Separation of Church and State Curricula? Examining Public and Religious Private School Textbooks*

Anjali Adukia† Emileigh Harrison‡

October 18, 2024
Active Working Draft

Abstract

Curricular materials not only impart knowledge but also instill values and shape collective memory. Growth in U.S. school choice programs has increased public funds directed to religious schools, but little is known about what is taught. We examine textbooks from religious educational settings and from public schools in Texas and California, applying and improving upon artificial intelligence tools to measure topics, values, representation, and portrayal over time. Political polarization suggests a narrative of divergence, but our analysis reveals meaningful parallels between the public school collections overall, while religious textbooks differ notably, featuring less female representation, characters with lighter skin, more famous White individuals, and differential portrayal of topics such as evolution and religion. Important similarities, however, also emerge: for example, each collection portrays females in contexts that are more positive but less active and powerful than males, and depicts the U.S. founding era and slavery in similar contexts.

Keywords: Curricula, Education Policy, Religious Education, Diversity and Inclusion in Education, Artificial Intelligence Tools

JEL Category: I20, I21, I28, J15, Z13

*For helpful feedback, we thank Ben Arold, Sam Bazzi, Marianne Bertrand, Dan Black, Ethan Bueno de Mesquita, Michael Dinerstein, Steven Durlauf, Alex Eble, Martin Fiszbein, Sendhil Mullainathan, Thomas Wollmann, David Yanagizawa-Drott, Hiro Yoshikawa, and seminar participants at AEFP, APPAM, CESifo, Columbia, Harvard, RAND, Stanford, SREE, UChicago, Univ. of Michigan, UPenn, BFI AI in Social Science conference, and the Zurich AI/Economics conference. For excellent research support, we thank Sam Abers, Matt Bonci, Elisa Xi Chen, Paula Dastres Gallardo, Benjamin Leiva, Celia Zhu, and other members of the MiiE Lab. For financial support, we thank the National Academy of Education, Spencer Foundation, and Becker Friedman Institute at UChicago.

† University of Chicago, MiiE Lab, IZA, and NBER, adukia@uchicago.edu

‡ University of Chicago, MiiE Lab, and the Inclusive Economy Lab, harrisone@uchicago.edu

Our collective memory and understanding of the world is shaped by a variety of forces including the stories we share (Michalopoulos and Xue, 2021; Adukia et al., 2023), the monuments we build and the history they represent (Bazzi et al., 2023), as well as narratives propagated through popular culture (Riley, 2022; Michalopoulos and Rauh, 2024). These elements not only influence what is remembered but also determine who is remembered and how. Education plays a pivotal role in this process with curricular materials serving as critical sites of learning. These materials do more than convey academic knowledge; they also instill children with societal values such as environmental stewardship, equality, and civic responsibility. Consequently, changes in educational content can profoundly impact children's worldviews, shaping their perspectives both explicitly through the discussion of certain topics and values, and indirectly through who is represented and how they are portrayed (Cantoni et al., 2017; Arold et al., 2022; Arold, 2024; Riley, 2022).

The role of education in shaping belief systems has historically empowered religious organizations to dominate educational sectors around the world, shaping both curricular content and the educational landscape (Bazzi, Hilmy and Marx, 2020). In the United States, this dynamic is evident in the recent explosion of school choice programs, which allow parents to use public funds to enroll their children in schools that better align with their values or religious beliefs. However, lack of regulation regarding curricula means we know very little about what children are taught in religious educational settings and how it compares to what children are taught in public schools. Given the growing allocation of tax dollars to these alternatives through vouchers and other school choice mechanisms, it becomes increasingly vital to understand the educational content that children are exposed to across educational settings to assess whether it aligns with the broader educational goals of our society and preserves our collective memory across generations.

In this study, we examine textbooks that children in the U.S. are either likely exposed to or increasingly exposed to. We assess state-adopted public school textbooks in California and Texas, states that serve the largest student populations in the U.S., which allows us to understand how educational materials vary across different political landscapes. We also examine curricular materials used in religious private and home school contexts, which helps us understand how both religious and political contexts influence the educational content across various educational environments. We examine three main categories of content analysis: First, we analyze *what* topics and values are being included. Second, we examine *who* are highlighted in these lessons. Third, we explore *how* different groups of people, topics, and values are discussed.

Specifically, we focus on Reading, Science, and Social Studies textbooks intended for

third and fifth grade students. We apply and improve upon computational content analysis methods to measure what topics, values, and groups of people from different race, gender, and sexuality backgrounds are present in educational materials and if present, how they are portrayed, and how this differs across educational settings over time.

To examine *what* is being communicated, we employ a large language model to summarize the topics and values present on a given page. We find similarities in the top topics present across public and religious private curricula which includes topics such as family, nature, friendship, biology, geography, agriculture, and astronomy. All three collections put a large emphasis on character values such as curiosity and perseverance with a slightly larger focus on these values in religious textbooks. In contrast to public school textbooks, religious textbooks place a significantly greater emphasis on faith and religious values, including in Science textbooks. We also see that religious textbooks are more likely to discuss creationism and less likely to discuss climate change.

When examining *who* is represented, we analyze the people shown in images and identity-specific words in the text. For images, we first use a face detection model described in Szasz et al. (2022) to identify the faces of all pictured characters (both in photographs as well as in illustrations). We then classify skin color by first segmenting the facial area of interest and then calculating a continuous measure of its darkness or lightness based on the color of each pixel in the segmented skin. Separately, we predict the putative race, gender, and age of each face using a feature classification model described in Szasz et al. (2022). From the text, we examine which famous individuals are highlighted and then examine the presence of different race and gender groups among famous people. Additionally, we assess text-based gender representation by examining the proportion of gender-specific words used, including gendered pronouns, gendered terms, famous individual names, and other mentioned character names. Within Religious private school textbooks, we see that pictured faces are depicted with lighter skin colors on average. We see similar presence of females and males in Texas and California public school textbooks. However, in Religious private school textbooks, we see less female representation in both the images and text. In particular, we observe unequal gender representation in the Science and Social Studies textbooks across all collections, likely due to historical underrepresentation of women in scientific fields and positions of power. Conversely, public school Reading textbooks show equal gender representation, presumably because these textbooks have greater flexibility in the stories they choose to present. However, this gender equality is not mirrored in the Reading textbooks from our religious collection, highlighting a significant disparity.

We also see very little LGBTQIA+ representation. Indeed, it is only in the California

collection that we see any mention of famous individuals who are identified as LGBTQIA+: Harvey Milk, a famous slain political pioneer and Chaz Bono, the transgender son of a musician. These two individuals appear in the same passage found within a 3rd grade Social Studies textbook.

To understand *how* different identities are portrayed, we use word embeddings to measure the context surrounding different groups of people. In particular we utilize word embedding association tests (WEAT) to gauge the proximity of these groups to various words or domains (groups of words related to specific topics including politics and family life). To facilitate this analysis, we introduce a novel placebo test to evaluate whether the word embedding associations we find are meaningfully large. Additionally, we apply semantic role labeling to discern and characterize the sentiment of actions attributed to different groups of people. Despite differences between who is represented between Religious and public school textbooks, we see similar patterns across the three collections when examining how different marginalized groups are portrayed when they are present. All three collections describe females more often in the context of the family and the home compared to males, and describe males more often in connection to politics and tools compared to females. The top words more associated with females and least associated with males predominantly revolve around appearance, family roles, or household duties such as sewing or groceries. In contrast, the top words associated with males and least associated with females revolve around military/fighting and leadership. In comparison to males, words and actions associated with females are more positive, but less active and less powerful.

When the two LGBTQIA+ individuals are mentioned in the California textbook, they are described as fighting for LGBTQIA+ rights, but there is little other accompanying discussion. However, in Religious textbooks, LGBTQIA+ rights are presented as a threat to religious rights.

Our study contributes to a vast literature focused on the content analysis of children's literature and educational materials (Wade, 1993; Allen, Allen and Sigler, 1993; Clark, 2007; Braden and Rodriguez, 2016; Bickford and Knoechel, 2017; Crisp, Gardner and Almeida, 2018; Koss, 2015; Koss, Johnson and Martinez, 2018; Koss and Paciga, 2020; Adukia et al., 2023). Most closely related to our approach is a study by Lucy et al. (2020) which uses natural language processing tools to examine who is present in the text of a small sample of 15 Texas high school textbooks used between 2015 and 2017. They find that Latine people are rarely discussed, the most common famous figures are nearly all White men, and that women tend to be discussed in the contexts of work and the home. Another related paper by Stevenson and Zlotnik (2018) focused on college level introductory economics textbooks

finds that fictionalized women more likely to be involved in food, fashion, or household tasks. In a qualitative analysis from Scaramanga and Reiss (2018) on curriculum published by Accelerated Christian Education (ACE) – a publisher in our sample of religious private school textbooks – revealed that churches and schools are often depicted as racially segregated in their materials. We expand on this work by applying computer-driven content analysis to examine both the images and text found in over two hundred K-12 textbooks used in a variety of contexts between 1980 to the present. Our work surpasses the number of books previously systematically analyzed and is able to both replicate qualitatively similar results found by previous studies in addition to documenting new findings. The larger sample and greater breadth of tools allows us to look at a broader set of questions such as how different skin colors are represented, which is an important dimension of representation not captured by the analysis of race alone (Keith and Monroe, 2016).

The innovative uses of computer vision and natural language processing tools we develop and apply throughout our analysis, contribute to the computational social science literature in a manner similar to Caliskan, Bryson and Narayanan (2017); Garg et al. (2018); Kozlowski, Taddy and Evans (2019). These methodological advancements have significant implications for the broader field of computational social science.

The paper proceeds as follows: Section I provides additional context to religious school choice in the United States, Section II describes the textbooks in our sample, Section III discusses our text and image processing tools, Section IV presents our results, and Section V concludes.

I Parental Educational Investment Decisions and School Choice

Changes in educational content and instruction have been shown to have a causal impact on children's worldviews and later life outcomes. In particular, changes in the coverage of evolution theory have been shown to affect students' probability of choosing a STEM career (Arold, 2024) and changes in high school politics curricula has been shown to affect students' views on democracy and free markets (Cantoni et al., 2017). Similarly, messages about the abilities of others with similar racial and gender identities have been shown to impact children's academic performance (Riley, 2022). Consequently, parents' decisions about their children's education are driven by the values, cultural norms, and particular sets of knowledge they want to instill in their children.

There are two main ways that values are transmitted to children, the first is horizontal transmission from society and peers, typically provided by formal institutions such as public schools and many private schools. The second is vertical transmission of values and knowl-

edge from family and parents. If parental values differ enough from those of mainstream society, they may choose to remove their children from public school and either homeschool them or place them in a private school whose values are perceived to be more aligned.

In recent years, there has been a notable increase in debates about what should be included or excluded from public school curricula, accompanied by significant shifts in public school content standards. Some places have embraced greater inclusion; a notable example is California's 2011 FAIR Education Act, which mandates instruction in History and Social Studies to include a variety of underrepresented ethnic and cultural populations in the teaching of California and United States history including both men and women, Native Americans, Mexican Americans, LGBTQIA+ individuals, and persons with disabilities. In other places, there has been a trend towards more restrictive content policies, often labeled as "educational gag orders." Between January and September 2021, for instance, legislatures in 24 U.S. states introduced 54 bills aimed at barring specific "prohibited" or "divisive" concepts from public K-12 and university curricula. These restrictions predominantly focus on limiting discussions around race, gender, and American history, as detailed in (Friedman and Tager, 2022). Notably, such restrictions are not imposed on private schools and homeschooling families. If parents disagree with what public schools are teaching, they have the option to remove their children from public school and find an alternative education setting. In tandem with the increasing discussion around what we allow to be taught in public schools, we have witnessed a surge in the adoption and expansion of school choice initiatives, including voucher programs, charter schools, and homeschooling (EdChoice, 2023).

Despite these increases in school choice options and subsequent public funding, curricular requirements in the U.S. are minimal for private schools and homeschooling families. Most states simply require private school curricula to be "similar" to standards prescribed for public schools or require that certain subjects be taught. There are 18 states that either have no requirements or only require that students be taught the state constitution and state history. While states usually have similar curricular requirements on paper for homeschoolers as they do for private schools, there is rarely any enforcement for homeschooling families Bartholet (2020). This lack of regulation is especially surprising given how much political attention is given to what can and cannot be included in public school curricula.

The evidence on the effect of school choice programs within the United States on children's academic outcomes is mixed. Deming et al. (2014) find that school choice across public high schools benefits applicants' academic outcomes only when they gain access to higher-quality schools. Research on the effectiveness of school voucher programs, which allow students to direct public funds toward tuition payments at private schools of their choice,

is also mixed. An analysis of the Milwaukee Parental Choice Program, one of the oldest voucher programs in the United States, finds modest test score gains for voucher users but Abdulkadiroglu, Pathak and Walters (2015) find that parents in more rural settings such as Louisiana are using these programs to move their children to lower quality schools as measured by students' academic performance.

This suggests that parents may be leaving the public school system for more than just better academic outcomes for their children; they may be looking for an educational setting that they perceive to better align with their values and beliefs. We see evidence of this in survey data. In a survey of parents with children in private schools, 80% listed character or values instruction as a top desired school quality (Catt and Rhinesmith, 2017). A desire for religious instruction is reflected in private enrollment patterns; most students educated outside the public school system are educated in a religious school environment. Indeed, In 2017, 76% of the 5.8 million private school students were enrolled in a religious school, with 49% enrolled in a conservative Christian school (PSS, 2015). Among those parents who chose to educate their children at home, or in informal settings, 75% reported a desire to provide moral instruction as a top reason they chose to homeschool (National Center for Education Statistics, 2022). Pre-COVID estimates of the number of homeschoolers who are religious, or for whom religion is a primary motivation for homeschooling, ranges from over half to 90% (Kunzman, 2010; Bartholet, 2020). While academic content certainly influences parents' school choice decisions, they are often not the sole or even most important factor.

Given the growing number of tax dollars being diverted from public schools to religious private schools and homeschooling families with little regulation on what they are taught and limited evidence that this diversion benefits students academically, it is important for to us to understand what children are being taught in these private educational settings and how it compares to what children are taught public schools.

II Data

We digitize a novel data set of Reading, Science, and Social Studies textbooks written for third and fifth grade. Our sample draws from three distinct sources, detailed in the subsequent subsections. As depicted in Table 1, our data set comprises over 75,000 pages originating from 261 different textbooks, amounting to a corpus of over 16 million words. To our knowledge, this is the largest data set of textbooks collected and digitized by an order of magnitude. Previous studies have typically digitized only a small number of textbooks, limiting their analyses to cross-sectional examinations. Our panel of textbooks affords a unique opportunity for longitudinal analysis across sets of curricula and over multiple decades, setting our study apart in both scope and depth within the academic literature.

II.A Religious (Conservative Christian) Textbooks

The religious textbooks in our sample include those published by three ideologically Christian textbook companies: Abeka, Bob Jones University Press (BJU), and Accelerated Christian Education (ACE). We refer to this set of books as our Religious collection. These textbooks are commonly used by homeschooling families and adopted by over a third of non-Catholic Christian schools receiving tax-payer funding (Klein, 2017). We focus on textbooks published by these three Christian publishers both because of their wide adoption, but also because there are many taxpayer dollars funding schools which use these curricula. Notably, in Indiana, about 4,240 students received over $16 million in scholarships to attend schools that use Abeka or Bob Jones curriculum, according to 2016-2017 figures from the Indiana Department of Education (Klein, 2017).

There are also few comprehensive curricular options that families can easily adopt when they choose to homeschool their children. These curricula conveniently provide such options for families. One can simply go online and purchase an entire grade's worth of curricular materials from one of these publishers. Using Google Trends search data as a measure of interest in these topics (Figure 1), we see that there was steady interest in Abeka over time, but after the onset of the COVID-19 lockdowns in March 2020, there was a marked increase in searches for Abeka. We also see that there was a steady decline in search interest for homeschool, but then it spiked once schools shut down and then matched the search interest for Abeka.

We digitize each textbook published by these publishers for the grades and subjects in our sample starting with books published in 1980 up until 2022. Note that most of the curriculum published by ACE consists of workbooks called PACEs instead of textbooks, so we only include books from their Reading curriculum in our analysis.

II.B Texas State-Adopted Textbooks

We include Texas textbooks as a comparator group because historically, Texas has driven much of the textbook procurement market in the U.S. due to their large share of the educational publishing market.

The state of Texas compiles a list of state adopted textbooks for all K-12 grades which is updated yearly. Before 2011, Texas public schools were only allowed to use funds to purchase books from this list. In 2011, this law changed to allow schools to use funds to purchase any textbooks, but most schools continued to purchase textbooks off the state adopted list. We digitized a convenience sample of approximately 50% of all textbooks on the yearly state adopted textbook list for textbooks in use between 1985-2010. For the years

2011 through 2021, we digitized the top three textbooks purchased by school districts each year for each relevant grade and subject (reliable textbook purchasing data was not collected before 2011).

II.C California State-Adopted Textbooks

California also has a large influence on the textbook publishing market due to the size of its student population, making it another useful comparator group. It is typically associated with more progressive policies and is often posed as the political or substantive converse of Texas, which is on the conservative end of the political spectrum in the United States.

The state of California compiles a list of state-adopted textbooks for grades K-8 which is updated every couple years for each subject.[1] We digitize the universe of books in the relevant grades and subjects which were adopted in California for all years in which the list of adopted materials are available. This includes adopted textbooks beginning in 1999 for Social Studies, 2000 for Science, and 2002 for Reading and ending with the books that are currently in use as of 2022.

III Methods

The majority of previous studies examining the content of textbooks have been conducted using manual content analysis which involves recording the presence of certain words or themes by hand. Because this process is time-consuming and therefore costly, resource constraints have limited the scope of such work. The need to analyze large corpora of educational materials led to our exploration of computer vision and natural language processing tools, which allow for a more cost-effective and systematic approach to content analysis. More recent studies which apply computational content analysis methods to examine the content of textbooks (Lucy et al., 2020) have only examined the text and not the images found in the textbooks. Our methods advance both image and text analysis in the computational social science literature along with recent papers by Adukia et al. (2023) which examines children's books, Ash et al. (2021) which examines newspapers, and Yanagizawa-Drott and Voth (2023) which examines yearbooks.

We focus on three main categories of content analysis: *What*, *Who*, and *How*.

***What* is being communicated.** To measure *what* these textbooks cover, we use a large language model to identify topics and values conveyed in the text on a given page. We also use token counts to measure whether and how often specific words are mentioned. Finally,

[1] Note that during the economic crisis of 2007, California did not make any changes to their state adopted textbook lists to save districts the expense of purchasing new textbooks.

use Named Entity Recognition to identify the geographies and historical events that are referenced.

***Who* is depicted.** We will measure *who* is highlighted in the images and text. To examine the photographs and illustrations in the textbooks, we use face detection to identify each pictured individual. We then classify skin color and predict features to measure how often people of different putative races, genders, and skin colors are represented in textbook images. To measure who is highlighted in the text of the textbooks, we use token counts and named entity recognition to measure how often gendered words (including names, pronouns, and terms) are mentioned as well as how often famous people of different races, genders, and nationalities are represented.

***How* content is portrayed.** To measure *how* different people, topics, and values are discussed, we use word embeddings to identify words found in similar contexts to them and semantic role labeling to identify the actions that people of different genders and races are described as taking. We then analyze the sentiment associated with these associated words and actions.

We describe the image and text analysis methods that we use for this exercise below and illustrate our overall methods pipeline in Figure 2.

III.A Image Analysis

The process that we use to convert images into data is illustrated in Figure 2a. In this section, we describe how we detect faces, classify skin color, and predict features such as putative race and gender.

III.A.1 *Image Analysis: Face Detection*

Our first step when measuring representation in images is to run face detection on every digitized page in our sample to identify all depicted faces. However, in the process of applying off-the-shelf face detection models, we discovered a significant drawback; while off-the-shelf methods are an excellent way to detect faces in photographs, they perform poorly when applied to illustrations. Since many of the images found in content targeted towards children contain illustrations, this presented a problem. To mitigate this issue, we trained a model which can better detect faces in both photographs and illustrations as described in Szasz et al. (2022) and Adukia et al. (2023). This model increased the number of detected faces by 250% in the sample of books. We apply this model to detect faces in the textbooks in our sample.

III.A.2 *Image Analysis: Skin Color*

We extract and classify skin color using a similar method as Adukia et al. (2023). We begin by identifying the facial area of interest in the pictured face by "segmenting" the skin. To do this, we use Google's MediaPipe Face Landmarker to detect facial landmarks and then to isolate the parts of the face which contain skin from other facial features. This helps us to eliminate hair, clothing, and other accessories. Then, we extract the dominant colors in the segmented skin by applying k-means clustering to group each pixel in the skin segmentation in RGB space and take the center of each cluster. Finally, we collapse the entire skin segmentation into a single "representative color" by taking a weighted average of the dominant skin colors (where the weights are determined by the size of the cluster each dominant color was drawn from). To construct a continuous measure of how light or dark the skin color of a detected face is, we convert the center of each cluster in RGB space and convert its color into the perceptually linear CIE L*a*b* color space. We take a weighted average the L* value of each center which ranges from 0 (darkest) to 100 (lightest). This results in a continuous measure of perceptual tint that serves as our measure of skin color. This process is illustrated in Figure 2a.

III.A.3 *Image Analysis: Feature Classification of Putative Race, Gender*

We then wish to classify different features of pictured characters, specifically putative race and gender. We use an AutoML feature classification model that we trained using a manually curated data set of both illustrations and photographs which can more accurately predict features of illustrated faces such as the race, age, and gender, as described in Szasz et al. (2022). We use this improved feature classification model for our analysis.

III.B Text Analysis

The process that we use to convert text into data is presented in Figure 2b.

III.B.1 *Text Analysis: Large Language Models*

Large language models (LLMs) are deep learning models that can recognize, summarize, translate, predict, and generate text. LLMs acquire these abilities by using massive amounts of data and computational resources during their training to achieve general-purpose language understanding and generation.

For our analysis we use OpenAI's GPT-4 model (from now on referred as GPT), which is an autoregressive pre-trained LLM that applies deep learning algorithms to generate what would be a human-like response to user input. This is done through the use of a transformer (i.e., neural networks) architecture, composed of encoding and decoding models. The encoding part takes the input given by the user, extracts its relevant features, "understands

it" through its context, and creates a high-level representation of it. This representation is then passed as an input to the decoding model, where the final response is generated using information learned in the pretraining process.

In particular, we ask GPT to analyze single pages of textbooks and identify topics and values present in them. OpenAI's API allows us to define a specific LLM, a system role, and a *temperature* for each of our prompts. In this paper we use the `gpt-4-1106-preview` model, which allows a 16,000 token limit for total input and response. The parameter *temperature* defines the freedom or creativity we allow the model to have for its responses, and is set up to its default median value, 1.[2] Finally, we assign a role to the system to set its behaviour. In our analysis, we instruct the model to take on the role of "a knowledgeable and unbiased judge."

The topics that are identified on a page by GPT are predictions based off the text on a page and do not report the ground truth, which then may lead to measurement error. One alternative way to measure substantive topics is to measure the instance of topic-related words on a given page. We validate the model's topic classification by counting how often a set of topic-related words appear on pages that were (or were not) classified as containing that topic by GPT. We find that pages which were classified as having a specific topic were much more likely to contain words related to that topic A.1. For example, pages that were not classified as containing the topic "Slavery" have on average 0.0 words related to slavery such as captivity, enslavement, slave, plantation, underground railroad, emancipation proclamation, and thirteenth amendment. On pages that GPT classified as having the topic "Slavery," however, we see 3-4 words per page on average. A notable exception is the "religious/religion" topic. There seem to be instances of words related to religion on pages that are not identified as having religious topics by GPT. Manual inspection of these pages yield examples such as a passage which discusses Martin Luther King, Junior and mentions that he was a pastor. In this page, there is a religious word, but religion is not a topic.

It is likely that we could just search for the prevalence of specific topics by looking for keywords instead of relying on a large language model to summarize the text for us, but then we would have to provide a list of topics ourselves as well as an exhaustive list of keywords for each subject. By using GPT to identify topics, we are better able to identify prevalent topics that we might not have thought to look for and test for differences between the prevalence of these topics across collections. This application of GPT is one example

[2]The optimal value for the temperature parameter will usually depend on the type of task we ask GPT to do, and to the best of our knowledge there is not a benchmark or standard for this. While higher temperature values give the model more room to hallucinate (create fake or made-up responses), lower ones make responses more deterministic, potentially limiting its flexibility to provide reasonable responses.

of using machine learning as a tool for hypothesis generation as discussed by Ludwig and Mullainathan (2024).

III.B.2 *Text Analysis: Token Counts*

Token counts are one of the simplest natural language processing tools which allow us to count the number of times specific groups or topics are mentioned. We can use these counts to see how often words related to females (such as "she", "queen", "sister", etc) are mentioned relative to words related to males. These words include pronouns and other gendered terms. We can also use these counts to examine whether changes in science education content standards surrounding the theory of evolution translate into more frequent mentions of the topic.

III.B.3 *Text Analysis: Named Entity Recognition*

Named Entity Recognition (NER) identifies "named entities" (nouns predicted to be names or people, locations, events, etc). And counts the number of times these names appear in the source text.

To identify names of famous people, we merge the "person" named entities with a pre-existing data set containing information from over 70,000 Wikipedia biographies (Yu et al., 2016) to identify famous people mentioned in the text such as Martin Luther King Jr. or Amelia Earhart. Using these data, we can examine the number of famous people mentioned by race, gender, occupation, birthplace, and sexual orientation. We use the remaining person named entities that were not matched to a famous person's name to measure the number of remaining female and male names mentioned in the text, which could be first names of famous individuals or names of other non-famous characters. To do this, we extract the first name of each remaining person named entity and estimate the probability that the name is female (or male) using data on the frequency of names by gender in the U.S. population from the Social Security Administration.

To understand what geography children may be exposed to, we examine how often different places in the world are mentioned. To do so, we merge the "location" named entities with a dataset that contains the shapefiles for all cities, countries, counties, bodies of waters, etc. We use these data to determine how often different places are mentioned.

Lastly, we use the named events identified to determine which historical events are discussed most often out of all events mentioned in textbooks and how this has changed over time.

III.B.4 *Text Analysis: Semantic Role Labeling*

Semantic role labeling (SRL) models identify the predicate-argument structure of a sentence. In other words, these models label the agent (active subject), main verb (predicate/action), and the patient (object or passive subject) in a sentence and the various elements that accompany it (e.g. "who" (agent) "did" (verb) "what" to "whom" (patient)).[3] For example, in the sentence "In the North, President Abraham Lincoln asked men to fight to preserve the union" the *agent* in the sentence is "President Abraham Lincoln," the *action* is "ask," and the *patient* is the statement, "to fight to preserve the union."

We use SRL to help us characterize the actions taken by agents of different races and genders in our text. To do this, we first apply SRL to all the sentences in each textbook to identify the agent and the action. Then we classify the agents' race and gender as well as the sentiment of each action.

To classify the race and gender of each agent, we start by using named entity recognition to extract any names in the text labeled as the agent. If any names are detected, we classify their associated gender and race (if possible) following the process in Section III.B.3. We identify if the name belongs to a famous person by matching it with data on 70,000 Wikipedia biographies. If the name belongs to a famous person, we can classify the agent by the famous person's known race and gender. If the name does not belong to a famous person, we cannot classify the race of the person, but we can predict the gender of the agent based off their first name. If no names are found, we apply part-of-speech tagging to identify if the agent of the sentence contains a gendered pronoun or term. We follow the same process as described in Section III.B.2 to classify gender. If the agent of the sentence contains a gendered pronoun (e.g. she), we use the gender of the pronoun to classify the gender of the agent. Similarly, if the agent of the sentence contains a gendered noun/term (such as brother or queen), we use it to classify the gender of the agent.

III.B.5 *Text Analysis: Sentiment Analysis using VAD*

To characterize the sentiment of the actions taken by agents of different races and genders, we measure their valence (V), arousal (A), and dominance (D). Associated scores provide measures of different attributes of each action. Valence measures range from happiness to unhappiness, or positive to negative. Arousal measures indicate a level of "affective activation,' or active/excitement to passive/sleep. Dominance measures reflect the level of control of the emotional state: dominant/powerful to submissive/weak. We use the NRC

[3]We implement Semantic Role Labeling using AllenNLP, a library for applying deep learning methods to NLP research (Gardner et al., 2018).

VAD Lexicon which includes a list of more than 20,000 English words and their valence, arousal, and dominance scores (Mohammad, 2018). Scores range from zero to one. For example, the words "teamwork" and "leader" have (V, A, D) scores of (0.92, 0.49, 0.77) and (0.83, 0.58, 0.93) respectively. Both have relatively positive valence, but "teamwork" has slightly more positive association to it. However, "leader" is more active and meaningfully more powerful compared to "teamwork."

III.B.6 *Text Analysis: Word Embeddings*

Word embeddings are vector representations of words which encode semantic meaning by incorporating information from the nearest neighbors (context) of that word. As a result, word embeddings can be used to examine not only whether a topic or group is present in our source text, but also the context in which it is discussed. Specifically, the closer two words are positioned together in high-dimensional space (as determined by the cosine similarity of two word vectors), the more often they are used within the same context.[4]

To examine the context in which females and males are portrayed, we apply a textual analysis variant of the Implicit Association Test (IAT) called the Word Embedding Association Test (WEAT) introduced by Caliskan, Bryson and Narayanan (2017). This word embeddings analogue of the IAT involves taking the estimated vectors for words related to a given gender group (e.g., "she" and "woman" belong to the "female" group) and calculating their average distance to words related to a given domain (e.g., "baby" and "marriage" belong to the "family" domain).We then estimate a domain-specific parameter of "gender-skew" by comparing the average distance to a domain between males and females. To determine where differences in domain association between genders are statistically significant, we use a novel application of placebo tests.

We also examine the Valence, Arousal, and Dominance scores of words most closely associated with females vs. males using the same VAD lexicon we use in our Semantic Role Labeling analysis (Mohammad, 2018) as described in Section III.B.5. Following Charlesworth, Caliskan and Banaji (2022), we do this by finding the average VAD score of words in the VAD lexicon that are mostly closely associated with female words relative to male words.

III.B.7 *Text Analysis: Document Embeddings*

Similar to word embeddings, document embeddings are vector representations of documents. Unlike word embeddings, document embeddings can capture the representations of input texts of variable length such as sentences, paragraphs, pages, or entire books. We

[4]We utilize word2vec from Python's Gensim library to estimate word embeddings. Our word2vec implementation employs the Skip-gram model architecture for training, which uses a given word to predict context words (words that appear within a certain window of the current word) in a sentence.

define a document to be either an entire textbook or a single page in a textbook. We train separate Doc2Vec models using the distributed memory architecture from which we estimate a set of textbook embeddings (one embedding per *textbook*), and separately, a set of page embeddings (one embedding per textbook *page*). In a similar manner to how to analyze word embeddings, we use cosine similarity to calculate how "close" or similar textbooks in our sample are to each other across collections, subjects, and time.

To visually assess similarities between textbooks, we need to apply techniques that reduce the dimensionality of these vectors without losing the detailed information that larger vector spaces provide. To do so, we implement t-SNE to reduce these 300-dimensional vectors into a 2D space; t-SNE is better suited for this task than the more common PCA, since it is more robust to outliers, less sensitive to the ordering of the data points, and is better at preserving the clusters in the data.

III.B.8 *Text Analysis: Topic Embeddings*

Finally, we integrate the output from GPT, which identifies topics on each page, with the page embeddings. We then compute the average page embedding for all pages containing a specific topic, producing a single topic embedding for each collection. We use word embeddings estimated in the same embedding space to describe the context in which topics are discussed. We assess the similarity in discussion across collections by calculating the cosine similarity between topic embeddings for each pair of collections.

IV Results

In this section, we discuss the content patterns that we uncover in the textbook collections. We begin by examining the *overall similarity* across the collections, by subject and over time. We then document *what* topics and values that are included and *how* similarly these topics are presented across collections. We follow this with a discussion of *who* is represented and *how* they are portrayed across collections.

IV.A Overall Comparison

First, we analyze the overall textual similarity between textbooks by using document embeddings to map distances between textbooks in embedding space. The closer two textbooks are positioned in embedding space, the more similar their textual content is, both in terms of the words used and the context in which they are used. Figure 3 plots our textbook embeddings in a two-dimensional space, with collections represented by colors and subjects by shapes.[5] As expected, textbooks cluster together by subject: Reading and Social Studies

[5]For ease of graphing, each textbook embedding is reduced to two-dimensional space without loss of significant information using t-distributed stochastic neighbor embedding (t-SNE).

textbooks are much closer to each other in the embedding space than to Science textbooks. This is likely because the vocabulary and topics found in Reading and Social Studies textbooks are more similar to each other than to those found in Science textbooks, which tend to be more technical.

Figure 3 also shows an overlap between California and Texas textbooks, which are clustered together within subject and distinct from the Religious textbooks in our sample. The similarities between California and Texas textbooks are particularly notable, given the widely accepted narrative of educational polarization between these states.

We quantify the similarity between textbooks across different collections by calculating the cosine similarities for every pair of textbook embeddings from distinct collections (e.g., California vs. Religious) that correspond to the same subject, grade, and year in use. Figure 4a displays histograms of these cosine similarities for pairs of textbooks from different collections (e.g. the blue histogram represents the distribution of cosine similarities between textbook pairs, one from the California collection and the other from the Texas collection, but both sharing the same subject, grade, and year in use). Additionally, in Figure 4b, we compute the average cosine similarity between textbooks from two different collections over time. The patterns in these plots are consistent patterns from our t-SNE plot in Figure 3. We see that the public school textbook collections (California and Texas) have higher cosine similarity with each other than either of them have with the Religious collection. When we disaggregate by subject (Figure 4c), the patterns largely hold, with textbooks from Texas and California consistently showing the highest similarities across all subjects. However, in the most recent decade, the similarity between Reading textbooks from Texas and California has decreased, reaching levels of similarity (or dissimilarity) comparable to those with textbooks from our Religious collection. Initially, Science textbooks across all collections exhibited the highest similarity, but this has declined over time. The similarity among Social Studies textbooks, on the other hand, has remained relatively stable throughout the study period.

In the following sections, we examine specific dimensions of content that contribute to these patterns of difference and similarity between and across textbook collections.

IV.B What: Textbook Topics, History, and Values

Parents choose schooling options, in part, because they want their children to be exposed (or not) to certain knowledge and values. In this section, we document what topics and values are being imparted to children, either implicitly or explicitly.

Topics. We first examine what topics are shared with children in each subject (Figure 5). We see that similar topics are discussed across collections within subjects with some notable exceptions unique to each collection. In Reading textbooks, the top three topics identified were the same for each collection: family, nature, and education. The Religious collection is more likely to include narratives related to family than the collections of public school textbooks (8% or 9% vs. 15%). In Science textbooks, astronomy, biology, physics, observation, and science are among the most common topics across all three collections and in Social Studies textbooks, geography, history, and agriculture are among the most common topics for all three collections (Figure 5). Note that we see religion as a common topic in Social Studies textbooks for our collection of Religious textbooks.

This suggests meaningful overlap in the most prevalent topics transmitted to children, which suggests that children are being exposed generally to similar topics with some exceptions. Much of the public debate, however, does not center on these identified topics. Rather, discussions often focus on topics that are considered to be more contentious such as those that address issues such as climate change, evolution, racism, and enslavement. We first examine the prevalence of these topics as identified by the LLM (Figure 6). This shows that California and Texas textbooks are much more likely to address topics related to climate change than the Religious collection. By contrast, the Religious collection is the only set of books that discusses creationism and is much more likely to address topics related to evolution (typically as a way of contrasting it to creationism). California is the most likely to include slavery-related topics. Racism and LGBTQIA+ related topics are rarely, if ever mentioned across all three collections.

Much of the debate surrounding what children learn involves the knowledge they are given about history and the world around them. Indeed, in each of the collections, geography and history were among the top Social Studies topics identified (Figure 5). When examining specific mentions of places around the world we see that children get exposure to different geographic regions similarly across all collections. Table 2 and Figure A.2 show the prevalence of geographic mentions. Figure 8 shows the most commonly mentioned historical events in our sample of textbooks. We see varying patterns across the collections, but in particular, we see an increase in mentions of the Civil War over time in California and the Religious collections, but a sharp decrease over the last two decades in Texas. By contrast, we see that Texas and California are more likely to mention the American Revolution than the Religious collection is.

Values. The majority of parents who choose private or home schooling options report a desire for "morals," "values," or "character" education as among the top most important

qualities when choosing a school (Catt and Rhinesmith, 2017). We accordingly examine the values that are being espoused on the pages of the textbooks.

Consistent with our analysis of top topics, the top values we identify are similar across collections with the exception of Faith and Religious Faith showing up as top values in the Religious collection (Figure 9). We explore this further by combining values related to religion such as Religious Faith, Religious Freedom, Religious Beliefs, Religious Diversity, Religiosity, Religious Tolerance, Religious Devotion, etc. Using this group of values we find that religious values are present on 2%-16% percent of pages on average in a given subject in our collection of Religious textbooks as compared to 0%-5% of pages in our collections of public school textbooks. Notably, in public school textbooks, religious values only appear in Reading and Social Studies textbooks, but they are present on approximately 2.1% of pages on average from Science textbooks in the Religious collection (Figure 10a).

Despite many parents emphasizing character and values education as a key factor when choosing a private school (Catt and Rhinesmith, 2017), we see that both Texas and California public school textbooks include a meaningful proportion of pages that espouse character values to children (Figure 10b). The Religious collection has a higher proportion of pages overall that include character values, but all three collections include character values on over 30% of textbook pages.

IV.C How: Examining Similarities Between Topics

In the preceding section, we describe how often certain topics are mentioned across textbook collections. However, just as important as whether certain topics are discussed is how they are discussed. Using topic embeddings as described in section III.B.8, we can examine differences in the context or sub-topics surrounding a given topic across collections. In figure **??** we shows the average cosine similarity of topic embeddings within a pair of collections. We find that compared to the overall similarity of pages, these specific topics are discussed less similar contexts across all three collections. first find the top 60 word embeddings closest to each topic embedding. Then we plot these top words along with the three topic embeddings (one for each collection) in 2-D space so we can visually examine the context surrounding each topic embedding. We also cluster the word embeddings into subtopics for each of interpretation.

In Figure **??**, we see that Climate Change and Slavery are talked about in similar contexts across collections which means the topic embeddings in each collection for Climate Change are close to each other in embedding space, have similar top words, and are a similar distance to each of their top words. For example, the topic embeddings for Climate

Change are similarly close to words related to natural disasters, heat, and pollutants across all three collections. However, evolution and religious topics are not discussed in similar contexts. For example, the topic embedding for evolution in the religious collection is closer to words related to creationism, genesis, adam, bible, and sinned than it is to words related to adaptation, survival, reproduction, and organism. Whereas the opposite is true for the evolution topic embedding in public school textbooks. Similarly, in the religious collection, the religious topic embedding is closer to words related to Christianity such as idols, sinned, jesus, and savior than it is to words related to the founding era such as colony, founding, freedom, and settlers. Whereas the opposite is true in public school textbooks.

IV.D Who: Historic Figures, Skin Color, Race, Gender, & Sexual Orientation

Historic Figures and Birthplaces Political and media discussions of curricula are commonly centered around which historical figures and events should be included in Social Studies textbooks and how they should be portrayed. This discourse reflects concerns expressed by parents as highlighted in a survey conducted by the Heritage Foundation which found that 59.5% of teachers and 56% of parents surveyed noted the importance of schools including content related to the Constitution and the founding era of the United States (Sailor et al., 2021).

We explore the inclusion of historical figures and events across the collections in Figure 11a and compare how often they are mentioned across educational settings. We see that each of the collections include a similar proportion of mentions of the Founding Fathers (e.g. George Washington, Thomas Jefferson, Benjamin Franklin, Alexander Hamilton) out of all famous people mentioned. However, California and Texas are just as likely to mention social activists (e.g. Martin Luther King Junior, Susan B. Anthony, Frederick Douglass, Rosa Parks), as they do Founding Fathers whereas we see the Religious collection is more likely to mention religious figures (e.g. Martin Luther, Hudson Taylor, Dwight L. Moody, Adoniram Judson) than Founding Fathers, suggesting the importance that they each place on these sets of historical figures.

When we examine the birthplaces of the famous individuals to whom children are exposed, we see that public school textbooks have more geographic heterogeneity in terms of female representation around the world than the Religious collection does (Figure 12).

Skin color. In Figure 13a, we show the distribution of detected faces in our textbook by skin tint (e.g. how light or dark their skin color is on a scale from 0 to 100). We plot these distributions over a scatterplot showing the skin color of every face in our sample. The x-axis for both the density and scatter plot is skint tint, but the y-axis for the scatterplot is the

vibrancy of each skin color.

Overall, we see that the distribution of faces in textbooks from our Religious collection are more likely to portray pictured characters with lighter skin than the Texas or California books. This is true even when conditioning on race; in the Religious collection the skin color of faces is lighter on average than faces in California and Texas (Table 3). This is true even when we condition on the predicted race of the face. (Table B.2) further separates faces by predicted race and finds that faces in the Religious collection are lighter for all races/ethnicities except for Latine. However, we do not see a similar clear pattern when comparing Texas and California. In California, for example, characters classified as being Asian or more likely to have lighter skin than characters classified as Asian in Texas; whereas this pattern is flipped for characters classified as being Black: they are shown with lighter skin in Texas compared to California.

To see how closely the distribution of skin colors in textbooks matches the true distribution of student skin colors, we compared faces from textbooks to those from student yearbooks. We extracted faces from a sample of high school yearbooks spanning from 1987 to 2016 in California and Texas. It is important to note that the sample of faces from textbooks, which includes both illustrations and photographs taken under varying lighting conditions and contexts, is quite different from the sample of faces from yearbooks, which consists exclusively of photographs taken under similar, well-lit conditions. This difference in sampling conditions may influence the comparability of these two samples. In Figure 13b we find that while the distribution of faces in textbooks has a wider range than faces in textbooks, the representation of skin colors in public school textbooks mirrors the actual skin color distribution of students within the respective state. Specifically, the skin colors of faces in textbooks from Texas are lighter on average compared to those from California, a trend that is consistent with the skin colors found in yearbooks from these states.

Gender. We find that females are less likely to be represented in both the text and the images of textbooks in the Religious collection than in our collections of public school textbooks (Table B.1). However, across all collections females are under-represented in relative to their share of the population. The disparity is much larger in the text than it is in the images. While females are depicted less often than males, female representation is close to parity in the images (Table 1); this is not the case for female representation in the text. The total proportion of female words over time ranges from approximately 20% in our collection of Religious textbooks to approximately 40% in our collections of public school textbooks, with these proportions being relatively stable over time (Figure A.3). Interestingly, when we disaggregate gender representation by subject, a distinct pattern emerges. Notably, in public

school Reading textbooks, females are mentioned as frequently as males (Figure 14). This pattern may be attributed to the greater flexibility publishers have in selecting the gender of characters for Reading textbooks. Unlike history textbooks, where the narrative often revolves around historical events and figures from which women were historically marginalized—such as the consistent male dominance in the U.S. presidency—Reading textbooks are not bound by these historical constraints. Consequently, discussions in Reading materials can feature a more balanced gender representation. We don't see this same pattern in the Religious collection of textbooks. As illustrated in Figure 14, females are consistently underrepresented in the text across all subjects within the Religious collection. The closest to gender parity is found in Reading textbooks, where female-related words constitute only 41% of all gendered words.

Race. A salient part of curricula are the famous individuals who are highlighted for children. History and science came from individuals, so in the conveyance of information, textbooks often include discussion of these individuals. They can serve as real-life characters in our histories, examples of leadership, and as role models to whom children may aspire to be similar. To measure how often famous individuals of different races/ethnicities are discussed, we apply Named Entity Recognition (NER) tools as discussed in Section III. We examine the incidence of famous individuals across the corpora and find that the vast majority of famous individuals mentioned are White individuals (between 81%-92% across all collections and subjects as shown in Table 1). This pattern is consistent over time with a slight decrease in famous White individuals represented in recent decades (Figure 15). Over time, the Religious collection is more likely to feature famous individuals who are White and less likely to feature famous individuals who are Black relative to the public school collections.

The most common famous people mentioned in each collection are shown in Table B.3. George Washington, Thomas Jefferson, Benjamin Franklin, and Abraham Lincoln – White males who are all Founding Fathers – are among the most commonly mentioned famous individuals across collections.

One caveat of focusing on famous individuals only to measure the representation of race is that we are unable to capture the race of non-famous individuals in the text. However, we can look at how often both famous and non-famous individuals of different races are pictured in textbooks by examining the race classifications of detected faces in our sample. We find that in all collections, most of faces depicted are White individuals. However, the difference between the proportion of White faces and the proportion of Black faces is much closer to parity in faces than it is when examining the race of famous individuals. In public school textbooks California textbooks have the largest percentage of Asian and Latine faces,

though this number is still very small; only about 5% each compared to 3% each in Texas textbooks and 2% each in Religious Textbooks (Figure A.4).

LGBTQIA+ Individuals. There were only two books in our sample that we identified as discussing LGBTQIA+ topics. These books were identified using token counts of the words lesbian, gay, bisexual, transgender, queer, intersex, and asexual.

The first book is a 5th grade Social Studies textbook from our Religious collection published in 2016. The relevant passage is taken from a section entitled "Gay Rights":

> "Another issue that troubled Americans was gay rights. During his time as president, Obama encouraged the growth of gay rights. He said that men should be able to marry men and that women should be able to marry women. When Obama first took office, many Americans disagreed with this view. By 2015, many agreed. Americans who disagreed faced problems. Many believed that God designed marriage to be between a man and a woman. They trusted that His plan benefited humanity. These Americans did not support gay marriage. They thought that it would harm society. For example, children in families where there is a same-sex marriage are denied having both a father and a mother. But in 2015, the Supreme Court ruled in favor of gay marriage. It had redefined marriage. Five justices voted in favor of gay marriage, but the other four justices voted against it and pointed out that this change threatened religious liberties. Some Christians who owned bakeries, flower shops, or photo studios were sued for not participating in a same-sex wedding. These businesses served all other customers, but refused to participate in same-sex weddings. Many of these business owners lost their court cases and were fined. This change in the United States raised other questions. Could religious schools continue to teach their views of marriage? Would schools and business owners come under more pressure to violate their beliefs on these issues? These questions remained unanswered."

The second book is a 3rd grade Social Studies textbook from our California collection that was published in 2019. The relevant passage is from a section entitled "Supporting Many Different Communities":

> "Many other Californians worked hard to make changes in their communities...
>
> ...Harvey Milk was a political leader and a gay activist in the 1970s. Gay men are attracted to other men. Milk was elected to a local government office in San Francisco in 1977. He was one of the first elected officials in history to tell the public that he was gay. Milk worked hard to protect gay and lesbian rights. Lesbians are women who are attracted to other women. Milk's work also made it easier for gay, lesbian, bisexual (those attracted to both men and women), and transgender groups to stand up for their rights.
>
> Chaz Bono is a transgender activist. The word gender refers to whether someone is male or female. Transgender people feel that their gender is the opposite of what it was at birth. Bono was born female, but always felt male. He changed his gender and today

lives as a man. Bono and other activists fight against discrimination and stereotypes of their community."

The passage in the Religious collection highlights the tension between LGBTQIA+ rights and religious freedoms. Given that this is the only representation of LGBTQIA+ individuals in our sample, children exposed to this curriculum are only seeing people in this community posed as a threat to the freedoms of others. The passage in the collection of California textbooks gives definitions of what it means to be a gay, lesbian, bisexual, and/or transgender individual and highlights two LGBTQIA+ activists. While the sentiment of this passage is positive overall, by not discussing any LGBTQIA+ individuals of color or who identify as female, this passage fails to take an intersectional lens. We did not identify any passages in the Texas state adopted curriculum that discuss LGBTQIA+ topics, rights, or individuals.

IV.E How: Examining Context & Actions

In the preceding section, we identified disparities in who is present (and absent) from textbooks. However, it is crucial to examine not only how often different groups are mentioned but also the context and manner of their portrayal. In our previous discussion, we alluded to variations in the portrayal of LGBTQIA+ individuals across curricula through comparative analysis of passages. In this section, however, we conduct a systematic examination of the differences in the actions and context surrounding different groups of individuals, specifically focusing on race and gender.

Actions. We apply semantic role labeling as detailed in Section III.B.4, to characterize the sentiment associated with actions attributed to various groups of people, we use the VAD lexicon as described in Section III.B.5 to measure the sentiment of actions on three dimensions: valence (positive vs. negative), arousal (active vs. passive), and dominance (powerful vs. weak). To investigate the presence of systematic differences in the portrayal of actions between females and males, we calculate the average valence, arousal, and dominance scores for actions associated with female agents and male agents in respective sentences. By subtracting these averages, we derive a gender skew metric for each sentiment dimension. In Figure 16a, we see that across all collections of textbooks, the actions taken by females are more positive, but less active and powerful than actions taken by males. This pattern remains even when we focus only on sentences where the gender of agents is identified using only gendered pronouns and not names and gendered terms (Figure A.6a) and is much more stark when we focus on actions that are only taken by female agents vs actions that are only taken by male agents (Figure A.6b).

Context. In Section III.B.6, we describe how we use word embeddings to analyze the contexts surrounding groups of people. We measure context in two different ways; through

the words most closely associated with different groups of people and by testing whether groups of people are more likely to be found in the context of specific domains (such as the home or politics).

First, we identify the top relative words most closely associated a target group (such as females) and least associated with a comparison group (such as males). This approach follows a similar methodology outlined in Manzini et al. (2019) and Charlesworth, Caliskan and Banaji (2022), where the "top relative words" are found by computing the group-to-word association between a single word and the target (or comparison) group, repeating for a given list of target words, and finding the top N words for which the difference in group-to-word associations between the target and comparison group is the greatest. More formally, group-to-word association is calculated using Equation 1 where the association between group G and word w can be calculated as $a(G, w)$:

$$a(G, w) = \frac{1}{|G|} \sum_{g \in G} cos(g, w) \tag{1}$$

where $cos(g, w)$ indicates the cosine similarity (e.g. distance in embedding space) between the two words embeddings g and w. We take the difference in word associations between the target and comparison groups to test whether a word w is "skewed" towards the target group (e.g. is more often found in the same context as the target group) as shown in Equation 2:

$$skew(w) = a(G_{target}, w) - a(G_{compare}, w) \tag{2}$$

Finally, the top N relative words most closely associated a target group and least associated with the comparison group are the N words $w \in W$ with the largest values of $skew(w)$. We restrict the set of target words to be the 20,000 English words in the VAD lexicon for which we have valence, arousal, and dominance scores (Mohammad, 2018).

Table 4 shows the top 10 relative words most closely associated with female and males relative to each other for each collection of textbooks in our sample along with the average VAD scores for the top words.

We see that females are more associated with words related to clothing and males are more associated with words relating to the military. The sentiment of the top 10 relative words for females are more positive, but less active and powerful on average than top relative male words. These findings are in line with the patterns of sentiment in actions taken by females versus males, as discussed in the previous section. This pattern is still present when we extend our analysis to include the top 50 relative words by gender as shown in Figure

16b.

We can test whether different groups have a greater association with different domains (such as politics or the home). To do this, we extend Equation 1 to calculate group-to-domain similarity between a group G and a domain D as shown in Equation 3 below:

$$A(G, D) = \frac{1}{|G| \times |D|} \sum_{g \in G} \sum_{d \in D} cos(g, d) \tag{3}$$

Similarly, we extend Equation 2 to test whether a domain D is "skewed" towards the target group (e.g. if words in domain D are more often found in the same context as the target group G) as shown in Equation 4:

$$Skew(D) = A(G_{target}, D) - A(G_{compare}, D) \tag{4}$$

In Figure 17, we show the gender centeredness of the several domains. For example, the Home domain (which includes words such as marriage, children, relatives, household, etc) is more skewed towards females and the Politics domain (which includes words such as government, army, democracy, capital, political, etc) is more skewed towards males. We test whether this measure of gender skew is statistically different from zero using a placebo test in which we randomly draw a set of words from the vocabulary of the textbook collection which is the same size and the list of words in the domain we are testing for gender skew in and calculate the skew for this "random" domain. We repeat this 1,000 times and plot the distribution of these random domain associations over the top of the bar graphs using a barplot where the edges of the barplot show the 10th and 90th percentile values of the possible gender skewness in a domain of random words. Compared to this distribution, we find that in all three textbook collections, the Family domain is skewed towards female, and this measure of skewness is statistically different than the skewness we would find in a random set of words. We find a similar (though smaller) gender skew in the Politics domain towards males for the Texas and Religious textbook collections.

Lastly, we see no meaningful differences by gender when examining the associations in the sports and performing arts domains in the Religious and Texas collections, but a slight skew towards females in these domains for California textbooks.

We also examine the context in which founding fathers are discussed across collections. Despite differences in how often founding fathers appear, Figure 11b shows that these individuals are depicted in relatively similar contexts across the three collections. Founding Fathers appear in contexts with similar valence and action but with greater dominance as

male pronouns generally. Even though there are differences in the proportion of religious leaders present in the texts, the contexts in which they appear have similar valence, arousal, and dominance scores across the collections. However, in the Religious collection, religious leaders and social activists appear in more positive contexts than male pronouns overall. We also use topic embeddings to find the top words that describe the context in which the founding fathers are discussed and see that the contexts are similar for all three collections (Figure 18).

V Conclusion

Previous literature has established the importance of what we teach our children on a variety of outcomes including their academic performance and career choices. As a result, it is vital that we understand the representation of people and topics in the educational content we present to our children. Given the growing polarization in our society and the increase in demand for school choice options that more closely align with parental values, it's necessary to examine curricula used by children across different geographic regions both within and outside the traditional public school setting.

Interestingly, despite widespread narratives of polarization in educational curricula across different states and religious affiliations, we find few meaningful differences in average portrayal of marginalized groups between our religious private school textbooks, Texas public school textbooks, and California public school textbooks. Across all collections, females are represented less often than males and more likely to be mentioned in contexts that reproduce societal norms and disparities. For example, females are more likely to be talked about in the context of the family and less likely to be talked about in the context of politics. Additionally, we see that words found in similar contexts to females are more positive but less active and powerful than words found in similar context to males. It is important to note that our methods are best at assessing overall portrayals rather than examining the nuances of individual representations, suggesting a potential gap that could be addressed through manual content analysis as evidenced by the differences in portrayal uncovered in our detailed examination of the two passages discussing LGTBQIA+ individuals.

However, despite similarities in *how* people are portrayed, we find meaningful disparities in *who* and *what* are present (or absent) within religious private school textbooks compared to those used in public schools. We find that our collection of Religious textbooks depict faces with lighter skin colors on average as compared to our sample of public school textbooks from Texas and California, more mentions of creationism than climate change, and a large proportion of pages dedicated to religious topics (even in Science textbooks). These Religious textbooks also have less representation of females in the text in comparison

to textbooks commonly used in public schools.

One important caveat when considering the implications of our work is that similarities in textbooks across different educational settings do not necessarily guarantee that children receive similar educational experiences in the classroom. Pedagogy plays a critical role in the educational process and differences in teaching norms and teacher beliefs across settings can drastically change the learning environment. Two teachers using the same textbook can deliver vastly different lessons depending on their interpretation of the material, the examples they choose to emphasize, or the supplementary materials they incorporate. Standardizing classroom delivery, however, requires a much more prescripted implementation approach while choosing and adopting textbooks that cover specific topics is far more feasible from a policy perspective.

This paper demonstrates how advances in artificial intelligence tools such as those from computer vision and natural language processing can be applied to measure how different topics and groups are portrayed in educational materials on a larger scale than was previously possible – such as within entire curricula. The methods introduced in this work also expand the set of tools available for education researchers, and social scientists more broadly, to measure representation in a variety of contexts. For example, advances in the measurement of skin color in images could be applied to measure the effect of student-teacher skin color congruence. In the midst of an increase in educational gag orders and statewide changes in the content standards of public school curriculum, our work shines a light on what messages children are receiving from informal and non-traditional curricula outside the public school system which could have long-term implications for children's beliefs and decision-making.

References

Abdulkadiroglu, Atila, Parag A Pathak, and Christopher R Walters. 2015. "School vouchers and student achievement: Evidence from the Louisiana Scholarship Program."

Adukia, Anjali, Alex Eble, Emileigh Harrison, Hakizumwami Birali Runesha, and Teodora Szasz. 2023. "What we teach about race and gender: Representation in images and text of children's books." The Quarterly Journal of Economics, 138(4): 2225–2285.

Allen, Ann M, Daniel N Allen, and Gary Sigler. 1993. "Changes in sex-role stereotyping in caldecott medal award picture books 1938—1988." Journal of Research in Childhood Education, 7(2): 67–73.

Arold, Benjamin W. 2024. "Evolution vs. creationism in the classroom: The lasting effects of science education." The Quarterly Journal of Economics, qjae019.

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

Ash, Elliott, Ruben Durante, Maria Grebenschikova, and Carlo Schwarz. 2021. "Visual representation and stereotypes in news media."

Bartholet, Elizabeth. 2020. "Homeschooling: Parent rights absolutism vs. child rights to education & protection." Ariz. L. Rev., 62: 1.

Bazzi, Samuel, Andreas Ferrara, Martin Fiszbein, Thomas P Pearson, and Patrick A Testa. 2023. "The Confederate Diaspora." , (w31331).

Bazzi, Samuel, Masyhur Hilmy, and Benjamin Marx. 2020. "Religion, Education, and the State."

Bickford, John Holden, and Lori A Knoechel. 2017. "The historical representation of Native Americans within Ppimary-and intermediate-level trade books." The Councilor: A Journal of the Social Studies, 78(2): 4.

Braden, Eliza G, and Sanjuana C Rodriguez. 2016. "Beyond mirrors and windows: A critical content analysis of Latinx children's books." Journal of Language and Literacy Education, 12(2): 56–83.

Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. 2017. "Semantics derived automatically from language corpora contain human-like biases." Science, 356(6334): 183–186. Publisher: American Association for the Advancement of Science.

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

Catt, Andrew D, and Evan Rhinesmith. 2017. "Why Indiana Parents Choose: A Cross-Sector Survey of Parents' Views in a Robust School Choice Environment." EdChoice.

Charlesworth, Tessa ES, Aylin Caliskan, and Mahzarin R Banaji. 2022. "Historical representations of social groups across 200 years of word embeddings from Google Books." Proceedings of the National Academy of Sciences, 119(28): e2121798119.

Clark, Roger. 2007. "From margin to margin? Females and minorities in Newbery and Caldecott Medal-winning and Honor books for children." International Journal of Sociology of the Family, 263–283.

Crisp, Thomas, Roberta Price Gardner, and Matheus Almeida. 2018. "The all-heterosexual world of children's nonfiction: A critical content analysis of LGBTQ identities in Orbis Pictus Award books, 1990–2017." Children's Literature in Education, 49(3): 246–263.

Deming, David J, Justine S Hastings, Thomas J Kane, and Douglas O Staiger. 2014. "School choice, school quality, and postsecondary attainment." American Economic Review, 104(3): 991–1013.

EdChoice. 2023. "Fast Facts on School Choice." Last modified April 17, 2024, Accessed on March 9, 2024.

Friedman, Jonathan, and James Tager. 2022. "Educational gag orders: Legislative restrictions on the freedom to read, learn, and teach."

Gardner, Matt, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson Liu, Matthew Peters, Michael Schmitz, and Luke Zettlemoyer. 2018. "AllenNLP: A deep semantic natural language processing platform." arXiv preprint arXiv:1803.07640.

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

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

Klein, Rebecca. 2017. "Voucher schools championed By Betsy DeVos can teach whatever they want. turns out they teach lies."

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

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

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

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

Kunzman, Robert. 2010. "Homeschooling and religious fundamentalism." International Electronic Journal of Elementary Education, 3(1): 17–28.

Lucy, Li, Dorottya Demszky, Patricia Bromley, and Dan Jurafsky. 2020. "Content analysis of textbooks via natural language processing: Findings on gender, race, and ethnicity in Texas US history textbooks." AERA Open, 6(3): 2332858420940312.

Ludwig, Jens, and Sendhil Mullainathan. 2024. "Machine learning as a tool for hypothesis generation." The Quarterly Journal of Economics, 139(2): 751–827.

Manzini, Thomas, Yao Chong Lim, Yulia Tsvetkov, and Alan W Black. 2019. "Black is to criminal as caucasian is to police: Detecting and removing multiclass bias in word embeddings." arXiv preprint arXiv:1904.04047.

Michalopoulos, Stelios, and Christopher Rauh. 2024. "Movies." , (Working Paper 32220). Accessed: 2024-08-06.

Michalopoulos, Stelios, and Melanie Meng Xue. 2021. "Folklore." The Quarterly Journal of Economics, 136(4): 1993–2046.

Mohammad, Saif M. 2018. "Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words." Melbourne, Australia.

National Center for Education Statistics. 2022. "Homeschooled Children and Reasons for Homeschooling." Condition of Education, Retrieved December 18th, 2023.

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

Sailor, Angela, Lindsey M Burke, Anne Segal, and Adam Kissel. 2021. "Civics Studies: Why They Matter, What Parents and Teachers Think, and How They Can Reclaim Truth." Heritage Foundation.

Scaramanga, Jonny, and Michael J Reiss. 2018. "Accelerated Christian Education: A case study of the use of race in voucher-funded private Christian schools." Journal of Curriculum Studies, 50(3): 333–351.

Stevenson, Betsey, and Hanna Zlotnik. 2018. "Representations of men and women in introductory economics textbooks." Vol. 108, 180–85.

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

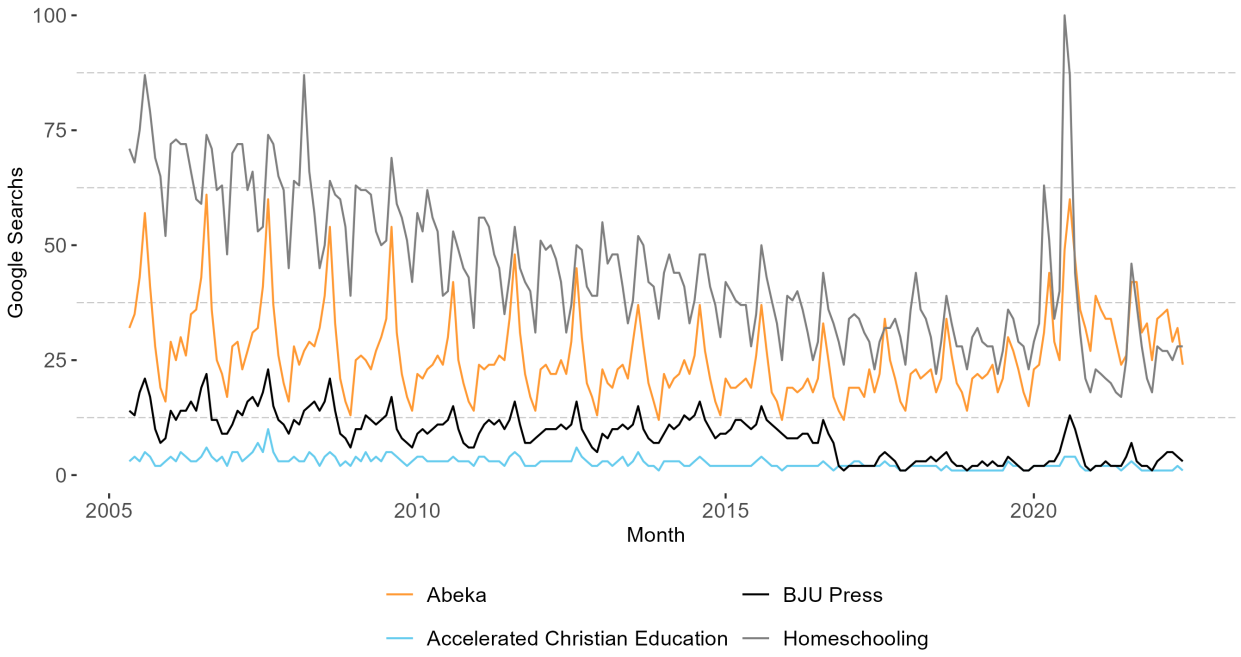
Wade, Rahima C. 1993. "Content analysis of social studies textbooks: A review of ten years of research." Theory & Research in Social Education, 21(3): 232–256.

Yanagizawa-Drott, David H., and Hans-Joachim Voth. 2023. "Image(s)." Working Paper, University of Zurich.

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

VI Main Figures

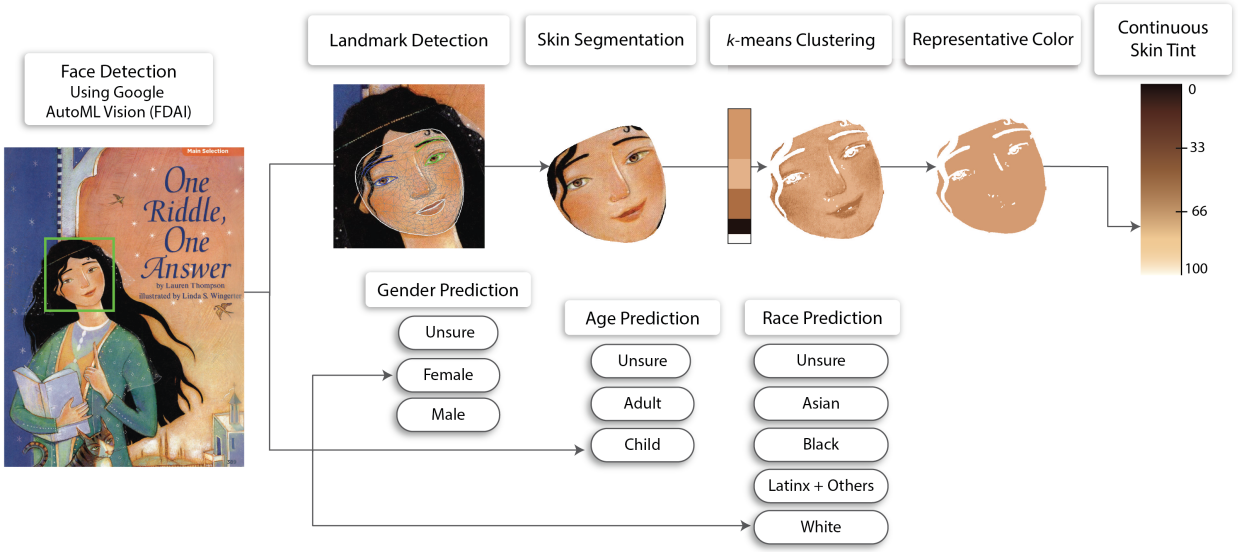
FIGURE 1
Google Searches Over Time: Homeschooling and Religious Collection



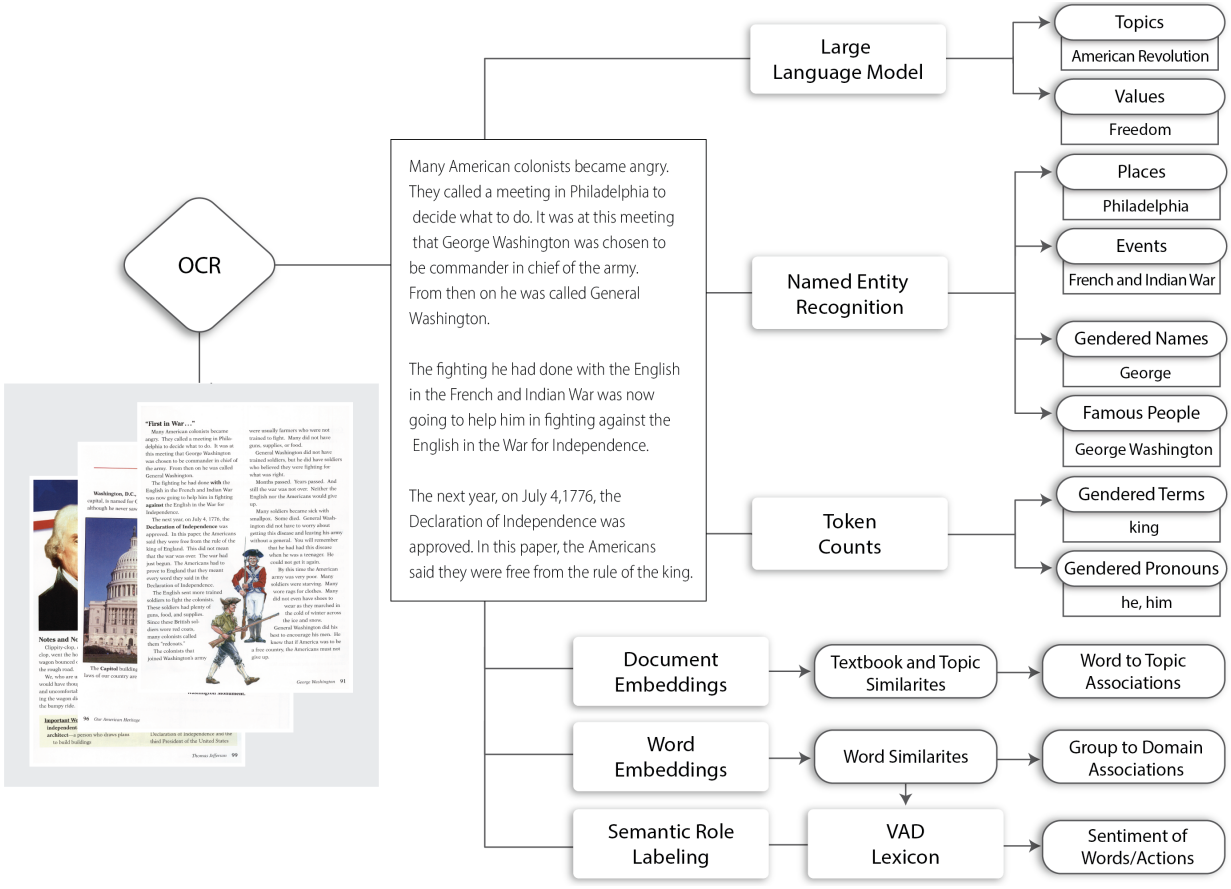
Note: This figure illustrates the Google search trends for the search term homeschool and each of the education publishing companies in the Religious textbook collections; Abeka, Acceleration Christian Education (ACE), and BJU Press (BJU). Note that we use the company topics created by Google Trends instead of the search terms when looking at these three publishers. This figure shows the increase in search interest for homeschooling and Abeka curriculum after the start of the COVID-19 crisis.

FIGURE 2
Data Extraction Pipelines: Image and Text Analysis

(a) Image Analysis Pipeline

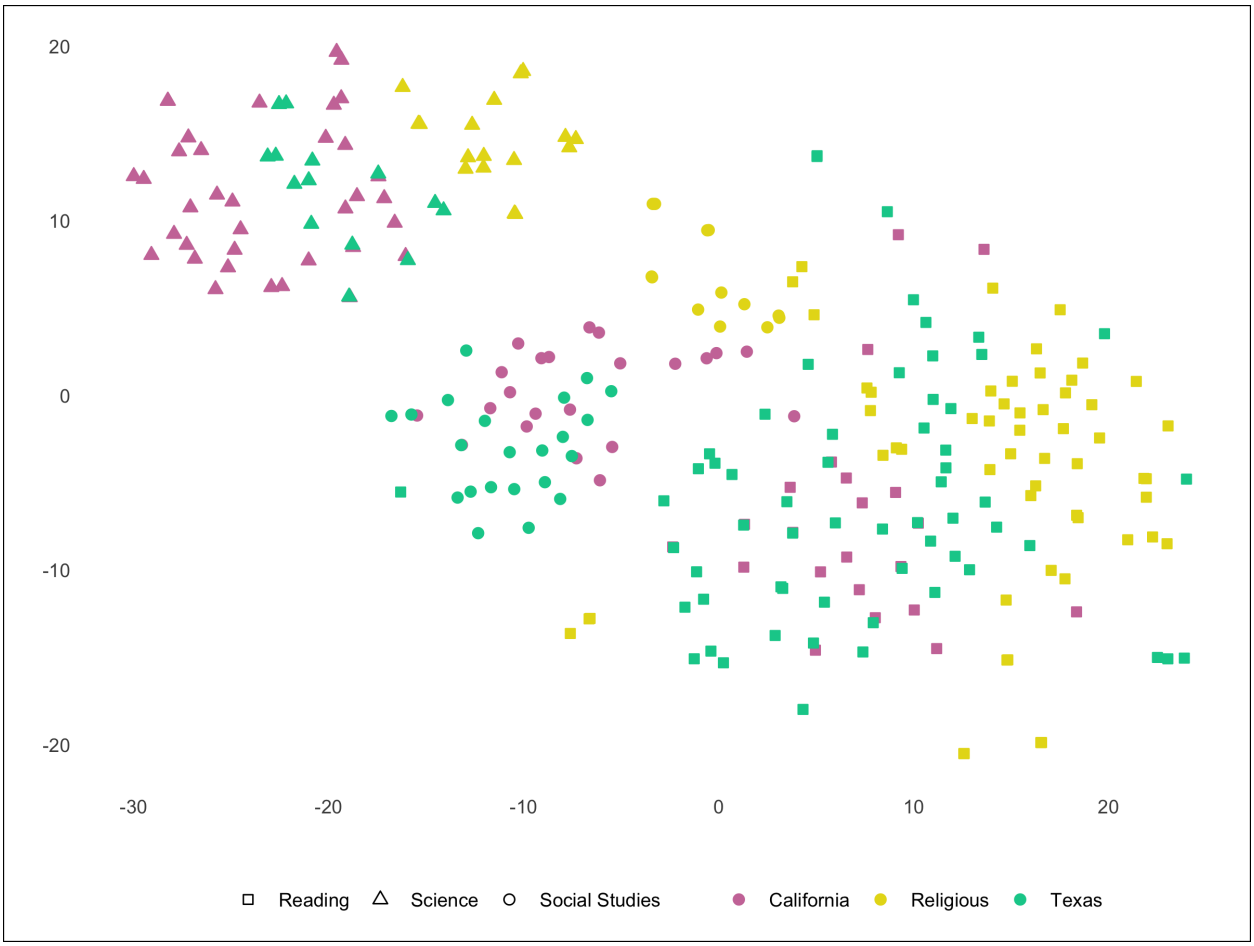


(b) Text Analysis Pipeline



Note: This figure illustrates the steps in our image analysis (top) and text analysis (bottom) pipelines beginning with the raw scan of each page in our sample of books.

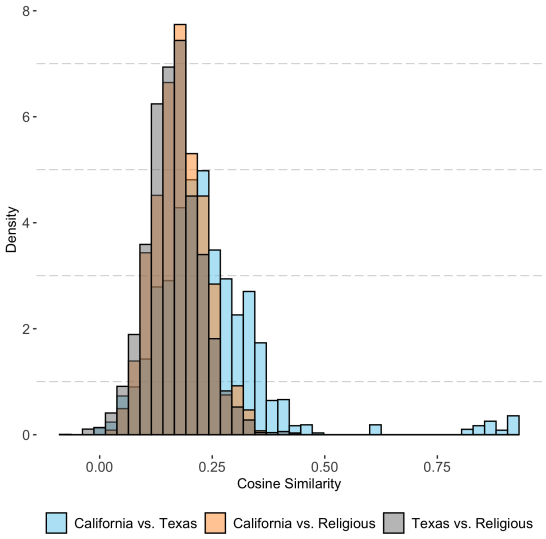
FIGURE 3
Book Embeddings: Similarities between Books across Collections and Subjects



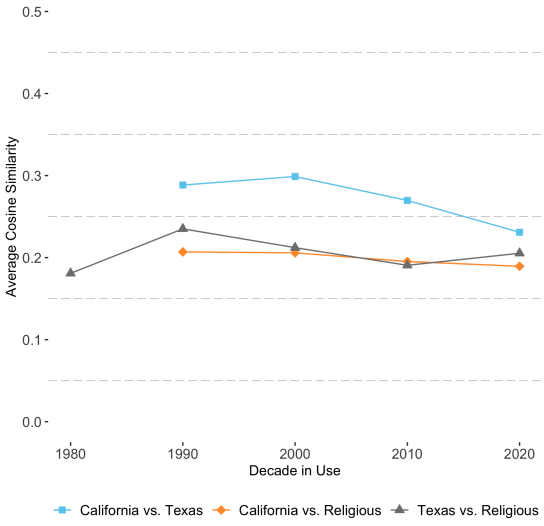
Note: Each observation represents a book from a specific collection (color) and subject (shape). Their content, originally represented in a 300-dimensional space, is here reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and book representations should be interpreted in terms of their relative position to other observations. Among collections, pink signifies California, yellow signifies Religious, and green signifies Texas. Among subjects, the squares (□) signify Reading, the triangles (△) signify Science, and the circles (○) signify Social Studies.

FIGURE 4
Average Cosine Similarity Between Book Embeddings by Collection Pairs

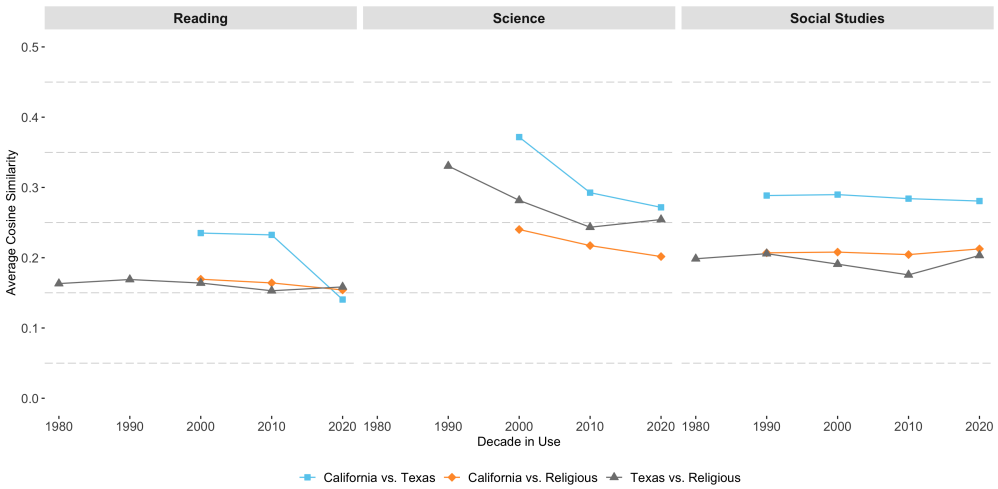
(a) Histogram of Cosine Similarities Between Book Embeddings by Collection Pairs



(b) Average Similarity Between Book Embeddings by Collection Pairs over Time

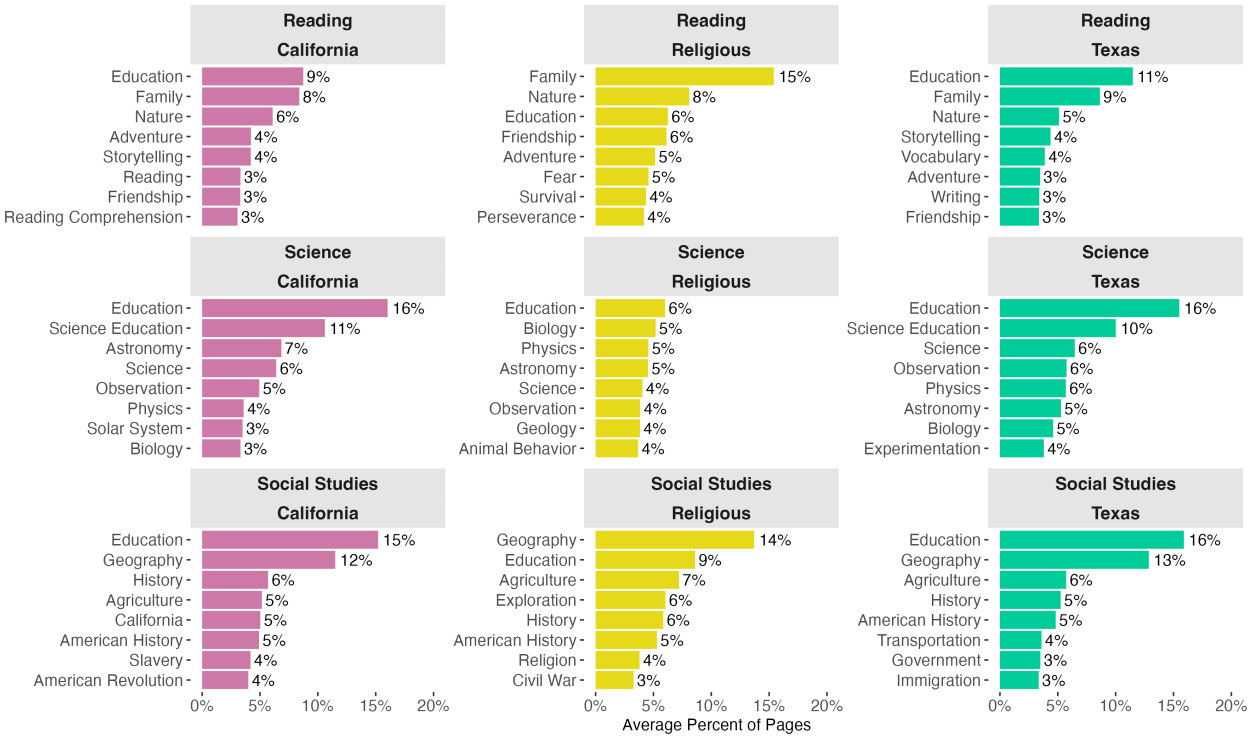


(c) Average Similarity Between Book Embeddings by Collection Pairs and Subject over Time



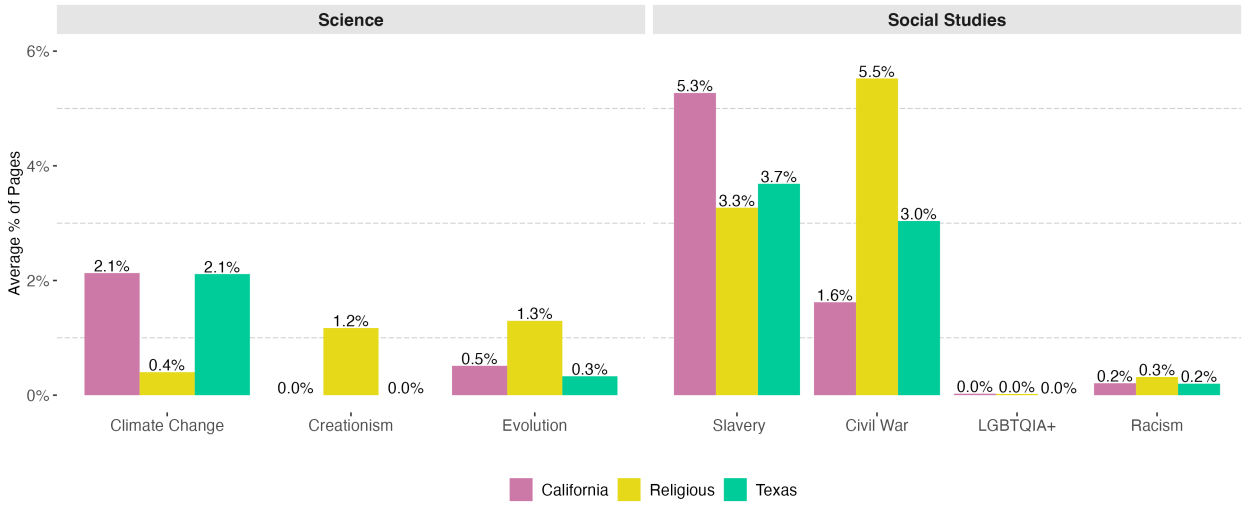
Notes: In Figure (a), we plot a histogram of the cosine similarities between any two books in different collections (e.g., California vs. Religious) that share the same subject, grade, and year in use. We separate the densities by collection pairs (e.g. one California textbook and one Religious textbook). Note that each book pair is repeated for every year those two books are both in use. In Figure (b), we average the cosine similarities of book pairs for each decade and collection pair; this average is weighted equally across subject, grade, and year in use. Figure (c) follows the same approach as Figure (b) except that we calculate the averages separately by subject.

FIGURE 5
Top 8 Topics by Collection



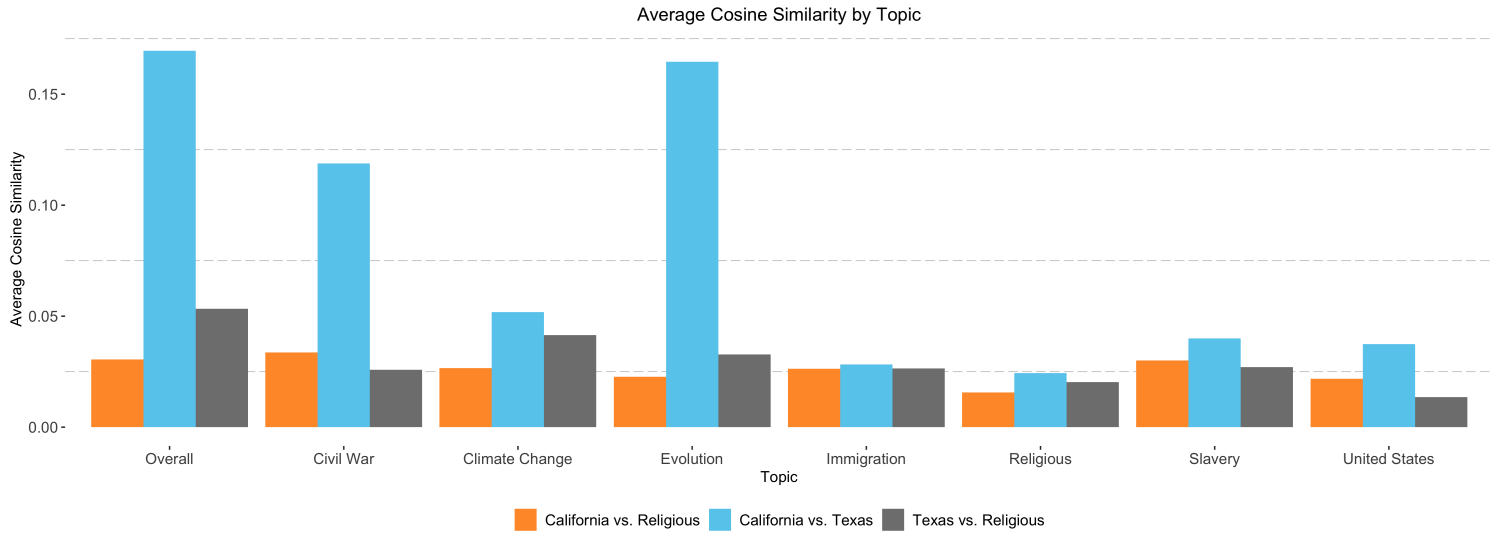
Notes: This figure shows the top topics (e.g. topics that appear on the highest average percentage of pages) within a collection. To calculate the average percentage of pages containing each topic, we first calculate the percentage of pages for each collection, subject, grade, and year in use. We then average these percentages at the collection level, to ensure that pages in each grade, subject, and year in use are weighted equally in the final averages.

FIGURE 6
Average Percent of Pages Containing Contentious Topics



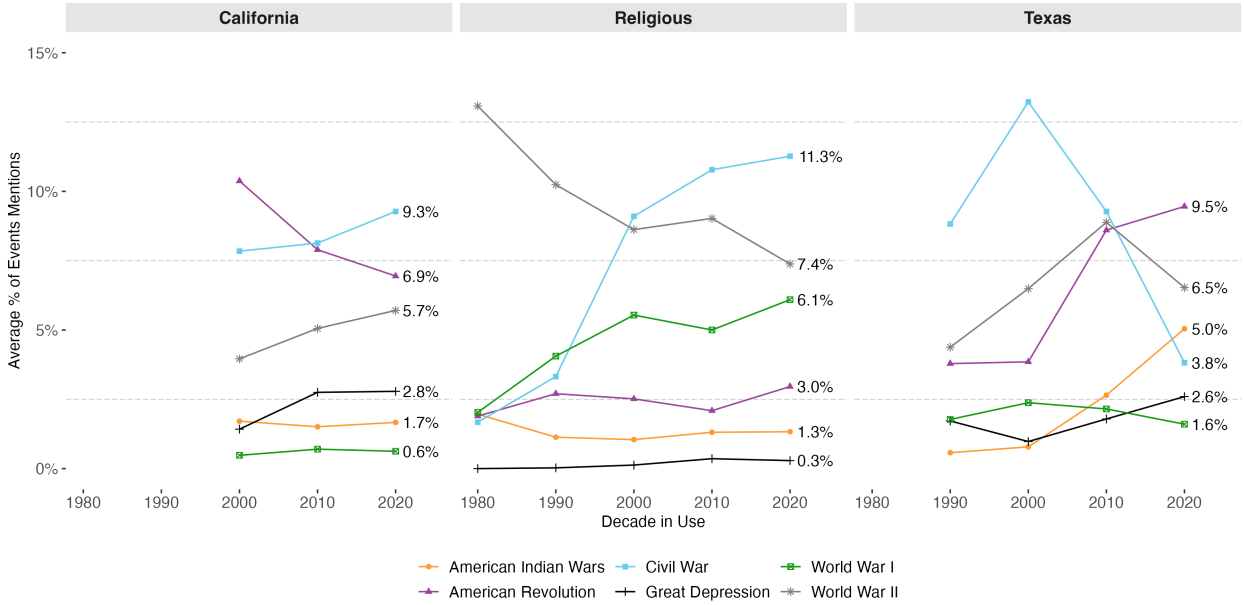
Notes: This figure illustrates the average percentage of pages within a specific subject that contain a specific topic. To calculate the average percentage of pages containing each topic, we first calculate the percentage of pages for each collection, subject, grade, and year in use. We then average these percentages at the collection level, to ensure that pages in each grade, subject, and year in use are weighted equally in the final averages.

FIGURE 7
Average Similarity Between Page Embeddings by Collection Pairs and Topics on Page



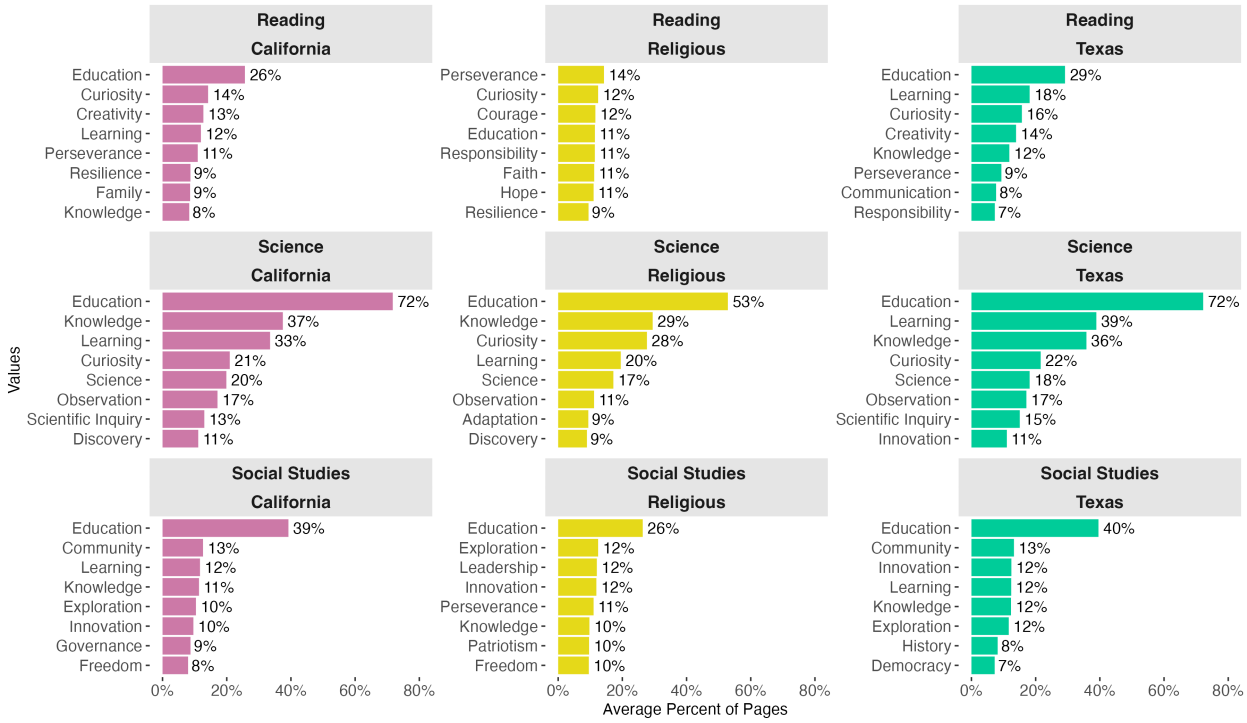
Notes:

FIGURE 8
Mentions of Historical Events



Notes: This figure illustrates the most frequently mentioned events within a collection over time. To determine the average percentage of mentions for each event, we first calculate the percentage of mentions across each collection, subject, grade, and year in use. We then average these percentages at the collection and decade level to ensure that pages from each grade and subject are weighted equally in the final averages.

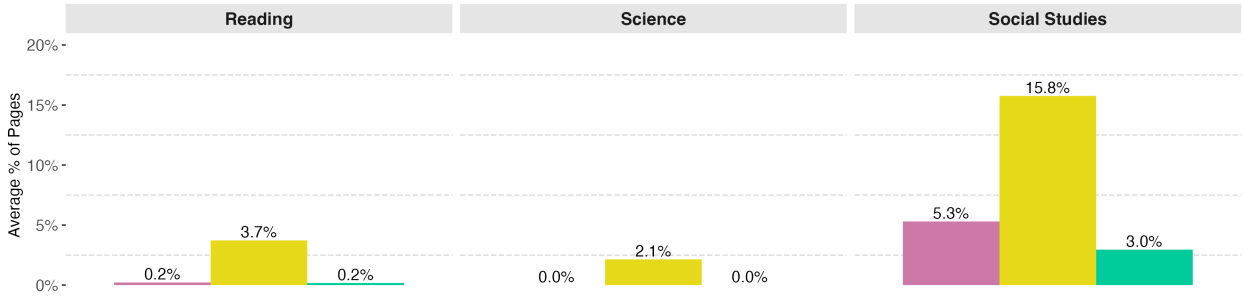
FIGURE 9
Top 8 Values by Collection



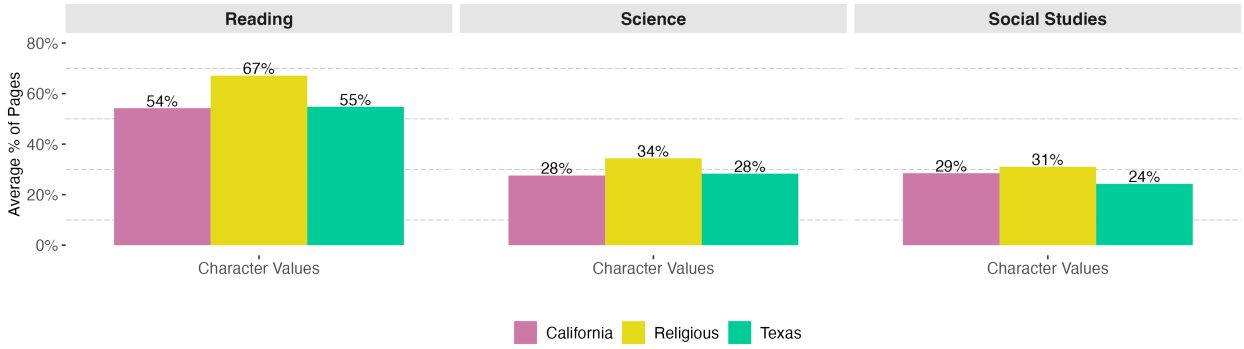
Notes: This figure shows the top values (e.g. values that appear on the highest average percentage of pages) within a collection. To calculate the average percentage of pages containing each value, we first calculate the percentage of pages for each collection, subject, grade, and year in use. We then average these percentages by collection, to ensure that pages in each grade, subject, and year in use are weighted equally in the final averages.

FIGURE 10
Pages Containing Specific Categories of Values

(a) Average Percent of Pages Containing Religious Values by Subject



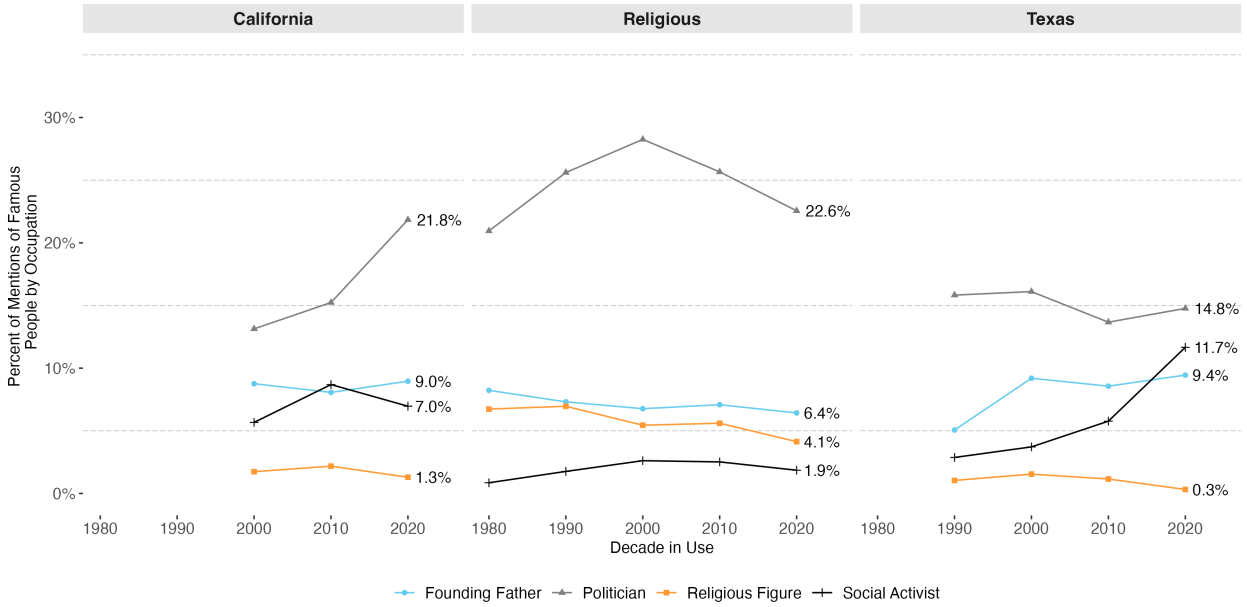
(b) Average Percent of Pages Containing Character Values by Subject



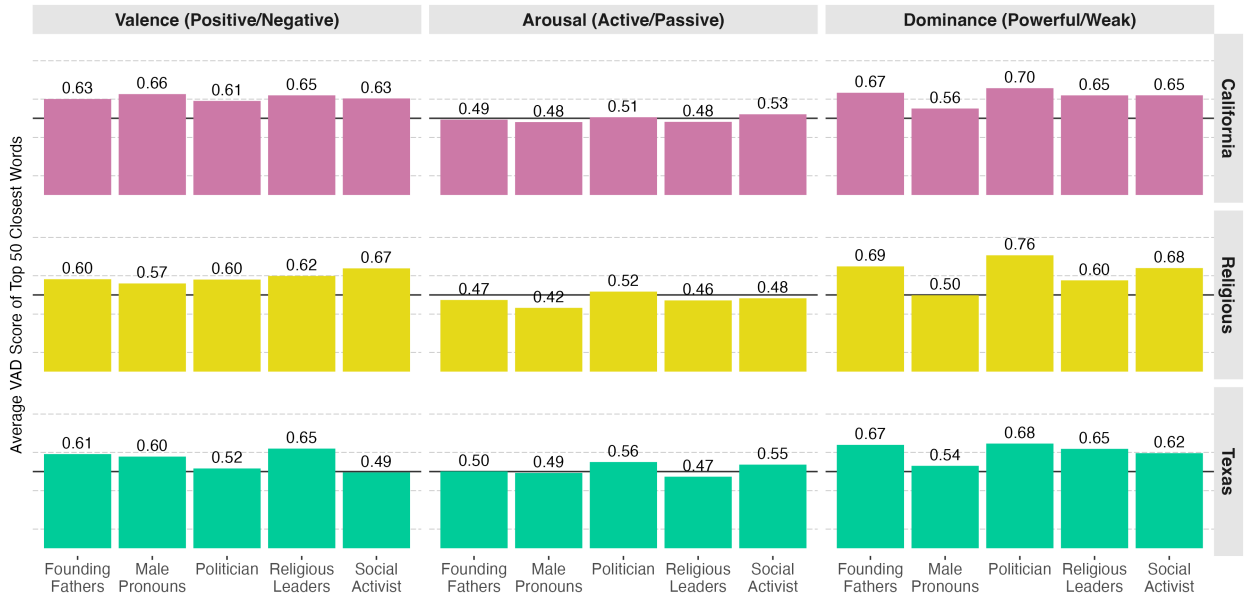
Notes: This figure shows the average percentage of pages within a specific subject that contain either a religious or a character value. To calculate the average percentage of pages containing each group of values, we first calculate the percentage of pages for each subject, grade, and year in use. We then average these percentages by collection, to ensure that pages in each grade, subject, and year in use are weighted equally in the final averages.

FIGURE 11
Historical Figures

(a) Mentions

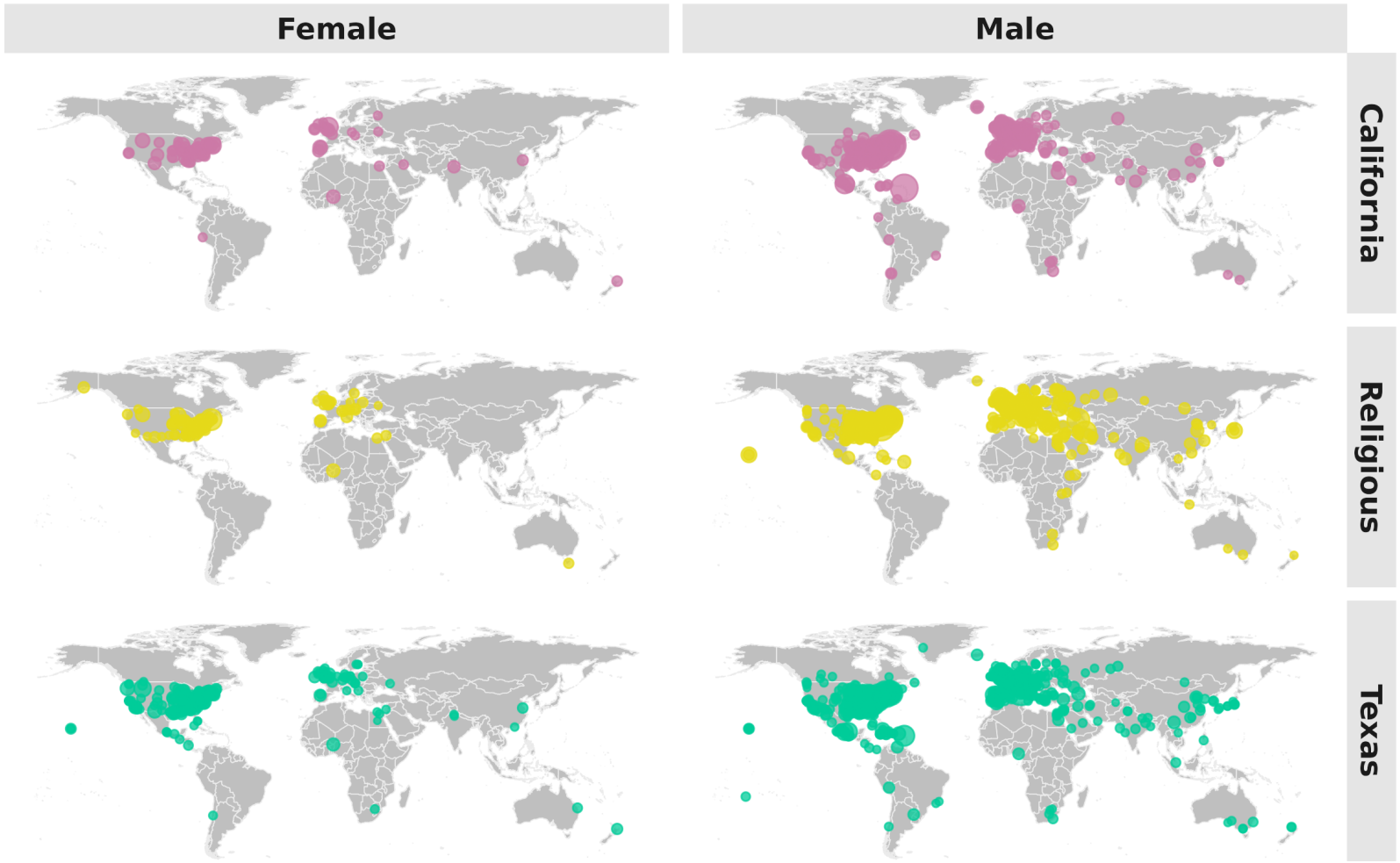


(b) Context



Notes: This figure shows our analysis of famous people by occupation. Panel (a) figure shows the number of times a famous person with a specific occupation is mentioned a collection over time. To determine the average percentage of mentions for each occupation, we first calculate the percentage of mentions across each subject, grade, and year in use. We then average these percentages at the collection and decade in use level to ensure that pages from each grade and subject are weighted equally in the final averages. Panel (b) figure shows the average Valence, Arousal, and Dominance scores of the top 50 words that are closest on average to the word embeddings of the names of famous people that have that occupation. We also show the same scores for the top 50 words closest on average to male pronouns for reference.

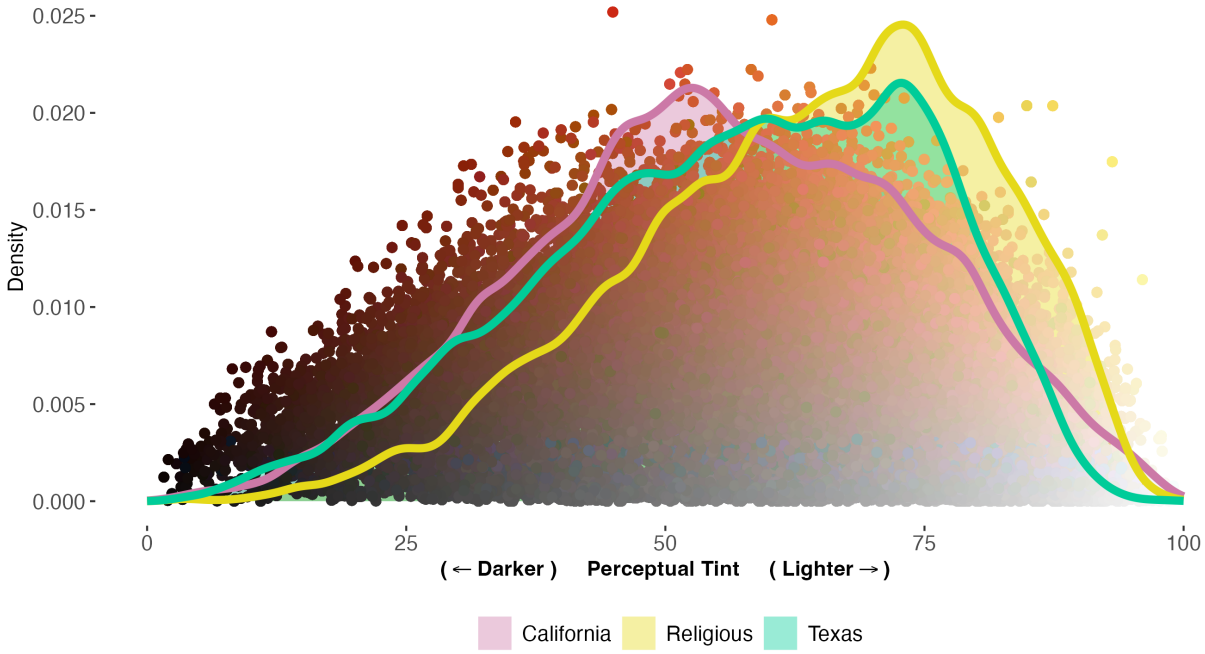
FIGURE 12
Birthplaces of Historical Figures



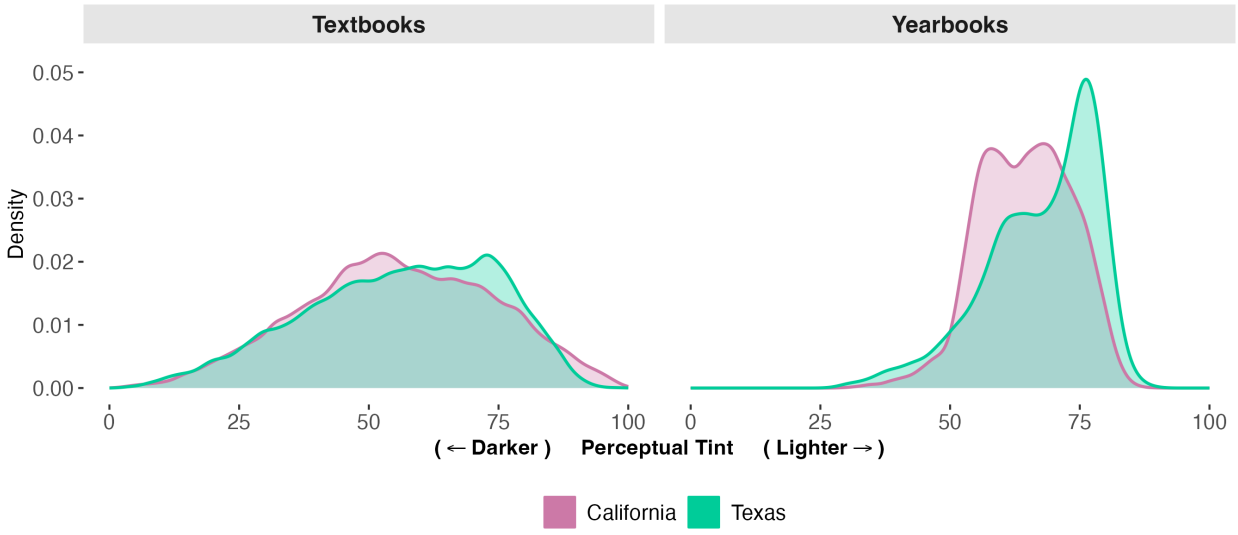
Notes: This figure shows the birthplaces of all famous people mentioned in a collection. The size of the point represent the number of mentions.

FIGURE 13
Distribution of Skin Tint of Pictured Faces

(a) Skin Tint by Textbook Collection

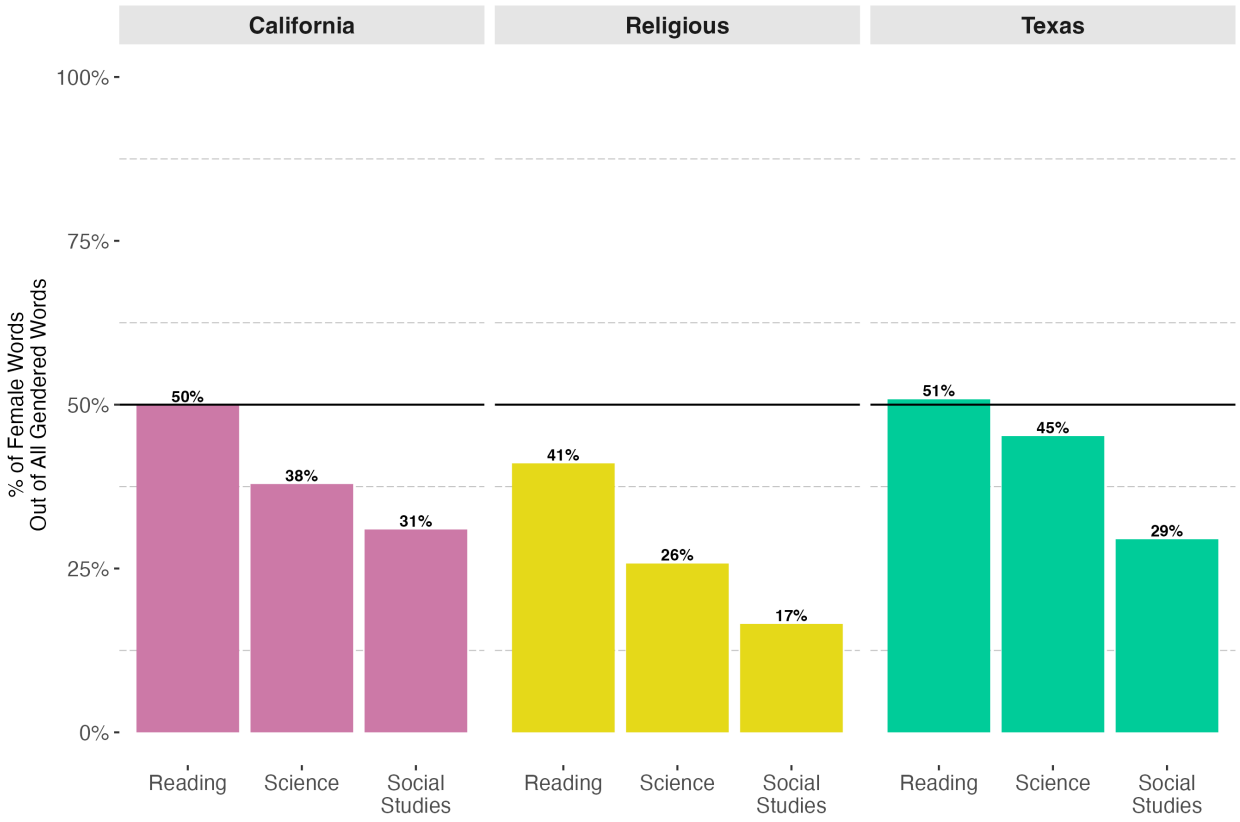


(b) Skin Tint from Textbooks and Yearbooks by State



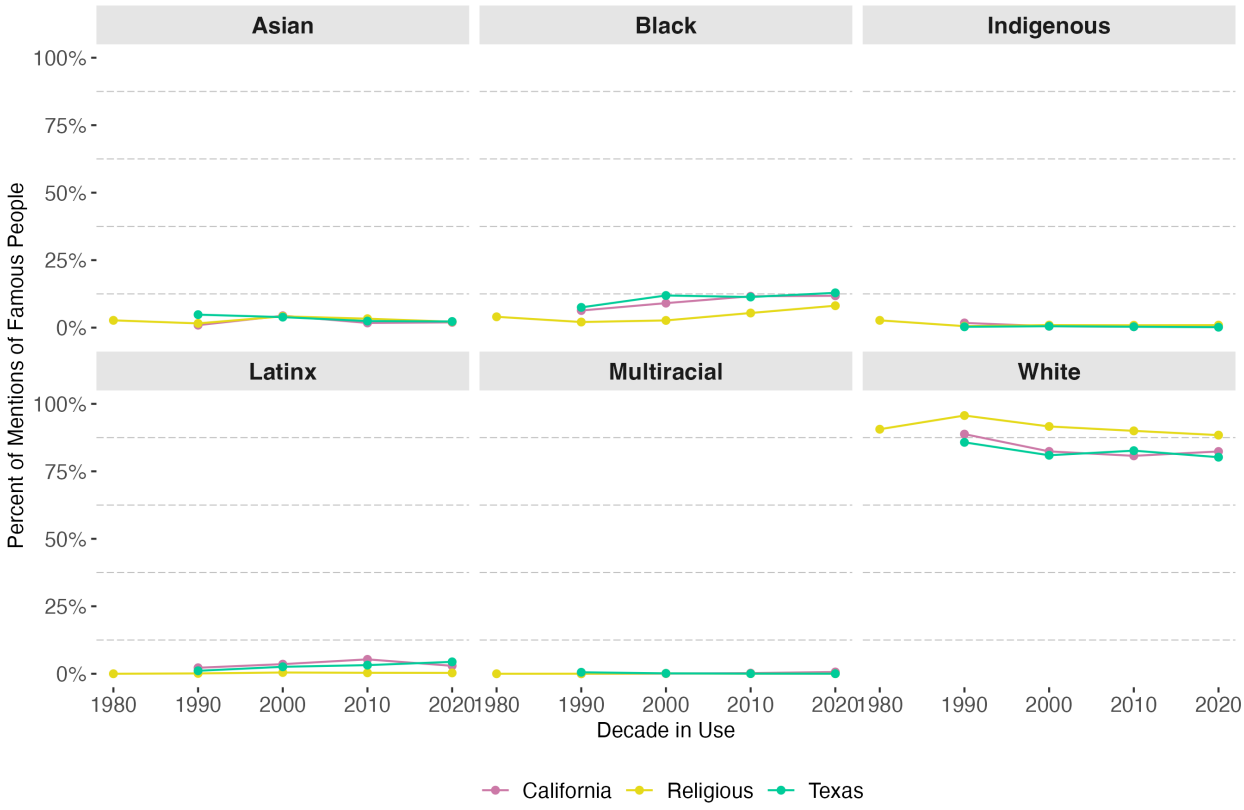
Notes: This figure compares the skin color of faces detected in textbooks and yearbooks. Panel (a) shows the distribution of skin color tint for faces detected in our collections of California, Texas, and Religious textbooks. Panel (b) shows the distribution of skin color tint for faces detected in California and Texas yearbooks.

FIGURE 14
Female Representation in Text By Subject



Notes: This figure shows the average percent of female words out of all gendered words by collection and subject. Note that female words includes female names, female pronouns, and female terms such as queen or brother. To calculate the average percentage of female words, we first calculate the percentage each collection, subject, grade, and year in use. We then average these percentages at the collection and subject level, to ensure that textbooks in each grade and year in use are weighted equally in the final averages.

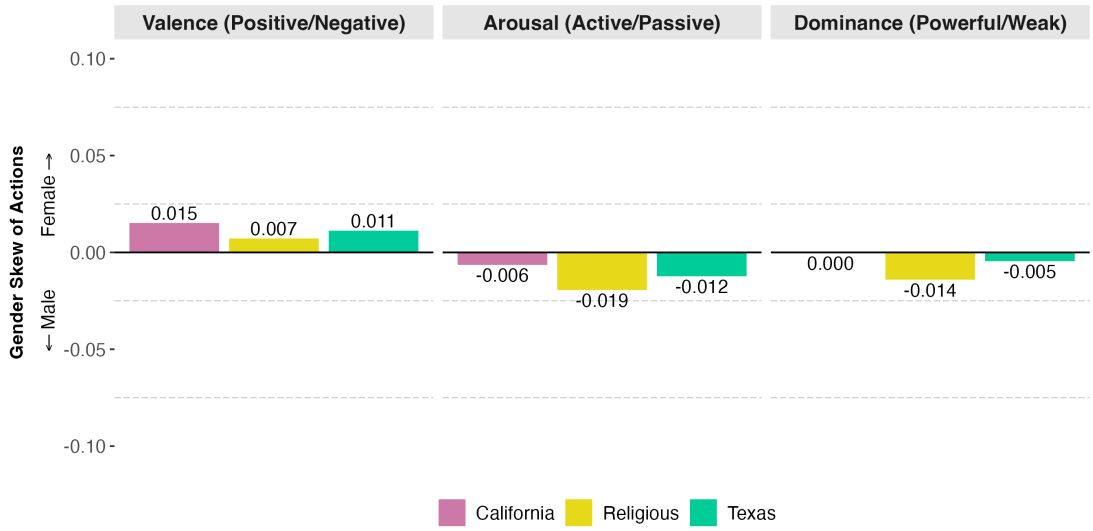
FIGURE 15
Percentage of Famous Individuals By Race in Each Collection Over Time



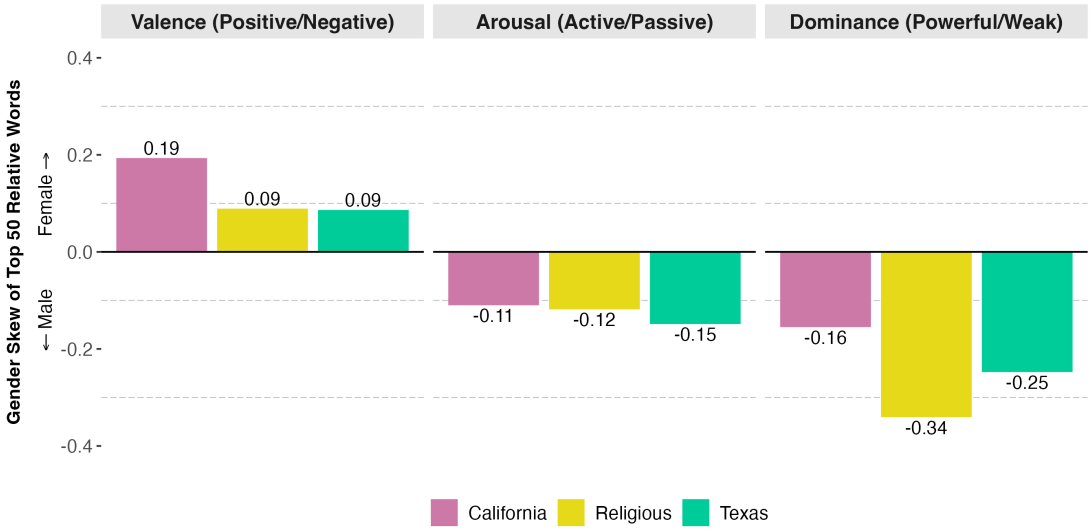
Notes: This figure shows the number of times a famous person of a specific race is mentioned a collection over time. To determine the average percentage of mentions for each race, we first calculate the percentage of mentions across each subject, grade, and year in use. We then average these percentages at the collection and decade in use level to ensure that pages from each grade and subject are weighted equally in the final averages.

FIGURE 16
Average VAD Scores by Gender Measures

(a) Average Difference in VAD Scores of Female vs. Male Actions

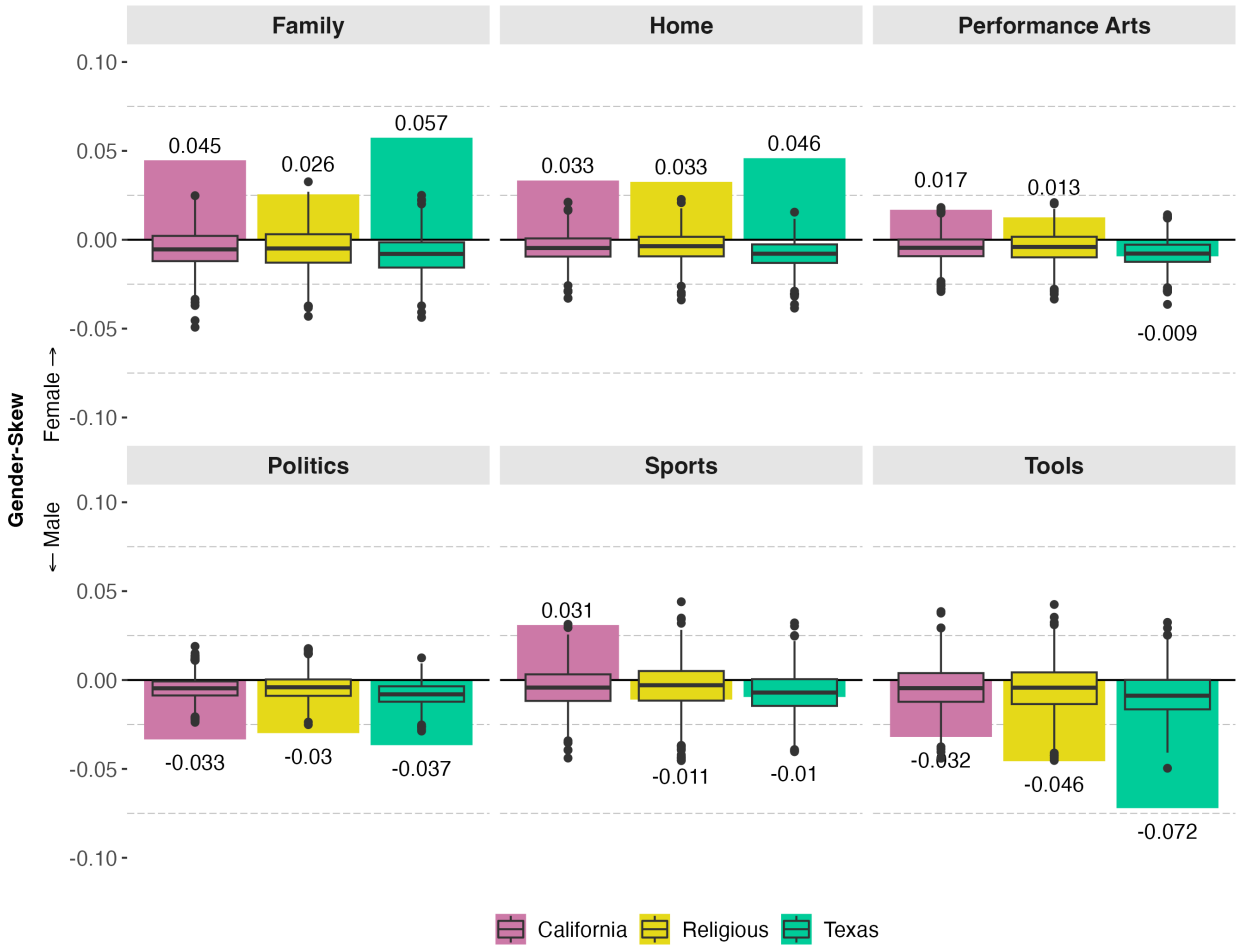


(b) Average VAD Scores of Top 50 Words by Gender



Notes: This figure shows the Valence, Arousal, and Dominance (VAD) scores of words and actions associated with males and females and then takes the differences of these score to find the gender skew of that dimension. Panel (a) shows the difference in the VAD scores of actions associated with female words and the actions associated with male words. If the skew is positive, it means that context in which females are discussed uses words that are higher on that Valence, Arousal, or Dominance dimension than the context that males are discussed in. We use Semantic Role Labeling to find the actions associated with males versus females. Panel (b) shows the difference in the VAD scores of the top 50 relative words closest to female words and relative words closest to male words. We call this difference in scores the gender skew of scores. If the skew is positive, it means that context in which females are discussed uses words that are higher on that Valence, Arousal, or Dominance dimension than the context that males are discussed in. To find the top relative words associated with females, we find the words that are closest to females in word embedding space and farthest from the males. We do the a similar exercise to find the top relative words associated with males. Note that we are limiting our analysis to words that are included in the VAD lexicon.

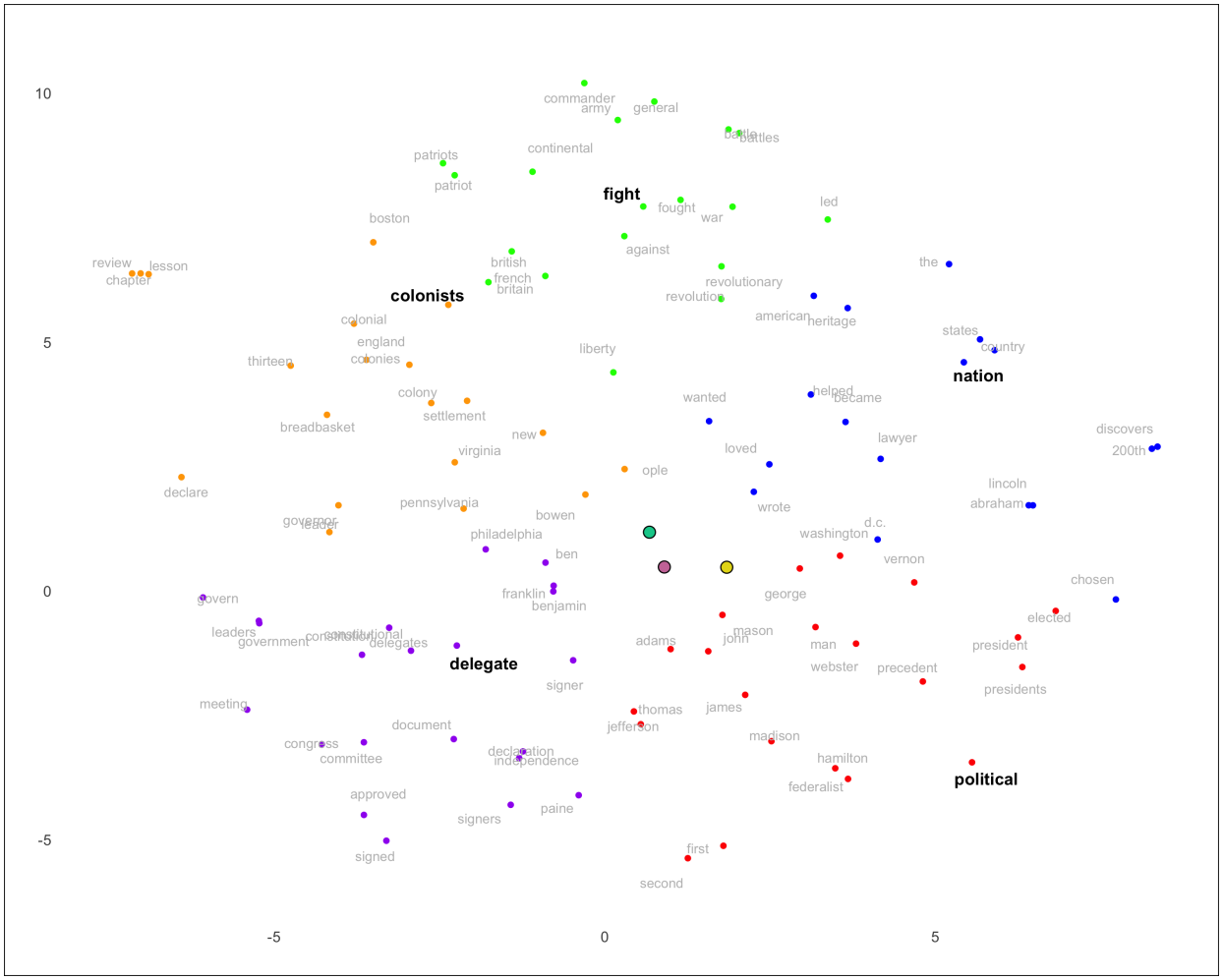
FIGURE 17
Gender Skew By Collection and Domain



Notes: This figure shows the gender skew of certain domains by calculating the difference in the association between female pronouns and a given domain with male pronouns and given domain. If the gender skew is positive, it means that words belonging to a specific domain are closer to female pronouns on average (e.g. meaning those words are more likely to be discussed in the same context as female pronouns). The boxplots show the distribution of gender skew in a random set of words the same size as the domain of interest. If the estimated gender skew is larger than the whiskers of the boxplot, this means that the gender skew for a collection is larger than that of a random set of words and therefore, likely to not have arisen by chance. Note that we are limiting our analysis to words that are included in the VAD lexicon.

FIGURE 18
Topic Embeddings For Famous Individuals by Occupation

(a) Founding Fathers



Notes: Each small dot represents one of the top 75 words closest to a collection topic embedding from a specific collection. We cluster the words into subtopics and denote these subtopic by color. The name of the subtopic is bolded and is chosen using the central word embedding of a cluster. Each large dot represents a topic embedding; pink signifies California, yellow signifies Religious, and green signifies Texas. These embeddings, originally represented in a 300-dimensional space, are here reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and book representations should be interpreted in terms of their relative position to other observations.

VII Main Tables

TABLE 1
Summary Statistics

	California (1)	Religious (2)	Texas (3)
Collection Totals			
Number of Reading Textbooks	22	55	60
Number of Science Textbooks	36	17	14
Number of Social Science Textbooks	22	13	22
Number of Grade 3 Textbooks	42	50	55
Number of Grade 5 Textbooks	38	35	41
Total Number of Textbooks	80	85	96
Total Number of Pages	29,262	19,814	39,235
Total Number of Faces	23,664	14,065	22,040
Sample Period			
Earliest Year in Use	1999	1980	1986
Latest Year in Use	2022	2022	2021
Average Number of Years Textbooks were in Use	7	13	8
Weighted Average: Page Content			
Average Number of Topics per Page	8	7	8
Average Number of Values per Page	6	6	6
Weighted Average: Skin Color			
Average Skin Tint of Faces	55	60	53
Weighted Average Percent: Putative Race of Faces			
Average Percent of Faces Classified as Asian	5%	2%	3%
Average Percent of Faces Classified as Black	31%	25%	35%
Average Percent of Faces Classified as Latine	5%	2%	3%
Average Percent of Faces Classified as White	48%	57%	43%
Average Percent of Faces Classified as Other Race	1%	1%	1%
Weighted Average Percent: Putative Race of Famous People			
Average Percent of Famous People Classified as Asian	3%	4%	3%
Average Percent of Famous People Classified as Black	11%	3%	11%
Average Percent of Famous People Classified as Indigenous	1%	1%	1%
Average Percent of Famous People Classified as Latine	5%	0%	3%
Average Percent of Famous People Classified as White	81%	92%	82%
Weighted Average Percent: Gender of Faces			
Average Percent of Faces Classified as Female	48%	41%	44%
Average Percent of Faces Classified as Male	50%	55%	53%
Average Percent of Faces Classified as Other Gender	3%	4%	4%
Weighted Average Percent: Gendered Words and Famous People			
Average Percent of Female Gendered Words	40%	27%	43%
Average Percent of Famous People Classified as Female	19%	7%	19%

Note: For all averages, we first compute the average for all textbooks within the same subject, grade, and year in use. We then average across these initial averages for each subject, grade, and year in use. This methodology ensures that textbooks used over longer periods are given greater weight and that equal emphasis is placed on each subject and grade. Average number of topics and values per page on conditional on having at least one topic/value identified on a page.

TABLE 2
Geographic Mentions by Region

Region	California	Religious	Texas
Africa	3.18	5.70	3.70
East Asia	1.42	4.91	2.52
Europe	12.88	32.57	12.35
Latin America	7.28	2.98	9.25
Middle East	0.37	4.31	1.01
North America	73.34	43.38	68.98
Oceania	0.39	2.06	0.63
South Asia	0.66	2.86	0.76
Southeast Asia	0.48	1.22	0.80

TABLE 3
Average Skin Tint by Collection

	Dependent variable:	
	Skin Tint	
	(1)	(2)
Religious	6.191***	5.013***
	(1.343)	(1.329)
Texas	−0.356	0.270
	(1.675)	(1.709)
Asian		−3.711***
		(1.078)
Black		−10.612***
		(0.500)
Latinx		−14.037***
		(0.801)
Other/Unsure		−6.586***
		(0.830)
Grade 5	−4.915***	−4.681***
	(1.290)	(1.303)
Science	−9.127***	−9.208***
	(2.120)	(2.182)
Social Studies	−2.687*	−2.367*
	(1.442)	(1.419)
Year in Use	−0.108*	−0.095
	(0.059)	(0.058)
Constant	64.022***	67.999***
	(2.305)	(2.298)
Observations	347,294	347,294
Adjusted R^2	0.084	0.159

In this table, we show a regression of the skin tint of a face detected in a given textbook on dummy variables indicating which collection, subject and grade the textbook belongs to. We also include the year a textbook was in use as a continuous variable centered around the first year a textbook was in use in our sample. In column 2, we also control for the predicted race of faces. Standard errors clustered at the textbook level. *p<0.1; **p<0.05; ***p<0.01

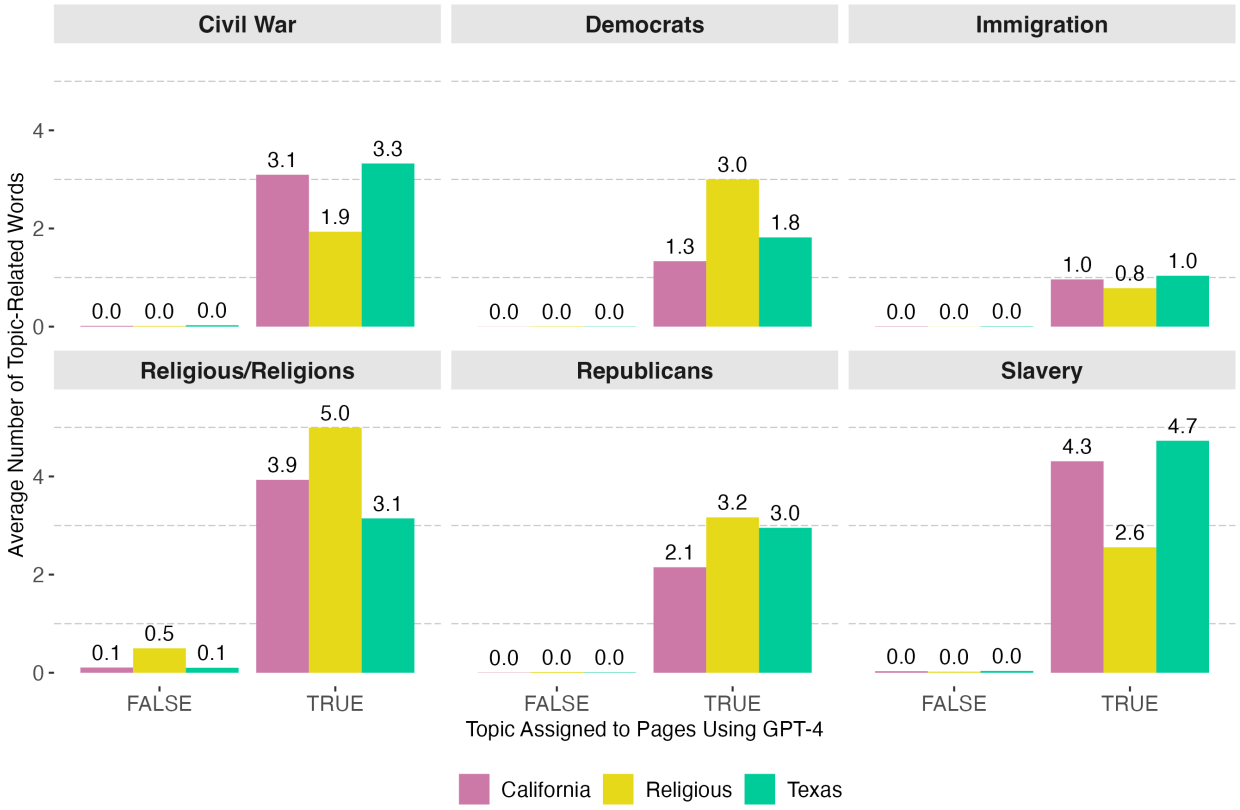
TABLE 4
Top 10 Relative Words by Gender

		Mean		
Top Words Associated with Female (vs. Male) Pronouns		V	A	D
California	husband, dress, sew, skater, cooking	0.67	0.48	0.49
Religious	doll, skirt, mama, shawl, grandmother	0.74	0.37	0.39
Texas	doll, dress, fabric, holly, skirt	0.65	0.31	0.34
Top Words Associated with Male (vs. Female) Pronouns		V	A	D
California	master, king, admiral, horse, lieutenant	0.69	0.56	0.81
Religious	harness, inventor, return, architect, horseback	0.68	0.53	0.64
Texas	sword, king, spear, stallion, believing	0.62	0.59	0.74

This figure shows the top 5 relative words associated with a specific gender as well as the average Valence, Arousal, and Dominance (VAD) scores of those words. To find the top relative words associated with females, we find the words that are closest to females in word embedding space and farthest from the males. We do the a similar exercise to find the top relative words associated with males. Note that we are limiting are analysis to words that are included in the VAD lexicon.

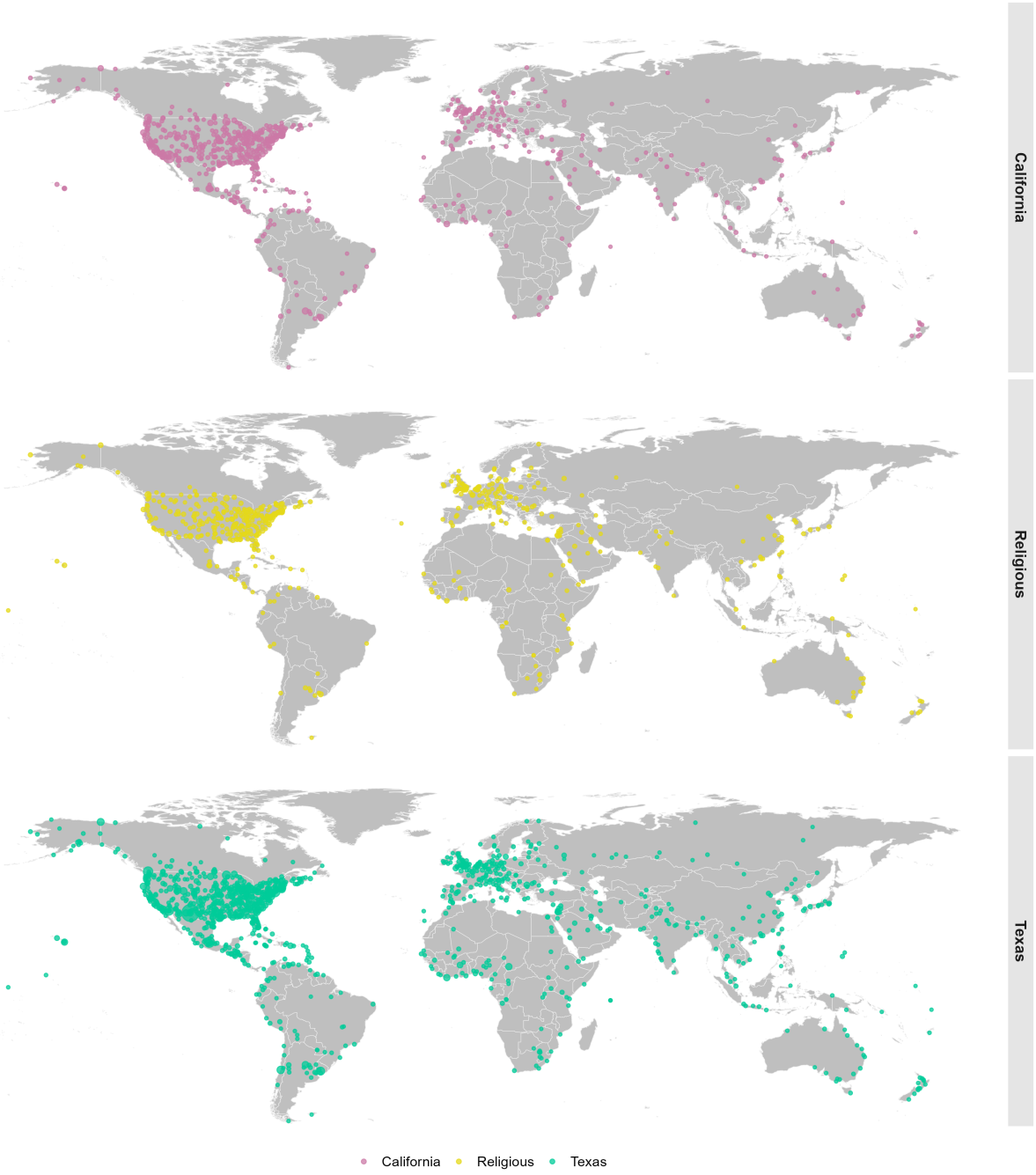
A Appendix Figures

FIGURE A.1
GPT-Identified Topics and Number of Topic-Related Tokens



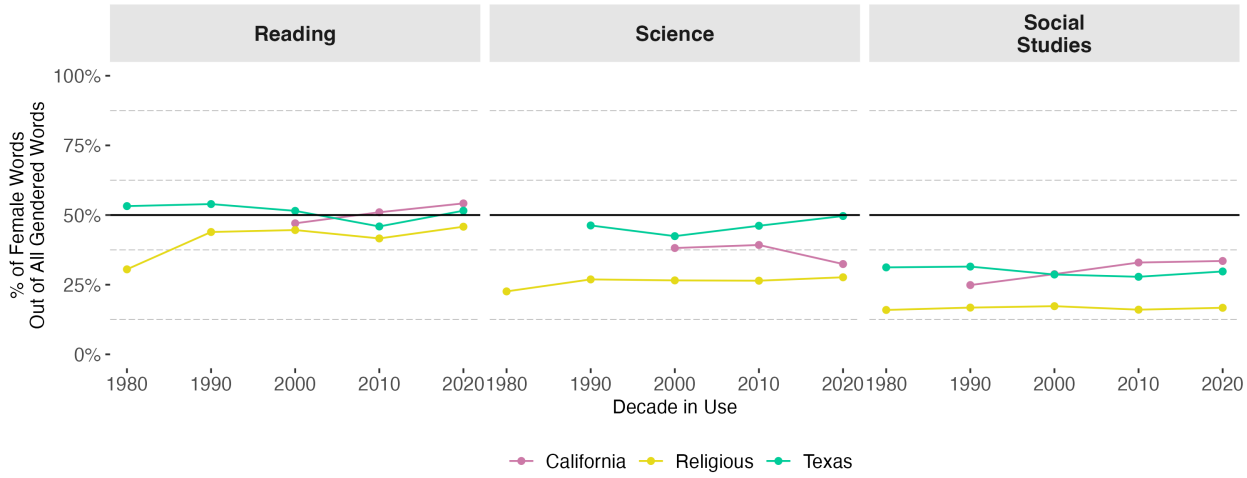
Notes: This figure shows the average number of words on a page that are related to a given topic that are then also assigned to a predicted GPT topic classification. "TRUE" indicates that a given page has a topic-related word is related to a topic, and GPT also classified it as connected to that topic. "FALSE" indicates that the word is related a topic, but GPT did not classify it as such.

FIGURE A.2
Mentions of Places



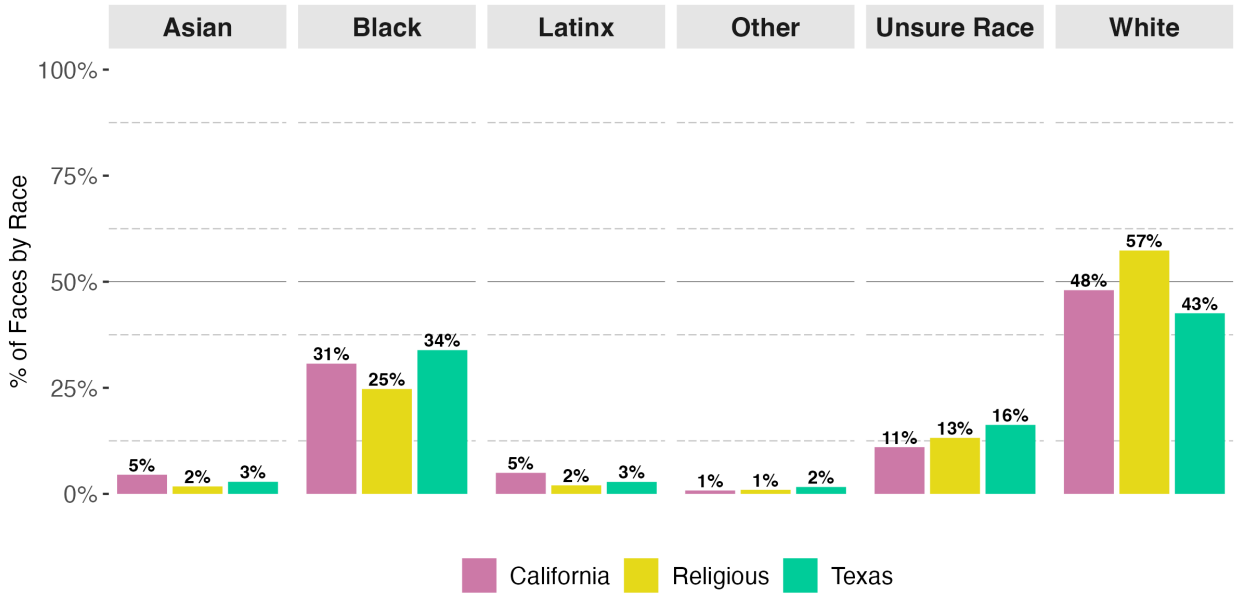
Notes: These maps depict dots for each location detected by Named Entity Recognition (NER). These do not include if a continent or country was named more generally.

FIGURE A.3
Percent of Female Words by Subject Over Time



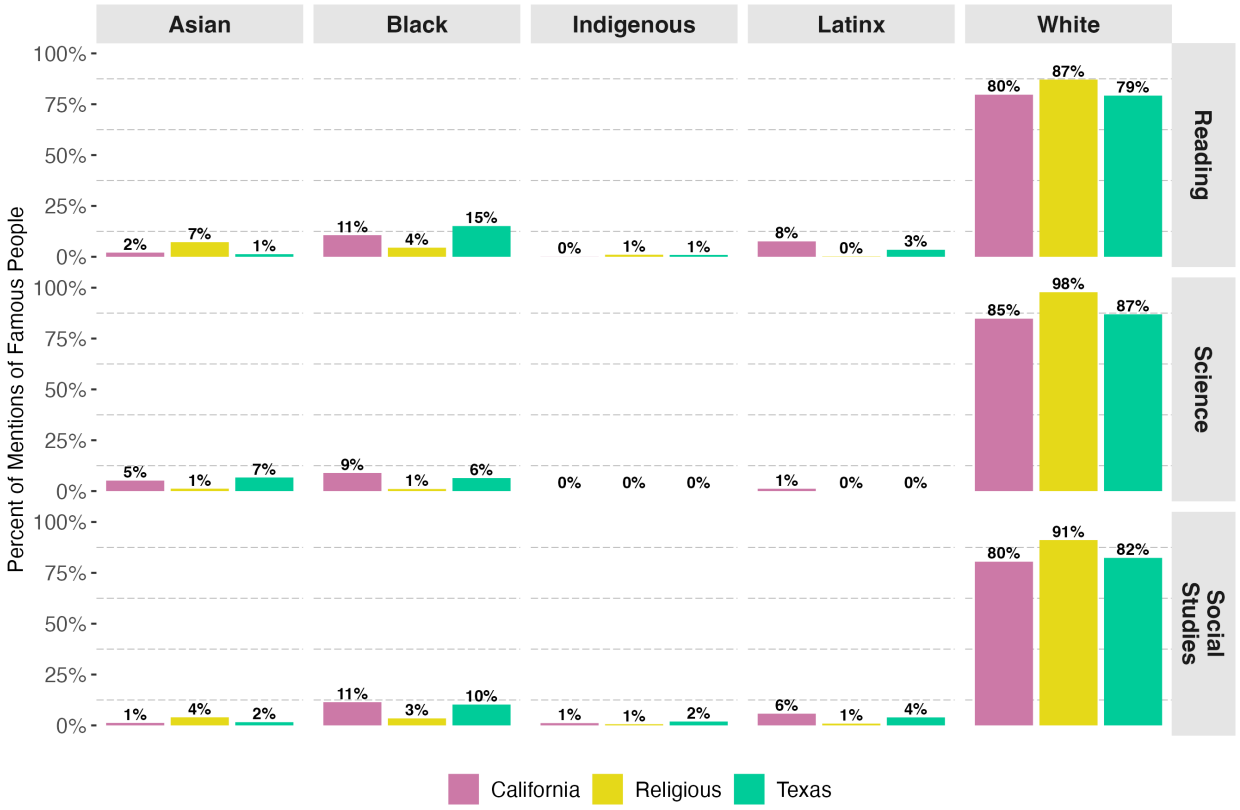
Notes: This figure shows the average percent of female words out of all gendered words by collection and subject over time. Note that female words includes female names, female pronouns, and female terms such as queen or brother. To calculate the average percentage of female words, we first calculate the percentage each collection, subject, grade, and year in use. We then average these percentages at the collection, subject and decade in use level, to ensure that textbooks in each grade and year in use are weighted equally in the final averages.

FIGURE A.4
Race/Ethnicity of Faces by Collection and Subject



Notes: This figure shows the average percentage of faces classified as a specific race by our feature classification model described in Section III.A.3 by collection and subject. To determine the average percentage, we first calculate the percentage of faces by race within each collection, subject, grade, and year in use. We then average these percentages at the collection level to ensure that pages from each grade, subject, and year in use are weighted equally.

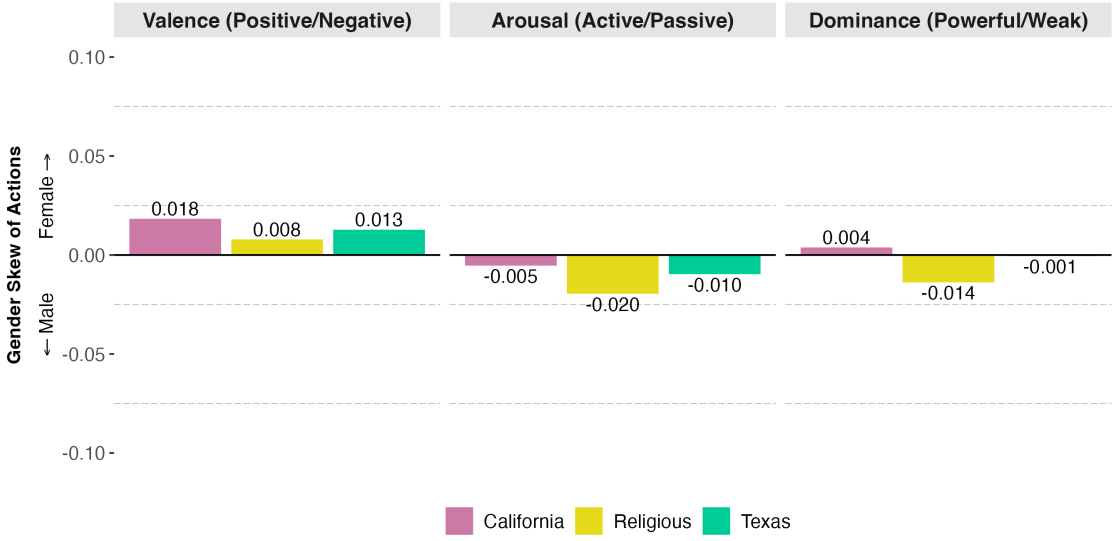
FIGURE A.5
Race/Ethnicity of Famous People by Collection and Subject



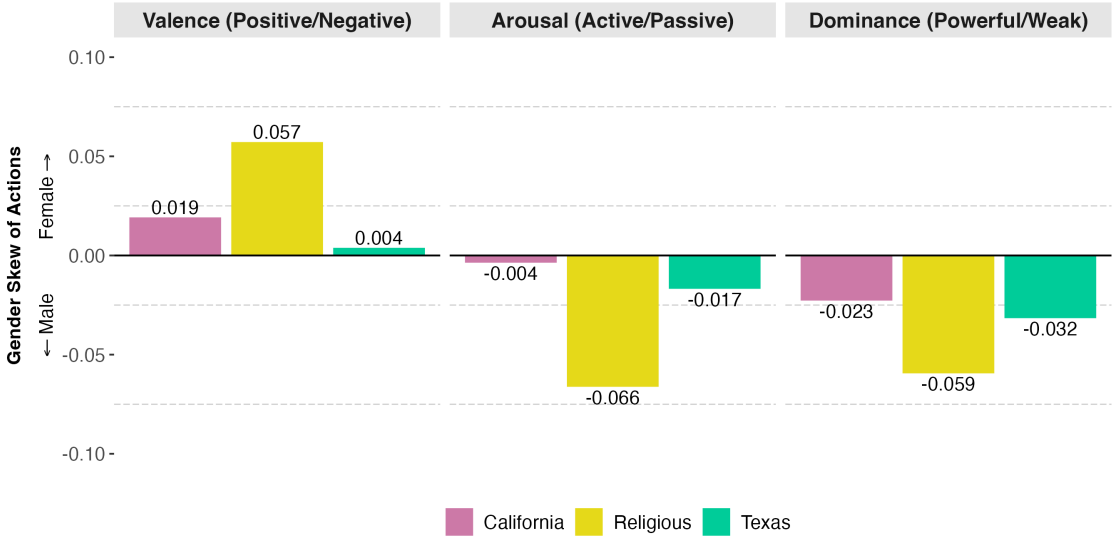
Notes: This figure shows the average percentage of famous people that identity as a specific race out of all mentions of famous people by collection and subject. To determine the average percentage of mentions, we first calculate the percentage of mentions of famous people by race across each subject, grade, and year in use. We then average these percentages at the decade level to ensure that pages from each grade, subject, and year in use are weighted equally in the final averages.

FIGURE A.6
Average Difference in VAD Scores of Gendered Actions

(a) Actions Associated with Male Pronouns vs. Actions Associated with Female Pronouns



(b) Actions Associated Only with Male Words vs. Actions Associated Only with Female Words



Notes: This figure shows the Valence, Arousal, and Dominance (VAD) scores of actions associated with males and females and then takes the differences of these score to find the gender skew of that dimension. If the skew is positive, it means that context in which females are discussed uses words that are higher on that Valence, Arousal, or Dominance dimension than the context that males are discussed in. We use Semantic Role Labeling to find the actions associated with males versus females. Panel (a) shows the difference in the VAD scores of actions associated with female pronouns and the actions associated with male pronouns. Panel (b) shows the difference in VAD scores of actions associated only with female words vs actions only associated with male words.

B Appendix Tables

TABLE B.1
Female Representation in Text and Images

	Dependent variable:	
	% Female Words	% Female Faces
	(1)	(2)
Religious	-0.103^{***}	-0.068^{***}
	(0.023)	(0.023)
Texas	0.022	-0.026
	(0.020)	(0.020)
Grade 5	-0.058^{***}	-0.008
	(0.022)	(0.024)
Science	-0.112^{***}	0.041^{*}
	(0.020)	(0.022)
Social Studies	-0.199^{***}	0.002
	(0.019)	(0.022)
Year in Use	0.0002	-0.0003
	(0.001)	(0.001)
Constant	0.523^{***}	0.449^{***}
	(0.033)	(0.033)
Observations	2,521	2,521
Adjusted R^2	0.322	0.054

Notes: Regression of two different measures of female representation in textbooks on indicator variables indicating which collection, subject, and grade the textbook belongs to. We also include the year a textbook was in use as a continuous variable centered around the first year a textbook was in use in our sample. Standard errors clustered at the textbook level. *p<0.1; **p<0.05; ***p<0.01

TABLE B.2
Average Skin Tint by Collection and Race

	Dependent variable:				
	Skin Tint				
	Asian	Black	Latine	White	Other/Unsure
	(1)	(2)	(3)	(4)	(5)
Religious	1.908	8.529***	1.750	4.476***	1.738
	(1.696)	(1.708)	(1.354)	(1.328)	(1.911)
Texas	−7.699***	3.403**	−2.177	0.457	−5.348**
	(2.617)	(1.544)	(2.296)	(1.770)	(2.559)
Grade 5	−7.929***	−2.786**	−2.147	−5.215***	−6.800***
	(2.300)	(1.177)	(1.442)	(1.287)	(2.424)
Science	−8.043***	−8.583***	−3.613	−9.289***	−12.744***
	(2.945)	(2.061)	(2.378)	(2.029)	(3.416)
Social Studies	−3.519	−2.609*	−0.947	−2.112	−4.436**
	(2.241)	(1.341)	(1.557)	(1.477)	(2.115)
Year in Use	−0.159*	−0.139**	0.160**	−0.080	−0.078
	(0.094)	(0.061)	(0.069)	(0.057)	(0.088)
Constant	70.976***	55.471***	44.536***	67.901***	66.186***
	(3.058)	(2.571)	(2.229)	(2.290)	(3.454)
Observations	12,574	99,455	13,084	182,639	39,542
Adjusted R^2	0.161	0.066	0.033	0.088	0.106

In this table, we show a regression of the skin tint of a face detected in a given textbook on dummy variables indicating which collection, subject and grade the textbook belongs to. In each column, we condition on a different predicted race of each face. Standard errors clustered at the textbook level. *p<0.1; **p<0.05; ***p<0.01

TABLE B.3
Top 10 Famous People by Collection

Collection	Name	Race	Gender	Rank	Mentions	Textbooks
California	Benjamin Franklin	White	Male	1	383	32
California	Thomas Jefferson	White	Male	2	322	31
California	George Washington	White	Male	3	417	30
California	Abraham Lincoln	White	Male	4	197	25
California	Martin Luther King Junior	Black	Male	5	133	24
California	John Adams	White	Male	6	233	18
California	Abigail Adams	White	Female	7	81	18
California	Frederick Douglass	Black	Male	8	95	17
California	Paul Revere	White	Male	9	194	16
California	James Madison	White	Male	10	139	16
Religious	George Washington	White	Male	1	310	33
Religious	Abraham Lincoln	White	Male	2	369	27
Religious	Benjamin Franklin	White	Male	3	158	25
Religious	Paul Revere	White	Male	4	185	19
Religious	Thomas Jefferson	White	Male	5	146	18
Religious	Noah Webster	White	Male	6	280	17
Religious	John F. Kennedy	White	Male	7	102	17
Religious	James Madison	White	Male	8	50	16
Religious	Robert E. Lee	White	Male	9	132	15
Religious	Daniel Boone	White	Male	10	103	14
Texas	George Washington	White	Male	1	531	45
Texas	Benjamin Franklin	White	Male	2	377	33
Texas	Abraham Lincoln	White	Male	3	422	31
Texas	Thomas Jefferson	White	Male	4	245	31
Texas	Alexander Hamilton	White	Male	5	245	27
Texas	Paul Revere	White	Male	6	268	25
Texas	Martin Luther King Junior	Black	Male	7	219	24
Texas	Neil Armstrong	White	Male	8	198	24
Texas	Harriet Tubman	Black	Female	9	127	22
Texas	John F. Kennedy	White	Male	10	214	21

Notes: This figure illustrates the most frequently famous people mentioned within a collection. We calculate both the number of times a person as mentioned and the number of textbooks they appear in. We then sort first by number of textbooks within a collection and then by number of mentions to find the top mentioned individuals.