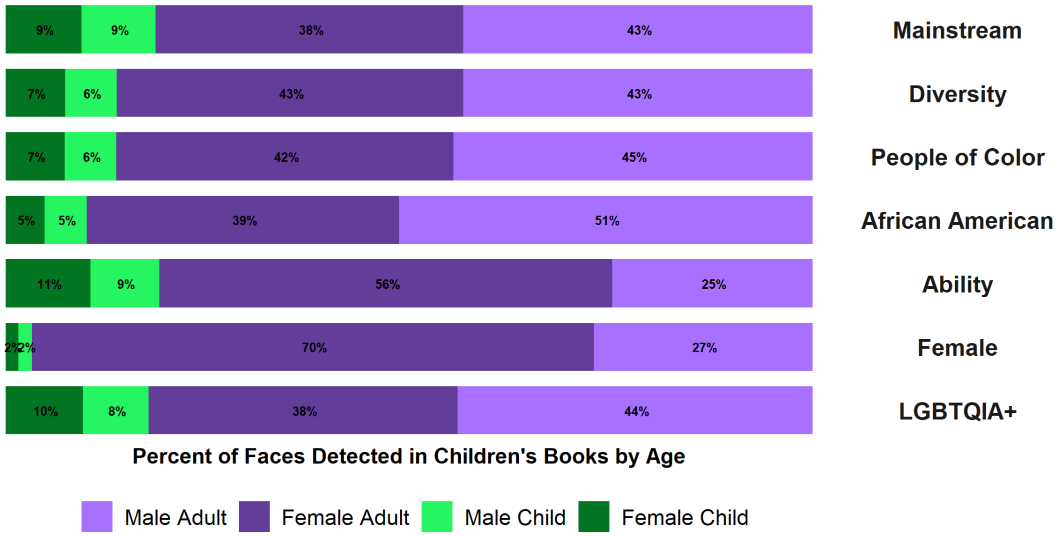
Figure B16. Race and Gender Representation in Images and Text



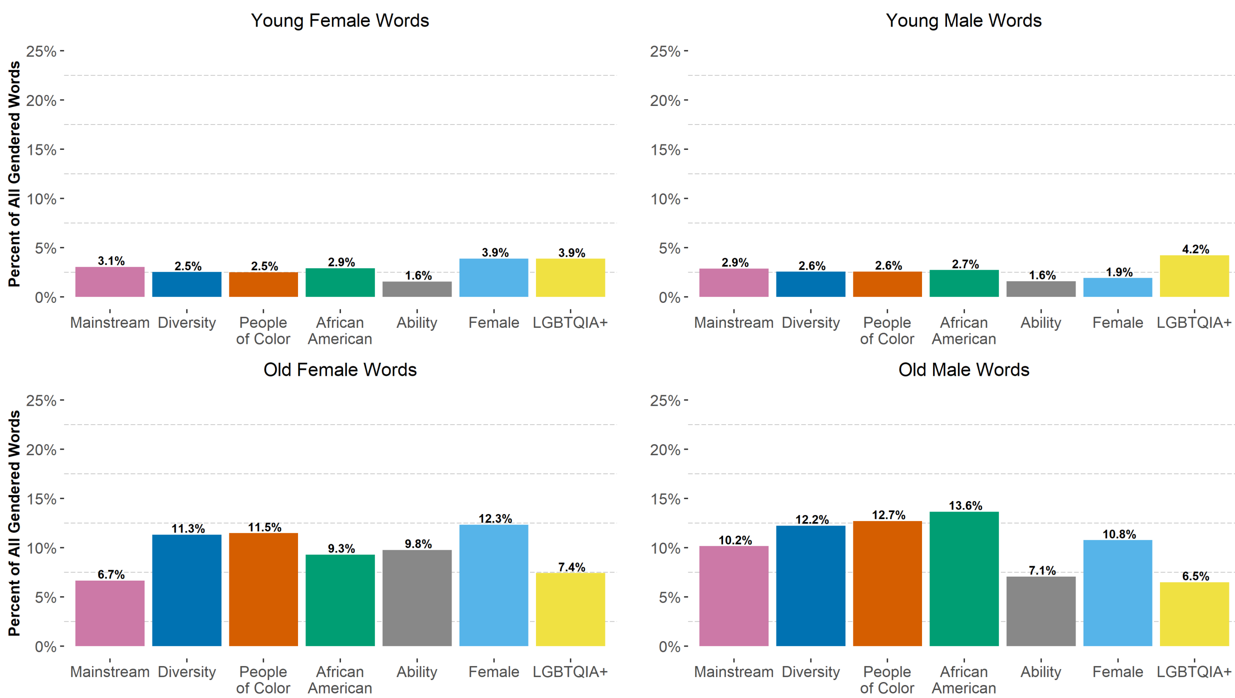
Note: In this figure, we plot female faces by race as a proportion of all faces with a given race classification on the y-axis and famous female characters by race as a proportion of all famous characters with a given race classification on the x-axis.

Figure B17. More Adults than Children in Images and Text

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

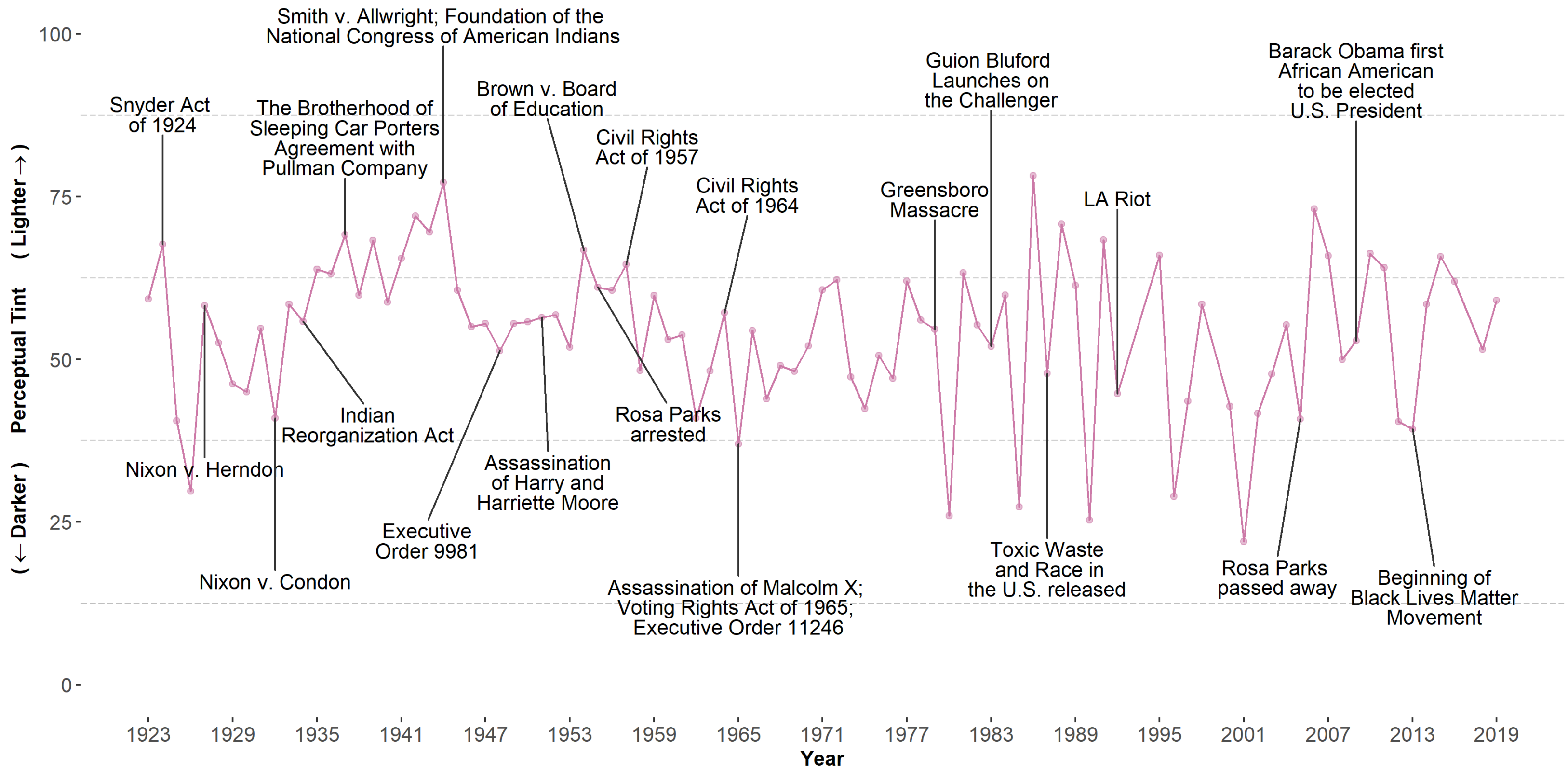


(b) Percent of Gendered Words by Age Group



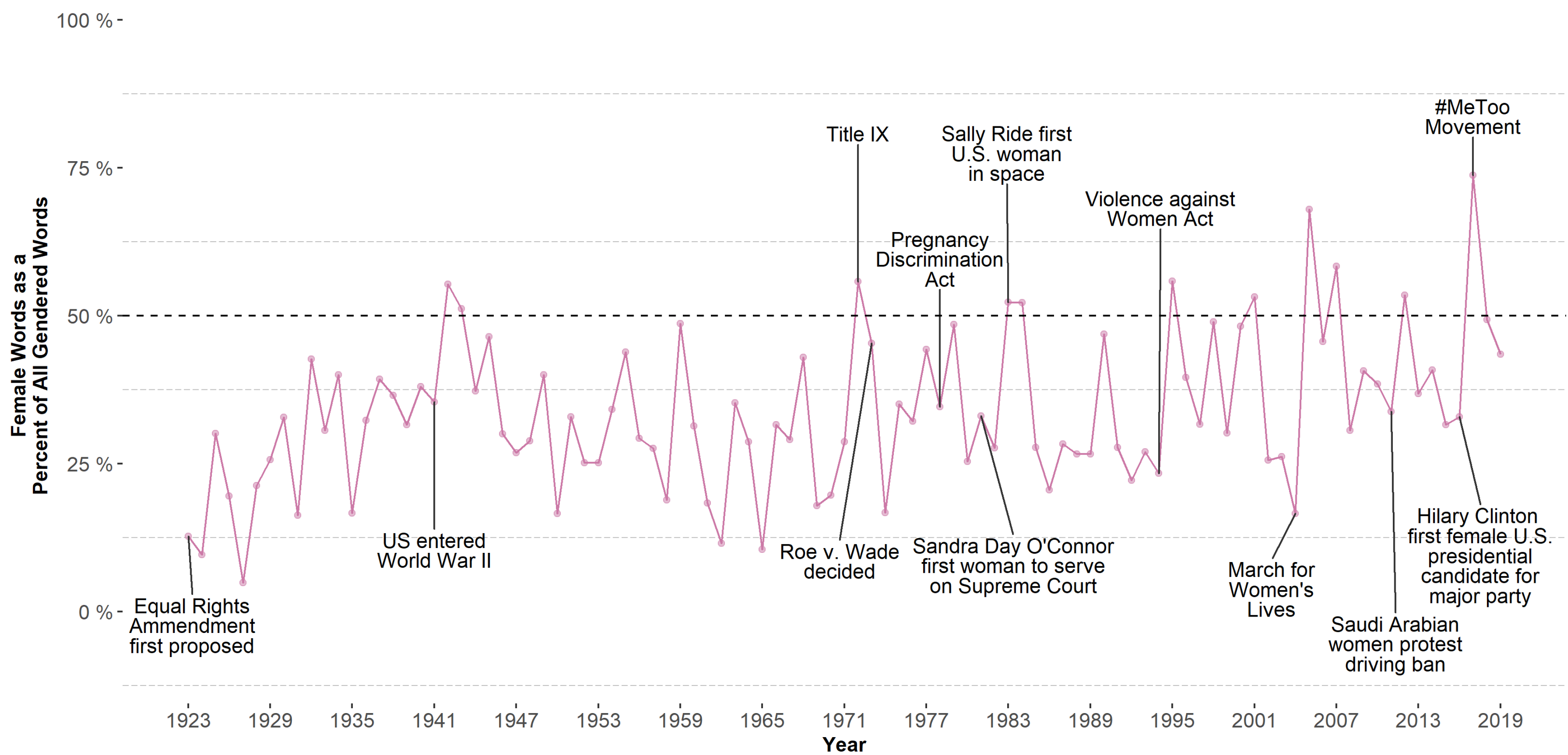
Note: In this figure, we show analysis of the representation of age and gender. In Panel A, we show analysis of predicted age and gender in the faces in images. Specifically, we plot the proportion of identified faces classified in each age (adult vs. child) and gender (female vs. male) category. In 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 tokens (e.g., queen, nephew). We list the pre-specified gendered tokens in the Data Appendix.

Figure B18. 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.

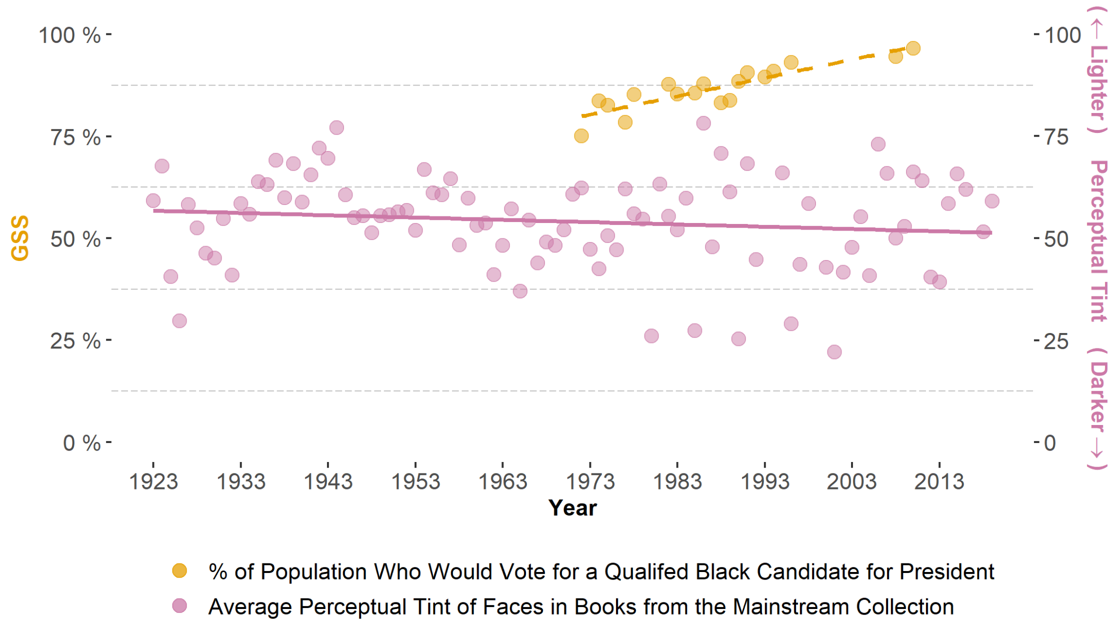
Figure B19. Mainstream Representation of Gender Throughout Historical Events



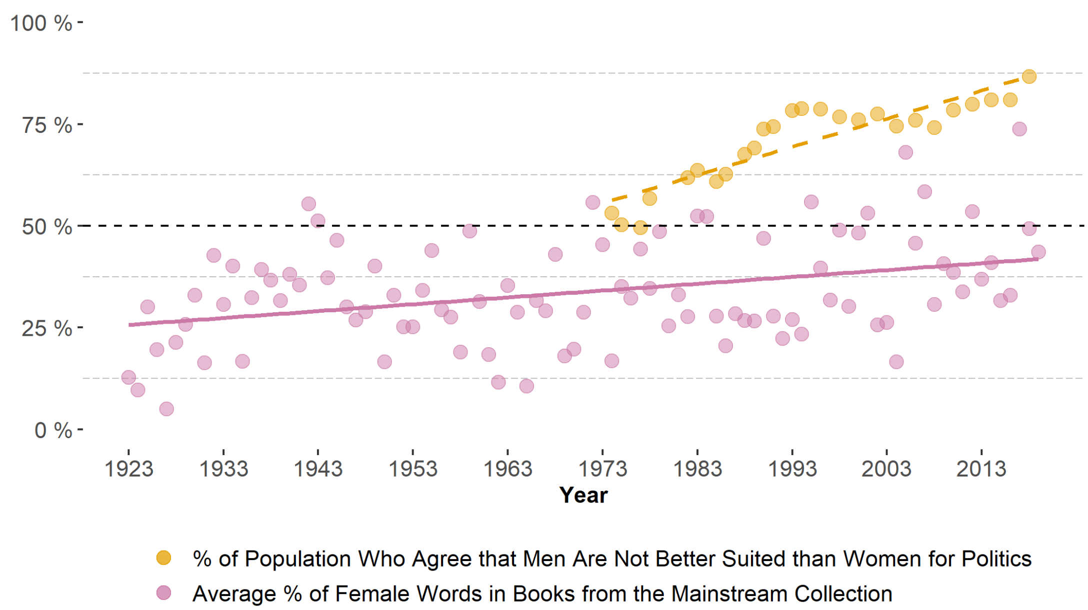
Note: In this figure we juxtapose textual measures of gender representation from the Mainstream collection with the timing of salient historical events.

Figure B20. Mainstream Representation and Social Attitudes Over Time

(a) Race



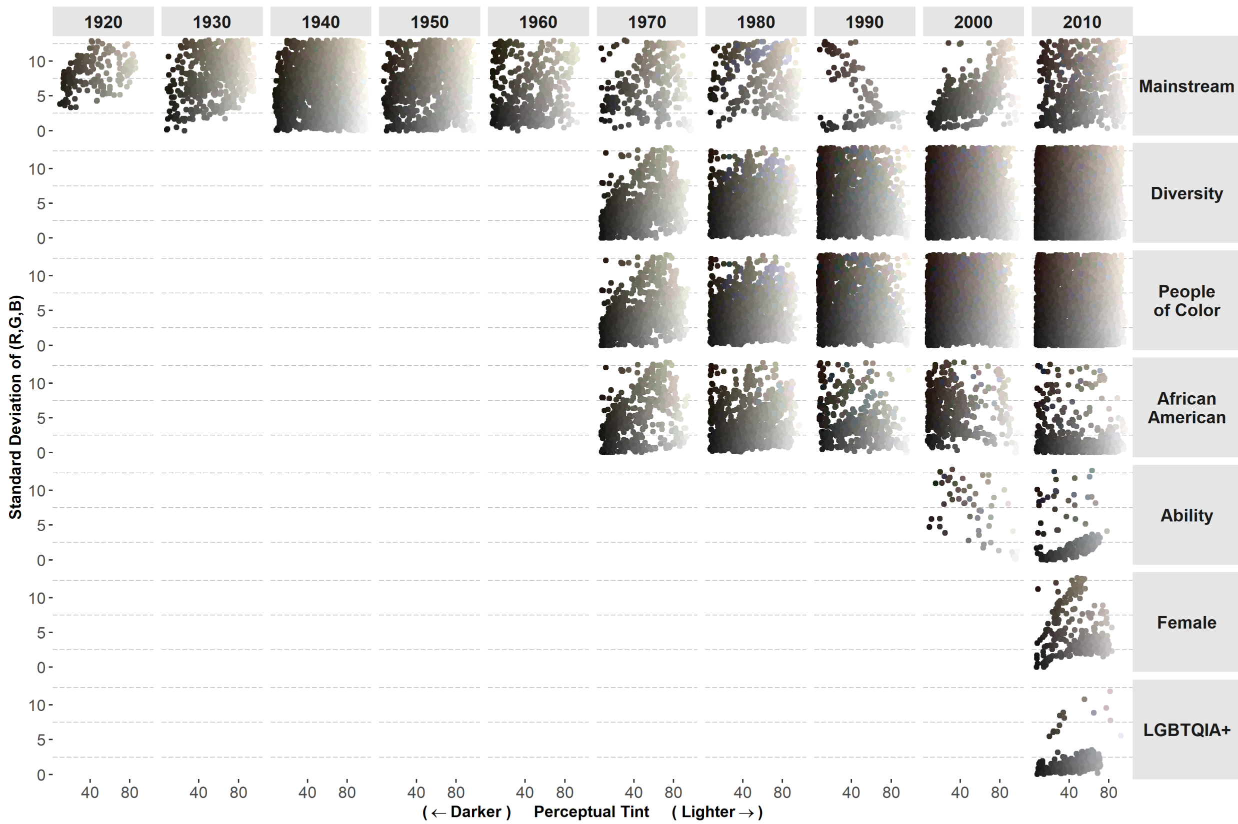
(b) Gender



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).

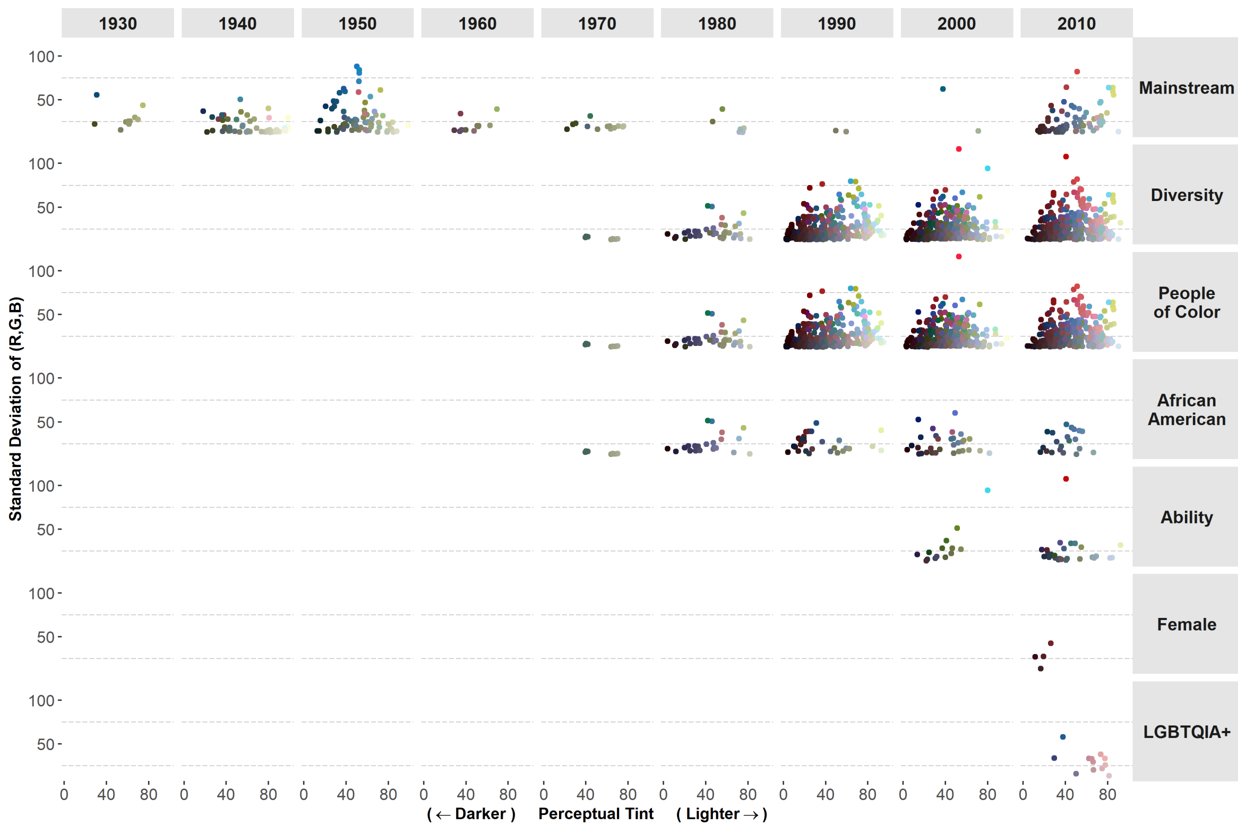
C Non-Typical Skin Color Appendix

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



Note: In this figure we show an analog to Figure B1, 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 III, 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 III.B.

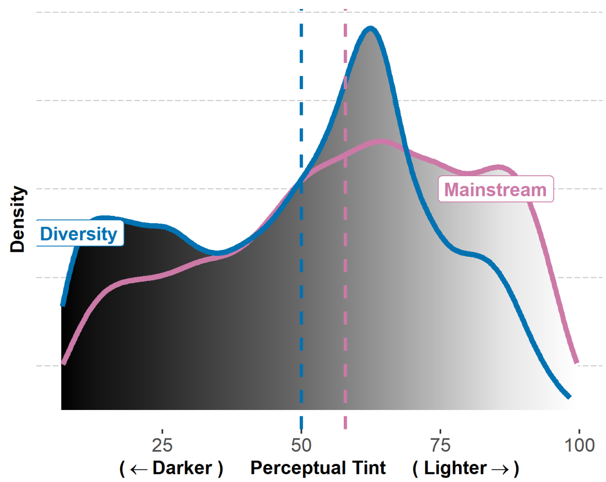
Figure C2. Skin Color Data Over Time, Polychromatic Non-Typical Skin Colors



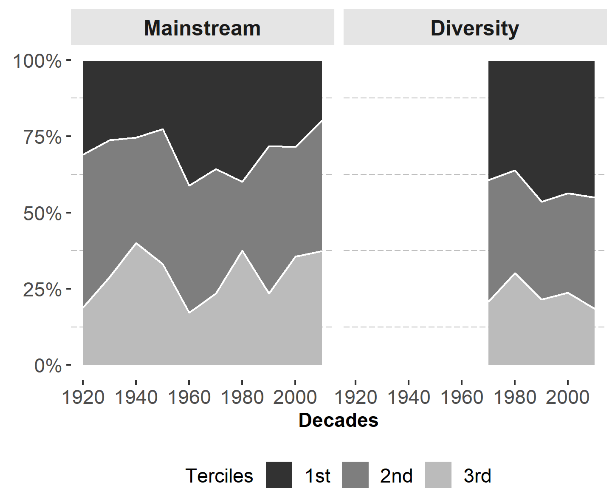
Note: In this figure, we show an analog to Figure B1, here focusing on the representative skin colors for all detected faces with non-typical skin colors (e.g., blue or green) in each collection-by-decade. As described in Section III, we use our face detection model (FDAI) trained on illustrations to classify faces in images. We determine a face's representative skin color using methods described in Section III.B. The data shown in this figure begin in the 1930s, as opposed to in the 1920s as in Figures B1 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

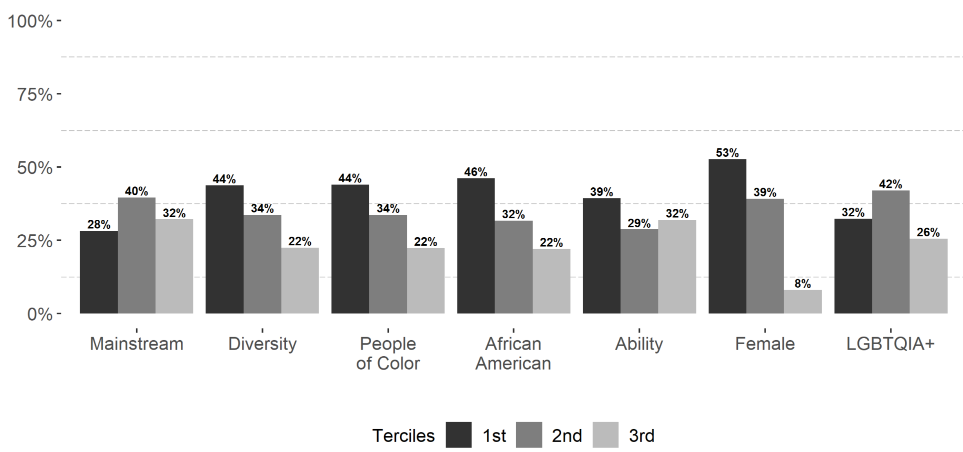
(a) Distribution of Skin Colors



(b) Mean Proportion in Each Tercile, Over Time



(c) Mean Proportion in Each Tercile, All Collections

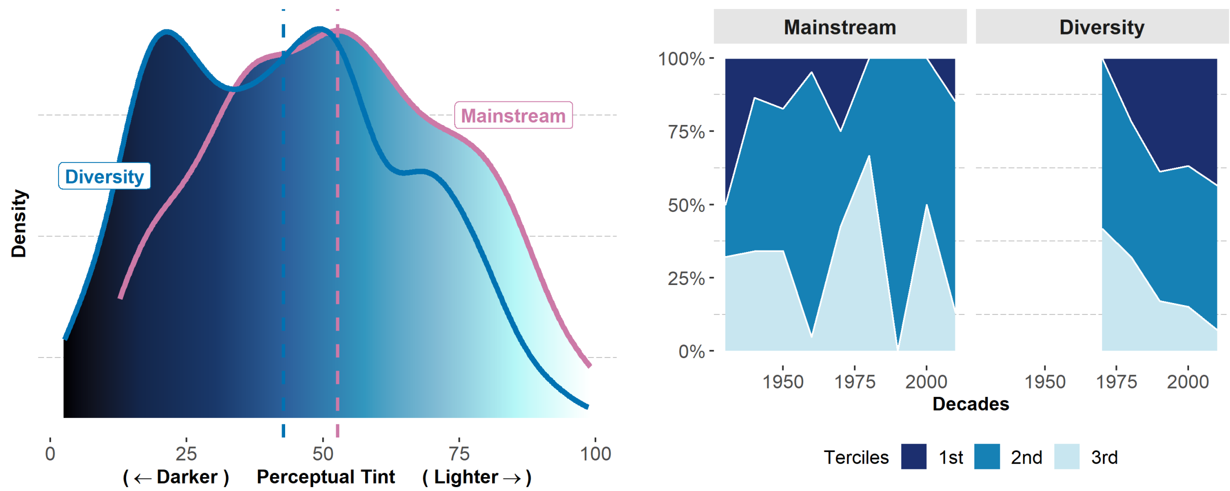


Note: This figure shows our analysis of the representative skin colors of the individual faces we detect in the images found in the books we analyze. This is an analog to Figure 5, only here we focus on monochromatic faces. Panel A shows the distribution of skin color tint for faces detected in books 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 III.

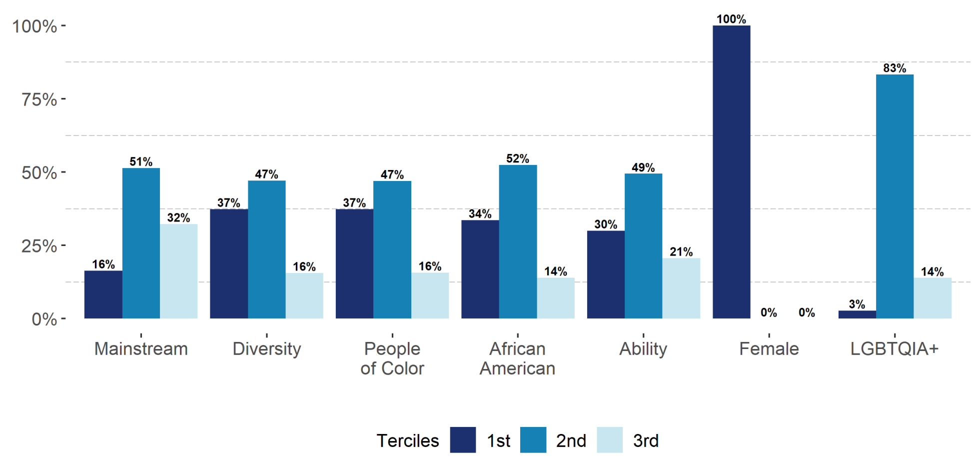
Figure C4. Skin Colors in Faces, by Collection: Polychromatic Non-Typical Skin Colors

(a) Distribution of Skin Colors

(b) Mean Proportion in Each Tercile, Over Time

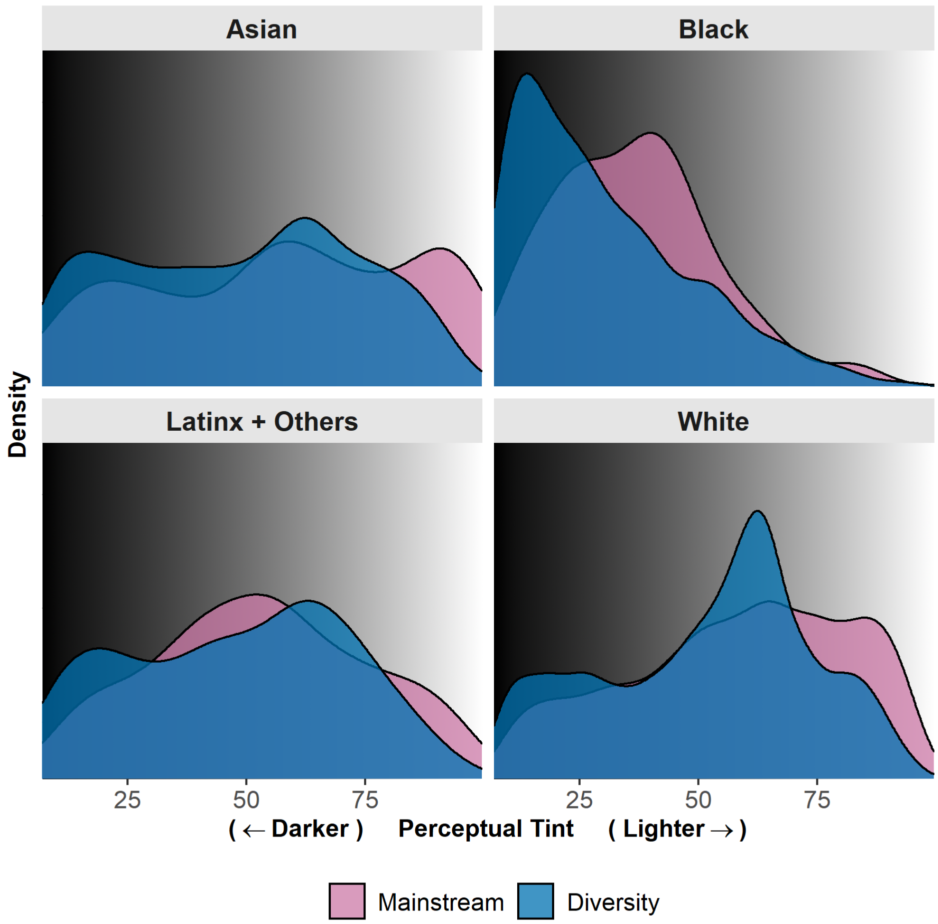


(c) Mean Proportion in Each Tercile, All Collections



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 5, 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 III.

Figure C5. Skin Color by Predicted Race of Pictured Characters: Monochromatic Faces



Note: This figure shows the distribution of skin color tint by predicted race of the detected faces in the Mainstream and Diversity collections. This is an analog to Figure 6, only here focusing on faces depicted in a monochromatic color scheme (e.g., black and white). Skin tint is determined by the L* value of a face's representative skin color in L*a*b* space. We extract a face's representative skin color using methods described in Section III.B. Race was classified by our trained AutoML model as described in Section III.C.

D Methods Appendix

D.A Comparing AI with Manual Content Analysis

In this section, we briefly describe the relationship our analysis has with manual content analysis. We first describe the key differences in scope and reach between our suite of computerized content analysis tools and the tools used by the field of manual 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.

Our use of automated content analysis provides a series of key advantages over manual content analysis. The first important advantage is speed. Our suite of computerized analysis tools can process a very large amount of content in a short period of time. While previous, well-resourced efforts to conduct manual content analysis study the content of between fifty and three hundred books (Weitzman et al., 1972; Davis, 1984; Crisp and Hiller, 2011; Koss, Johnson and Martinez, 2018), our analysis included over one thousand books. Furthermore, in this analysis the most binding constraint was acquiring digitized versions of book content. Were our tools to be used by publishers, libraries, or other entities with access to large bodies or universes of relevant digitized books and other curricular materials, these analyses could be performed even more quickly and inexpensively on collections of books that were previously entirely infeasible to analyze because of speed and cost.

The second important advantage is scope. Our suite of tools is able to analyze all characters contained in the image and text. This is in important contrast to the vast majority of manual content analysis we are aware of, particularly those with larger sample sizes (e.g., in the low hundreds of books). These analyses are able to study a greater number of books by focusing on a smaller number of prominent features, such as the book's title, the images on its cover, and the identities of only the main characters, and such studies often explicitly mention making such restrictions in order to keep the costs of content analysis manageable (Kortenhaus and Demarest, 1993; McCabe et al., 2011; Koss, 2015; Koss, Johnson and Martinez, 2018).

In manual content analysis, the marginal cost of coder time increases with the addition of new dimensions of study. Furthermore, if re-analysis or new analysis is required after the initial analysis appears again, the fixed costs of identifying, hiring, and training coders are again incurred. For automated content analysis, the only cost is the computational power and the cost of adjusting the code, allowing for far greater flexibility and scope in addition to substantially lower costs.

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

The key historical advantage of manual content analysis has been its superiority in measuring more complex and nuanced understanding than those we capture here. Our focus in this study is primarily on measuring the presence of different identities, a domain for which computer-driven analysis is particularly suitable. Furthermore, recent advances in natural language processing and computer vision have begun to make progress in incorporating the analysis of more complex features into the toolkit of computerized content analysis (Nenkova and McKeown, 2012; Ouyang and McKeown, 2014; Caliskan, Bryson and Narayanan, 2017; Garg et al., 2018).

Next, we describe our work to validate our tools using manual content analysis. 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.

Finally, we estimate a rough measure of the cost-effectiveness of our tools. It took approximately 40 hours to code the entire book (400 pages at an average of 6 minutes per page).[55] 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).[56]

The manually labeled test data are kept separate from the training and hyper-parameter tuning algorithms.[57] The models compare results from the algorithms to the manual labels in the test data to evaluate the accuracy of the algorithms.

[55] 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.

[56] 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.

[57] 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.

We use two specific parameters that are commonly used to evaluate the performance of this class of model: "precision" and "recall."[58] 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.[59] 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 (6.6 percent of identified faces may not be true faces) and 76.8 percent recall (1 in 4 true faces may not be identified).

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, geometric transformations, and partial occlusions (Lumini and Nanni, 2020). To deal with these issues, we isolate skin from non-skin parts of the detected face using a deep learning approach called Fully-Connected Convolutional Neural Network Continuous Conditional Random Field (FC-CNN CRF). An equivalent term for this is Fully-Convolutional Contin-

[58] 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.

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

uous Conditional Random Field. 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., greyscale, 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 tones (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.[60] 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

Race Classification (Images). The model assigns the probability that a detected face is of a given race category: Asian, Black, Latinx + Others, or White.[61] Each identified face is assigned the race category which the model gives the highest predicted probability to.[62,63]

Gender Classification (Images). For each face detected, we predict the probability

[60] 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).

[61] 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.

[62] 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)

[63] 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.)

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

In this section, we describe the tools we use to measure representation in the text of books. Social scientists 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 automated 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 3b, we show our process of extracting text from digitized books and then analyzing it; we refer to this as our "Text-to-Data Pipeline."

The first step in conducting this analysis is to use optical character recognition (OCR) to extract text from digital scans of books. We use the Google Vision Optical Character Recognition (GVOCR) tool for this task. 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.[64]

[64] 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

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").

From these text data, we then derive several features. These features include: token (single word) counts[65], the presence of famous people, and the first names of characters. In the rest of this section, we describe how we use these features to construct measures of the representation of gender, race, and age in the text of each book.

D.C.1 Text Analysis: Token Counts

One branch of traditional content analysis consists of enumerating words that represent a particular attribute (Krippendorff, 2018). This process generates counts of different "tokens," which comprise a maximal sequence of non-delimiting consecutive characters.[66] In our context, a token is an individual word. We generated a set of tokens associated with identities related to gender, race, or age. The vocabulary used for each of these lists is shown in Section D.C.6 below. We aggregate counts of these words by their respective identity category (such as female or male) by book, generating our "token count" measures of the representation of each identity in each book (Neuendorf, 2016).

Gender (Token Counts). To calculate gender representation in token counts, we calculate the number of words with a gendered meaning. For example, female gendered tokens consist of titles and pronouns such as queen, aunt, girl, she, etc. Similar examples for male gendered tokens include husband, prince, son, himself, etc. [67]

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 (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.

[65]This differs slightly from the notion of "lemma" counts, which are measures of the count of word stems. To understand the difference, take the words "father," "fatherly," and "fathered." If each word appeared once in a book, it would generate a token count of one for each word, but a lemma count of three for the lemma "father."

[66]We use the spaCy library to generate these counts, but we see similar patterns in our findings when we use NLTK instead.

[67]Traditional content analysis often restricts gendered words to pronoun counts. We show the sensitivity of our findings related to this construct by restricting the analysis to gendered pronouns only in Appendix

We show how gender representation varies on three additional dimensions: one, whether the gendered identity is represented by individuals (singular) or groups (plural); two, whether the character is placed as the subject or object of a sentence; and three, by the age of the gendered word. To analyze singular and plural representation separately, we separate gendered tokens into those referring to singular cases (e.g., daughter) and plural cases (e.g., daughters). To analyze whether the character is the subject or object of a sentence, we generate counts of the number of gendered pronouns that are capitalized versus lowercase, under the assumption that an individual who is the subject of a sentence is in a position of more active importance than the same character when used as the object and thus occupying a more passive role. To analyze representation of gender by age, we generate a list of "younger" gendered words (e.g., princess, boy) and "older" gendered words (e.g., queen, man).

Color (Token Counts). As another proxy for the analysis of race, we calculate the proportion of all words that refer to colors. For parsimony, in the paper we only show the words black, blacks, white, whites, blue, and blues.

Nationality (Token Counts). To calculate nationality representation in token counts, we calculate the proportion of all words that refer to nationalities. Note that we converted all multi-token nationalities to a single token analog.

D.C.2 Text Analysis: Named Entity Recognition

We also measure the representation of gender and race among named characters in these stories, be they fictional or historical. A series of studies show that exposure to salient examples of historical figures or celebrities from historically marginalized identities can lead to meaningful change in social attitudes towards people who hold that identity, and potentially are associated with changes in beliefs and academic performance among children who share that identity (Marx, Ko and Friedman, 2009; Plant et al., 2009; Alrababah et al., 2021). To identify characters in our text, we use a tool called Named Entity Recognition (NER).[68] NER identifies and segments "named entities," or proper nouns, starting with a pre-defined library of such entities and also identifying new entities through the application of neural nets. NER recognizes these entities in strings of text; applying NER to our data, we identify these entities and count how many times each specific named entity is mentioned

Figure B11. Our results are robust to this alternate specification.

[68] We run our NER analysis using the open-source software library spaCy, which employs convolutional neural networks for both text categorization and NER. Another commonly-used library for NER is NLTK, but it only recognizes single words for NER, whereas spaCy can recognize strings of words as a distinct entity. For example, "Martin Luther King" would be recognized as one entity in SpaCy but as three entities with NLTK ("Martin," "Luther," and "King").

in a given book. We then associate these frequency counts with identifiable traits of the people identified by NER, such as their race, gender, or place of birth. There are two types of named entities that we identify: (1) famous characters and (2) first names of characters.

D.C.3 Text Analysis: Famous Characters

To identify the instances of famous characters represented in books, such as Martin Luther King Junior or Amelia Earhart, we match all of 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 have a presence in more than 15 language editions of Wikipedia (Yu et al., 2016). This generates a data set of 2,697 famous people. To examine the race of the famous figures mentioned in the text, we count the number of famous people mentioned at least once in a book and sum over all books 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 Black. 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.

Gender and Birthplace (Famous People). The Pantheon 2.0 data set contains information on the gender and birthplace of these famous people. We match these data to each famous figure identified from the NER in our data.[69]

Race (Famous People). We then manually code race for each identified person. This was conducted based on a manual internet search for each person, starting with Wikipedia.[70] 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.

D.C.4 Text Analysis: Character First Names

We also study the representation of gender among people who are named but not identified as "famous" using the methods described above. Using the named entities identified by the spaCy NER engine, we limit the sample to those entities categorized as a person and

[69] The Pantheon 2.0 curators run a classifier over the English text of the Wikipedia biographies to extract information such as place of birth and gender from each biography. Their classifier was trained on a data set called Pantheon 1.0 (Yu et al., 2016) which contains a subset of manually curated biographies.

[70] Note that coding of putative race is subject to the individual biases and perceptions of each human coder and may be classified with error.

remove the famous characters we found by applying the process described in Section D.C.3.[71] We then categorize the remaining named entities and construct a data set containing the name of each unique character and the number of times that character is mentioned in a given book.

Gender (Character First Names). To identify the gender of characters not identified as famous, we extract the first name of each remaining named entity and estimate the probability that the character is female using data on the frequency of names by gender in the U.S. population from the Social Security Administration. For example, if a character's first name is "Cameron," our estimated probability that the character is female is 9.16 percent because that is the proportion of people named "Cameron" in relevant Social Security data who are female. Our sample of "relevant" Social Security data include only data from years which overlap with the years in our sample of children's data.

If the predicted probability that a character is female is greater than 50 percent, we label that character as female. Otherwise, the character is labeled as male.[72] Using this method, we are able to make gender predictions for approximately 60,000 characters. 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.[73]

We are not able to make a prediction for the remaining named entities. For example, characters such as "New Yorker" which the spaCy NER engine identified and labeled as a person will not receive a prediction because "New" does not appear as a first name in Social Security data.

D.C.5 Text Analysis: All Gendered Words

We aggregate all gendered mentions (gendered tokens (e.g., titles, pronouns, specific gender terms such as queen and husband), predicted gender of character first names, and gender of famous characters) to generate a composite measure of gender representation in text. We refer to this aggregate measure as "gendered words," or "words with a gender association."

[71] NER tags each entity with a different category: people, locations, currency, and more. This entity categorization (e.g., person, location) is not always correct, so there may be entities misclassified or missed overall. We do not use this categorization when identifying famous characters.

[72] We predict gender with the *gender* package available in R which uses Social Security Administration data (Mullen, 2020).

[73] 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 accurately distinguish between Black people and White people (Garg et al., 2018).

The total number of female gendered words in book i is calculated as follows:

$$\begin{aligned}(\text{female words})_i = &(\text{total number of female-specific tokens})_i\\ &+ (\text{total number of mentions of famous female characters})_i\\ &+ (\text{total number of characters with female first names})_i\end{aligned}$$

D.C.6 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 Tokens. The gendered tokens we enumerate are as follows. Subset lists are used for the specific gendered token counts, gendered pronouns, singular/plural gendered token counts, younger/older gendered token counts and uppercase/lowercase pronouns.

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

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

Racial Proxy Tokens. The tokens we use as proxies for race are as follows.

Colors. The color word tokens used as proxies for race and falsification words are the following: black, blue, brown, gold, golden, green, orange, pink, purple, red, silver, violet, white, yellow. For parsimony, in the paper we only show the words black, blacks, white, whites, blue, and blues.

Nationalities. afghan, african, albanian, algerian, american, andorran, angolan, antiguans, apache, argentinean, armenian, asian, australian, austrian, azerbaijani, bahamian, bahraini, bangladeshi, barbadian, barbudans, batswana, belarusian, belgian, belizean, beninese, bhutanese, bolivian, bosnian, brazilian, british, bruneian, bulgarian, burkinabe, burmese, burundian, cambodian, cameroonian, canadian, cape verdean, chadian, cherokee, chicana, chicano, chicanx, chilean, chinese, choctaw, colombian, comoran, congolese, croatian, cuban, cypriot, czech, danish, djibouti, dominican, dutch, dutchman, dutchwoman, ecuadorean, egyptian, emirian, english, eritrean, estonian, ethiopian, fijian, filipino, finnish, french, gabonese, gambian, georgian, german, ghanaian, greek, grenadian, guatemalan, guinean, guinean, guyanese, haitian, herzegovinian, hispanic, honduran, hungarian, icelander, i-kiribati, indian, indonesian, iranian, iraqi, irish, irish, iroquois, israeli, italian, ivorian, jamaican, japanese, jordanian, kazakhstani, kenyan, kittian, korean, kuwaiti, kyrgyz, laotian, latina, latino, latinx, latvian, lebanese, leonean, liberian, libyan, liechtensteiner, lithuanian, lucian, luxembourger, macedonian, malagasy, malawian, malaysian, maldivan, malian, maltese, marinese, marshallese, mauritanian, mauritian, mexican, micronesian, moldovan, monacan, mongolian, mongols, moroccan, mosotho, motswana, mozambican, namibian, nauruan, navajo, nepalese, netherlander, nevisian, nicaraguan, nigerian, nigerien, ni-vanuatu, norwegian, ojibwe, omani, pakistani, palauan, panamanian, paraguayan, persian, peruvian, polish, portuguese, qatari, rican, romanian, russian, rwandan, salvadoran, samoan, saudi, scottish, senegalese, serbian, seychellois, singaporean, sioux, slovakian, slovenian, somali, spanish, sri-lankan, sudanese, surinamer, swazi, swedish, swiss, syrian, taiwanese, tajik, tanzanian, thai, timorese, tobagonian, togolese, tomean, tongan, trinidadian, tunisian, turkish, tuvaluan, ugandan, ukrainian, uruguayan, uzbekistani, venezuelan, vietnamese, welsh, yemenite, zambian, zealander, zimbabwean.

E Seattle Public Library Checkouts Data

To study the impact of being honored by the children's book awards we examine, we analyze data from the Seattle Public Library system on all public checkouts from the library between April 2005 and September 2017.[74] 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 of collection-by-day checkout likelihoods scaled by the number of books in the collection to generate a measure of the number of checkouts per book, per day, in each of the three collections. We limit checkout data for each book to the calendar year before the award was given and the two following calendar years.

To generate Figure 2, 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

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

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 2 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 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 $Diversity$ collection variables. Our main coefficients of interest are β_2 and β_3.

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 constant (the $Diversity$ 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 0.483. 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, the constant, and the coefficient on *Post*, which sums to approximately 0.120.

An alternate interpretation is that after winning the award, the Mainstream collection books are approximately 2.6 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*, the constant, 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

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

Parameter	**Estimate**
Non-Recognized Children's Books in Library Fixed Effect	0.019***
	(0.004)
Mainstream Collection Fixed Effect	0.107***
	(0.006)
Diversity Collection Fixed Effect (constant)	0.075***
	(0.004)
Post	0.026***
	(0.005)
Post × Mainstream Collection	0.274***
	(0.007)
Post × Diversity Collection	0.017**
	(0.007)
Observations	3,375
Adjusted R^2	0.773

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

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 1.7 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 0.118. 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 0.120. Prior to receipt of the award, they were approximately 20 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 2 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 and Honors

Parameter	**Estimate**
Non-Recognized Children's Books in Library Fixed Effect	0.026***
	(0.006)
Mainstream Winner Fixed Effect	0.117***
	(0.008)
Mainstream Honoree Fixed Effect	0.101***
	(0.008)
Diversity Winner Fixed Effect	0.001
	(0.008)
Diversity Honoree Fixed Effect (Constant)	0.064***
	(0.008)
Post	0.031***
	(0.007)
Post × Mainstream Winner	0.500***
	(0.010)
Post × Mainstream Honoree	0.205***
	(0.010)
Post × Diversity Winner	0.014**
	(0.010)
Post × Diversity Honoree	0.006**
	(0.010)
Observations	5,620
Adjusted R^2	0.775

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. 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.[75] 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.

[75] 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 has previously won 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.[76] 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.

[76] 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 has previously won 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