WHAT WE TEACH ABOUT RACE AND GENDER: REPRESENTATION IN IMAGES AND TEXT OF CHILDREN'S BOOKS*

Anjali Adukia
Alex Eble
Emileigh Harrison
Hakizumwami Birali Runesha
Teodora Szasz

*For helpful feedback, we thank two editors, four anonymous referees, Milena Almagro, Barbara Atkin, Karen Baicker, Chris Blattman, Anna Brailovsky, Tom Brock, Mary Gen Davies, Steven Durlauf, Lisa Englehardt, James Evans, James Fenske, Alex Frankel, Adam Gamoran, Claudia Goldin, Josh Goodman, Austan Goolsbee, Jon Guryan, Andrew Ho, Rick Hornbeck, Caroline Hoxby, Anders Humlum, Larry Katz, Dennis LeLoup, Ro'ee Levy, Susanna Loeb, Jens Ludwig, Jonathan Meer, Martha Minow, Sendhil Mullainathan, Shirley Mullin, Derek Neal, Matt Notowidigdo, Steve Raudenbush, Cybele Raver, Heather Sarsons, Fred Stafford, David Uminsky, Miguel Urquiola, Angelino Viceisza, Alessandra Voena, Melanie Xue, and others, including many seminar audiences. For financial support, we thank UChicago Becker Friedman Institute, UChicago Center for Data and Applied Computing, National Academy of Education/Spencer Foundation Postdoctoral Fellowship, 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. We thank Callista Christ, Anjali Das, Jarvis Lam, Erica Lin, Ping-Jung Liu, Simon Mahns, and Celia Zhu for their invaluable contributions. We also thank Fabiola Alba-Vivar, Marliese Dalton, Saloni Gupta, Zipporah Klain, Juan Miguel Jimenez, Noah McLean, Karla Monteiro, Jan Morgan, Raj Shukla, and other MiiE Lab Research Assistants for excellent research assistance. We thank Ashiyana and Kairav Adukia-Hornbeck for manual coding assistance. For access to important resources, we thank UChicago Research Computing Center, Heather McGuire, Bridget Madden, and Kalli Mathios. The research reported here was also supported by Google Cloud Research through research credits. Researchers' 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.

The Quarterly Journal of Economics (2023), 1–61. https://doi.org/10.1093/qje/qjad028.

Books shape how children learn about society and norms, in part through representation of different characters. We use computational tools to characterize representation in children's books widely read in homes, classrooms, and libraries over the past century and describe economic forces that may contribute to these patterns. We introduce new artificial intelligence methods for systematically converting images into data. We apply these tools, alongside text analysis methods, to measure skin color, race, gender, and age in the content of these books, documenting what has changed and what has endured over time. We find underrepresentation of Black and Latinx people in the most influential books, relative to their population shares, though representation of Black individuals increases over time. Females are also increasingly present but appear less often in text than in images, suggesting greater symbolic inclusion in pictures than substantive inclusion in stories. Characters in these influential books have lighter average skin color than in other books, even after conditioning on race, and children are depicted with lighter skin color than adults on average. We present empirical analysis of related economic behavior to better understand the representation we find in these books. On the demand side, we show that people consume books that center their own identities and that the types of children's books purchased correlate with local political beliefs. On the supply side, we document higher prices for books that center nondominant social identities and fewer copies of these books in libraries that serve predominantly White communities. *JEL codes:* I24, I21, Z1, J15, J16.

I. Introduction

Education teaches children about the world, its people, and their place in it. Much of this happens through the books society presents to children in school and at home (Giroux 1981; Cantoni et al. 2017). These lessons can be conveyed in part through messages transmitted about identity—for example, by the presence or absence of different identities. Such messages, in books and beyond, can influence children's beliefs about themselves and others, their effort, and their learning (Plant et al. 2009; Fuchs-Schündeln and Masella 2016; Riley forthcoming). Given persistent racial and gender inequality, better understanding of the representation contained in the images and text of books may help us better understand and address these and related structural inequities.

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

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

We report a series of results from applying our computational tools to measure representation in these books. We begin by describing our results measuring the representation of skin color, race, and age over time and across collections. We find that over time, these books include more characters with darker skin, but those in the Mainstream collection are significantly more likely to depict lighter-skinned characters than those in the Diversity collection. This pattern remains even when comparing pictured characters with the same predicted race classification. In both collections, children are more likely than adults to be shown with lighter skin, despite there not being a definitive biological foundation for any systematic difference. Regardless of the reasons behind this difference, our estimates show that lighter-skinned children see themselves represented more often in these books than do darker-skinned children. In addition, we show that Black and Latinx people have been historically underrepresented relative to their share of the U.S. population, corroborating prior work on the representation of race in smaller subsets of these collections of books (e.g., Valadez, Sutterby, and Donaldson 2013; Koss 2015). Our analysis of age reveals a surprising result: even though these books are targeted to children, adults are depicted more often than children in both images and text.

We characterize the representation of gender across collections and over time. Comparing the presence of females in text and in images, we find that females are consistently more likely to be visualized (seen) in images than mentioned (heard) in the text. This suggests there may be symbolic inclusion of females in pictures without their substantive inclusion in the actual stories. Looking over time, we find that females are persistently less likely than males to be represented in the text of books in our sample overall and over time. This finding is consistent across all of the measures we use: pronoun counts, specific gendered terms, gender of famous individuals, and predicted gender of character first names. This generalizes results from prior analysis of the representation of gender in studies focusing on smaller subsets, or a small number of specific features contained in these books (Weitzman et al. 1972; Crisp and Hiller 2011).

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

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

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

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

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

In summary, this article makes three key contributions. First, we develop and hone a series of tools from the field of computer vision to systematically process images into analyzable measures of representation; this includes introducing a novel computational method to measure skin color. Second, we apply these image-to-data tools alongside established natural language processing tools to measure the representation of skin color, race, gender, and age in the images and text contained in a century of influential children's books and document how this changes over time. Third, we describe economic forces on the supply side and demand side that may contribute to these levels of representation, and then present empirical evidence showing how the pressures from these forces may contribute to persistent overrepresentation of historically dominant identities. Using data on local book consumption and local consumer beliefs, we show that the levels of representation contained in the books people buy are highly correlated with their views on race and immigration. Given that the books used to teach children shape the beliefs these people hold when they are adults (Cantoni et al. 2017; Arold, Woessmann, and Zierow 2022), the patterns in children's book purchases that we document may help explain the persistence and intergenerational transmission of related beliefs (Dhar, Jain, and Jayachandran 2019; Eble and Hu 2022).

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

1. The Online Appendix includes further analysis; details on award criteria; a discussion of the benefits, limitations, and validity of computational content analysis; further information on the methods and supplementary data; limitations of the economic analysis; and qualitative interviews with suppliers of children's books.

II. Background: The Importance of Representation

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

One mechanism through which inequality of representation may contribute to inequality in outcomes is through its potential to instill beliefs about who belongs in which societal domains (Bian, Leslie, and Cimpian 2017; Rodríguez-Planas and Nollenberger 2018). In particular, 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 potentially self-reinforcing loop: not seeing such examples may diminish a child's expected return to effort. If that change in expectation reduces actual effort, it may lower performance, thus reinforcing the message behind the (once-erroneous) message. This highlights the importance of addressing inequality in representation in educational content.

Curricular materials are designed and used with the intent to shape children's development and their views of the world and are likely to make important contributions to the formation of children's social preferences (Cappelen et al. 2020; Alan et al. 2021). Exposure to variation in content among textbooks, ranging from subjects as diverse as history and religion, can also lead to variations in later-life beliefs (Fuchs-Schündeln and Masella 2016; Arold, Woessmann, and Zierow 2022). Evidence from psychology shows that deliberately manipulated exposure to content can, but does not always, shape child beliefs (Hughes, Bigler, and Levy 2007). 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 (Levstik and Tyson 2010).

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

Broadening representation to be more inclusive also has been shown to influence the beliefs, actions, and learning of children. In economics alone, changes in representation have been shown to influence these outcomes for women (Stout et al. 2011; Porter and Serra 2020), and, separately, people of underrepresented racial and ethnic identities regardless of gender (Kearney and Levine 2020; Riley forthcoming). While not a panacea, such "subject-object identity match"—for example, teacher-student identity match, or content-reader identity match—can help improve academic performance for students by changing their own and others' beliefs, among other potential channels. This may function via a wide range of potential pathways, such as by reducing stereotype threat, changing one's own beliefs, and by changing others' beliefs (Steele and Aronson 1995; Wilson 2012; Alrababah et al. 2021).

We also draw on a central insight from the study of intersectionality. Different aspects of identity—such as race, gender identity, class, sexual orientation, and disability—do not exist separately from each other but are inextricably linked (Crenshaw 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. We highlight that intersectionality does not merely refer to an "interaction effect" (e.g., between race and gender), but the distinct experiences of individuals whose identities exist at intersections of multiple dimensions of marginalization.

III. Context: Award-Winning Children's Books

We focus on the content of a series of books that are particularly likely to appear in the homes, schools, and libraries of a large proportion of children in the United States. Specifically, we study the representation contained in the images and text of books recognized by any of 19 awards administered or featured by the Association for Library Service to Children (ALSC), a division of the American Library Association (ALA). These groups began honoring children's books in 1922, and continue to the present.

In this section, we describe these books and how we group them by award type. We provide descriptive analyses quantifying changes in book consumption associated with being recognized by these awards.

III.A. Collections of Books

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

1. *Mainstream Collection.* The Mainstream collection comprises books recognized by either the Newbery or Caldecott awards, the two oldest children's book awards in the United States. 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." Books receiving these awards are considered to be of general interest to all children. We provide further evidence demonstrating the importance of these books in Section III.B. We use the term

"Mainstream" to capture the influence of these awards on book consumption (Smith 2013). We do not assert any centrality or default for this collection beyond the historical prominence of these books. The primary goal for studying these books is to understand the representation contained in a set of books to which a large proportion of children in the United States are exposed.

2. *Diversity Collection.* The Diversity collection comprises book awards featured by the ALSC that center the experiences of excluded or marginalized identities. These books are also likely to be found on "diversity lists" during events such as Black History Month or Women's History Month. We study the representation contained in these books for multiple reasons: one, to estimate a potential upper bound on representation in children's books in the market; two, to measure the efficacy of these books in highlighting the identity on which they focus; and three, to measure the levels of representation of historically excluded identities beyond the identity on which a given award focuses. We use this last feature to assess the extent to which these books have greater, similar, or less representation of identities that exist at the intersection of multiple sites of exclusion.

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

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

Each award has a single "winner" or "medalist" of the award. Many awards also recognize a set of other leading contenders for the award in a given year; these are often called "honorees." In our main analysis we refer to the superset of these two groups as those "recognized" by the award. In some analysis in Section III.B, we examine trends in consumption separately by winners and honorees. In Online Appendix D, we describe the criteria used by each award for recognizing books in greater detail.

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

III.B. Quantifying the Importance of Mainstream Awards

Mainstream awards are considered to be highly influential, with recognition by either the Newbery or Caldecott Medals placing books into the "canon" of children's literature and making 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 catalogs to school curricula and summer reading lists (Knowles and Smith 1997). Publishers in the industry take cues from winners for guidance in what to publish, given the large boost in sales that the award stimulates, and many children's librarians ensure award-winning books' presence in their inventories (Nilsen 1971; Cockcroft 2018).

We further establish the importance of these awards in children's experiences by estimating the relationship between receipt of these awards and book popularity. Our analyses use data on three measures of book consumption: (i) library checkouts, (ii) book purchases, and (iii) internet searches. Each measure captures a different—but not mutually exclusive—set of consumer preferences. We describe the data below and go into further detail in Online Appendix E.

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

TABLE I
SUMMARY STATISTICS

	Mainstream (1)	Diversity (2)	People of color (3)	African American (4)	Ability (5)	Female (6)	LGBTQIA+ (7)
Collection totals							
Total number of books	495	635	577	130	29	14	15
Range of years in our sample	1923–2019	1971–2019	1971–2019	1971–2017	2000–2014	2013–2017	2010–2017
Book-level averages: book attributes							
Number of pages	139	148	137	147	213	314	268
Number of words	24,362	26,520	23,816	26,328	35,273	87,411	56,771
Number of faces	44	59	60	41	30	30	79
Number of famous people	3	8	8	10	5	40	13
% faces - monochromatic skin color	58	47	47	52	45	55	45
Book-level averages: skin color							
Perceptual skin tint of all faces	55	44	44	41	46	34	47
Book-level averages: putative race							
% faces classified as Asian	6	16	16	11	6	9	4
% faces classified as Black	2	13	13	22	8	21	3
% faces classified as Latinx + Others	4	3	3	3	4	1	5
% faces classified as White	88	68	67	64	82	69	88
% famous people classified as Asian	3	7	7	1	3	8	5
% famous people classified as Black	5	22	23	55	8	21	8
% famous people classified as Indigenous	0	1	1	0	0	1	0
% famous people classified as Latinx	1	9	9	0	1	0	2
% famous people classified as multiracial	0	2	2	1	0	1	3
% famous people classified as White	92	59	56	43	87	68	81
Book-level averages: gender							
% faces classified as female	48	50	49	43	67	71	48
% female gendered words	34	43	42	40	42	56	45
% famous people classified as female	15	22	20	24	28	37	41
Book-level averages: age							
% faces classified as children	19	14	14	10	19	3	18
% young gendered terms	26	20	20	21	17	21	32

Notes. We present summary statistics (described in the row titles) for each collection of books that we analyze (named in the column titles). Percentages may not sum to 100 due to rounding error.

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

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

We present three event studies that show average daily library checkouts (Figure I, Panel A), average daily purchases (Figure I, Panel B), and average weekly search interest by collection (Figure I, Panel C), centered around the time when awards are announced. In Figure I, Panels A and B, we disaggregate the data by Mainstream winners (medalists) or honorees in that year, Diversity winners or honorees in that year, and all other children's books.

First, we find that library checkouts of books selected for Mainstream awards increase substantially after announcement of awards. Further, we estimate an even larger increase for award winners relative to books receiving an honorable mention.[2] This persists for at least two years after the award announcement, during which time average daily checkouts of books in the Mainstream collection plateau at a rate approximately four times that of the comparator groups. The increase in library checkout rates

2. 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.

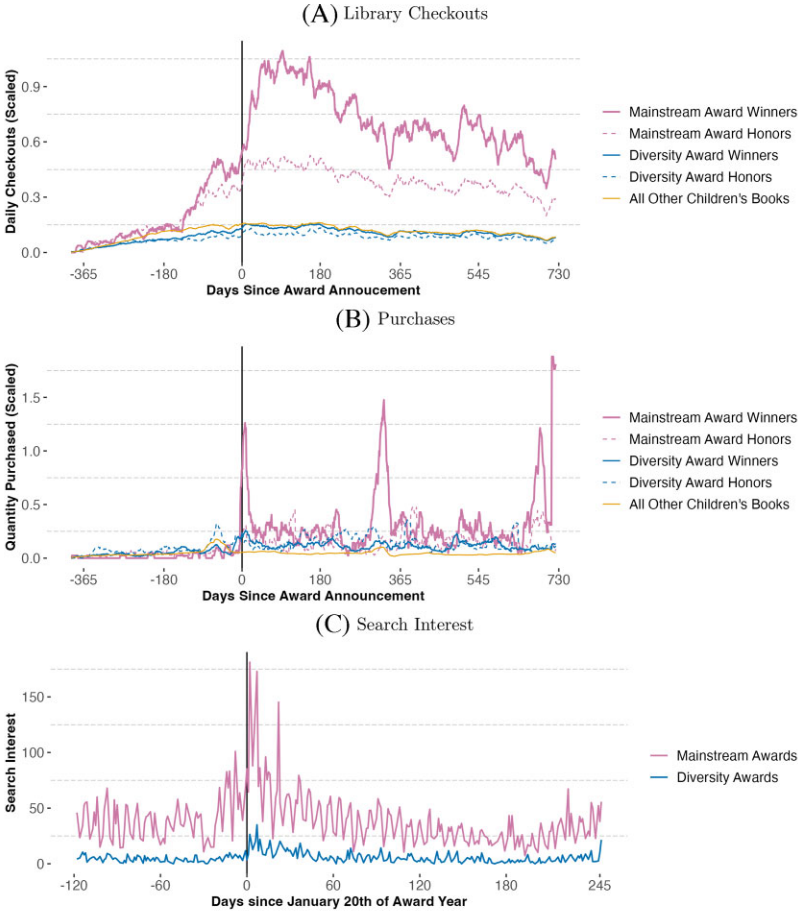

FIGURE I

Children's Book Readership Centered around Award Announcements

Panel A shows average daily checkouts of children's books in 2005–2017 from the Seattle Public Library. Panel B shows average daily children's book purchases in 2017–2020 from the Numerator OmniPanel. We scale daily checkouts and purchases by the number of unique titles in each collection and smooth the data with a 14-day moving average. Panel C shows average weekly search interest in the United States in December 2016 – December 2021 from Google Trends. Further information can be found in Online Appendix E. Panels are centered around the time of award announcements each year. Panel C's x-axis label and centering differ from Panels A and B because its data are measured at a weekly, rather than daily, frequency.

for books in the Diversity collection after the award announcements is substantially smaller in magnitude and, as expected, we see no change around the award announcement in checkout rates for other children's books. We discuss this analysis in greater detail in Online Appendix F.

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

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

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

IV. Prior Work and the Need for Computational Measurement Tools

The field of content analysis studies the content of books, including the representations contained in them. Historically, content analysis has been conducted primarily by humans reading carefully through images, text, or other media while coding the presence of certain words, themes, or concepts by hand (Neuendorf 2016; Krippendorff 2018). Prior work has studied the content of some of these award-winning books, including the representation of gender and, more recently, race. An influential study by Weitzman et al. (1972) examined the gender representation throughout the text of 18 Caldecott Medal recipients published over a five-year period, documenting that females were less likely than males to be represented in the content of the books; when they were depicted, these portrayals often reinforced traditional gender roles. Many studies since have measured the representation of race, gender, and other identities in various, smaller subsets of these books published in specific time windows (e.g., Williams et al. 1987; Koss 2015). They find that the books in their samples often underrepresent women and people of color, relative to males and White people, though there is not consensus as to whether these patterns attenuate or persist over time. These differences often coincide with differences in focus, choice of sample, or time period.[3]

There are strengths and limitations to manual content analysis. A key strength is its ability to capture narrative structure, societal norms, and other complex messages that content may contain. On the other hand, the time and other costs it takes to perform manual content analyses constrain the sample size and scope of the analysis that can be performed in a study. The sample sizes of most studies range from between a few dozen books to—with rare exceptions—at most 100 or 200. The few studies with a larger scope (500 to several thousand books) focus only on one or a small number of sites of representation—for example, the title of the book, the illustrations on its cover, or the main character instead of the representations contained in the full content of the book.

The work of content analysis can also be approached using computational methods. In this approach, scholars use tools from computer science to analyze content, drawing on fields such as computer vision and natural language processing, which involves leveraging machines to read and parse messages contained in the images and text of printed material. In this study, we apply and develop computational methods to measure representation in both the images and the text of these books, building on the rich content analysis literature.

There are a set of key advantages—and thus advances—of a computational approach. These advantages include, but are not limited to, improved speed and reduced cost, which allow for the study of more books; greater scope for measurement in each book; greater flexibility and scalability; increased reliability; and greater cost-effectiveness. We discuss these advantages in more detail in Online Appendix G. In that section, we also discuss two important dimensions of our work. First, we explain

3. We provide a bibliographic list of a selection of these studies in Online Appendix Table AII.

how manual and computational content analysis reflect human-introduced biases in measurement and describe how these biases can be minimized. Second, we describe how we use manual content analysis to validate our computational measures of representation. We note that given the strengths and limitations of each approach, computational content analysis and manual content analysis should be seen as complementary rather than substitute approaches to understanding the messages in any given book.

Our computational approach also allows us to advance and expand the scope of analysis exploring whether there is differential representation of identities at the intersections of multiple sites of marginalization in dimensions of skin color, race, gender, and age. This analysis draws on a central insight from the large body of work on intersectionality: when analyzing representation of different dimensions of identity, such as race and gender, it is critical to characterize the power imbalances and their manifestations that lead to greater disadvantage among people at the intersection of multiple marginalized identities.[4] The inclusion or exclusion of identity groups in the content we study is a fundamental expression of power for two reasons. One, it signals to the reader the spaces that these identities do or do not occupy in society (Crenshaw 1990). Two, it has the potential to shape the beliefs, norms, and conceptions of history that the next generation will adopt (Fuchs-Schündeln and Masella 2016; Cantoni et al. 2017; Arold, Woessmann, and Zierow 2022).

V. Methods and Data

In this section, we describe the methods we use to create data from the images and text in books.[5]

V.A. Methods: Images as Data

Currently, images are neither widely nor systematically analyzed in social science research despite the richness of information they contain, as alluded to by the maxim "a picture is worth a thousand words." This leaves an important data source

4. We acknowledge that a more developed intersectional analysis requires a wide-reaching analysis of norms, rules, laws, and history that is beyond the scope of our study.

5. We include shareable code and relevant resources at https://github.com/miielab/replication_qje_whatweteach.

"on the table" (i.e., unused), in contrast to the use of text as data, which has seen growing attention from social science in the past 15 years (Gentzkow, Shapiro, and Taddy 2019; Kozlowski, Taddy, and Evans 2019). Images may be particularly important in children's books, especially for those who are not yet textually literate (Sadoski and Paivio 2013). Relatedly, the use of curricular materials with pictures and text can lead to better comprehension, as compared to those with text only (Fletcher and Tobias 2005; Eitel et al. 2013).

We introduce and develop tools for computational analysis of the content of images. These tools first identify pictured faces of characters and then classify their skin color, "putative" race (defined as the race that society assigns to a person), gender, and age. We depict this process in Figure II, Panel A and refer to it as our "Image-to-Data Pipeline."

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

To address the potential undercounting of characters in illustrations, we trained a custom transfer learning model to detect and classify both illustrated and photographic faces using Google's AutoML Vision (Zoph and Le 2017).[6] Transfer learning

6. At the time of writing, Google was in the process of migrating the relevant workflows from AutoML to Vertex AI. The two have similar functionality, but our models in this article used AutoML. People who wish to use these approaches in future will use Vertex AI.

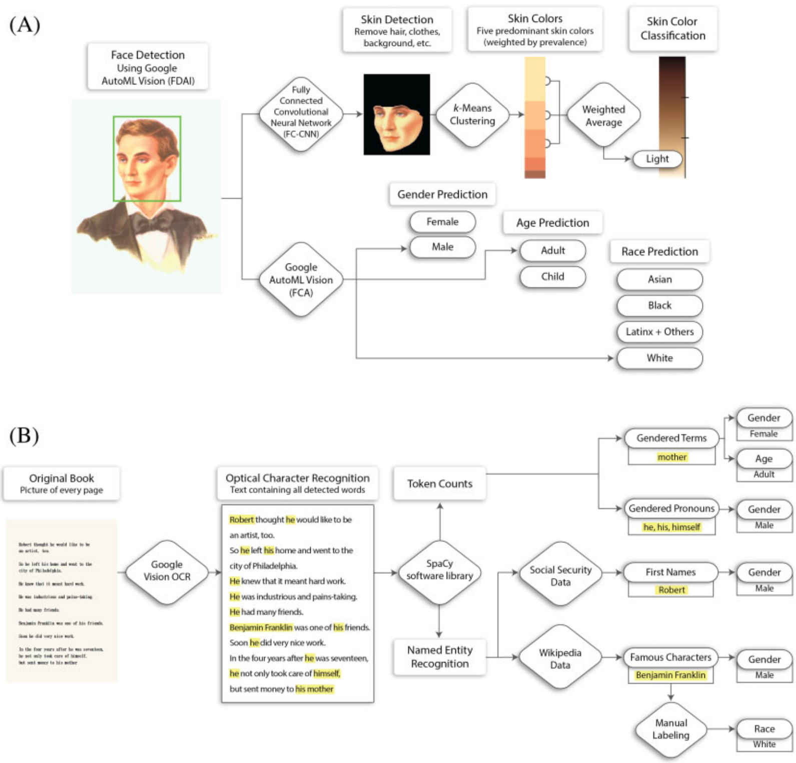

FIGURE II

Converting Images and Text into Data

In this figure, we show how we process scanned book pages into image and text data. In Panel A, we show how we extract image data to construct image measures of skin color, race, gender, and age. The image in Panel A is from Whipple (1915). In Panel B, we show how we extract and isolate various dimensions of text to construct textual measures of gender, race, and age. The text in Panel B is from Macomber (1897).

is a process that facilitates the use of a pretrained model as a "shortcut" to learn patterns from data on which it was not originally trained. This mitigates concerns around having a sufficiently large amount of manually labeled data necessary to train deep-learning models, particularly in the absence of public data sets using illustrations. We trained our face detection model using a manually labeled data set of 5,403 illustrated faces from our sample, which contains a wide variety of illustrated characters.[7] This process is described in greater depth in Szasz et al. (2022), and we present further detail on it in Online Appendix H.

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

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

i. *Skin segmentation.* We begin by isolating skin components from nonskin components of each detected face using a deep-learning approach called Fully Connected Convolutional Neural Network Continuous Conditional Random Field (FC-CNN CRF).[8] This process of "skin segmentation" comprises three steps (Jackson, Valstar, and Tzimiropoulos 2016; Zhou, Liu, and He 2017; Beyer 2018; Lu 2018). First, we apply a fully connected convolutional neural network (FC-CNN).[9] This allows us to

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

8. Further information about how our skin segmentation approach improves on traditional approaches can be found in Online Appendix H.A.2.

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

predict periphery landmarks such as the edges of the facial skin area, eyes, nose, and mouth. Second, we then use these predicted landmarks to extract a convex hull "mask" for the targeted facial region. Third, we refine this mask by applying a continuous conditional random field (CRF) module, which predicts the labels of neighboring pixels (i.e., whether they are predicted to be skin or not skin) to produce a more fine-grained segmentation result. We measure skin color using the resulting face mask.

ii. *Representative skin color.* We identify the predominant colors in this face mask (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. 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 nonskin portions of the detected face. We take the centroid of the remaining three largest clusters—which provide the dominant skin colors in the segmented skin—and use a linear mapping to convert these values from RGB color space into the CIELAB, or L*a*b*, color space.[10] 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 the top three dominant skin colors came. This weighted average provides our measure of each face's representative skin color.

iii. *Skin color classification: Perceptual tint and skin color type.* Once we have a representative skin color, we can measure how light or dark the skin color of each face is on a scale of 0–100 (where 0 is the darkest and 100 is the lightest) using the L* value from the representation of each face's representative skin color in L*a*b* color space. This measure reduces the dimensionality of skin color to a single value and provides our main skin color

10. We convert colors from RGB space to L*a*b* space before averaging because L*a*b* color space—unlike RGB color space—is perceptually linear.

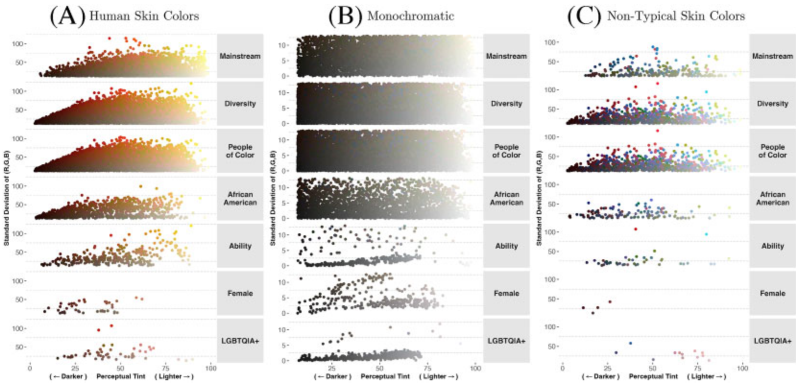

FIGURE III

Skin Color Data, by Color Type

This figure shows the representative skin colors of the individual faces we detect in the images found in the books from each collection (color version available online). We show these by the three color "types" present in these images: human skin colors (polychromatic skin colors where R $\geq$ G $\geq$ B), monochromatic skin colors (e.g., black and white, sepia), and nontypical polychromatic skin colors (e.g., blue, green). We discuss how we separate skin colors into these three types in Online Appendix H.A.3. The y-axis indicates the standard deviation of the RGB values of each face. The higher the standard deviation, the more vibrant the color.

measure of interest, which we call "perceptual skin tint."[11] A given numerical change in the skin tint value can be interpreted as a similar perceived change in the darkness/lightness of a color. We also divide this continuous measure of skin tint into three terciles (darker, medium, or lighter) for a coarser, but more intuitive, skin color classification.

We separate the representative skin colors into three types: (i) polychromatic human skin colors (e.g., brown, beige), (ii) monochromatic skin colors (e.g., grayscale), and (iii) polychromatic nontypical skin colors (e.g., blue, green). We discuss how we separate skin colors into these three types in Online Appendix H.A.3. In Figure III, we show the representative skin colors of over 44,000 individual faces detected in each collection by the

11. A more common term for L* is "perceptual lightness," but to decenter and deemphasize "lightness" or "brightness" relative to "darkness," we refer to the concept as "perceptual tint" or "skin tint."

three skin color types present in these images.[12] The x-axis indicates perceptual tint, and the y-axis indicates vibrancy of each representative skin color.

3. *Image Feature Classification: Race, Gender, and Age.* To classify putative race, gender, and age of detected faces in images, we trained a multilabel classification transfer learning model using Google's AutoML Vision platform. This model was trained on the UTKFace public data set, which contains over 20,000 faces manually labeled with race, gender, and age (Zhang, Song, and Qi 2017).[13] Our model assigns probabilities that a detected face is of a given race, gender, and age, respectively. In each dimension, we classify a face with the identity to which the model gives the highest predicted probability.

There are various limitations of this model. First, it was trained on photographs, which means that the predictions will be more accurate for photographs of faces than for illustrated faces.[14] Second, previously, many existing artificial intelligence models that classified putative race had a high error rate, 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 nonhuman (Fu, He, and Hou 2014; Krishnan, Almadan, and Rattani 2020). Ongoing work attempts to recognize and address these disparities (Buolamwini and Gebru 2018; Mitchell et al. 2019). Third, we acknowledge that race is a human-made construct that exists for political and economic purposes (Roberts 2011; Logan 2022)—and as a result, any attempt to classify race with either a human or a computer is an imperfect exercise that will yield imperfect results. Conditional on the imperfect nature of this

12. We show these for each collection by decade for human skin colors (Online Appendix Figure BII), monochromatic skin colors (Online Appendix Figure CI), and nontypical skin colors (Online Appendix Figure CII).

13. 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)). The resulting model has 90.6% precision and 89.0% recall in our testing data. We provide additional detail in Online Appendix H.

14. 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.

enterprise, however, classifying race using a computer rather than humans has a key advantage: its classification rules—and any error therein—are consistent across all content that we measure (i.e., racial categories are classified in the same manner in both the Mainstream and Diversity collections). Fourth, when labeling gender, we recognize that our classifications are binary and therefore incomplete. They also focus only on the performative aspect of gender presentation, as they are trained based on how humans classify images. Furthermore, because we are classifying character gender based on the character's appearance, our measurements use the same binarized gender classification to assess the perceived presentation of gender, that is, whether the character is female-presenting or male-presenting, rather than female or male per se. Future work should incorporate the classification of fluid and nonbinary gender identities.

V.B. Methods: Text as Data

In this section, we describe the tools we use to measure representation in the text of books. Researchers have manually analyzed the messages contained in text of printed material for centuries, a process that is highly resource intensive in terms of labor and time (Neuendorf 2016; Krippendorff 2018). Recent work by economists and sociologists showcases how the computational speed and power of (super)computers can be harnessed to conduct computational text analysis, greatly accelerating the speed of work that 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—for example, from a book—and extract various features of interest. In Figure II, Panel B, 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 Online Appendix H.B.

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

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

2. *Text Analysis: Token Counts (Gender and Age).* We generate counts of different "tokens"—maximal sequences of non-delimiting consecutive characters; in our context, individual words—associated with gender and age. To calculate gender representation in text, we calculate the number of female and male pronouns along with a list of other gendered terms such as "queen" and "husband". To measure representation of age in text, we generate lists of gendered terms associated with children or "younger" individuals (e.g., girl, son) and gendered terms associated with adults or "older" individuals (e.g., woman, dad). The vocabulary used for each of these lists is shown in Online Appendix H.B.3.

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

i. *Famous individuals.* Exposure to salient examples of historical figures or celebrities from marginalized backgrounds can lead to meaningful changes in social attitudes toward 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 Jr. 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 preexisting data set, Pantheon 2.0, that contains data from over 70,000 Wikipedia biographies (Yu et al. 2016). This provides information on gender for each famous individual. We then manually code putative race for each identified person.

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, we classify them as multiracial. 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.

ii. *Character first names.* We measure the gender of characters names who are identified via NER and tagged by the NER model to be a person but are not identified as "famous." We extract the first word (name) of each of these named entities and estimate the probability that it is female (or male) using data on the frequency of names by gender in the U.S. population from the Social Security Administration (SSA). This yields an estimate of the probability that a name is associated with a given gender over the whole time period (as opposed to in each time period). Using this method, we are able to make gender predictions for approximately 60,000 names. If the predicted probability that a name is female is greater than 50%, we classify the name as female. Otherwise, we classify the name as male. For example, in the SSA data, the proportion of people named Cameron who identify as female is 9.16%. We therefore assign a probability of 90.84% that the name Cameron is male, and classify it as male.

V.C. Data Collection, Aggregation, and Analysis

To analyze representation, we collected and digitized the books recognized by the awards in our sample, using both library and online sources. Our final sample comprises 1,130 books recognized by at least one award over the period 1923–2019. This includes books that are award winners (sometimes called medalists) and books receiving an honorable mention from an award.[15] We divide these books into different collections, as described in Section III.A. We transform digitized page scans into data on the images and text in these books using the methods described in this section.

15. The 19 award corpora comprise 3,447 total books that either won an award or received an honorable mention. Our sample contains all but 16 Mainstream medalists: 3 Newbery and 13 Caldecott winners.

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

To generate our estimates of representation, we first summarize each measure at the book level, and then calculate the average across all books in a given collection, overall and over time. For example, to estimate the average percent of female faces in a collection, we first calculate the percent of female faces in each book in the collection and then take the average across books. This ensures that each book contributes equally to our collection-level measures of skin color, race, gender, and age representation, regardless of book length. We generate these estimates at the book level and aggregate them to the collection level, overall and, separately, over time. While different awards commence in different years, we study all books ever recognized by these awards, rather than limiting the analysis to years in which all awards are active.

VI. Results

In this section, we describe patterns of representation of skin color, race, gender, and age in the images and text of these books across collections and over time.[16]

VI.A. Skin Color

We begin by characterizing patterns, across collections and over time, in the skin color of the characters pictured in images. We focus our discussion on characters with human skin colors. Results for characters with monochromatic or nontypical skin colors can be found in Online Appendix C; these show patterns

16. A previous version of this article (available here: https://doi.org/10.3386/w29123) includes some results that were removed in the revision process.

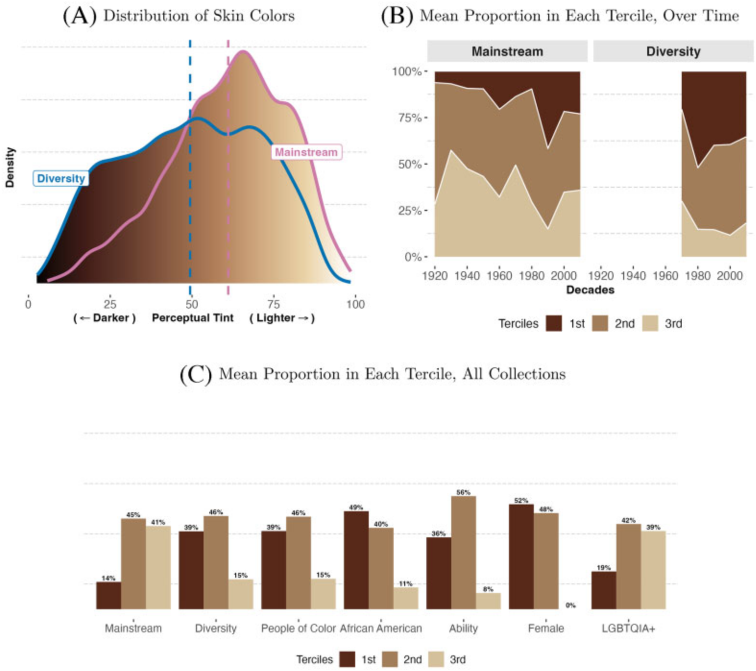

FIGURE IV

Skin Colors in Faces, by Collection: Human Skin Colors

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 the seven collections. Skin classification methods are described in Section V.A.

similar to those for characters with human skin colors. Figure IV, Panel A shows the distribution of perceptual tint for detected faces in the Mainstream and Diversity collections. This figure shows that the faces in the Diversity collection have darker skin tints, on average, than those in the Mainstream collection.[17]

17. Online Appendix Figures CIII and CIV demonstrate that this result holds regardless of image color type: monochromatic or nontypical skin colors.

A Kolmogorov-Smirnov test rejects the equality of the two distributions ($p < .001$); in other words, the distributions of skin colors in pictured characters in the two collections are statistically distinct. Furthermore, the distribution of skin color tint in the Mainstream collection has a much smaller variance than that of the Diversity collection: a test of the null hypothesis that the two variances are equal also rejects equality with $p < .001$. This implies that there is a greater diversity of skin color tint shown in the Diversity collection.

We examine the proportion of character faces in each skin color tercile—darker, medium, or lighter. Over time, the proportion of characters with skin colors in the darker and medium skin color terciles increases relative to those in the lighter skin color tercile in both the Mainstream and the Diversity collections (Figure IV, Panel B). Change is slower for Mainstream than Diversity: the distribution of skin color across the three terciles in books in the Mainstream collection 2010–2019 is similar to that in the Diversity collection 1970–1979. A related but distinct parameter of interest is the mean value of perceptual skin tint. Unlike our result for the distribution of skin color in faces across terciles, we find that average perceptual tint has changed less over time (Online Appendix Figure CV).

Figure IV, Panel C shows the proportion of faces in each skin color tercile 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 40% of faces), whereas in the Diversity collection, darker skin is the second most common skin color tercile (approximately 40% of faces). This suggests that the Diversity collection is more representative of characters that have darker skin tints. Of the seven collections, the Mainstream collection has the lowest proportion of faces falling in the darker skin color tercile and the Female collection has the greatest proportion.[18]

18. Online Appendix Figure CII shows that the method of classifying “human” versus “nontypical” polychromatic skin colors may underestimate the number of darker-skinned faces if the browns that are used do not follow the polychromatic $R \geq G \geq B$ rule as described in Online Appendix H. However, Online Appendix Figure CIV shows that this does not change the patterns in skin color representation by collection over time.

We then explore how skin color representation varies by race, gender, and age. We see that the Mainstream collection is more likely to show characters within a given race as lighter than their counterparts in the Diversity collection (Figure V, Panel A).[19] This finding shows that even when the Mainstream collection includes more Black, Latinx, or other characters, the reader sees these representations refracted through the lens of lighter skin color. Given the minoritization of females and those with darker skin color, we test for a difference in representation at the intersection of female gender identity and darker skin tint. We find no significant difference in skin color between males and females in the Mainstream collection (Figure V, Panel B). However we do find evidence that female adults are slightly lighter than male adults on average in the Diversity collection (Online Appendix Table AIII).

Examining representation of age and skin color, we find that children depicted in images are shown with lighter skin color on average than adults (Figure V, Panel C). This difference in mean skin color between children and adults is statistically significant (Online Appendix Table AIII).[20] We are aware of no definitive biological justification for this systematic difference in the representation of skin colors by age, although 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 who are pictured are more likely than adults to be products of mixed-race couples, which may lead to children having lighter skin, on average. 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 antimiscegenation laws in many contexts for a substantial portion of our study period and their incidence remains low. On the other hand, children could be depicted as having darker skin, on

19. We see the same result for monochromatic faces in Online Appendix Figure CVIa.

20. 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. In Online Appendix Figure BIII, we present the representation of skin color and age by the percentage presence in each of the coarser categories.

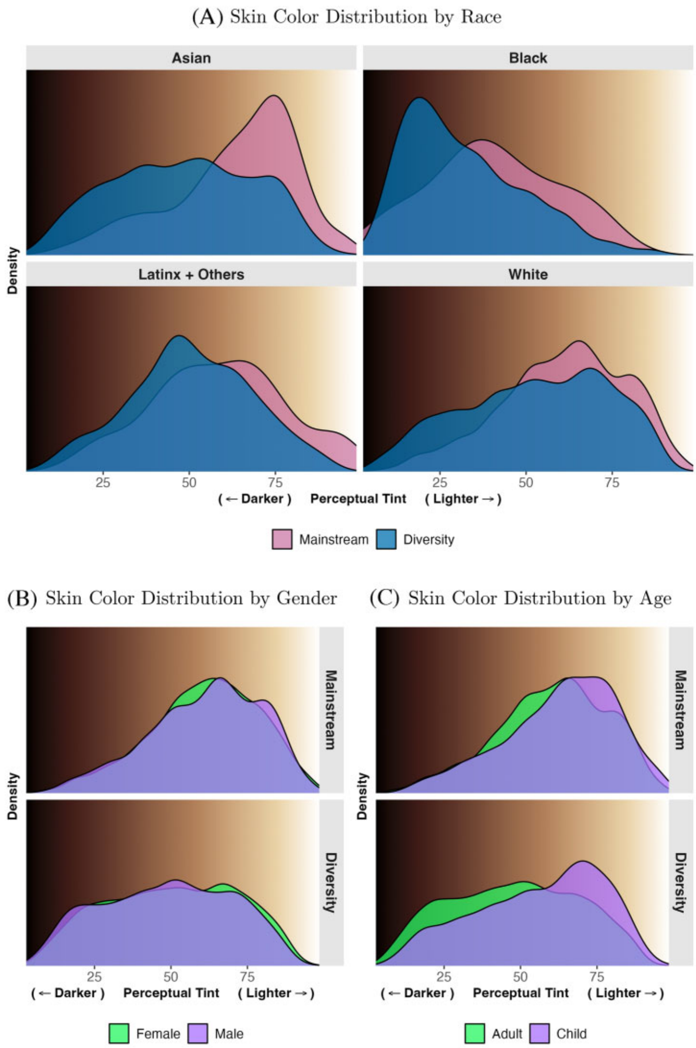

FIGURE V

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

This figure shows the distribution of skin color tint by predicted features of the detected faces in the Mainstream and Diversity collections. Panel A shows differences in the skin tint distributions between collections, conditional on predicted race. Panel B shows differences in the skin tint distributions between faces predicted to be male and faces predicted to be female, conditional on collection. Panel C shows differences in the skin tint distributions between faces predicted to be adults and faces predicted to be children, conditional on collection.

average, for a number of other potential reasons. For example, evidence of the breakdown of melanin over the life course suggests that there may be reason to expect the skin tint of adults to be lighter than that of children (Sarna et al. 2003). Nonetheless, the pattern we find of children being represented with lighter skin than adults is consistent in both the Mainstream and Diversity collections. Although there are many potential interpretations of this pattern, some include brightness being used to connote innocence (e.g., of childhood), supernatural features (e.g., of angels), or another type of emphasis that separates the character from the rest of the context. Exploration of the reasons behind this phenomenon merits further work beyond the scope of our study.

VI.B. Putative Race

We next examine racial representation of famous people. In the Mainstream collection, over 90% of famous figures are White (Online Appendix Figure BIV). Prior conventional content analyses studying the race of the main characters in Caldecott and Newbery award-winning books find qualitatively similar results (Koss, Johnson, and Martinez 2018; Koss and Paciga 2020). In the African American collection, Black people are the most represented, making up 50% of the famous people in that collection. In other collections, Black people make up 7%–29% of famous figures mentioned. Other groups appear far less frequently. Famous people of Asian, Latinx, Indigenous, and multiracial identities account for 3%–11% of famous people combined, a high level of inequality in representation relative to population averages: the U.S. Census (2019) estimates that only 60% of the population is non-Latinx White.

When we explore trends in racial representation of famous people over time compared with estimated population shares by race (Figure VI),[21] we see that in the Mainstream collection, relative to their population shares, Black people and Latinx people have been historically underrepresented while White people have been overrepresented. The past three decades, however, have shown increasing parity in representation of Black famous people. We see that despite increases in the diversity of

21. Online Appendix Figure BV shows a similar version of this graph with nonstandard axes to more clearly view changes in groups with small population proportions.

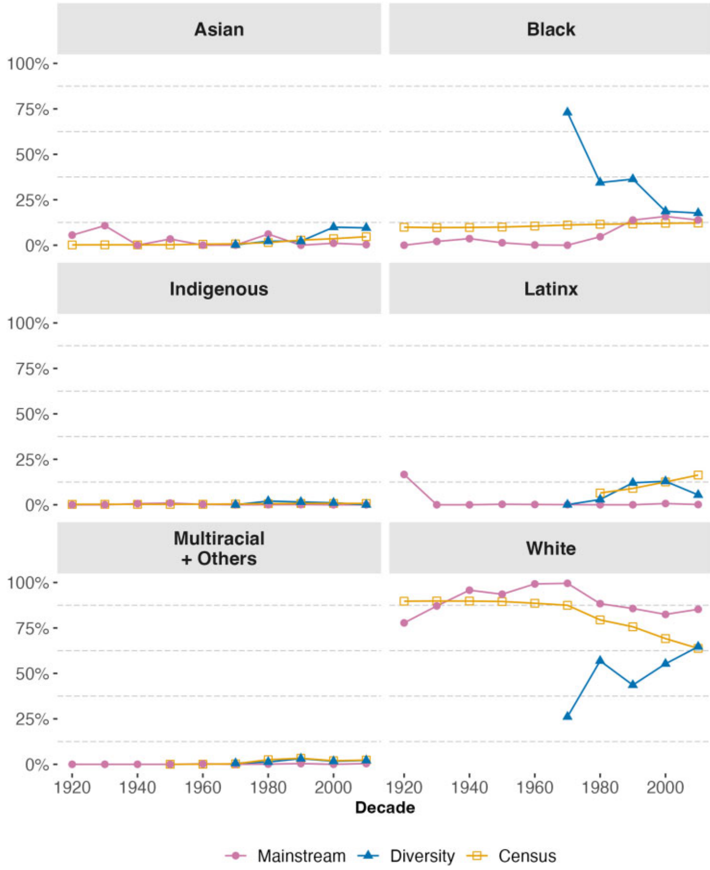

FIGURE VI

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

In this figure, we show the percent breakdown of famous people mentioned in a given book by race/ethnicity. For example, if Aretha Franklin were mentioned three times in a book and Jimmy Carter were mentioned two times (and if these were the only famous individuals mentioned), then 60% of the mentions of famous people in that book would be Black. We 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 V.B. We collapse the following identities: East Asian, Middle Eastern, and South Asian into the Asian category; North American Indigenous peoples and South American Indigenous peoples into the Indigenous category. If an individual was coded as having more than one race, we classify them as multiracial. See Online Appendix Figure BV for a similar version of this graph with nonstandard axes to better see changes in groups with small population proportions.

representation over time, the average individual included—whether a famous person or a pictured character—is a White person, regardless of collection.

In images, most pictured characters are classified as being White (Online Appendix Figure BVI).[22] Both White adults and children are more likely to be pictured than adults and children of any other racial category across all collections (Online Appendix Figure BVIII). Juvenile ageism, a relevant term coined in Westman (1991), refers to the notion that social systems ignore the interests of children. From an intersectional perspective, this also means that children of color, whose identities fall at the intersection of at least two sites of societal marginalization, are least likely to be seen by readers.

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

Among famous figures, after White males and females, Black males comprise the next most represented group (5%–37% of famous people). The representation of Black females (between 2% and 8% of famous people, except in the African American collection, where they make up 13%) is consistently less than that of Black males, despite their approximately equal shares in the

22. We also show how the share of faces by predicted race tracks with the share of the U.S. population over time (Online Appendix Figure BVII).

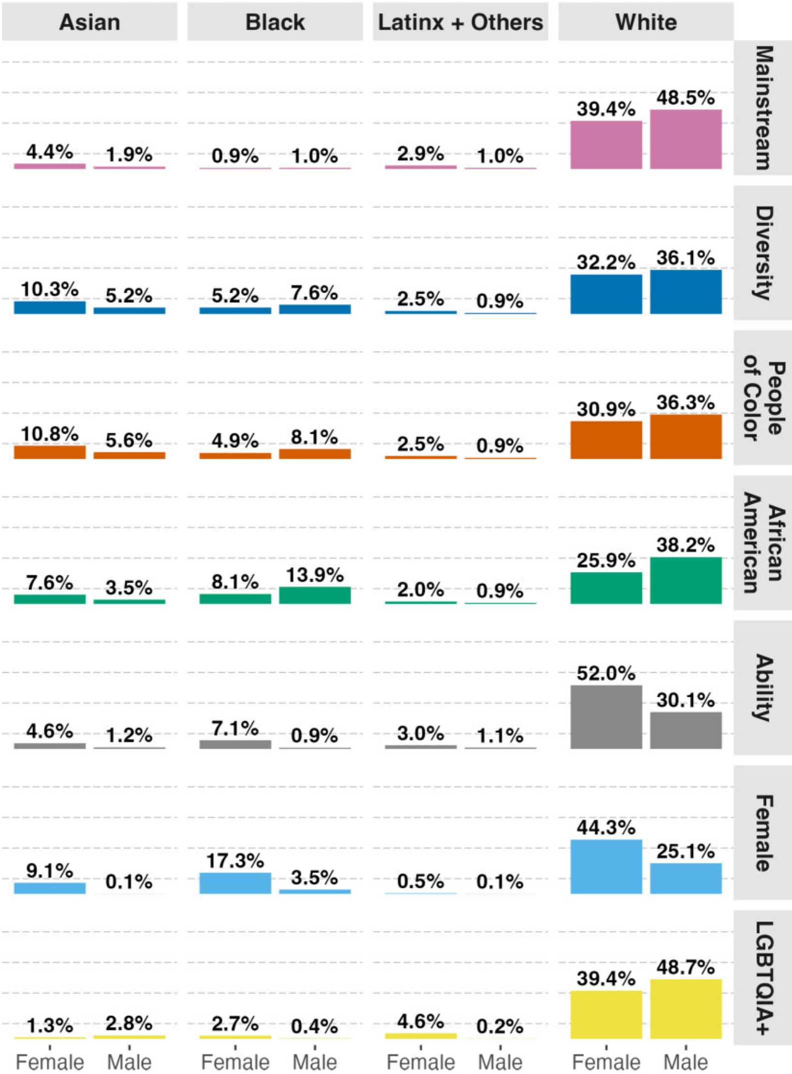

FIGURE VII

Race and Gender Predictions of Pictured Characters

In this figure, we show the average proportion of detected faces in all collections by race and gender predictions. We first find the proportion of faces in each race and gender category for every book; then we average across all books in a collection. Race and gender were classified by our trained AutoML model as described in Section V.A. See Online Appendix Figure BVI for the same figure broken down by race alone.

population (Online Appendix Figure BIX).[23] Conditional on the famous person being Black, we see greater representation of famous females in the Mainstream and Female collections than in the Diversity or African American collections (the representation of Asian and Latinx people is often close to zero for this measure, making comparison difficult). This highlights that even in collections of books curated to highlight a given racial identity, there is less representation of people at the intersection of multiple dimensions of marginalization than of those who occupy only one such dimension.

In Online Appendix Table AIV, 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 Jr. For the Mainstream collection, all five of the most commonly mentioned people are White males. For the Diversity collection, all five are males, three of whom are Black (Martin Luther King Jr., Frederick Douglass, and Langston Hughes) and two of whom are White (Abraham Lincoln, George Washington). In the Female collection, where one might anticipate the presence of more females, the three most uniquely mentioned people are men (John F. Kennedy, Martin Luther King Jr., and Jimmy Carter) and the fourth is a women (Betty Friedan).[24]

VI.C. Gender

We explore the representation of gender. We first measure the incidence of words with any gender association, which includes pronouns and other gendered terms, the gender of the famous people mentioned in the text, and the gender classifications for character first names. In Table I and Figure VIII, we present average book-level proportions of female words out of all gendered words. For all collections except those books specifically recognized for highlighting girls and women, we observe fewer female words than male words. Table I shows that the proportion of

23. We show how race-gender representation in images and text vary over time in Online Appendix Figure BX.

24. Online Appendix Tables AV and AVI show this for the top five females and top five males, respectively, uniquely mentioned in each collection. Online Appendix Table AVII shows the most uniquely mentioned famous figure by collection for each decade.

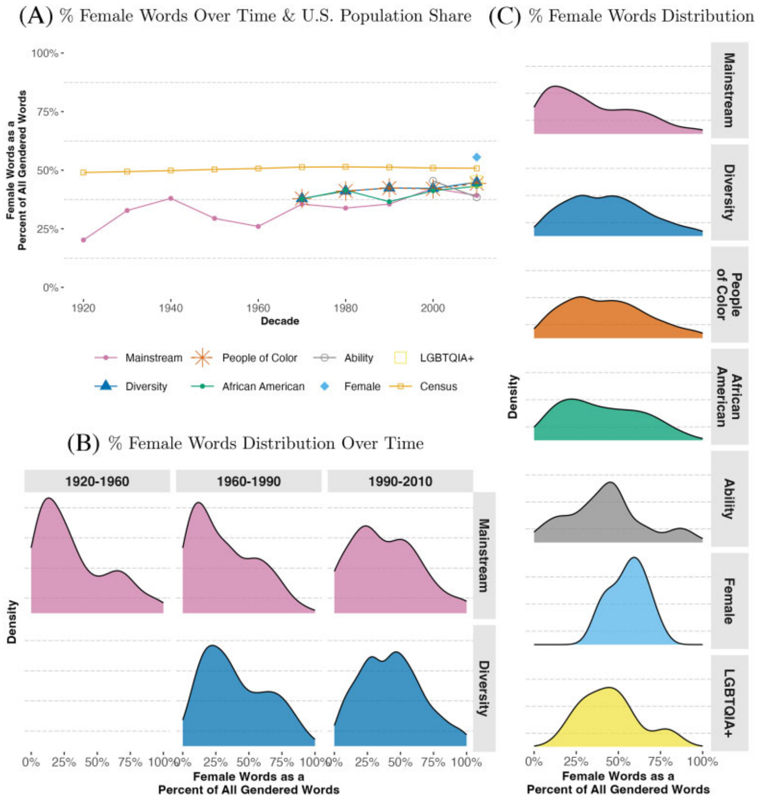

FIGURE VIII

Female Words as a Percent of All Gendered Words

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

gendered words that are female in these collections is between 34% and 45%, as opposed to 56% in the Female collection. Figure VIII, Panel A shows that this proportion increases gradually over time, but remains below the U.S. population share of females for all collections in every decade, except for the Female collection.

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

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

Our results are robust to restricting analysis to each type of gendered word: gendered pronouns, gendered terms, or first names (Online Appendix Figure BXI). This addresses the concern that we could be misattributing changes in gender representation to changes in the historical grammatical convention to use what were then considered "gender-neutral" pronouns (e.g. he, his). For example, if an author writing in an earlier era wanted to include more female representation, we would see this reflected in the proportion of named female characters but not in the proportion of female pronouns. We do not see this skewed pattern in our results. The robustness of our results to this sample restriction demonstrates that our results are not driven by measurement error stemming from changes over time in this historical convention. Our results are also robust to restricting analysis of gender representation to gender of famous figures. Famous figures transmit more implicit information to a child than generic terms or characters by virtue of their identity in society. This can occur through any of a number of channels, for example, via role model effects (Porter and Serra 2020) or via effects on more general social preferences and beliefs (Plant et al. 2009; Alrababah et al. 2021). In Table I, we show that on average over 85% of the famous figures mentioned in books belonging to the Mainstream collection were male, for example, and even books in the Female collection included more unique famous males than females on average. Overall, less than one-third of famous figures in the books we study are female (Online Appendix Figure BXI).

Next we describe the representation of gender in the images of these books.[25] We show the proportion of faces in each collection identified as female in Figure VII and Online Appendix Figure BXIIa. 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, our model classifies 71% and 67% of the faces as female. Online Appendix Figure BXIIb 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 make up 39%–51% of all detected character faces over time.[26]

We compare representation of gender across images and text. In Figure IX, we show a scatterplot of collection-by-decade average proportions of female words on the x-axis and the average proportion of female-presenting faces on the y-axis. It shows that females are more likely to appear in images rather than text, which means that females are more likely to be visualized (seen) than mentioned in the story (heard). One interpretation of this pattern is that authors or illustrators may perfunctorily include additional females in pictures, giving the appearance of equity while not actually having them play an important role in the story. It also highlights that on average, females are represented less than half of the time in both images and text.[27]

VI.D. Age

Finally we describe the representation of people by age in the images and text of our books. In Table I, we show that across all collections, adults are more likely to be present in both images and text. Three percent to 19% of characters presented in images are classified as children, and 17%–32% of age-specific gendered words refer to children. In Online Appendix Figure BXVa, 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

25. This exercise demonstrates the limitations of existing AI approaches. Compared to the state of the art, a human would be better able to more accurately classify individuals who identify as transgender or nonbinary.

26. We show a similar pattern when using a continuous measure of the average probability that a face is classified as being female in Online Appendix Figure BXIII.

27. In Online Appendix Figure BXIV, we show these results for females by race in which we see Black and Latinx females less represented.

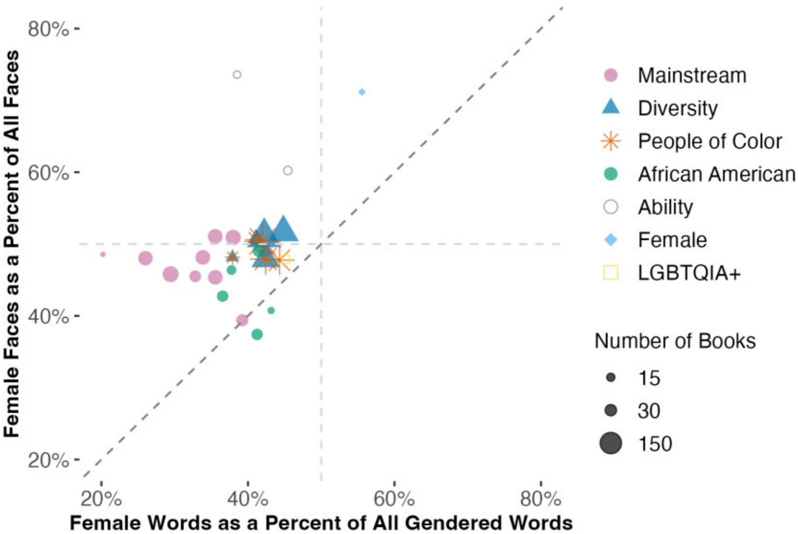

FIGURE IX

Female Representation in Images and Text of Children's Books

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

each collection.[28] We also see in Online Appendix Figure BVIIc that adults are overrepresented relative to their U.S. population share, meaning that adult depictions are more common than child depictions in books targeted to children. Children of color are the least likely to be pictured, even in the People of Color or African American collections (Online Appendix Figure BVIII).

In Online Appendix Figure BXVb, we show the age classifications of gendered words (e.g., girl versus woman). Similar to images, we see that older people are more likely to be mentioned than younger people. In most books, the distribution of young people by gender is similar, although in the Female collection, girls are approximately twice as likely to appear as boys. In most

28. 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 adults' presence to children's presence in images and in text.

collections, men appear more often than women in gendered terms specific to adults.

VII. Economic and Social Factors Underlying Representation in Books

In this section, we investigate a series of economic and social factors that may contribute to the patterns of representation of skin color, race, gender, and age in prominent children's books that we document in Section VI. First, we discuss relevant prior theoretical and empirical research related to the economics of the media and, separately, the economics of identity, to conceptually characterize a set of market forces that may influence the patterns of representation in children's books. For clarity, we separate these into demand- and supply-side forces. Second, we generate a series of stylized facts that relate the patterns in representation to this series of demand and supply forces suggested by prior literature. We estimate the relationship between historical trends—first historical events, followed by changes over time in social mores and, separately, in market shares of consumers of different identities—and the representation we see in books. Finally, we explore how local political beliefs relate to the consumption of books with different levels of representation. In Online Appendix I, we discuss some limitations to these analyses.

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

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

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

The second is informed by the notion that deviating from social norms is costly (Akerlof and Kranton 2000; Shayo 2020). This force can lead to demand for representation that hews closely to the (perceived) status quo. Applied to our setting, this suggests that consumers who have identities that have been historically overrepresented in media have been socialized to suffer greater disutility from consuming content that does not center their (socially dominant) identities than historically underrepresented consumers, because consuming such content deviates from the perceived status quo or social norm. For example, men might suffer greater disutility than women from reading a book with a female main character than women would from reading a book with a male main character. Furthermore, this force of "status quo bias" in consumption of books would push consumers of all identities to be more likely to consume children's books containing characters with socially dominant identities than those containing characters with other identities. This is reflected in a result from Bernheim (1994) showing that under certain conditions, people will adapt their preferences to match broader societal preferences.

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

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

A second supply-side force in such markets is a "pricing-in of representation." This refers to the notion that books which deliberately elevate nondominant 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).[29]

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

VII.B. Empirical Analysis of Economic Forces

In this section, we present a series of empirical analyses probing the economic forces related to the supply of and demand for representation in children's books. Our analyses use a range of data, including data on book purchases and purchaser demographics, and library branch–level data on library acquisitions linked to neighborhood demographic characteristics.

We present analyses of book consumption that document patterns suggesting demand-side utility from homophily. Using book consumption data from the Numerator OmniPanel, we estimate the correlations between book purchaser identity and book-level female representation in images and text. We present results in Table II; these relate to the representation findings we report in Table I and Figure VIII. In Table II, Panel A, we show that purchasers who have a daughter purchase books with 2 percentage points more female names as a proportion of all gendered names, and 3 percentage points more female words as a proportion of all

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

TABLE II

GENDER REPRESENTATION IN BOOK CONTENT BY PURCHASER IDENTITIES

	Words (1)	Names (2)	Faces (3)	Images vs. text (4)
Panel A: Gender of purchaser child				
Purchaser has a daughter	0.031***	0.019**	−0.003	−0.042***
	(0.008)	(0.009)	(0.010)	(0.012)
Purchaser has a son	−0.013	−0.020**	0.003	0.012
	(0.008)	(0.009)	(0.010)	(0.012)
Constant	0.386***	0.363***	0.417***	0.058***
(baseline group: no children)	(0.003)	(0.003)	(0.004)	(0.005)
Observations	9,716	9,477	6,737	6,696
Adjusted R^2	0.001	0.001	−0.0003	0.001
Panel B: Purchaser gender				
Male	−0.015***	−0.017***	−0.019***	−0.005
	(0.005)	(0.006)	(0.006)	(0.008)
Other	−0.006	−0.038**	0.024	0.030
	(0.016)	(0.019)	(0.022)	(0.027)
Constant	0.389***	0.370***	0.434***	0.080***
(baseline group: female)	(0.002)	(0.002)	(0.002)	(0.003)
Observations	28,760	28,235	18,848	18,753
Adjusted R^2	0.0003	0.0004	0.0004	−0.00002

Notes. We regress four different measures of female representation contained in a purchased book on indicator variables for whether the purchaser has a daughter or son (Panel A) and purchaser gender (Panel B). The dependent variable in column (1) is the percent of female words out of all gendered words where gendered words include all gendered names, gendered pronouns, and gendered terms. The dependent variable in column (2) is the percent of female names out of all gendered names. The dependent variable in column (3) is the percent of female faces out of all faces detected. The dependent variable in column (4) is the difference between the third and first columns' dependent variables. We obtain book-level purchasing data from the Numerator OmniPanel, which contains data on purchases made in 2017–2020, and we merge it with our curated data on representation in award-winning children's books. We subset purchasing data to include purchases of award-winning children's books that we have digitized that contain at least one gendered word, name, or face. * $p < .10$; ** $p < .05$; *** $p < .01$.

gendered words, compared with purchasers who have no children (baseline rates 36.3% and 38.6%, respectively). We see symmetric preferences for purchasers who have a son in terms of books purchased with a lower proportion of female names and female gendered words, as compared with purchasers who have no children. Finally, we see that consumers with daughters purchase books that, on average, have more similar proportions of gender representation in images and text (i.e., books in which females are more equally "seen" and "heard") (column (4)), despite books overall skewing toward having more female representation in images compared with text (as shown in Figure IX). In Table II, Panel B, we see that males' purchasing patterns exhibit a slight revealed

TABLE III
SKIN COLOR AND RACE REPRESENTATION IN BOOK CONTENT BY PURCHASER IDENTITIES

	Dependent variable				
		Percent of famous mentions by race			
Purchaser ethnicity	Average skin tint (1)	Asian (2)	Black (3)	Latinx (4)	White (5)
Asian	−0.086	0.005***	−0.003	0.002	−0.005
	(0.704)	(0.002)	(0.007)	(0.002)	(0.008)
Black	−6.405***	−0.001	0.126***	0.004**	−0.130***
	(0.712)	(0.002)	(0.007)	(0.002)	(0.008)
Latinx	−3.287***	0.001	0.022***	0.013***	−0.035***
	(0.640)	(0.001)	(0.007)	(0.002)	(0.007)
Other	−2.341**	0.003	0.021**	−0.002	−0.023**
	(1.025)	(0.002)	(0.010)	(0.003)	(0.011)
Constant	59.283***	0.008***	0.082***	0.007***	0.900***
(baseline group: White)	(0.189)	(0.0005)	(0.002)	(0.001)	(0.002)
Observations	14,219	18,330	18,330	18,330	18,330
Adjusted R^2	0.007	0.0004	0.017	0.003	0.015

Notes. We regress five different measures of racial representation contained in a purchased book on indicator variables indicating the race or ethnicity of the purchaser. 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 in 2017–2020 and merge it with our curated data on representation in award-winning children's books. We subset purchasing data to include purchases of award-winning children's books that 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). * $p < .10$; ** $p < .05$; *** $p < .01$.

preference for books with more male words, names, and faces. Specifically, compared with female purchasers, males purchase books with 1 to 2 percentage points less female representation in images and text.

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

We explore how the representation of age in these books might relate to the purchaser behavior we observe. If we assume that adults are making the majority of purchasing decisions, then the overrepresentation of adults and underrepresentation of children as shown in Online Appendix Figure BXV (even in these books targeted to children) is consistent with utility from homophily.[30]

Next we explore the relationship between specific purchaser identities and consumption of books that were recognized for highlighting the experiences of people with those specific identities (Online Appendix Table AVIII). We see, for example, that purchasers who are Black are more likely to purchase books from the African American collection, purchasers who are Asian are more likely to purchase books that received awards for highlighting the experiences of Asian individuals, purchasers who are Latinx are more likely to purchase books that received awards for highlighting the experiences of Latinx individuals, and purchasers who identify as LGBTQIA+ are more likely to buy books that are in the LGBTQIA+ collection.

We characterize the relationship between library holdings and local characteristics using inventory data from branches of the Seattle Public Library system. These findings also suggest behavior consistent with utility from homophily. In Table IV, 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 in a community. These results relate to recent work

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

TABLE IV

THE NUMBER OF MAINSTREAM AND DIVERSITY BOOKS IN LIBRARY COLLECTIONS BY COMMUNITY CHARACTERISTICS

	Dependent variable: Number of award-winning children's books by collection			
	Mainstream (1)	Diversity (2)	Mainstream (3)	Diversity (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

Notes. Each observation in the data used to make this table corresponds to a community reporting area (CRA). Each community area is manually matched to its closest Seattle Public Library branch. Each Seattle Public Library branch is matched to at least one CRA. We regress the number of Mainstream books (columns (1) and (3)) and Diversity books (columns (2) and (4)) available in a community's library on community characteristics. Population demographics are taken from the American Community Survey, five-year series 2013–2017 accessed through Seattle's Data Portal. Seattle Public Library inventory data, as reported on October 1, 2017, was 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 to 100. * $p < .10$; ** $p < .05$; *** $p < .01$.

showing that holdings of school library collections reflect the beliefs held by those in the surrounding area (Mumma 2022).

We describe purchaser behavior that relates to the demand-side force of status quo bias. We find that a majority of the books purchased in our data have predominantly male-focused content, even though most of the purchasers in our sample are women (Online Appendix Table AI). In addition, results in Table II indicate that while purchasers with daughters purchase books with more female words and names than purchasers with sons, they are still purchasing books with less than 50% female words and names on average. This implies that parents' preference for purchasing books with male characters for their son is stronger than their preference for purchasing books with female characters for their daughter. Put differently, these results suggest that many parents' book-buying preferences may reflect the notion that boys should read about boys but girls can read about anyone. While only suggestive, this pattern is consistent with the phenomenon of status quo bias.

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

VII.C. Historical Trends and Representation

We explore how changes in representation over time in the Mainstream collection may be associated with historical events, trends in societal attitudes toward issues related to race and gender, and changes in market share of various identity groups.

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

TABLE V
READERSHIP BY COLLECTION

Collection	Number of checkouts (1)	Mean checkouts per title (2)	Number of unique titles (3)	Mean library copies per title (4)
Panel A: Seattle Public Library Inventory and Checkouts				
Mainstream	388,357	991	392	14.0
Diversity	248,860	212	1,176	7.0
All other children's books	17,027,557	238	71,590	5.6
People of color	225,851	216	1,045	7.0
African American	37,367	217	172	8.3
Female	7,272	97	75	6.5
Ability	14,170	301	47	7.7
LGBTQIA+	8,295	251	33	9.3
Panel B: Average price and copies purchased in Numerator OmniPanel				
	Number of copies sold (1)	Mean book price (2)	Number of unique titles (3)	Mean copies sold per title (4)
Mainstream	40,854	$7.66	493	83
Diversity	35,553	$9.34	1,067	33
All other children's books	1,683,406	$7.42	97,866	17
People of color	26,899	$9.51	880	31
African American	9,081	$9.95	149	61
Female	4,892	$8.68	120	41
Ability	2,834	$8.70	55	52
LGBTQIA+	2,838	$9.07	34	83

Notes. In Panel A, 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 2017–2020. Panel B presents 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 and 2017.

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

We can also characterize the correlation between changes in market share and the representation of race and gender in books over time. Following existing studies estimating this type of relationship (Acemoglu and Linn 2004; DellaVigna and Pollet 2007), we calculate the market share of various race and gender groups and use this to estimate whether there is a statistically detectable relationship between market share and representation of the group in the books we study. For race, we use the share of racial groups in the U.S. population according to the decadal census. For gender, although the share of females in the census is relatively stable, we can instead use the female labor force participation rate as a measure of market share. We conceive of this as capturing the (relative) consumer power of women relative to men.[31]

We find a positive and significant relationship between the market share of Asian, Black, and White people in a decade and their representation in books from the Mainstream collection published in that decade (Online Appendix Table AXI). We find no evidence of a correlation between market share and representation of Latinx people and their representation in books, but we believe this is primarily an artifact of the very low representation of this group in the books we study.[32] Also, census data on Latinx people are only available beginning in 1970, and we are only able to predict whether the race of a detected face is "Latinx + Others," both of which lead to noisier estimates. For gender, we find a positive and significant association. The female labor force participation rate is strongly related to the proportion of gendered words contained in books, which increases over time as shown in Figure VIII, Panel A. Although we find no such correlation with the representation of gender in images, we suspect this is primarily because, throughout our period of study,

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

32. These patterns are shown in Figure VI, which plots the relationship over time between population share and representation by race and ethnicity in text.

representation of gender in images is closer to parity than it is in text.[33]

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

VII.D. Local Beliefs and Book Consumption

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

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

In Table VI, we show that a greater number of purchases of books from the Diversity collection in a given ZIP code is associated with a smaller proportion of individuals who believe that undocumented immigrants should be deported (column (1)), a

33. Based on these correlations and population projections from the U.S. Census Bureau made in 2020, we would we expect to see increases in representation of Black and Hispanic people, but not women.

TABLE VI
LOCAL BELIEFS AND CHILDREN'S BOOK PURCHASES IN ZIP CODES

	Dependent variable:			
	% of respondents who think the U.S. government should		% of respondents somewhat or strongly agree	
	Identify and deport undocumented immigrants (1)	Withhold federal funds from localities that do not follow federal immigration laws (2)	White people in the U.S. have certain advantages because of the color of their skin (3)	I am angry that racism exists (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

Notes. 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 that were recognized by an award in our Mainstream collection and/or Diversity collection. Data on beliefs at the ZIP code level are drawn from the 2017 Cooperative Election Study Common Content Survey (Schaffner and Ansolabhere 2019). Data on children's book purchases at the ZIP code level are drawn from the Numerator OmniPanel data from 2017–2020. Variables containing percentages are scaled so that potential values range from 0 to 100. In the CCES, the wording of the question on undocumented people referred to "illegal" immigrants. * $p < .10$; ** $p < .05$; *** $p < .01$.

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 United States 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%) answer yes to this question, as opposed to only 37% who believe that undocumented immigrants should be deported.

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

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

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

Our approach has a few key limitations. First, although we focus on representation in light of its important role in the processes we describe, it is only one component of the complex, larger societal processes we are trying to describe. Second, our focus on representation is limited to estimating the presence of identities, not their depiction. Measuring how people are portrayed has historically been a key strength of manual content analysis (Rosenberg, Schnurr, and Oxman 1990; Linderman 2001), and this limitation of computational tools highlights how manual and computational approaches complement each other.[34] Third, there are many other child-specific media—for example, television, movies, and computer applications and websites—that are equally or more influential than the books we study. Fourth, the measures of representation we use are imperfect: our measures of gender identity neglect measurement of nonbinary and genderfluid identities, and race is a multifaceted construct of human categorization that is ill-defined, making any effort to measure it inherently fraught. Finally, although we acquired 91% of all books that won an award in our Mainstream collection (as opposed to being honored by an award), we were only able to access and analyze roughly one-third of all the books ever recognized by the awards in our sample. We argue that our ability to access these books is most likely to be positively correlated with

34. Understanding patterns in the manner in which characters are represented is also important, and we are pursuing this work in separate projects (e.g., Adukia et al. 2022a, 2022b).

consumers' ability to access them, such that our estimates are likely to closely track the levels of representation in the books to which children are actually exposed.

The image-to-data tools we introduce allow for the systematic measurement of characteristics in visual data that were previously beyond the reach of empirical researchers. This contribution is in the spirit of other recent work introducing new sources of data to the economic study of social phenomena, such as text (Gentzkow, Shapiro, and Taddy 2019), geospatial imagery (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 (Kearney and Levine 2019), advertising (Bertrand et al. 2010), and textbooks (Cantoni et al. 2017). Indeed, recent scholarship has begun to use them to study stereotypes in news media (Ash et al. 2021).

The findings in this study—and the power of the tools we use to generate them—generate hypotheses that can motivate and inform subsequent research on the causes and consequences of representation in children's books. Measurements such as those we generate could be paired with causal inference tools to advance prior work on the effect of book content on children's beliefs and later life outcomes (Fuchs-Schündeln and Masella 2016; Cantoni et al. 2017; Arold, Woessmann, and Zierow 2022). For example, such work could precisely measure childhood exposure to different levels of representation and link it to the formation of beliefs, preferences, and societal outcomes. These same measurements could also be used to better understand the objective functions of different publishers, and how these change over time and in response to societal events.

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

University of Chicago and National Bureau of Economic Research, United States
Columbia University, United States
University of Chicago, United States
University of Chicago, United States
University of Chicago, United States

Supplementary Material

An Online Appendix for this article can be found at *The Quarterly Journal of Economics* online.

Data Availability

The data underlying this article are available in the Harvard Dataverse, https://doi.org/10.7910/DVN/OR72L1 (Adukia et al. 2023).

References

Acemoglu, Daron, and Joshua Linn, "Market size in Innovation: Theory and Evidence from the Pharmaceutical Industry," *Quarterly Journal of Economics*, 119 (2004), 1049–1090. https://doi.org/10.1162/0033553041502144

Adukia, Anjali, Patricia Chiril, Callista Christ, Anjali Das, Alex Eble, Emileigh Harrison, and Hakizumwami Birali Runesha, "Tales and Tropes: Gender Roles from Word Embeddings in a Century of Children's Books," *Proceedings of the 28th International Conference on Computational Linguistics (COLING)* (2022a), 3086–3097. https://doi.org/10.2139/ssrn.4221213

Adukia, Anjali, Callista Christ, Anjali Das, and Ayush Raj, "Portrayals of Race and Gender: Sentiment in 100 Years of Children's Literature," *Proceedings of the ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies*, (2022b), 20–28. https://doi.org/10.1145/3530190.3534811

Adukia, Anjali, Alex Eble, Emileigh Harrison, Hakizumwami Birali Runesha, and Teodora Szasz, "Replication Data for: 'What We Teach about Race and Gender: Representation in Images and Text of Children's Books," (2023), Harvard Dataverse, https://doi.org/10.7910/DVN/OR72L1.

Ajzen, Icek, Martin Fishbein, Sophie Lohmann, and Dolores Albarracín, "The Influence of Attitudes on Behavior," in *The Handbook of Attitudes, vol. 1: Basic Principles, 2nd ed.,* Dolores Albarracín and Blair T. Johnson, eds. (New York: Routledge, (2018), 197–255. https://doi.org/10.4324/9781315178103

Akerlof, George A., and Rachel E. Kranton, "Economics and Identity," *Quarterly Journal of Economics*, 115 (2000), 715–753. https://doi.org/10.1162/003355300554881

Alan, Sule, Ceren Baysan, Mert Gumren, and Elif Kubilay, "Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking," *Quarterly Journal of Economics*, 136 (2021), 2147–2194. https://doi.org/10.1093/qje/qjab009

Alrababa′h, Ala′, William Marble, Salma Mousa, and Alexandra A. Siegel, "Can Exposure to Celebrities Reduce Prejudice? The Effect of Mohamed Salah on Islamophobic Behaviors and Attitudes," *American Political Science Review*, 115 (2021), 1111–1128. https://doi.org/10.1017/S0003055421000423

Arold, Benjamin W., Ludger Woessmann, and Larissa Zierow, "Religious Education in School Affects Students' Lives in the Long Run," *CESifo Forum*, 23 (2022), 40–43.
Ash, Elliott, Ruben Durante, Maria Grebenshchikova, and Carlo Schwarz, "Visual Representation and Stereotypes in News Media," SSRN Working Paper, 2021. https://dx.doi.org/10.2139/ssrn.3934858
Bernheim, B. Douglas, "A Theory of Conformity," *Journal of Political Economy*, 102 (1994), 841–877. https://doi.org/10.1086/261957
Berry, Steven, and Joel Waldfogel, "Product Quality and Market Size," *The Journal of Industrial Economics*, 58 (2010), 1–31. https://doi.org/10.1111/j.1467-6451.2010.00404.x
Bertrand, Marianne, Dean Karlan, Sendhil Mullainathan, Eldar Shafir, and Jonathan Zinman, "What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment," *Quarterly Journal of Economics*, 125 (2010), 263–306. https://doi.org/10.1162/qjec.2010.125.1.263
Beyer, Lucas, "PyDenseCRF, GitHub, 2018. https://github.com/lucasb-eyer/pydensecrf/blob/master/setup.py.
Bian, Lin, Sarah-Jane Leslie, and Andrei Cimpian, "Gender Stereotypes about Intellectual Ability Emerge Early and Influence Children's Interests," *Science*, 355 (2017), 389–391. https://doi.org/10.1126/science.aah6524
Buolamwini, Joy, and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," *Conference on Fairness, Accountability and Transparency*, 81 (2018), 77–91.
Cantoni, Davide, Yuyu Chen, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang, "Curriculum and Ideology," *Journal of Political Economy*, 125 (2017), 338–392. https://doi.org/10.1086/690951
Cappelen, Alexander, John List, Anya Samek, and Bertil Tungodden, "The Effect of Early-Childhood Education on Social Preferences," *Journal of Political Economy*, 128 (2020), 2739–2758. https://doi.org/10.1086/706858
Clark, Heather D., George A. Wells, Charlotte Huët, Finlay A. McAlister, L. Rachid Salmi, Dean Fergusson, and Andreas Laupacis, "Assessing the Quality of Randomized Trials: Reliability of the Jadad Scale," *Controlled Clinical Trials*, 20 (1999), 448–452. https://doi.org/10.1016/S0197-2456(99)00026-4
Cockcroft, Marlaina, "Caldecott and Newbery Medal Wins Bring Instant Boost to Book Sales," *School Library Journal*, 64 (February 10, 2018).
Crenshaw, Kimberl é, "Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color," *Stanford Law Review*, 43 (1990), 1241.
Crisp, Thomas, and Brittany Hiller, "Telling Tales about Gender: A Critical Analysis of Caldecott Medal-Winning Picturebooks, 1938–2011," *Journal of Children's Literature*, 37 (2011), 18–29.
Daniels, Elizabeth A., Marlee C. Layh, and Linda K. Porzelius, "Grooming Ten-Year-Olds with Gender Stereotypes? A Content Analysis of Preteen and Teen Girl Magazines," *Body Image*, 19 (2016), 57–67. https://doi.org/10.1016/j.bodyim.2016.08.011
DellaVigna, Stefano, and Joshua M. Pollet, "Demographics and Industry Returns," *American Economic Review*, 97 (2007), 1667–1702. https://doi.org/10.1257/aer.97.5.1667
Dhar, Diva, Tarun Jain, and Seema Jayachandran, "Intergenerational Transmission of Gender Attitudes: Evidence from India," *Journal of Development Studies*, 55 (2019), 2572–2592.
Eble, Alex, and Feng Hu, "Child Beliefs, Societal Beliefs, and Teacher-Student Identity Match," *Economics of Education Review*, 77 (2020), 101994. https://doi.org/10.1016/j.econedurev.2020.101994
———, "Gendered Beliefs about Mathematics Ability Transmit across Generations through Children's Peers," *Nature Human Behaviour*, 6 (2022), 868–879.
Eitel, Alexander, Katharina Scheiter, Anne Schüler, Marcus Nyström, and Kenneth Holmqvist, "How a Picture Facilitates the Process of Learning from Text: Evidence for Scaffolding," *Learning and Instruction*, 28 (2013), 48–63. https://doi.org/10.1016/j.learninstruc.2013.05.002

Fletcher, J. D., and Sigmund Tobias, “The Multimedia Principle,” in *The Cambridge Handbook of Multimedia Learning*, R. E. Mayer, ed. (New York: Cambridge University Press, 2005), 117–134. https://doi.org/10.1017/CBO9780511816819.008

Fu, Siyao, Haibo He, and Zeng-Guang Hou, “Learning Race from Face: A Survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36 (2014), 2483–2509. https://doi.org/10.1109/TPAMI.2014.2321570.

Fuchs-Schündeln, Nicola, and Paolo Masella, “Long-Lasting Effects of Socialist Education,” *Review of Economics and Statistics*, 98 (2016), 428–441. https://doi.org/10.1162/REST_a_00583

Genicot, Garance, and Debraj Ray, “Aspirations and Inequality,” *Econometrica*, 85 (2017), 489–519. https://doi.org/10.3982/ECTA13865

Gentzkow, Matthew, Bryan T. Kelly, and Matt Taddy, “Text as Data,” *Journal of Economic Literature*, 57 (2019), 535–574. https://doi.org/10.1257/jel.20181020

Gentzkow, Matthew, Jesse M. Shapiro, and Matt Taddy, “Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech,” *Econometrica*, 87 (2019), 1307–1340. https://doi.org/10.3982/ECTA16566

Ghavami, Negin, Dalal Katsiaficas, and Leoandra Onnie Rogers, “Toward an Intersectional Approach in Developmental Science: The Role of Race, Gender, Sexual Orientation, and Immigrant Status,” Advances in Child Development and Behavior, 50 (2016), 31–73.

Giroux, Henry A., *Ideology, Culture, and the Process of Schooling* (Philadelphia, PA: Temple University Press, 1981).

Henderson, J. Vernon, Adam Storeygard, and David N. Weil, “Measuring Economic Growth from Outer Space,” *American Economic Review*, 102 (2012), 994–1028. https://doi.org/10.1257/aer.102.2.994

Hersch, Joni, “Profiling the New Immigrant Worker: The Effects of Skin Color and Height,” *Journal of Labor Economics*, 26 (2008), 345–386. https://doi.org/10.1086/587428

Horrigan, John, “Who Uses Libraries and What They Do at Their Libraries,” in *Libraries at the Crossroads* (Pew Research Center, 2015).

Hughes, Julie M., Rebecca S. Bigler, and Sheri R. Levy, “Consequences of Learning about Historical Racism among European American and African American Children,” *Child Development*, 78 (2007), 1689–1705. https://doi.org/10.1111/j.1467-8624.2007.01096.x

Jackson, Aaron S., Michel Valstar, and Georgios Tzimiropoulos, “A CNN Cascade for Landmark Guided Semantic Part Segmentation,” *European Conference on Computer Vision* (2016), 143–155. https://doi.org/10.1007/978-3-319-49409-8_14

Jackson, Matthew O., *Social and Economic Networks* (Princeton, NJ: Princeton University Press, 2010).

Jakiela, Pamela, and Owen Ozier, “Gendered Language,” World Bank Policy Research Working Paper No. 8464, 2018. https://doi.org/10.1596/1813-9450-8464

Jayadev, Arjun, and Robert Johnson, “Tides and Prejudice: Racial Attitudes during Downturns in the United States 1979–2014,” *Review of Black Political Economy*, 44 (2017), 379–392. https://doi.org/10.1007/s12114-017-9264-y

Kearney, Melissa S., and Phillip B. Levine, “Early Childhood Education by Television: Lessons from Sesame Street,” *American Economic Journal: Applied Economics*, 11 (2019), 318–350. https://doi.org/10.1257/app.20170300

———, “Role Models, Mentors, and Media Influences,” *Future of Children*, 30 (2020), 83–106. https://doi.org/10.1353/foc.2020.0006

Knowles, Elizabeth, and Martha Smith, *The Reading Connection: Bringing Parents, Teachers, and Librarians Together* (Westport, CT: Libraries Unlimited, 1997).

Koss, Melanie D., “Diversity in Contemporary Picturebooks: A Content Analysis,” *Journal of Children’s Literature*, 41 (2015), 32.

Koss, Melanie D., Nancy J. Johnson, and Miriam Martinez, "Mapping the Diversity in Caldecott Books from 1938 to 2017: The Changing Topography," *Journal of Children's Literature*, 44 (2018), 4–20.
Koss, Melanie D., and Kathleen A. Paciga, "Diversity in Newbery Medal-Winning Titles: A Content Analysis," *Journal of Language and Literacy Education*, 16 (2020), n2.
Kozlowski, Austin C., Matt Taddy, and James A. Evans, "The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings," *American Sociological Review*, 84 (2019), 905–949. https://doi.org/10.1177/0003122419877135
Krippendorff, Klaus, *Content Analysis: An Introduction to its Methodology* (Beverly Hills, CA: Sage Publications, 2018).
Krishnan, Anoop, Ali Almadan, and Ajita Rattani, "Understanding Fairness of Gender Classification Algorithms Across Gender-Race Groups," *arXiv*, 2020. https://doi.org/10.1109/ICMLA51294.2020.00167
Levstik, Linda S., and Cynthia A. Tyson, *Handbook of Research in Social Studies Education* (New York: Routledge, 2010).
Linderman, Alf, "Computer Content Analysis and Manual Coding Techniques: A Comparative Analysis," in *Theory, Method, and Practice in Computer Content Analysis,* M. D. West, ed. (Connecticut: Ablex, 2001), 97–109.
Logan, Trevon D., "American Enslavement and the Recovery of Black Economic History," *Journal of Economic Perspectives*, 36 (2022), 81–98. https://doi.org/10.1257/jep.36.2.81
Lu, Conny, "Face Segmentation with CNN and CRF," *GitHub*, 2018. https://github.com/criminalking/face_segmentation
Macomber, Hattie E., *Stories of Great Inventors: Fulton, Whitney, Morse, Cooper, Edison*, (Boston, MA: Educational Pub. Co., 1897).
Martin, Ardis C., "Television Media as a Potential Negative Factor in the Racial Identity Development of African American Youth," *Academic Psychiatry*, 32 (2008), 338–342. https://doi.org/10.1176/appi.ap.32.4.338
Marx, David M., Sei Jin Ko, and Ray A. Friedman, "The 'Obama Effect': How a Salient Role Model Reduces Race-Based Performance Differences," *Journal of Experimental Social Psychology*, 45 (2009), 953–956. https://doi.org/10.1016/j.jesp.2009.03.012
Michalopoulos, Stelios, and Melanie Meng Xue, "Folklore," *Quarterly Journal of Economics*, 136 (2021), 1993–2046. https://doi.org/10.1093/qje/qjab003
Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru, "Model Cards for Model Reporting," *Proceedings of the Conference on Fairness, Accountability, and Transparency* (2019), 220–229. https://doi.org/10.1145/3287560.3287596
Monk, Ellis P., Jr, "The Cost of Color: Skin Color, Discrimination, and Health among African-Americans," *American Journal of Sociology*, 121 (2015), 396–444. https://doi.org/10.1086/682162
Mumma, Kirsten Slungaard, "Politics and Children's Books: Evidence from School Library Collections," Ed Working Paper no. 22–693, 2022.
Neuendorf, Kimberly A., *The Content Analysis Guidebook* (Beverly Hills, CA: Sage Publications, 2016).
Nilsen, Alleen Pace, "Women in Children's Literature," *College English*, 32 (1971), 918–926. https://doi.org/10.2307/375631
Plant, E. Ashby, Patricia G. Devine, William T. L. Cox, Corey Columb, Saul L. Miller, Joanna Goplen, and B. Michelle Peruche, "The Obama Effect: Decreasing Implicit Prejudice and Stereotyping," *Journal of Experimental Social Psychology*, 45 (2009), 961–964. https://doi.org/10.1016/j.jesp.2009.04.018
Porter, Catherine, and Danila Serra, "Gender Differences in the Choice of Major: The Importance of Female Role Models," *American Economic Journal: Applied Economics*, 12 (2020), 226–254. https://doi.org/10.1257/app.20180426
Riley, Emma, "Role Models in Movies: The Impact of Queen of Katwe on Students' Educational Attainment," *Review of Economics and Statistics* (forthcoming). https://doi.org/10.1162/rest_a_01153

Roberts, Dorothy, *Fatal Invention: How Science, Politics, and Big Business Re-create Race in the Twenty-first Century* (New York: New Press/ORIM, 2011).
Rodríguez-Planas, Núria, and Natalia Nollenberger, “Let the Girls Learn! It Is Not Only about Math... It’s about Gender Social Norms,” *Economics of Education Review*, 62 (2018), 230–253. https://doi.org/10.1016/j.econedurev.2017.11.006
Rosenberg, Stanley D., Paula P. Schnurr, and Thomas E. Oxman, “Content Analysis: A Comparison of Manual and Computerized Systems,” *Journal of Personality Assessment*, 54 (1990), 298–310.
Sadoski, Mark, and Allan Paivio. *Imagery and Text: A Dual Coding Theory of Reading and Writing* (New York: Routledge, 2013). https://doi.org/10.4324/9780203801932
Sarna, Tadeusz, Janice M. Burke, Witold Korytowski, Małgorzata Różanowska, Christine M. B. Skumatz, Agnieszka Zaręba, and Mariusz Zaręba, “Loss of Melanin from Human RPE with Aging: Possible Role of Melanin Photooxidation,” *Experimental Eye Research*, 76 (2003), 89–98. https://doi.org/10.1016/S0014-4835(02)00247-6
Schaffner, Brian, and Stephen Ansolabhere, “2017 CCES Common Content,” Harvard Dataverse, 2019. https://doi.org/10.7910/DVN/3STEZY
Schuman, Howard, Charlotte Steeh, Lawrence Bobo, and Maria Krysan, *Racial Attitudes in America: Trends and Interpretations*, rev. ed. (Cambridge, MA: Harvard University Press, 1997).
Shayo, Moses, “Social Identity and Economic Policy,” *Annual Review of Economics*, 12 (2020), 355–389. https://doi.org/10.1146/annurev-economics-082019-110313
Smith, Tom W., Michael Davern, Jeremy Freese, and Stephen Morgan, “General Social Surveys Cross-Sectional Cumulative Data, 1972–2021,”, NORC University of Chicago Working Paper, 2021.
Smith, Vicky, “The ‘Caldecott Effect’,” *Children and Libraries: The Journal of the Association for Library Service to Children*, 1 (2013), 9–13.
Sobel, Joel, “Interdependent Preferences and Reciprocity,” *Journal of Economic Literature*, 43 (2005), 392–436. https://doi.org/10.1257/0022051054661530
Steele, Claude M., and Joshua Aronson, “Stereotype Threat and the Intellectual Test Performance of African Americans,” *Journal of Personality and Social Psychology*, 69 (1995), 797–811. https://doi.org/10.1037/0022-3514.69.5.797
Stout, Jane G., Nilanjana Dasgupta, Matthew Hunsinger, and Melissa A. McManus, “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 (2011), 255–270. https://doi.org/10.1037/a0021385
Szasz, Teodora, Emileigh Harrison, Ping-Jung Liu, Ping-Chang Lin, Hakizumwami Birali Runesha, and Anjali Adukia, “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* (2022). https://doi.org/10.1109/WACV51458.2022.00343
U.S. Census, “Census QuickFacts,” 2019. https://www.census.gov/programs-surveys/sis/resources/data-tools/quickfacts.html, accessed July 12, 2022.
Valadez, Corinne, Sandra Murillo Sutterby, and Tammy Francis Donaldson, “Content Analysis of Latino Award-Winning Children’s Literature,” in “Preparing Effective Leaders for Tomorrow’s Schools, Norma Zunker, ed. (Corpus Christi, TX: Consortium for Educational Development, Evaluation, and Research, 2013, 77–104.
Waldfogel, Joel, “Preference Externalities: An Empirical Study of Who Benefits Whom in Differentiated-Product Markets,” *RAND Journal of Economics*, 34 (2003), 557–569.
———, *The Tyranny of the Market* (Cambridge, MA: Harvard University Press, 2007).
Weitzman, Lenore J., Deborah Eifler, Elizabeth Hokada, and Catherine Ross, “Sex-Role Socialization in Picture Books for Preschool Children,” *American Journal of Sociology*, 77 (1972), 1125–1150. https://doi.org/10.1086/225261

Westman, Jack C., "Juvenile Ageism: Unrecognized Prejudice and Discrimination against the Young," *Child Psychiatry and Human Development*, 21 (1991), 237–256. https://doi.org/10.1007/BF00705929
Whipple, Wayne, *The Story of Young Abraham Lincoln*, (Philadelphia, PA: Henry Altemus Co., 1915).
Williams, J. Allen, Jr., JoEtta Vernon, Martha C. Williams, and Karen Malecha, "Sex Role Socialization in Picture Books: An Update," *Social Science Quarterly*, 68 (1987), 148–156.
Wilson, William J., *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy* (Chicago: University of Chicago Press, 2012).
Zhao Yu, Amy, Shahar Ronen, Kevin Hu, Tiffany Lu, and César A. Hidalgo, "Pantheon 1.0, a Manually Verified Dataset of Globally Famous Biographies," *Scientific Data*, 3 (2016), 1–16. https://doi.org/10.1038/sdata.2015.75.
Zhang, Zhifei, Yang Song, and Hairong Qi, "Age Progression/Regression by Conditional Adversarial Autoencoder," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017).
Zhou, Lei, Zhi Liu, and Xiangjian He, "Face Parsing via a Fully-Convolutional Continuous CRF Neural Network," *CoRR*, abs/1708.03736, 2017.
Zoph, Barret, and Quoc V. Le, "Neural Architecture Search with Reinforcement Learning," arXiv, 2017.