Portrayals of Race and Gender: Sentiment in 100 Years of Children's Literature

Anjali Adukia[*]
University of Chicago
Chicago, Illinois, USA
adukia@uchicago.edu

Callista Christ
University of Chicago
Chicago, Illinois, USA
callistac@uchicago.edu

Anjali Das
Columbia University
New York, New York, USA
ad3888@columbia.edu

Ayush Raj
Saint Francis High School
Mountain View, California, USA
raj.ayush@gmail.com

ABSTRACT

The way that people of different identities are portrayed in children's books can send subconscious messages about how positively or negatively children should think about people with those identities. These messages can then shape the next generation's perceptions and attitudes about people, which can have important implications for belief formation and resource allocation. In this paper, we make two contributions: (1) we examine the depiction of race and gender in award-winning children's books from the last century, and (2) we examine how consumption of these books relates to local beliefs. First, we analyze the sentiment associated with the famous individuals mentioned in these books. While the sentiment surrounding women is positive overall, on average, we see that Black women are more often portrayed with negative sentiment in Mainstream books, while White women are more often portrayed with positive sentiment. Because children's books in the United States depict more White women overall, this disguises the more negative intersectional portrayals of Black women. Books that center underrepresented identities are more likely to portray all characters with more positive sentiment. A century ago, women were much less positively spoken about than men, but the average sentiment of females and males has converged over time. The difference in sentiment connected with Black people and White people has also decreased over time, but there still remains a substantial gap. Second, we then analyze the relationship between book purchases and local beliefs to understand the potential messages being transmitted to children in different parts of society. We see that more purchases of books with positive sentiment towards Black characters are associated with a larger proportion of individuals who believe that White people in the United States have certain advantages because of the color of their skin and who are angry that racism exists. Understanding the messages that may be implicitly – or explicitly – sent to children through highly influential books can lend insight into the factors that may shape children's beliefs and attitudes.

[*]All authors contributed equally to this research.

CCS CONCEPTS

• **Social and professional topics → Race and ethnicity**; **Gender**; • **Applied computing → Law, social and behavioral sciences**; *Education*.

KEYWORDS

sentiment analysis, children's books, race, gender, education

ACM Reference Format:
Anjali Adukia, Callista Christ, Anjali Das, and Ayush Raj. 2022. Portrayals of Race and Gender: Sentiment in 100 Years of Children's Literature. In *ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS) (COMPASS '22), June 29-July 1, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3530190.3534811

1 INTRODUCTION

Children's books transmit messages about society that can shape and influence the next generation's perceptions, assumptions, and attitudes about people. In particular, the way that people of different identities are talked and written about can send subconscious messages about how positively or negatively children should think about people with those identities.

In this paper, we examine how race and gender are portrayed by analyzing how "positively" or "negatively" famous individuals in children's books from the last century are portrayed. We term this portrayal "sentiment" and use it because it is an indicator of an author's or book's attitude towards characters of different racial groups or genders. We trained a BERT model in order to make predictions of the sentiment of sentences that mention famous individuals. We then analyzed the varying sentiments towards characters of different identities.

The books we examine have received recognition from the American Library Association. We divide the books into two collections: "Mainstream" books, which are a common feature in the lives of children in the United States, whether in schools or homes, and are more likely to be bought or checked out than other children's books, and "Diversity" books, which explicitly center underrepresented identities. Because children are more likely to be exposed to

Mainstream books, it is important to understand the messages being transmitted in these books, and thus the potential relationship between exposure and beliefs.

We find that books that center underrepresented identities are more likely to portray characters of all identities with positive sentiment. While the sentiment surrounding women is positive overall, on average, we see that Black women are more often portrayed with negative sentiment in Mainstream books, while White women are more often portrayed with positive sentiment. Because children's books in the United States depict more White women overall, this disguises the more negative intersectional portrayals of Black women. A century ago, women were much less positively written about than men, but the average sentiment of females and males has converged over time.

Additionally, we analyze the relationship between book purchases and local beliefs to understand the potential messages being transmitted to children in different parts of society. We see that more purchases of books with positive sentiment towards Black characters are associated with a larger proportion of individuals who believe that White people in the United States have certain advantages because of the color of their skin and, separately, who are angry that racism exists.

How people are differentially portrayed in children's literature sends implicit and explicit messages to children about different identities. These messages can then shape beliefs and attitudes which may influence long-term decisions that have important equity and justice implications. This paper helps to shed light on some of the disparate messages contained in highly influential children's books in the United States over the last century.

The paper is organized as follows. In Section 2, we present some related background. In Section 3, we describe our data. In Section 4, we present our methods for our language model and then for analyzing sentiment. In Section 5, we summarize the results of our analysis, and in Section 6, we conclude.

2 BACKGROUND

Early exposure to stimuli shape children's beliefs about the world, others, and themselves [6, 11, 21, 37, 43]. Recent studies of books in Germany and China show that childhood exposure to textbooks with different content yields correspondingly different beliefs in adulthood, about history and self, respectively [10, 25]. Abstracting their estimates to the U.S. context, [2] find that children exposed to influential books in the United States are presented with a version of the world in which White people and males are the dominant groups. This message is transmitted in both text and images; people from these groups are represented far more often than explained by their share of the population, with all other groups being represented far less. This suggests that the millions of children exposed to these books are far more likely to believe in a world where White people and, separately, males, occupy dominant social roles, than they would were the books used to teach them more equitable in their representation of race and gender.[1]

Skewed representation of social categories such as race and gender are particularly salient to children [6, 22] and could contribute to inequality in personal and societal outcomes, ranging from labor force participation and pay to incarceration and mortality [8, 16, 38]. However, race and gender are often treated as mutually exclusive categories of experience that do not appropriately account for the intersectionality of identities [13–15].[2] A consequence of this is that individuals with multiple marginalized or excluded identities are often then excluded from theoretical, analytical, and practical consideration.[3]

Children may process identity categories – and intersectionality – subconsciously and depend less on focused cognitive skills when categorizing different identities. They may develop their heuristics and implicit biases through the feeling that is associated with different identities. For example, if Black females are depicted as angry, but White females are depicted as benevolent, this may shape a child's perceptions of how different individuals interact with the world [26].

Moreover, prior research has found that natural human language, and thus literature, has a positivity bias [9, 19, 27, 32, 48]. However, these studies examine sentiment in aggregate. Analyses that average across all individuals, thus potentially overweighting those who are more represented, may conceal the differential ways that people from underrepresented identities may be portrayed and neglect to consider intersectional portrayals.

When presenting content to children, educators not only want to know *whether* characters of different backgrounds are represented; they also want to know *how* characters of different backgrounds are represented. In this study, we examine the sentiment typically attached to characters with different race and gender backgrounds to shed light on one aspect of the way that identities are portrayed, which can have important implications for shaping children's beliefs.

3 DATA

We are interested in understanding the representation to which children are exposed regularly. We therefore draw from a set of children's books with which children are likely to interact. Specifically, we examine books that received awards administered or featured by the Association for Library Service to Children, a division of the American Library Association. We divide them into two collections: (1) a "Mainstream" collection and (2) a "Diversity" collection.

The Mainstream collections comprises books that have received recognition through the Newbery Medal or Caldecott Medal, the two oldest children's book awards in the United States starting in 1922. The books in the Mainstream collection are more likely to be purchased and two to four times as likely to be checked out from libraries than other children's books [2, 12]. This increased exposure of these books may then lead to them having greater influence than other children's books, which is consistent with

[1] Studies also show or suggest that exposure to influencing factors such as books, peers, and teachers and other adults can influence child learning outcomes such as aspirations, academic performance, and persistence in school [6, 24, 39, 40].

[2] [17] defines intersectionality as "the interaction of multiple identities and experiences of exclusion and subordination," thus accounting not only for multiple identities but also how those identities interact with experiences of exclusion and subordination.

[3] For example, the experiences of women of color may be minimized or ignored even in content designed to be deliberately equitable in their representation, if representation of women is restricted to White women, and representation of people of color is restricted to males.

research on the central role these books have in U.S. children's literature [33, 44].

The Diversity collection comprises books that have received recognition for explicitly centering underrepresented identities. The first award in this collection began in 1970. We examine books from the following awards: Rise Feminist (formerly known as Amelia Bloomer), American Indian Youth Literature Award, Américas Award, Arab American Book Award, Asian/Pacific American Award for Literature, Carter G. Woodson Book Awards, Coretta Scott King Book Awards, Dolly Gray Award, Ezra Jack Keats Book Award, Middle East Book Award, National Jewish Book Award, Notable Books for a Global Society, Pura Belpré Award, Rainbow Book List, Schneider Family Book Award, Skipping Stones Honor Awards, South Asia Book Awards, Stonewall Book Awards, Sydney Taylor Book Award, and Tomas Rivera Mexican American Award.

To characterize how race-gender identities are portrayed in our set of books overall and over time, we combine our data at two levels: (1) at the collection level, so that we can measure the overall representation in the Diversity and Mainstream collections, and (2) at the collection-by-decade level, so that we can measure changes in sentiment over time.

4 METHODS

4.1 Data Pre-Processing

We use Google Vision Optical Character Recognition to detect and extract text from the scanned pages of each children's book.[4] We then use the pre-trained Punkt tokenizer from Python's NLTK library [7] to split the corpus into sentences. For each sentence, we remove line breaks, non-ASCII characters, and punctuation. We then perform the famous figure extraction detailed in Section 4.2 and finally add the "[CLS]" and "[SEP]" tokens to the start and end of sentences, which is necessary for our sentiment models.

4.2 Famous Figure Extraction

We are interested in understanding how the sentiment associated with different identities may differ in children's books. Because children's books often have fictional characters without verifiable identities, we focus our analysis on the famous individuals mentioned in the book text who have known race and gender. This enables us to classify the race and gender of characters with high accuracy.

We use SpaCy's Named Entity Recognition (NER) pipeline to extract all "PERSON" entities from the Mainstream and Diversity collections [28].[5] To determine which entities identified by NER refer to a famous individual, we compare each entity that has at least two names (i.e., a first and last name) with a publicly available data set, Pantheon 2.0, that contains data from over 70,000 Wikipedia biographies that have a presence in more than 15 language editions of Wikipedia [50]. If the entity exists in the Pantheon 2.0 data set, we classify them as famous. We apply a fuzzy matching algorithm[6] to determine whether a name matches a famous figure in this data set.

[4]This process is restricted to the conversion of scanned text into ASCII characters only.

[5]We test a variety of SpaCy NER pipelines, and found that the en_core_web_lg pipeline performed best on our data set.

[6]We use a Levenshtein's ratio threshold of 0.9 for the fuzzy matching to capture slight variations in a name.

The Pantheon 2.0 data set contains some demographic information about each famous individual such as their gender but does not include race; we manually code the race of each famous individual that appears in our sample of children's books. If a famous figure is detected in our text, we replace their name in the text with their corresponding race-gender category, as in [1]. For example, "Rosa Parks" would be replaced with "black_woman" in the text. In the Mainstream collection, 24,365 unique and 272,776 total "PERSON" counts were detected from NER. Of these, 925 unique and 5,526 total counts were mapped to the Pantheon dataset and labeled as famous figures. In the Diversity collection, there were 45,892 unique "PERSON" entities and 379,074 total counts. From this, 2,267 unique individuals and 11,061 counts were determined as famous figures. While this method can not perfectly identify all famous individuals, we found no evidence that errors differed by race or gender.[7]

In Figure 1, Panel A, we show the number of instances of famous people who are classified as a given race-gender identity in the Mainstream and Diversity collections. A majority of the famous individuals mentioned in both collections are White males, with the Diversity collection highlighting many more Black individuals than the Mainstream collection. Black females are the least likely to be included. Note that we limit our analysis to Black and White individuals who identify as part of a binarized gender categorization (female, male), because our sample of other race-gender identities is too small to have sufficient power to estimate sentiment using our computational models. This underscores how other identities are even further underrepresented.[8]

4.3 Sentiment Analysis Model Selection

We are interested in analyzing *how* different identities are portrayed to children in the text of their books. One way to measure attitudes or feelings towards a group in text data is by analyzing sentiment. Sentiment analysis has been used to assess text from various sources such as news reports [5, 51], social media [3, 41], literature [32, 36], consumer reviews [29, 41, 49], and academic writing [34, 46]. We apply this natural language processing tool to analyze sentiment of race and gender in our corpora of children's books.

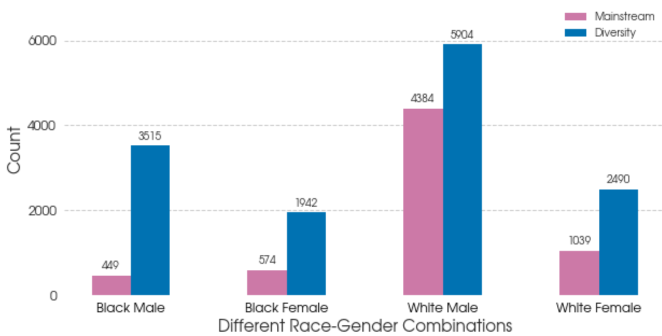
Many sentiment models are based off of Python's TextBlob [35] and VADER (Valence Aware Dictionary and sEntiment Reasoner) models [31]. We do not use these as our primary models, because they perform with lower accuracy in our corpora. This is likely because they are dictionary-based and rule-based sentiment analysis tools specifically attuned to sentiments expressed in social media, and therefore fail to understand more complex expressions of sentiment such as double negatives. We therefore use TextBlob and Vader to establish a baseline.

In order to measure sentiment, we use a BERT (Bidirectional Encoder Representations from Transformers) language model [18]. BERT has several advantages over dictionary-based approaches because it is able to establish context around text rather than scoring sentiment based on individual words, and can be fine-tuned easily.

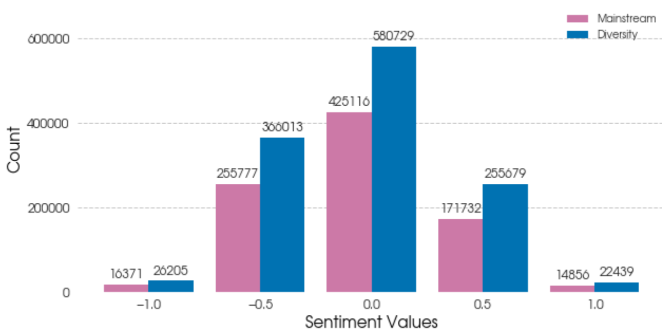
[7]One important limitation of this method is that NER does not perfectly capture names. For example, Martin Luther King Junior is mentioned in our text but in some instances only "Martin Luther" is recognized, resulting in an under count of Martin Luther King Jr. and an over count of Martin Luther.

[8]Traditional content analysis methods would be better able to assess sentiment of other under-represented race-gender identities in these corpora.

Figure 1: Sample Characteristics



(a) Number of Mentions, by Race-Gender



(b) Distributions of Predicted Sentiment Labels

Note: **Panel A shows the number of famous figures mentioned in the Mainstream and Diversity collections for each race-gender identity category. The total number of famous figures in the Mainstream collection is 5,526 and in the Diversity collection is 11,061. Panel B shows the number of sentences classified with each predicted sentiment labels by collection.**

This allows for pre-training of contextual representations from unlabeled text by considering context from both sides of the text of interest.

We tested various BERT models on our data to determine the optimal model for our task of sentiment analysis. We evaluated how well each model understood our data by using a masked language modeling task. We input 380 randomly selected sentences from our data into each model and asked the model to predict the last word of each sentence. We found that the bert-large-cased model marginally outperformed other models on this task. However, all of the BERT models we tested achieved between 36% and 42% accuracy on this masked language modeling task. So, all of the models were fine-tuned and evaluated on a custom-created test set.

4.4 Fine-tuning BERT models for Sentiment Analysis

We then fine-tune our BERT models to account for our specific context. Because BERT is a general-purpose language model, it is not specifically made for sentiment analysis. Therefore, we must fine-tune the language model on a separate data set in order to help the model develop its ability to classify the sentiment of sentences.

We used the Stanford Sentiment Treebank (SST-5) data set to fine-tune our model [45]. These data contain 11,855 sentences extracted from movie reviews, each labeled on a scale from 1-5 (1 for very negative, 2 for negative, 3 for neutral, 4 for positive, and 5 for very positive). We chose this dataset because its sentences are more formal and therefore more similar to sentences found in children's books than tweets or other social media posts as used in VADER. We convert the range of the labels to between -1 (Very Negative) and 1 (Very Positive) in order to be interpreted by our PyTorch dataloaders.

To fine-tune the model, we apply a one-layer feedforward classifier on top of the BERT model itself, which consists of a Linear layer, a ReLU activation function, and final Linear layer. We then train this model on the SST-5 dataset. This then predicts the sentiment associated with each sentence in both collections. In our analysis, we use average sentiment across collections, or collections-by-decade, for each race and gender identity.

In Figure 1, Panel B, we show the distributions of the predicted sentiment for each collection. There are overall more sentences with predicted negative sentiment than predicted positive sentiment in each collection.

4.5 Evaluating BERT Models

To evaluate our models, we create a test set containing 400 randomly chosen sentences from both the Mainstream and Diversity collections (cf. Section 4.1). Because we are more interested in non-neutral sentiments and a majority of sentences in our data have neutral sentiment, we selected sentences that VADER predicted to have non-neutral sentiment to create the test set. We had three hand-coders manually label the sentiment for each of the 400 sentences on the same scale as the SST-5 data set, and used the majority score for each sentence as our final sentiment score.

We predict the sentiment score for each sentence in our test set for all models. We then compute the accuracies and average F-scores for the test data. Table 1 organizes our findings for accuracy and F-scores for the models we evaluated.

We use the bert-base-uncased model as it had the highest accuracy score on the non-neutral sentences.[9] While the BERT model is accurate on the SST-5 dataset, one potential limitation is that the performance may not be completely replicated when the model makes predictions on our custom test data set, which is limited to sentences that contain a famous figure.

5 RESULTS

We are interested in understanding how children's books depict characters of different race-gender identities. Specifically, we examine how positively or negatively characters are portrayed using sentiment analysis. We then examine how local book purchases relate to local consumer beliefs.

[9]The bert-base-uncased model that we used had approximately 47 percent accuracy, and the highest accuracy a BERT model has achieved on the SST-5 dataset is approximately 59 percent, or 80 percent of the maximum accuracy we may have hoped to have obtained [47].

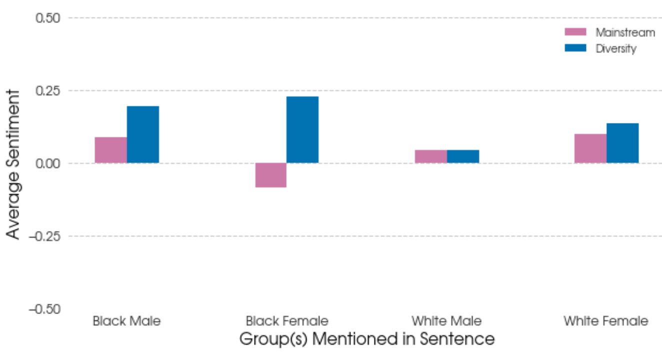
Table 1: Model Performance

Model	Accuracy	F-Score
bert-base-uncased	**0.47**	**0.46**
best-bert-model	0.43	0.41
bert-base-uncased-pytorch-trainer	0.45	0.42
bert-base-uncased-new-layers	0.46	0.44
roberta-base	0.39	0.37
vader	0.38	0.28
textblob	0.42	0.37
nlptown	0.26	0.30

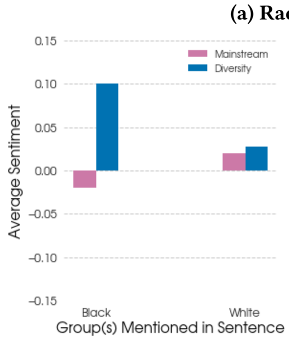
***Note:* This table reports accuracy and F-scores for non-neutral sentences in our test set. The bert-base-uncased model has the architecture described in Section 4.4. The best-bert-model has a Softmax layer added to this baseline architecture. The bert-base-uncased-new-layers model contains some layer experimentation, which includes new Linear layers.**

5.1 Sentiment of Race and Gender

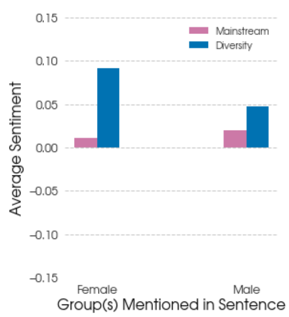
Figure 2: Sentiment of Characters, by Race and Gender



(a) Race-Gender



(b) Race



(c) Gender

***Note:* This figure shows the average sentiment of famous figures in different race-gender (Panel A), race (Panel B), and gender (Panel C) identity groups in the Mainstream and Diversity collections for non-neutral sentences.**

In Figure 2, we show the average sentiment by race-gender, race, and gender in the Mainstream and Diversity collections. We see that the average sentiment in the books in the Diversity collection is more positive than the average sentiment in the books in the Mainstream collection for each identity.

We show in Figure 2, Panel A that the difference in average sentiment between Diversity and Mainstream books is greatest for Black females and for Black males. There is little difference in average sentiment of White males between each collection. Black males, White females, and White males are portrayed positively on average in both Mainstream and Diversity collections. However, Black females are more likely to be depicted with negative sentiment in the Mainstream collection. These reflect the net sentiment associated with different identities, and each set of identities have sentences associated with them that are classified as being positive and as being negative.

Examples of sentences classified as having positive sentiment include:

> "Closing her eyes, Marian Anderson began to sing, and her thrilling contralto voice carried across the Mall, touching every person who had come to hear her. [23]"

> "Anne Hutchinson's fearless stand for truth as she saw it-this, together with her remarkable gifts, set her apart from all the women of her day. [20]"

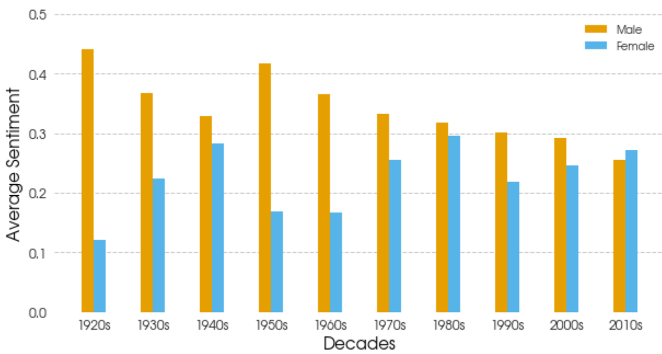
Examples of sentences classified as having negative sentiment include:

> "Interrupting another's speech was almost as horrendous an act in Aunt Cordelia's code as neglecting one's homework or calling Jane Austen a bore. [30]"

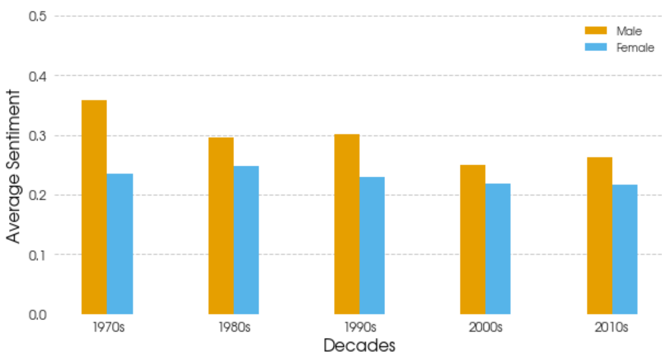
> "As in: I wanted a pair of Stephon Marbury's sneakers (Starburys), but Dad called him a selfish millionaire with a bad attitude, and why would I want to be associated with such a churlish choke artist. [4]"

We see in Figure 2, Panels B and C that the average sentiment surrounding women, men, and White people is positive overall. However, we see that Mainstream books are more likely to depict Black people with negative sentiment. As we see in Panel A, this average negative sentiment may be driven by the average negative depiction of Black females. Similarly, we see that females are more likely to be depicted with positive sentiment, with much less positive sentiment overall in the Mainstream collection. This smaller positive sentiment of females disguises the disparity between how Black women and White women are portrayed. We see that Black females are more likely to be depicted with negative sentiment in the Mainstream books, whereas White women are more often portrayed with positive sentiment. Because U.S. children's books include more White women than Black women as characters, this may conceal the more negative intersectional portrayals of Black women in society.

Figure 3: Sentiment of Characters by Gender, Over Time



(a) Mainstream

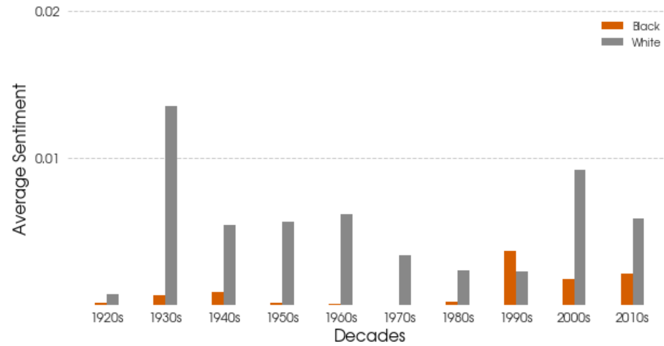


(b) Diversity

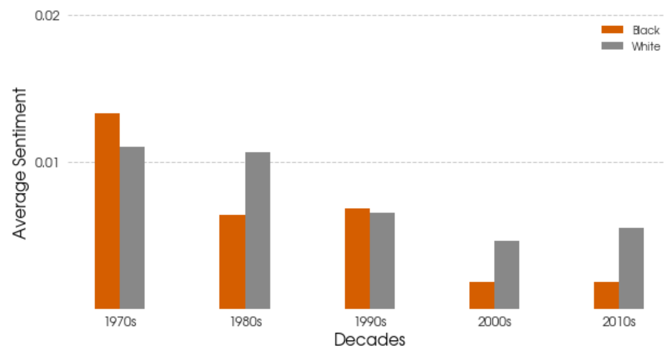
***Note:* This figure shows the average sentiment of female and male famous figures over time across all books.**

We show how average sentiment of females and males has changed over time in each collection in Figure 3. A century ago, women were much less positively spoken about than men, but the difference in the average sentiment of females and males has decreased over the decades. In the Mainstream collection (cf. Panel A), the average sentiments between females and males have converged over time. In the Diversity collection (cf. Panel B), average sentiment towards females has remained similar, but the average sentiment of how males have been depicted has decreased over time. Both females and males have been depicted positively on average in each decade.

Figure 4: Sentiment of Characters by Race, Over Time



(a) Mainstream



(b) Diversity

***Note:* This figure shows the average sentiment of Black and White famous figures over time across all books.**

In Figure 4, we show how average sentiment of females and males has changed over time in each collection. The patterns differ across the Mainstream and Diversity collections. Average sentiment of Black characters in the Mainstream collection was extremely low 100 years ago, especially relative to average sentiment of White characters, but it has increased over time, especially in the last three decades (cf. Panel A). The difference in sentiment connected with Black people and White people has also decreased over time, but there still remains a substantial gap.

By contrast, the average sentiment of Black and White characters was more similar at the advent of the Diversity collection. The average sentiment of each group has decreased over the decades, but the difference between the groups has increased (cf. Panel B). In both collections, White characters are generally more likely to be positively depicted than Black characters.

5.2 Local Beliefs and Book Consumption

The messages in books purchased reflect the beliefs of consumers. These individuals – and their associated beliefs – comprise their local societies. We examine the relationship between book purchases and local beliefs to understand the potential messages being transmitted to children in different parts of society. In particular, we draw from the 2017 Cooperative Election Study (CCES), a nationally representative survey administered by YouGov.[10] The survey collects information about general political attitudes and demographics. We merge these data with the number of books from the Mainstream and Diversity collections purchased by zip code between 2017-2020 using data from the Numerator Omnipanel data, a consumer panel data set. These data include information from over 1 billion shopping trips from over 44,000 retailers.

Table 2: Beliefs and Book Consumption, Sentiment by Race

	Dependent variable:	
	% of Respondents who somewhat or strongly agree	
	White people in the U.S. have certain advantages because of the color of their skin	I am angry that racism exists
	(1)	(2)
Sentiment of Black characters	0.131^{*} (0.076)	0.176^{***} (0.060)
Sentiment of White characters	-0.077 (0.076)	-0.087 (0.060)
Constant	55.871^{***} (0.293)	80.234^{***} (0.232)
Observations	17,930	17,930
Adjusted R^2	0.0001	0.0005

***Note:* Data on beliefs at the zip code level are drawn from the 2017 Cooperative Election Study Common Content Survey [42]. Data on children's book purchases at the zip code level are drawn from the 2017-2020 Numerator Omnipanel. Variables containing percentages are scaled so that potential values range from $0 - 100$. *p<0.1; ***p<0.01**

In Table 2, we show that more purchases of books with positive sentiment towards Black characters are associated with 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 1) and who are angry that racism exists (column 2). When examining this pattern by race-gender, we see that a larger proportion of books that depict Black males positively are purchased in areas where a larger proportion of individuals report being angry that racism exists. As a placebo check, we separately examine this pattern by gender. and we see no differences between sentiment of females versus sentiment of males.

These relationships may be driven by several potential explanations. For example, higher consumption of books with positive sentiment towards Black characters might be driven by demographic composition or other contemporaneous factors that may have affected beliefs. These patterns reflect correlations and do not represent causal evidence.

[10]CCES was formerly known as the Cooperative Congressional Election Study.

6 CONCLUSION

The way that people from different backgrounds are portrayed in content given to children can shape and influence children's beliefs about people from those backgrounds. We measure the sentiment associated with the sentences in which race-gender identities are depicted in award-winning children's books from the last century in the United States. The measure of "sentiment" captures a feeling of positivity or negativity towards characters of different racial groups or genders.

We examine the books in two different collections: Mainstream books, those that are commonly found in schools or homes, and Diversity books, which explicitly center underrepresented identities. Mainstream books are more likely to be bought or checked out than other children's books, which suggests that children are more likely to be exposed to the messages in them. Therefore, it is useful to understand the messages being transmitted in these books, and thus the potential relationship between exposure and beliefs.

Books that center underrepresented identities are more likely to portray characters of all identities with positive sentiment. The sentiment associated with women is positive on average; however, Black women are more often portrayed with negative sentiment in Mainstream books, while White women are more often portrayed with positive sentiment. Children's books in the U.S. depict more White women overall, so this average positive sentiment for women overall conceal the more negative intersectional portrayals of Black women. Women were much less positively written about than men 100 years ago, but the average sentiment of females and males has converged over time.

We then analyze the relationship between book purchases and local beliefs to understand the potential messages being transmitted to children in different parts of society. We see suggestive evidence that more purchases of books with positive sentiment towards Black characters are associated with a larger proportion of individuals who believe that White people in the United States have certain advantages because of the color of their skin and who are angry that racism exists. These associations may be driven by various contemporaneous factors that may have affected beliefs.

This work is particularly relevant to approaches and research that address challenges faced by marginalized, underrepresented, or misrepresented communities. This work brings together insights from multiple disciplines including computer science, the social sciences, and the humanities in order to address key challenges for sustainable societies such as equity, education, poverty, accessibility, and economic growth. Understanding the messages that may be implicitly – or explicitly – sent to children through highly influential books can lend insight into the factors that may shape children's beliefs and attitudes.

ACKNOWLEDGMENTS

For financial support, we thank Becker-Friedman Institute at UChicago, Center for Data and Applied Computing at UChicago, and UChicago Career Advancement. The research reported here was also supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A200478 to the University of Chicago. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. For excellent research assistance, we thank Juan Miguel Jimenez, Avery Schoen, Celia Zhu, and Umama Zillur. For very helpful feedback, we thank Patricia Chiril, Alex Eble, Emileigh Harrison, Hakizumwami Birali Runesha, and David Uminsky. We thank the UChicago Research Computing Center for access to computational resources. Researchers' own analyses calculated (or derived) based in part on data from Market Track, LLC dba Numerator and marketing databases provided through the Numerator Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Numerator data are those of the researchers and do not reflect the views of Numerator. Numerator is not responsible for and had no role in analyzing or preparing the results reported herein.

REFERENCES

[1] Anjali Adukia, Callista Christ, Anjali Das, Alex Eble, Emileigh Harrison, and Hakizumwami Birali Runesha. 2022. Category Embeddings Measure Intersectional Portrayals of Race and Gender. *Working Paper* (2022).

[2] Anjali Adukia, Alex Eble, Emileigh Harrison, Hakizumwami Birali Runesha, and Teodora Szasz. 2021. What We Teach About Race and Gender: Representation in Images and Text of Children's Books. *National Bureau of Economic Research Working Paper 29123* (2021).

[3] Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca J Passonneau. 2011. Sentiment analysis of twitter data. *Proceedings of the Workshop on Language in Social Media (LSM)* (2011), 30–38.

[4] Kwame Alexander. 2019. *The crossover.* Houghton Mifflin Harcourt.

[5] Alexandra Balahur, Ralf Steinberger, Mijail Kabadjov, Vanni Zavarella, Erik Van Der Goot, Matina Halkia, Bruno Pouliquen, and Jenya Belyaeva. 2013. Sentiment analysis in the news. *arXiv preprint arXiv:1309.6202* (2013).

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

[7] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: Analyzing text with the natural language toolkit.* O'Reilly Media, Inc.

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

[9] Jerry Boucher and Charles E Osgood. 1969. The pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behavior* 8, 1 (1969), 1–8.

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

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

[12] Marlaina Cockcroft. 2018. Caldecott and Newbery Medal wins bring instant boost to book sales. *School Library Journal* 64, 2 (2018). https://www.slj.com/?detailStory=caldecott-newbery-medals-translate-bigger-short-term-sales

[13] Combahee River Collective. 1983. The Combahee River Collective Statement. *Home Girls: A Black Feminist Anthology* 1 (1983), 264–274.

[14] Kimberlé Crenshaw. 1989. Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum* 1989, 8 (1989), 139.

[15] Kimberlé Crenshaw. 1990. Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review* 43 (1990), 1241.

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

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

[18] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv e-prints*, Article arXiv:1810.04805 (Oct. 2018), arXiv:1810.04805 pages. arXiv:1810.04805 [cs.CL]

[19] Peter Sheridan Dodds, Eric M Clark, Suma Desu, Morgan R Frank, Andrew J Reagan, Jake Ryland Williams, Lewis Mitchell, Kameron Decker Harris, Isabel M Kloumann, James P Bagrow, et al. 2015. Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences* 112, 8 (2015), 2389–2394.

[20] Jeanette Eaton. 1944. *Lone journey: The life of Roger Williams.* Harcourt Brace.

[21] Alex Eble and Feng Hu. 2019. Does primary school duration matter? Evaluating the consequences of a large Chinese policy experiment. *Economics of Education Review* 70 (2019), 61–74.

[22] Joe R. Feagin and Debra Van Ausdale. 2001. *The first R: How children learn race and racism.* Rowman & Littlefield Publishers.

[23] Russell Freedman. 2011. *The voice that challenged a nation: Marian Anderson and the struggle for equal rights.* Sandpiper/Houghton Mifflin Harcourt.

[24] Jane Cooley Fruehwirth and Jessica Gagete-Miranda. 2019. Your peers' parents: Spillovers from parental education. *Economics of Education Review* 73 (Dec. 2019), 101910. http://www.sciencedirect.com/science/article/pii/S0272775719301219

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

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

[27] Clarence Green. 2017. Introducing the Corpus of the Canon of Western Literature: A corpus for culturomics and stylistics. *Language and Literature* 26, 4 (2017), 282–299.

[28] M Honnibal and I Montani. 2017. Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. *Unpublished software application* (2017). https://spacy.io

[29] Yi Hu and Wenjie Li. 2011. Document sentiment classification by exploring description model of topical terms. *Computer Speech & Language* 25, 2 (2011), 386–403.

[30] Irene Hunt. 1994. *Up a road slowly.* Library of Congress, NLS/BPH.

[31] Clayton J. Hutto and Eric Gilbert. 2014. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *ICWSM* (2014).

[32] Arthur M. Jacobs, Berenike Herrmann, Gerhard Lauer, Jana Lüdtke, and Sascha Schroeder. 2020. Sentiment Analysis of Children and Youth Literature: Is There a Pollyanna Effect? *Frontiers in Psychology* 11 (2020), 2310. https://doi.org/10.3389/fpsyg.2020.574746

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

[34] Robert J Lennox, Diogo Veríssimo, William M Twardek, Colin R Davis, and Ivan Jarić. 2020. Sentiment analysis as a measure of conservation culture in scientific literature. *Conservation Biology* 34, 2 (2020), 462–471.

[35] Steven Loria. 2018. textblob Documentation. *Release 0.15* 2 (2018).

[36] Saif M Mohammad. 2012. From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems* 53, 4 (2012), 730–741.

[37] Natalia Nollenberger, Núria Rodríguez-Planas, and Almudena Sevilla. 2016. The math gender gap: The role of culture. *American Economic Review* 106, 5 (2016), 257–61.

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

[39] Claudia Olivetti, Eleonora Patacchini, and Yves Zenou. 2020. Mothers, peers, and gender-role identity. *Journal of the European Economic Association* 18, 1 (2020), 266–301.

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

[41] Huaxia Rui, Yizao Liu, and Andrew Whinston. 2013. Whose and what chatter matters? The effect of tweets on movie sales. *Decision Support Systems* 55, 4 (2013), 863–870.

[42] Brian Schaffner and Stephen Ansolabhere. 2019. 2017 CCES Common Content. https://doi.org/10.7910/DVN/3STEZY

[43] Robert S Siegler, Judy S DeLoache, and Nancy Eisenberg. 2003. *How children develop.* Macmillan.

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

[45] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. 2013. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. 1631–1642.

[46] Valery Solovyev, Marina Solnyshkina, Elzara Gafiyatova, Danielle McNamara, and Vladimir Ivanov. 2019. Sentiment in academic texts. *2019 24th Conference of Open Innovations Association (FRUCT), IEEE* (2019), 408–414.

[47] Zijun Sun, Chun Fan, Qinghong Han, Xiaofei Sun, Yuxian Meng, Fei Wu, and Jiwei Li. 2020. Self-explaining structures improve nlp models. *arXiv preprint arXiv:2012.01786* (2020).

[48] Teresa Sylvester, Mario Braun, David Schmidtke, and Arthur M Jacobs. 2016. The Berlin affective word list for children (kidBAWL): exploring processing of affective lexical semantics in the visual and auditory modalities. *Frontiers in Psychology* 7 (2016), 969.

[49] Fu Xianghua, Liu Guo, Guo Yanyan, and Wang Zhiqiang. 2013. Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling

and HowNet lexicon. *Knowledge-Based Systems* 37 (2013), 186–195.

[50] Amy Zhao Yu, Shahar Ronen, Kevin Hu, Tiffany Lu, and César A Hidalgo. 2016. Pantheon 1.0, a manually verified dataset of globally famous biographies. *Scientific Data* 3, 1 (2016), 1–16. https://doi.org/10.1038/sdata.2015.75

[51] Liang-Chih Yu, Jheng-Long Wu, Pei-Chann Chang, and Hsuan-Shou Chu. 2013. Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news. *Knowledge-Based Systems* 41 (2013), 89–97.