Separation of Church and State Curricula? Public Standards, Private Values, and Textbook Content*

Anjali Adukia† Emileigh Harrison‡

June 10, 2026

Abstract

Curricula are a critical site of cultural transmission, yet we know little about the values conveyed in textbooks across educational settings or the forces that shape them. We examine textbooks from Texas and California public schools and religious-private and home schools spanning 1980-2022, using computational and AI tools to measure presence and portrayal of people, topics, and values over time. Despite narratives of political polarization, Texas and California textbooks show substantial similarity. In contrast, religious private-school textbooks place greater emphasis on religious and character values and depict characters with lighter skin colors; we show these differences align with preferences and demographics of target markets. To explain public-school convergence, we develop a framework in which publishers serve a national market constrained by state standards and provide evidence consistent with cross-state spillovers in textbook content following standards revisions. Conversely, private-school publishers serve a more segmented, less regulated market, leading to greater differentiation.

*For helpful feedback, we thank Ben Arold, Elliott Ash, Sam Bazzi, Marianne Bertrand, Dan Black, Sandy Black, Terence Chau, Judy Chevalier, Michael Dinerstein, Jeff Denning, Steven Durlauf, Alex Eble, Martin Fiszbein, Devika Lakhote, Sendhil Mullainathan, Elias Papaioannou, Hakizumwami Birali Runesha, Heather Sarsons, Teodora Szasz, Jon Valant, Mythili Vinnakota, David Yanagizawa-Drott, Hiro Yoshikawa, and seminar participants at AEFP, APPAM, Boston University, CESifo, Columbia, FGV, Harvard, Johns Hopkins, London Business School, RAND, Stanford, SREE, UChicago, UCLA, Univ. of Michigan, Université Paris Dauphine-PSL, UPenn, UT-Austin LBJ School, Warwick, BFI AI in Social Science conference, Yale, and the Zurich AI/Economics conference. For exemplary research support, we especially thank Matt Bonci and Jake Nicoll. We also thank Sam Abers, Elisa Xi Chen, Paula Dastres Gallardo, Benjamin Leiva, Celia Zhu, and other members of the MiiE Lab for excellent research assistance. For manual labeling assistance, we thank Sara Basheer, Rodrigo Chaves, Jiwon Choi, Rain Hu, Arthur Langlois, and Yongjae Lee. For financial support, we thank the National Academy of Education, Spencer Foundation, and Becker Friedman Institute at UChicago. The research reported here was also supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A200478 to the University of Chicago. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

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

‡ University of Texas at Austin, MiiE Lab, and the University of Chicago Inclusive Economy Lab, emileigh.harrison@austin.utexas.edu

Our collective memory and understanding of the world are shaped by a variety of forces, including the stories we share (Michalopoulos and Xue, 2021; Adukia et al., 2023), the narratives propagated through popular culture (Riley, 2024; Michalopoulos and Rauh, 2024), as well as the monuments we build and the history they represent (Bazzi et al., 2026a). These elements not only influence *what* is remembered but also shape *who* is remembered and *how*, thereby reflecting what is valued both explicitly and implicitly in society. Education plays a pivotal role in this process with curricular materials serving as critical sites of cultural transmission. These materials do more than convey academic knowledge; they also shape children's worldviews and values (Cantoni et al., 2017; Bazzi et al., 2026b). Consequently, the topics and values to which children are exposed in educational materials can have long-run consequences for shaping beliefs, cultural transmission, and nation building.

Parents recognize the role of education in transmitting values and seek out learning environments that reflect the values, cultural norms, or beliefs they want to instill in their children (Catt and Rhinesmith, 2016; Bisin and Verdier, 2001; Wheaton, 2020; Williamson and Withrow, 2025). However, despite intensifying debates over public-school curricula, increasing educational gag orders (Friedman and Tager, 2022), and growing support for tax-funded school choice programs that expand access to private alternatives – including religious private schools (EdChoice, 2023), little is known about how curricular content varies across political and religious educational settings in the United States (U.S.).

Our study addresses this gap by systematically analyzing values, topics, and people embedded in an influential set of U.S. school textbooks used between 1980 and 2022. We examine public-school textbooks adopted in California and Texas, the two states with the largest K-12 student populations and major textbook markets, enabling us to understand how educational materials vary across distinct political contexts for a substantial share of students across the U.S. We also examine textbooks commonly adopted in religious private and home schools, focusing on materials produced by conservative Christian publishers Abeka, Accelerated Christian Education (ACE), and Bob Jones University Press (BJU Press), which allows us to compare religious-school curricula with state-adopted public-school materials. Our analyses focus on third- and fifth-grade textbooks in three foundational subjects: Reading, Science, and Social Studies. We use computational social science and artificial intelligence methods to extract measures of presence and portrayal across these textbooks.

Three main patterns emerge. First, Texas and California public-school textbooks are more similar than prevailing narratives imply (Goldstein, 2020). Second, religious textbooks differ notably from public-school materials. They place greater emphasis on religious values and character values such as caring and obedience, contain less female representation, depict

lighter-skinned and more White characters, and treat topics such as evolution and religion differently, including in science textbooks. Religious topics are present in all three collections, but in public-school textbooks they are largely confined to contexts such as the nation's founding, whereas in religious textbooks they appear in explicitly theological discussions such as heaven or prayer. Third, important similarities emerge across all collections: for example, each collection portrays females in contexts that are more positive but less active and less powerful than males, discusses Christianity more than other religions, and gives limited attention to many groups that are rarely represented in mainstream historical narratives.

These descriptive findings pose a puzzle: if public-school curricula are politically salient, why are Texas and California public-school textbooks relatively similar, while religious private-school textbooks diverge? To reconcile these findings, we develop a conceptual framework in which mainstream publishers serve a national public-school market shaped by state standards and costly customization. These forces help to explain why public-school textbook content may converge toward a common baseline, even across politically distinct states. Religious publishers, in contrast, serve a more differentiated demand segment that is not subject to the same state standards, allowing schools and families to select materials they perceive to be closely aligned with their values. Consistent with this framework, we find that revisions to one state's content standards spill over into textbook content across states. This suggests that state standards do not just shape in-state content; through publishers' national production decisions, they can diffuse across markets. We further show, using public opinion data from the General Social Survey, that some differences between public-school and religious-school textbooks align with the demographics and survey-measured beliefs of populations that proxy for their relevant target markets.

Taken together, these findings suggest textbook markets shape cultural transmission differently across sectors. In the public school sector, state standards and customization costs encourage convergence in public-school textbooks; in the religious private school sector, weaker regulatory constraints leave more room for differentiation. This distinction has an important policy implication: expanding public funding for private schooling could increase curricular divergence, and, in turn, long-run divergence in beliefs across student populations, unless such funding is accompanied by curricular standards or other shared expectations about instructional content.

Our study contributes to several literatures. First, we contribute to the literature on education and cultural transmission. Existing work shows that schooling and curricular content can shape beliefs, political and religious attitudes, and identity, and that education is often a central instrument of state-building and cultural transmission (Cantoni et al., 2017;

Alesina et al., 2021; Paglayan, 2022; Ansell and Lindvall, 2013; Bazzi et al., 2026b; Bisin and Verdier, 2001; Wheaton, 2020; Arold et al., 2025). We extend this literature by examining the content of curricular materials themselves, documenting how the values children encounter in school vary across different political and religious educational environments.

Second, we contribute to the literature on content analysis of curricula and children's educational materials (Wade, 1993; Clark, 2007; Koss et al., 2018; Stevenson and Zlotnik, 2018; Lucy et al., 2020; Crawfurd et al., 2024; Lucy et al., 2025; Adukia et al., 2026). By developing and testing a conceptual framework of publisher competition under state standards to organize our descriptive findings, we extend the literature beyond content analysis toward explaining the market and regulatory forces that shape curricular content.

Finally, the computer vision and natural language processing tools that we develop and apply throughout our analysis contribute to the growing body of computational social science literature in a manner similar to Caliskan et al. (2017); Garg et al. (2018); Kozlowski et al. (2019); Adukia et al. (2022a,b, 2023, 2026); and Voth and Yanagizawa-Drott (2024). For example, we introduce a novel approach that combines generative AI with document embeddings to analyze the contexts in which characters are portrayed in images. This allows us to capture not only how often characters of different identities appear, but also the roles, activities, and settings in which they are shown. These methodological advancements can have meaningful implications for expanding the broader field of computational social science and the scope of research inquiry that can be pursued.

I Parental Values, Curriculum, and School Choice

Knowledge and values are transmitted to children through two primary pathways: vertical transmission from parents and family, and horizontal transmission from society and peers. Vertical transmission occurs from the beginning of a child's life in their home. Schools have traditionally served as key vehicles for horizontal transmission, imparting not only academic knowledge but also broader societal values. A growing body of evidence shows that educational content and instruction have causal impacts not only on children's academic outcomes but also on their worldviews. Changes in high school politics curricula can affect students' views on democracy and free markets (Cantoni et al., 2017) and changes in public-school content standards related to evolution can affect students' probability of choosing a STEM career (Arold, 2024). Messages about the abilities of others with similar racial and gender identities can impact academic performance (Riley, 2024). The removal of compulsory religious education can reduce the religiosity of students in adulthood (Arold et al., 2025).

Many stakeholders have long recognized the importance of schooling as a vehicle for

transmitting knowledge and values to children. Consequently, debates about what should be taught – and what should be excluded – from public-school curricula have persisted for centuries. In U.S. colonial times, schools were deeply intertwined with Christian instruction. However, as the nation formed, tensions between religious influence in public institutions and the ideal of a secular state motivated figures like Thomas Jefferson to advocate for a "wall of separation between church and state." This principle has continued to shape legal battles in education, including debates over school-sponsored prayer (e.g. *Engel v. Vitale*) and the teaching of evolution (e.g. *Epperson v. Arkansas*, *Tennessee v. Scopes*).

More recently, curricular debates have expanded beyond religion to encompass how history and identity are taught in schools. In the U.S., authority over curricular decisions lies primarily at the state or local level rather than nationally. While federal legislation such as No Child Left Behind (NCLB) and the Every Student Succeeds Act (ESSA) promoted standards-based reforms, these laws did not impose national standards, instead leaving states to define their own.[1] These differences are evident in specific state policies: California, for example, has adopted measures such as the 2011 FAIR Education Act to promote more inclusive representation in history instruction. Conversely, other states have adopted "educational gag orders" that limit classroom discussions of race, gender, and aspects of American history (Friedman and Tager, 2022). Similar debates occur globally, as national and regional governments adopt curricular standards that reflect their political, cultural, and ideological priorities, producing notable variation in what and how subjects are taught – sometimes shifting in response to world events or changes in national leadership. Such variations in curricular standards or restrictions across jurisdictions with distinct political or ideological orientations may be reflected in their respective curricular materials, so students may encounter different messages even within public-school settings. Qualitative accounts of curricular content in states such as California and Texas – frequently understood as lying at opposite ends of the political spectrum – often highlight meaningful differences in the messages to which students are exposed (Goldstein, 2020).

A majority of parents choose public schools, but a growing share are turning to alternative options such as private and home schools, often citing dissatisfaction with academic standards or perceived misalignment between their values and those emphasized in public education. This shift toward non-public options, often facilitated by vouchers (EdChoice, 2023), occurs despite mixed evidence regarding the academic benefits of school choice programs. Deming et al. (2014), for example, find that academic improvements can occur, but only when students gain access to higher-quality schools; while Abdulkadiroğlu et al.

[1] These standards apply only to public schools, excluding private institutions and homeschooling families.

(2018) find that many rural parents use private-school vouchers to move their children to academically lower-quality schools. These patterns suggest that some parents are motivated by factors beyond academics, such as a desire for educational environments more closely aligned with their preferences, such as an emphasis on religious values. Indeed, most private-school students attend religious schools, and religion is a common motivation among homeschooling families.[2] This behavior aligns with theories of ideological threat and models of cultural transmission, in which individuals respond to perceived ideological threats by selecting institutions that reinforce preferred beliefs across generations (Bisin and Verdier, 2001; Tabellini, 2008; Wheaton, 2020). More broadly, parents' prioritization of social values in school choice is not a new phenomenon. In the 1960s and 1970s, some parents responded to public-school desegregation by creating all-White private "segregation academies," which reduced academic outcomes of White students attending these schools and had no effect on Black students (Williamson and Withrow, 2025). Together, this evidence indicates that some parents are willing to trade off measured academic quality in favor of school environments or curricula that more closely align with their social, moral, or religious values.

Public policy has rapidly evolved to accommodate parents' demand for more educational options. School choice programs such as vouchers, tax-credit scholarships, and education savings accounts have proliferated in the U.S., especially in the past few years. These programs make it easier than ever for parents to direct public education funds toward private and home schools that better align with their preferences. Recent U.S. Supreme Court decisions (e.g. *Espinoza v. Montana Department of Revenue* and *Carson v. Makin*) have also struck down barriers to using public funds at religious schools, ruling that once a state subsidizes private education, it cannot exclude faith-based schools from participation. Additionally, Congress recently authorized the nation's first federal tax-credit scholarship program, signaling growing federal momentum behind publicly supported school choice. Parents therefore have increased opportunities to opt out of the public-school system and still obtain public funding for the private school environments they perceive as better aligned with their values. However, this movement also raises questions about what is being taught in these alternative settings and how it compares to the content of public education.

Despite the growing public investment in private and home schooling, there is surprisingly little oversight or transparency regarding the curricula and values being taught outside

[2]In 2017, among the 5.8 million children enrolled in private school, 76% attended a religious school, with 49% enrolled in a conservative Christian school (National Center for Education Statistics, 2019). Pre-COVID estimates of the share of homeschoolers who identify as religious, or for whom religion is a primary motivation for homeschooling, ranged from over 50% to 90% (Kunzman, 2010; Bartholet, 2020).

the public-school system.[3] Consequently, we lack basic knowledge of the curricula and values being transmitted outside the public-school system. Given the role of education in cultural transmission and nation building, it is crucial to understand what different schools teach. By empirically examining the values embedded in public, private, and home school curricula, we aim to shed light on how curricular content varies across political and religious educational settings in the U.S. as well as the market and regulatory forces that shape it.

II Prior Analyses of Textbook Content

Our study complements existing qualitative and quantitative work examining curricular content in the U.S. and around the world. Much of this work focuses on the representation of gender and race in educational materials. For example, Crawfurd et al. (2024) analyze 1,255 English-language textbooks from 34 countries and find that women tend to be portrayed in stereotypical gendered roles. Lucy et al. (2020) also use NLP tools to examine who is present in the text of a sample of 15 Texas high school textbooks used between 2015 and 2017 and 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 most commonly purchased in the U.S., finds that fictionalized women are more likely to be involved in food, fashion, or household tasks.

A smaller body of research examines values transmitted through textbooks (c.f. Sallabaş 2013 who conducted a qualitative analysis of character values in Turkish textbooks). Even fewer studies examine textbooks commonly used in religious settings. The most relevant is a qualitative analysis by Scaramanga and Reiss (2018) on curricula published by Accelerated Christian Education (ACE) – one of the textbook publishers we include in our study – which found that churches and schools are often depicted as racially segregated in their materials.

Finally, most prior textbook analyses focus on text and largely overlook images. One exception is Deckman et al. (2018), who manually coded 1,468 images from U.S. elementary- and middle-school health textbooks to study portrayals of race, gender, and sexuality. They find that, although women and people of color appeared frequently, they were often depicted in stereotypical roles, and no images depicted same-sex couples. We extend this line of work by introducing a scalable approach to measuring visual portrayal. Using generative AI, we produce descriptions of pictured characters' features, actions, and contexts; we then estimate

[3]Most states require private school curricula to be "similar" to public standards, but 18 states either have no requirements or only require that private schools teach students the state constitution and state history. While states often have similar curricular requirements on paper for homeschoolers as they do for private schools, there is rarely any enforcement for homeschooling families (Bartholet, 2020).

document embeddings from these descriptions to measure visual portrayals of characters at scale. This enables us to examine not only how often characters with different identities appear, but also the roles, activities, and settings in which they are shown.

Our study extends the textbook-analysis literature by applying computational tools to both the text and images of 261 elementary-school textbooks used in public and religious private educational settings from 1980 to 2022. Our panel allows for longitudinal analyses across different educational settings and over more than four decades, rather than a cross-sectional examination within a single setting or point in time as prior studies have typically done. Our methods also make it possible to measure dimensions that are difficult to study manually at scale, such as skin color of pictured characters in images, a salient but often overlooked dimension of human categorization (Keith and Monroe, 2016). More broadly, we measure both presence and portrayal of people, topics, and values over time in a unified framework that incorporates visual as well as textual content. This approach enables us to replicate several qualitative patterns documented in prior work while also uncovering new ones. By pairing our findings with a conceptual framework of publisher competition under state standards and empirical tests of cross-state spillovers in textbook content following state-standards revisions, we move beyond descriptive content analysis to help explain the market and regulatory forces that shape curricular content and generate spillovers across states. Our methods also contribute to a growing social-science literature using computational tools to analyze text and images, including work on children's books, newspapers, yearbooks, and sports statistics (Adukia et al., 2022a,b, 2023; Ash et al., 2021; Voth and Yanagizawa-Drott, 2024; Kamel and Woo-Mora, 2023).

III Data Sources

Our sample includes three distinct collections of elementary school textbooks used in: (1) Texas public schools, (2) California public schools, and (3) religious private and home school settings. We digitize a novel data set of Reading, Science, and Social Studies textbooks written for the third and fifth grades. Our data set comprises over 88 thousand pages originating from 261 different textbooks, amounting to a corpus of approximately 16 million words and 60 thousand detected faces (Table B.1).[4]

[4]Note that our sample is not perfectly balanced across subjects, grades, and year in use across collections. Because of this, when comparing average measures across collections, we make sure to first calculate the measure for each collection, subject, grade, and year in use and then average these measures within a collection. This approach allows us to attribute observed differences across collections to actual differences in content, rather than to shifts in the composition of the sample over time.

III.A Texas State-Adopted Textbooks

We study Texas public-school textbooks for four primary reasons. First, Texas adopts textbooks at the state level, allowing us to construct a panel of textbooks to which children attending public schools were likely to have been exposed over time. Second, Texas serves one of the largest student populations in the U.S., so the messages in these books are likely to reach a broad audience. Third, because they are such a large share of the textbook publishing market, Texas has historically influenced the content of textbooks used across the U.S. Finally, Texas is often portrayed as reflecting conservative values, both politically and in the media, which provides a view into how one side of the ideological spectrum presents education in the public sphere.

The Texas State Board of Education adopts instructional materials and maintains an annual state-adopted list for K–12 grades. Before 2011, funding for textbooks was tied to this list, so the adopted list is a reasonable proxy for books in use. In 2011, Texas shifted to a more flexible funding system, allowing districts to purchase either state-adopted or other TEKS-aligned materials. Accordingly, we digitized a convenience sample of about 50% of textbooks on the annual state-adopted list from 1985–2010, and for 2011–2021, we digitized the top two textbooks purchased each year for each relevant grade and subject.[5]

III.B California State-Adopted Textbooks

Similar to Texas, California adopts elementary-school textbooks at the state level and exerts meaningful influence on the textbook publishing market due to the size of its student population. California is typically associated with more progressive policies and is often cast as the political or substantive counterpart to Texas. Including these two political extremes in our sample allows us to bound what are often regarded as the most conservative and the most liberal implementations of traditional public-school curricula.[6]

California curates a list of state-adopted textbooks for grades K-8 and updates it every couple of years for each subject.[7] We digitize the universe of books in the relevant grades and subjects that were adopted in California for all years in which the list of adopted materials is available. This includes adopted textbooks beginning in 1999 for Social Studies, 2000 for Science, and 2002 for Reading and ending with the books that were in use as of 2022.

[5]Reliable textbook purchasing data were not collected before 2011.

[6]Indeed, because they are widely perceived as occupying opposite ends of an ideological continuum, California and Texas are frequently studied in tandem or comparison (c.f., Martinez and Todorov 2023).

[7]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.

III.C Textbooks by Religious (Conservative Christian) Publishers

A third corpus that we examine comprises religious-school textbooks published by three ideologically conservative Christian textbook companies: Abeka, Bob Jones University Press (BJU Press), and Accelerated Christian Education (ACE). We refer to this set of books as our Religious collection. These textbooks are among the most commonly used by homeschooling families (Kunzman and Gaither, 2020) and adopted by over a third of non-Catholic Christian schools receiving taxpayer funding (Klein, 2017). We focus on textbooks published by these three Christian publishers both because they are widely adopted and because public funds support many schools that use these curricula.[8] We digitize the universe of textbooks published between 1980 and 2022 by these three publishers for the grades and subjects in our sample, including updated editions of the same textbooks.[9]

Google Trends search data provides another window into the importance of this collection of books (Figure A.1). While search interest in *homeschool* had been steadily declining over time, it spiked sharply when schools closed in March 2020. This is consistent with evidence that the number of parents choosing to homeschool their children increased during the COVID lockdowns (Musaddiq et al., 2022; Jamison et al., 2023). Alongside this surge, searches for *Abeka* also rose and began to track closely with those for *homeschool*. Although the increase in homeschooling in March 2020 was not primarily driven by religious motivations, many families turned to religious curricula because few comprehensive, turnkey non-religious options were readily available to families (Kaleem, 2016).

IV Data Extraction Methods

We leverage tools from the computational social sciences, including AI methods, which allow for a scalable, cost-effective, and systematic approach to content analysis. We treat each scanned textbook page as both (i) an image and (ii) a source of printed text. We show the process used to convert images into data in Figure 1a, and the process used to convert text into data in Figure 1b. Additional details for all methods including complete prompt specifications, pre-processing steps, and regression details are provided in Appendix C.

IV.A Images as Data

From images, we extract information about the characters and their contexts. This involves character detection and feature classification for attributes such as race, gender, and skin color, in addition to classification of their actions and settings.

[8] In 2016-17, approximately 4,240 students in Indiana received over $16 million in scholarships to attend schools that use the Abeka or BJU Press curricula (Klein, 2017).

[9] The ACE curriculum primarily consists of self-guided workbooks known as PACEs (Packets of Accelerated Christian Education). Therefore, our analysis includes only the books from their Reading curriculum.

IV.A.1 *Character Detection, Face Detection, and Feature Classification*

For our primary results, we extract data from images using a Large Multimodal Model (LMM), specifically Gemini 2.5 Pro Preview, which enables large-scale analysis of visual content, including the detection of pictured characters, their characteristics, setting, and actions. Importantly, unlike typical face detection models, this allows us to detect characters whose faces are not visible, as well as a larger number of non-humanoid characters. We extract two types of data from each textbook page containing an image: scene-level metadata that capture contextual information including setting descriptions and depicted activities and character-level data that contains demographic attributes including race, gender, and human versus non-human classification for each pictured character we detect.

For robustness, we supplement this character detection and feature classification with a Convolutional Neural Network (CNN)-based face classifier deployed on Google's Vertex AI, trained to detect faces in both photographs and illustrations (Szasz et al., 2022; Adukia et al., 2023), and predict the putative race and gender of each detected face using a CNN feature classification model trained on a manually curated data set of both illustrations and photographs (FDAI) (Szasz et al., 2022). In the sample of LMM-detected characters that can be matched to a CNN-detected face, the race and gender predictions of the LMM outperformed our CNN-based predictions; validation exercises are discussed in Appendix D.

IV.A.2 *Skin Color*

To measure skin color, we segment skin pixels in faces detected using our CNN-based face classifier, apply k-means clustering to partition the skin color, and construct a continuous measure of skin tint based on the L* value of a representative skin color in CIE L*a*b* space (0 = darkest; 100 = lightest), following Adukia et al. (2023).

IV.A.3 *Image Description Embeddings*

To compare visual portrayals at scale, we construct embeddings from the LMM-generated text describing each pictured character and the context they are depicted in. For each character, we concatenate the model's descriptions of actions, location, and what is happening into a single short document and estimate a 300-dimensional document embedding model (using Doc2Vec) on these character-documents. We use these embeddings in three ways: (i) we estimate embedding regressions that compare image-description embeddings across textbook collections; (ii) we identify the words with the highest associations with image-description embeddings containing characters of different demographics and examine the valence, arousal, and dominance of those words;[10] and (iii) we measure the similarity between

[10] A detailed description of the VAD lexicon we use in this analysis can be found in the Methods Appendix.

image embeddings containing characters of different demographics and words belonging to pre-specified semantic domains such as family, sports, home, and politics.

IV.B Text as Data

From text, we measure information related to topics, values, geography, history, and people. To do so, we begin by applying optical character recognition (OCR) to extract machine-readable text from the scanned pages.

IV.B.1 *Large Language Models*

To measure the topics and values present in the text, we employ a large language model (LLM), specifically OpenAI's GPT-4, to analyze individual pages of textbooks and identify relevant topics and values. LLMs are autoregressive models that extract and summarize content by leveraging patterns learned from vast training data to predict and categorize the presence of specific themes and values within the text; however, their outputs are not direct measurements of ground truth, but instead reflect probabilistic assessments based on model inference. We therefore validate these AI-generated measures in two ways: first by comparing model output to human annotations and second by benchmarking the GPT topic labels against topic-related keyword counts. These validation exercises along with additional prompt details are provided in Appendices C and D. Our use of an LLM is also supported by recent evidence that, for some free-form coding tasks, LLM-generated annotations can match or exceed the quality of human annotations (Ziems et al., 2024). By using an LLM for free-form annotation rather than classifying prespecified topic lists, we are better able to generate hypotheses about unanticipated topics and systematically test for differences in topic prevalence across collections (Ludwig and Mullainathan, 2024).

IV.B.2 *Token Counts and Named Entity Recognition*

We extract information about geography, history, and identities by analyzing words identified through Named Entity Recognition (NER) and token counts. NER is an NLP method used to identify and categorize named entities – such as locations, events, and people – within text. To identify geographies, we use the "GPE" and "LOC" tags; to identify historical events, we use the "EVENT" tag. To identify famous individuals, we merge entities with the Pantheon 2.0 dataset (Yu et al., 2016), which contains information on gender, occupation, sexuality, and birthplace, and we manually code their race.[11] We use remaining "PERSON" entities to measure gender of other named individuals by extracting first names

[11] In the text, we cannot reliably infer the race of fictional characters from names alone, so we restrict our analysis to famous individuals identified using Named Entity Recognition. We measure sexuality mentions from the LLM predictions and verify them using token counts of the words *lesbian, gay, bisexual, transgender, queer, intersex,* and *asexual.*

and estimating their gender probability using Social Security Administration name-frequency data. We use token counts to measure the prevalence of gendered terms.

IV.B.3 *Word, Page, and Textbook Embeddings*

Text embeddings represent words or documents as high-dimensional vectors that capture semantic relationships learned from context, allowing us to compare their similarity using measures such as cosine similarity. In our setting, we use text embeddings to characterize differences in semantic contexts across people, topics, and textbooks. We construct pronoun-instance à la carte (ALC) word embeddings using the `conText` framework, following Khodak et al. (2018) and Rodriguez et al. (2023). We use these embeddings to (i) identify top relative words by gender and their VAD scores and (ii) measure gender skew in similarity to pre-specified domains.[12] We also estimate document embeddings at the page and textbook level. Finally, we integrate GPT topic labels with page embeddings to construct topic embeddings by averaging over the normalized page embeddings for all pages that contain a given topic. We use embedding regressions to formally test for differences across collections (Appendix C).

V Results

In this section, we examine the content patterns across our textbook collections, beginning with the similarities between collections overall and over time.

V.A Overall Patterns in Textbooks

We analyze overall similarity of textbook content (in both text and images) using document embeddings which allow us to measure similarity in semantic content in both the text and the described settings and actions of pictured characters between textbooks using cosine similarity between embeddings. We begin with a descriptive view of the embedding space. Figures 2a and 2b plot book-level text and image embeddings (based on image descriptions) in two dimensions using t-SNE, with colors denoting collections and shapes denoting subjects.[13] Two patterns stand out. First, books cluster strongly by subject, suggesting that the embeddings recover meaningful differences in content. Second, within subject, California and Texas textbooks tend to occupy similar regions of the space, whereas the Religious collection is more clearly separated from both.

We formally test these visual impressions using normed coefficients from page- or image-level embedding regressions in the spirit of Rodriguez et al. (2023), which provide measures of the overall magnitude of semantic difference in embedding space between groups, aggregated across all embedding dimensions (Figures 2c and 2d). Larger normed coefficients indicate

[12] A detailed description of the VAD lexicon we use in this analysis can be found in the Methods Appendix.
[13] For visualization, we project the 300-dimensional embeddings into two dimensions using t-SNE.

greater semantic differences between groups in the text or the described settings and actions of pictured characters. These regressions show relatively large differences across textbook subjects in both text and image embedding space; differences between collections are smaller than differences between subjects and show that while California and Texas textbooks are not identical, they are substantially more similar to one another than either is to the Religious collection.[14] Figure A.3 reaches the same conclusion by calculating cosine similarities for textbook pairs from different collections that share the same subject, grade, and year in use. The histogram in Figure A.3a shows that some textbook pairs have near-perfect similarity scores, consistent with publishers reusing closely related content across state editions (see Section VI for a discussion of the textbook publishing process). Averaging similarities over time yields the same ranking in every decade: California and Texas textbooks are consistently more similar to each other than either is to the Religious collection (Figure A.3b). Over time, however, Texas Science textbooks have become less similar to California and more similar to Religious textbooks, while California and Texas Social Studies textbooks have grown more similar to each other.[15]

In the next sections, we further unpack these patterns to explore and understand similarities and differences by specific dimensions of content.

V.B Values

We first examine values in each collection. The top values are similar across collections, with the three most common values centered on learning. We see this overlap especially between California and Texas textbooks (Figure A.4). The main exception is *faith*, which shows up as a top value in the Religious collection; this is consistent with the large proportion of religious values found in the Religious collection which we discuss below.

Character Values. Character values are an important part of early education. Texas codifies character education in state law, requiring instruction in four broad domains: trustworthiness, responsibility, caring, and citizenship (Texas State Board of Education, 2020). In contrast, California provides non-binding guidance that supports six core values: trustworthiness, respect, responsibility, fairness, caring, and citizenship (California Department of Education, n.d.).[16] Despite the importance that public schools place on character values through their standards, the majority of parents who choose private or home schooling

[14] Differences between the coefficient estimating the difference between California and Texas and the coefficient estimating the difference between California and Religious are statistically significant with a p-value < 0.001 in the text embedding regressions and $p \leq 0.002$ in the image embedding regressions.

[15] One possible explanation is that California adopted the Next Generation Science Standards in 2013, as did many other U.S. states, while Texas did not. We return to this discussion in Section VI.C.1.

[16] These *Six Pillars of Character* align with national models like CHARACTER COUNTS! and mirror the four Texas pillars, though Texas bundles fairness and respect into citizenship.

options report "morals," "values," or "character" education as among the most important qualities when choosing a school (Catt and Rhinesmith, 2016). We next examine whether the Religious collection places greater emphasis on character values than public-school textbooks, and whether the proportion of character values in Texas and California differs in ways that reflect the relative strictness of their curriculum standards.[17]

Character values are prevalent in all collections (Figure 3a). Each collection exhibits a similar share of pages with character values in Science (17–18%) and Social Studies (40–48%) textbooks. Reading textbooks in the Religious collection contain a higher share of character values (71%) than those in California (59%) or Texas (55%). We find no statistically significant differences between Texas and California in the prevalence of character value categories outlined in Texas state standards and California's guidance, despite differences in whether these values are mandatory (Figure A.5a). Likewise, when we examine the most frequently mentioned character values across collections, we observe no meaningful differences between Texas and California. Although religious-school Reading textbooks contain a higher share of character values overall, some individual values appear more often in public-school textbooks (Figure A.5b); for example, creativity appears on a larger share of pages in public-school textbooks (11%–13%) than in the Religious collection (5%).

Religious Values. Religion is a frequently cited reason why some parents opt for private or home schooling (Catt and Rhinesmith, 2016; NCES, 2022; Bartholet, 2020). Given this and the fact that textbooks in the Religious collection are published by Christian publishers, we would expect these textbooks to contain a larger proportion of religious values with an emphasis on Christianity. We test the extent to which this is true by combining values related to religion such as *religious faith, religious freedom, religious beliefs, religious diversity, religiosity, religious tolerance*, and *religious devotion*. Using this group of values, we find that religious values appear substantially and significantly more often in textbooks from the Religious collection across all subjects (Figure 3b). In Reading textbooks, religious values appear on about 4.0% of pages in the Religious collection compared to 0.4% in California and 0.2% in Texas. In Science textbooks, where we may least expect to see reference to religion, religious values are present on roughly 2.1% of pages in the Religious collection and are essentially absent in both California and Texas textbooks. The largest differences appear

[17]We focus on the four character value pillars explicitly mentioned in the Texas standards and referenced in California guidance (we bundle respect and fairness into citizenship, following the Texas framework). Publishers in our Religious collection do not adhere to a predefined list of character traits but Abeka's K5 development materials include additional traits not specified by Texas or California, such as obedience, self-control, tolerance, and modesty. To incorporate these, we add two categories: an "Attitude" category for values related to disposition or temperament (e.g., cheerful, joyful) and an "Other" category for all remaining uncategorized traits. A full list of values and their groupings appears in Table B.2.

in Social Studies textbooks, where religious values occur on 15.9% of pages in the Religious collection compared to 5.6% in California and 2.9% in Texas. Together, these patterns show that while religious values appear occasionally in public-school textbooks, they are far more prevalent in textbooks from the Religious collection, particularly in Social Studies. In the next section, we show a similar prevalence of religious topics in our Religious collection.

V.C Topics

While the topics that are covered in textbooks are not explicit values, which topics are discussed, how often they are discussed, and the context in which they are discussed reflect the values of society. Here, we document topics emphasized in each of our collections.

We first show the overall top topics that are addressed by subject (Figure A.4). We see that in general, similar topics are discussed across collections, but there are some notable exceptions. In Reading textbooks, many of the top topics (e.g. *family, education, nature, friendship,* and *adventure*) are the same across all collections. However, only the Religious collection includes *prayer* and *faith* among its top topics. In Science textbooks, *astronomy, biology, physics, observation*, and *science* are among the most common topics across all three collections. In Social Studies textbooks, *geography, exploration, history* (including *American history*), and *agriculture* are among the most common topics. Notably, *religion* appears as a top topic in Social Studies textbooks in the Religious collection but not in public-school textbooks. Despite some differences, these overall patterns in the top topics indicate meaningful overlap in the topics most commonly taught to children.

High-Profile Topics. While the most common topics comprise a majority of what students learn, much of the public debate and discourse related to curricula centers on particularly high-profile topics such as those related to evolution and creationism, slavery, racism, sexuality, and climate change. We first examine the prevalence of these high-profile topics both overall and over time (Figures 4a and A.6). We see that California and Texas textbooks are much more likely to address topics related to climate change than textbooks in our Religious collection. In contrast, only textbooks by religious publishers discuss creationism and are more likely to address topics related to evolution (typically as a way of contrasting it to creationism). California textbooks are the most likely to include slavery-related topics overall, though patterns are mixed over time; whereas the Religious collection is more likely to include topics related to the Civil War, and increasingly so over time. Both California and Texas discuss immigration more often than religious-school textbooks, though in Texas textbooks we see an increase in the proportion of pages that discuss immigration over our sample period, and a decrease in California. Racism and LGBTQIA+ related topics are rarely, if ever, mentioned.

Just as important as *whether* a topic is discussed is *how* it is discussed. By integrating page-level data on the presence of specific topics with our page-level embeddings, we can examine whether collections discuss the same topic in different semantic contexts. For each topic, we restrict the sample to pages classified as discussing that topic and estimate embedding regressions on the corresponding page embeddings. Figure 4b reports the resulting normed coefficients which measure the magnitude of systematic separation in embedding space between collections and indicate whether pages discussing the same topic differ semantically across collections. To provide suggestive evidence as to which particular words or themes may be driving any differences between the Religious collection and the public-school collections, figures in Appendix E show a descriptive visualization of the words with the highest similarity between each collection-specific topic embedding using t-SNE to project high-dimensional embeddings into two dimensions.

Figure 4b shows that the overall similarities across Texas and California we observe in Figure 2c also hold for most high-profile topics. None of the differences between Texas and California in these regressions differ significantly at the 5% level. By contrast, the Religious collection differs significantly from the public-school collections for all high-profile topics, suggesting that public-school textbooks in Texas and California discuss high-profile topics more similarly to each other than either does to religious-school textbooks. Evolution is the main exception, which has descriptively distinct patterns of discussion in all three collections (albeit insignificant between Texas and California, likely because we are underpowered to detect differences between evolution topics in public school textbooks). Unlike many of our other topics which appear on hundreds to thousands of pages, the topic of evolution only appears on fewer than 100 pages in each collection (Figure 4a). We therefore manually examined these pages and extracted representative passages to illustrate qualitative similarities and differences in the treatment of evolution across all three collections.

Pages in the religious-school textbooks that contain the topic of evolution fall under three main categories: (1) concepts related to evolution are discussed (such as natural selection), but evolution is not named or mentioned; (2) evolution is explicitly addressed and portrayed as false; and (3) evolution is discussed in the context of Social Darwinism: "The idea that only the fittest people can survive in society is known as Social Darwinism. Because people are naturally sinful and selfish this belief was dangerous." Examples of (2) are common. One such passage from a religious school 5th grade Science textbook reads:

> "Evolutionists interpret the fact that dogs and cats are similar as being an evidence for evolution. Creationists on the other hand explain that God created living things with similar structures because they would need the structures to do similar jobs. Cats and dogs for example have similar teeth because they enjoy similar eating habits."

In our public-school textbooks, pages predicted to contain the topic of evolution discuss a wide variety of concepts including Neanderthals, fossils, natural selection, prehistoric animals, the Ice Age, and behavioral or structural adaptations. One difference between Texas and California textbooks is that in the Texas collection, even though there are discussions of natural selection, adaptation, and the Earth being millions of years old, passages that we identify as discussing evolution do not explicitly mention the Big Bang Theory or Charles Darwin. An example from a Texas 3rd grade Reading textbook reads:

> "If eohippus does not look like a horse of today, how do we know that eohippus was a horse? How do we know that eohippus was not a member of the dog family? We know that eohippus is a member of the horse family because we have found skeletons of other horses that lived long ago. When we put the skeletons in a row we can see small changes. Picture 2 shows just how horses changed over millions of years. It shows that eohippus did not change into a dog. It changed into the horses of today."

The California collection, however, includes explicit mentions of the Big Bang Theory and Charles Darwin, in addition to a passage in a fifth grade Social Studies textbook called "Stories of How the World Began," which recounts creation stories from different cultures.

Our takeaways from this qualitative analysis are also reflected in Figure E.2 which shows that the evolution topic embedding in the Religious collection is often associated with words such as *creation, theories,* and *Genesis*, but this is not the case in public-school textbooks.

Overall, in the grades and subjects we study, many topics central to curriculum debates appear rarely. This suggests that parents, teachers, and policymakers care not only about the intensive margin of textbook content, but also about the extensive margin of whether contentious topics appear at all. Given the importance of this extensive margin, we next examine other low-frequency topics exclusive to a particular collection.

Exclusive Topics. Although we document many overall similarities across public-school textbooks, there are informative dimensions on which textbooks from Texas and California differ. In Figure A.8, we show the topics that are exclusively identified in specific collections and not in other collections. We see that only California textbooks are identified as having topics related to California history and geography, including discussion related to *California national parks* and the *Kumeyaay tribe.*[18] California is also the only collection that mentions the Next Generation Science Standards, which were adopted in California but not in Texas. Only Texas textbooks include topics related to the practice and knowledge of *Texas Essential Knowledge and Skills (TEKS) state standards*, some places in Texas such as *Fort Worth*, in addition to topics such as *prewriting, correctness, the Loch Ness Monster, the Berlin Airlift, the Special Olympics,* and *palindromes.* Only the religious-school textbooks have topics

[18] An Indigenous people from what is now Southern California.

related to *spiritual growth, Christian values, evangelism,* and *creationism*.

Religious Topics. We further explore the extent to which textbooks in the Religious collection discuss religious topics. Religious topics appear on a much larger share of Social Studies textbook pages in the Religious collection (12%, compared with 4.4% in California and 2.8% in Texas) (Figure A.7), consistent with the patterns for religious values discussed in Section V.B. Moreover, when a specific religion is mentioned, it is overwhelmingly Christianity. These differences show that religion appears more often in the Religious collection, but they do not indicate whether religious topics are discussed in different semantic contexts across collections. We examine this question using topic-specific embedding regressions. Figure 4b shows that pages discussing religious topics in the Religious collection show significant semantic separation from those discussing religious topics in both California and Texas, while the difference between pages in California and Texas is smaller and statistically insignificant. Figure E.3 suggests that in the Religious collection, the religious topic embedding is more associated with religion-specific words such as *savior, pray, sins,* and *heaven*. In contrast, the religious topic embeddings in the California and Texas collections are more associated with words related to other religions and the U.S. founding era, including *Quaker, Islam, pilgrims, colony*, and *settlers*.

Geography and History. Political and media discussions of curricula are commonly centered around which historical events and figures should be highlighted and how they should be portrayed, in addition to which geographies should be prioritized (Friedman and Tager, 2022). Here, we examine which receive the most attention in textbooks.

Figure 5a shows the most commonly mentioned historical events. Most of these events center on U.S. history; the remainder are major wars involving the U.S., namely World War I and World War II. Although the overall set of events is similar across collections, the frequency with which they are discussed varies. In particular, mentions of the Civil War increase over time in the California and Religious collections but decline sharply in Texas textbooks over the past two decades.

Geography is one of the most prominent Social Studies topics across all collections (Figure A.4). Yet the geographic scope of U.S. textbook content is relatively narrow. Students are exposed primarily to references to the Earth in general and to the U.S. in particular, with comparatively little attention to other parts of the world (Figure A.12a). State-specific differences do appear at the margin: California and Texas textbooks are more likely to mention California and Texas geography, respectively. Outside the U.S., the most frequently mentioned places are in outer space or Europe (Figure 5b).

In addition to geography and historical events, we examine representation of historical figures and other prominent individuals. Historical and scientific narratives are often organized around individuals, and textbooks therefore commonly use notable figures to convey information, illustrate leadership, and present role models whom children may be encouraged to emulate. Table B.3 reports each collection's most frequently mentioned famous people. George Washington, Thomas Jefferson, and Benjamin Franklin – all Founding Fathers – are among the most commonly included famous individuals across all collections. Their prominent inclusion is consistent with a survey conducted by the Heritage Foundation which found that 59.5% of teachers and 56% of parents surveyed considered it important for schools to include content on the Constitution and the U.S. founding era (Sailor et al., 2021).

V.D People: Representation of Birthplace, Race, Skin Color, Sexuality, Gender

The people who are highlighted in the narratives and history presented to children reflect societal views about whose stories are told and the roles different people are expected to occupy in society. In this section, we discuss the representation of identity dimensions including birthplace, race, skin color, sexuality, and gender of people included in these textbooks.

Birthplace. Examining the birthplaces of famous individuals (Figures 5c and A.12b), we see broadly similar patterns across public-school textbooks overall and over time; most mentioned individuals were born in North America, followed by Europe. The majority are male, and the females who do appear are predominantly from North America. The Religious collection exhibits somewhat greater geographic diversity with near-parity representation of individuals born in North America and Europe. Across all collections, however, individuals from Africa, Asia, the Middle East, and South America are scarcely represented.

Race. We measure racial representation separately for famous individuals mentioned in text and in the pictured characters in images. We find that across collections and subjects, the vast majority of famous individuals mentioned are White, accounting for 83% to 92% of famous individuals (Table B.1). This pattern is also evident in images and persists over time (Figure A.10). Relative to public-school textbooks, textbooks in the Religious collection include more White and fewer Black famous individuals and pictured characters. Across all collections, Asian, Latine, and Indigenous individuals are rarely represented in either text or images. California public-school textbooks contain the highest shares of pictured Asian and Latine characters, though these remain modest at about 10–14%, compared with 7–11% in Texas and 6% in the Religious collection (Table B.1).

Skin Color. Skin color is one of the first ways that children learn to distinguish humans, and is distinct from (though highly related to) measures of putative race (Kessen and Weiskopf, 1976; Katz and Kofkin, 1997; Kinzler and Spelke, 2011). In Figure A.10c, we

present the distribution of detected faces according to skin tint, measured on a scale from 0 to 100 representing how dark or light their skin color appears. We plot these distributions over a scatterplot showing the skin color of every face in our sample. Patterns are relatively stable over time, with a slight trend toward increased lightening of character skin color in the Religious collection over time (Figure A.10d).

Pictured characters in Texas and California textbooks are more likely to be depicted with darker skin colors on average than pictured characters in the Religious collection. This pattern holds even when conditioning on race (Table 1). We also see that older children (children in grade 5 versus grade 3) are more likely to be shown characters with darker skin.

Sexuality. To examine content related to sexuality, we look at both the presence of individuals who openly identified as lesbian, gay, or bisexual (LGB) as well as the presence of pages that discuss LGBTQIA+-related topics. We found fewer than 20 famous individuals who identified as LGB mentioned in each of our public-school textbook collections and fewer than 10 individuals mentioned within our Religious collection of textbooks. In all three collections, the total mentions of LGB famous individuals comprise less than 1% of all mentions. Note that only Harvey Milk was mentioned as identifying as gay and Billie Jean King was mentioned as identifying as lesbian, while the other famous individuals (such as Jane Addams) were not discussed in the context of their sexuality.

There were only two pages in our sample that discussed LGBTQIA+ topics. The first is in a 5th grade Social Studies religious-school textbook in a section entitled "Gay Rights":

> "These Americans did not support gay marriage. They thought that it would harm society... 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."

The second is from a 3rd grade Social Studies textbook in our California collection:

> "Many other Californians worked hard to make changes in their communities... Harvey Milk was a political leader and a gay activist in the 1970s... 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... Bono was born female, but always felt male... Bono and other activists fight against discrimination and stereotypes of their community... Billie Jean King is one of the greatest female athletes of all time... She believes in equal pay for female athletes. She is also a lesbian."

The passage in the Religious collection highlights a potential tension between LGBTQIA+ rights and religious freedoms. Because this is the only LGBTQIA+-related passage in the

Religious collection, children exposed to this curriculum encounter LGBTQIA+ people solely as figures framed as threatening the freedoms of others. By contrast, California included passages defining what it means to be gay, lesbian, bisexual, and/or transgender and highlighting LGBTQIA+ activists. We identified no LGBTQIA+-related passages in Texas.

Gender. In looking at gender presence in textbooks,[19] we see the public-school collections have higher female representation in both text and images than the Religious collection (Table 1 and Figure A.13). This pattern holds when examining gender representation within racial groups (Figure A.11). Across all collections, however, females are underrepresented relative to their share of the population. These patterns are consistent over time. In each collection, at most one of the top ten famous people and top ten first names is female; Abigail Adams appears in California and Harriet Tubman appears in Texas (Table B.3). The most common first name in each collection is *John*.[20] This suggests that children are more likely to read about males rather than females.

One explanation for lower female representation is that women have been historically excluded from many of the fields and roles emphasized in Social Studies and Science textbooks, such as the U.S. presidency. Consistent with this, Science and Social Studies textbooks contain fewer textual references to females than Reading textbooks across all collections, whereas Reading textbooks show near parity between females and males. This pattern may reflect greater perceived flexibility in assigning character gender in Reading textbooks, in contrast to history-oriented texts. However, we do not see similar textual gender parity in religious-school Reading textbooks. The highest female representation in the religious-school textbooks is in Reading, where female-related words still only constitute 41% of all gendered words (Figure A.13c). However, this text-based gender parity in public-school Reading textbooks does not extend to images: disaggregating pictured characters by subject reveals no gender parity in any collection (Figure A.13d).

Next, we assess the contexts in which females and males are portrayed in text and images using two approaches: (1) identifying the top relative words most closely associated with females versus males, and vice versa, and (2) examining the gender representation within specific domains, such as family or politics. Table B.4 shows the top relative words most closely associated with females vs. males and, separately, with males vs. females in both the text and image descriptions for each collection, along with the average VAD scores for the top words. We see that in both text and images, females are more associated with words related to the home and family such as *mother, wife,* and *gardening*; males are more

[19]Note that we only look at binary definitions of gender.

[20]The origins and genders of generic names can implicitly signal who represents the societal default.

associated with words relating to the military conflict such as *army, conquest,* and *troops*. The sentiment of the top relative words for females is more or equally positive, but less active and powerful on average than the top relative male words. This pattern remains when we extend our analyses to include the top 50 relative words by gender (Figures 6a and 6b).

Figures 6c and 6d show gender skew across six domains in text and images: *family, home, performance arts, politics, sports,* and *tools*. Across all collections, *family* and *home* are consistently skewed toward females, while *politics* and *tools* are generally skewed toward males. These patterns appear in both text and images, though the magnitude of the skew differs across collections and is generally larger in images. The *sports* domain is more male-skewed in the Religious collection of textbooks than in the public-school textbooks.

VI Market and Regulatory Forces Driving Textbook Content

This section provides a conceptual framework connecting textbook-market structure to curriculum content. The goal is to rationalize two empirical regularities in our data: (1) public-school textbooks adopted in California and Texas are relatively similar to each other on a variety of dimensions, and (2) textbooks from religious publishers systematically differ along specific dimensions. Before presenting our conceptual framework, we provide some background on the setting in which these textbooks are published and adopted.

VI.A Institutional Background

Textbook adoption and purchasing in the U.S. are decentralized at the school-district level, but the set of materials districts may choose from depends on whether they operate in an adoption state or an open-adoption state. Adoption states, including California and Texas for the grades and subjects in our sample, maintain state-approved lists and run formal review cycles to screen instructional materials for alignment with state standards. In these states, districts select textbooks from the approved list or seek additional approval to use non-listed textbooks (Doan and Kaufman, 2024). In open-adoption states such as Illinois and Arizona, local education agencies select textbooks without being constrained by a state-adopted list, but their selections are still subject to state curriculum standards.

Large textbook publishers competing in the national public-school textbook market, such as McGraw Hill or Pearson, start by creating a baseline or "national" edition of a textbook and only customize this baseline textbook into a state-specific version when expected sales justify the fixed costs of customization (Doan and Kaufman, 2024).[21] In adoption states, publishers must submit materials, along with documentation of standards alignment,

[21] Industry reports similarly note: "Generally, publishers say it is uneconomical for them to create separate versions of a textbook for the 'big three' states and for other states. As a result, the Texas, California, or Florida version of a book usually is the version published nationally (House Research Organization, 2002)."

during the state's formal review cycle in order to have their textbook considered for adoption in that state. States then screen submissions for compliance, provide feedback, and may allow publishers to revise or withdraw any materials that fail compliance screens. After adoption, publishers have limited scope to compete for district sales on price, because state procurement rules often constrain pricing. California, for example, imposes "lowest price" requirements on adopted materials and restricts post-adoption price increases (California Legislature, 2026). As a result, conditional on passing state compliance screens, publishers primarily differentiate on textbook content rather than on price when competing for district sales within a state.

Although states screen materials, the relevant public-school market is not purely state-specific: large publishers market to districts in multiple states. Consistent with this, Figure A.14 shows substantial cross-state overlap in publishers of textbooks adopted by Texas and California: in any given year, 60–80% of Texas adoptions and 90–100% of California adoptions come from publishers that were adopted in both states at any point in the sample.[22] While it is uncommon to have adoption of identical textbooks (as shown in Figure 2), it indicates that publishers are competing for district purchases in both states' textbook markets. This cross-state overlap in supply matters because publishers are not responding to a single state-level preference, but to heterogeneous district preferences that vary within states and overlap across them. Recent events illustrate the meaningful heterogeneity in preferences within states. In Temecula, California, the school board rejected a state-approved elementary social-studies curriculum aligned with the FAIR Education Act due to objections tied to Harvey Milk, before ultimately adopting revised materials under state pressure. In Coppell, Texas, the attorney general sued the district after an undercover video appeared to show an administrator describing ways to work around state restrictions on critical race theory and unapproved science materials (Newsom and Bonta, 2023; Office of the Attorney General of Texas, 2025). These events illustrate that district-level preferences are heterogeneous and do not always align with state standards.

The religious publishers in our sample operate in a distinct market. They explicitly target religious private schools and homeschooling families rather than public-school districts. Since their materials are not subject to the same state regulatory standards, they can sustain curricular positions that are not designed for mainstream public-school approval even when those positions do not align with state standards. This lack of alignment with mainstream learning trajectories extends beyond the K–12 context. The University of California system,

[22]Note that Texas adopts a larger set of textbooks than California, so the lower proportion of overlapping adoptions in Texas could be somewhat mechanical.

for example, deemed some coursework from Abeka and BJU Press inadequate preparation for college. This decision faced legal challenges, and in *Association of Christian Schools International v. Stearns (2010)*, the Ninth Circuit upheld UC's decision, noting that some course content was "not consistent with the empirical historical knowledge generally accepted in the collegiate community." The court noted, however, that religious schools and homeschooling families remained free to teach this coursework even when UC denied approval. This illustrates that curricula designed for religious schools need not be built to satisfy the same screening criteria that shape mainstream public-school materials.

These institutional features matter for how competition shapes textbook content. Mainstream publishers serve a national public-school market in which adoption standards constrain the feasible set of textbooks, while religious publishers serve a more segmented market.

VI.B Conceptual Framework

In this section, we develop a conceptual framework that outlines the market and regulatory forces which jointly shape the content of textbooks. Our framework is intentionally reduced-form: we do not model a publisher's decision to enter the national public-school market or to specialize in a niche market, as is the case for the Religious textbook publishers in our sample, since such decisions may reflect ideological motives in addition to profit maximization. Instead, we take market participation as given and distinguish between two markets. In the public-school market, publishers compete for textbook sales in schools across multiple states while also satisfying state standards. In the religious private school and homeschool market, publishers serve a more distinct set of purchasers and are not constrained by the same state standards. The goal of this framework is to organize the empirical patterns and generate testable predictions rather than to provide a full structural model.

Suppose purchasers are schools (or school districts) and publishers are firms. Let a textbook's overall content be summarized by a one-dimensional position c (for example, political slant). Schools differ in their preferred content position c and choose the textbook closest to their preferred position. Prices are fixed and publishers compete on content position. Under standard assumptions in spatial-competition models, publishers have incentives to locate near the preferences of the median purchaser rather than to supply highly differentiated products (Hotelling, 1929; Downs, 1957). We use this as a benchmark intuition rather than as a description of the full textbook market. The key implication is that, absent additional frictions or segmentation, competition for a broad public-school market pushes publishers toward minimum differentiation. However, when firms serve distinct consumer segments with different preferences and fewer shared constraints, standard models of product differentiation imply that equilibrium products can diverge rather than converge (Shaked and Sutton,

1982; Anderson et al., 1992; Fournier et al., 2020). In our context, this segmented-market logic implies that religious publishers, who serve religious private schools and homeschooling families with distinct preferences and are not constrained by state adoption rules, will differentiate content to meet the demands of their target market.

In principle, the same logic could also generate substantial heterogeneity within the public-school market, since public schools are not homogeneous and some districts may prefer content that departs materially from the national mainstream. However, one key additional force affecting the public-school market is that their textbooks must satisfy state standards in order to be adopted and sold in major state markets. State standards therefore restrict the feasible set of content choices. As a result, state standards – especially when combined with the fixed costs of customization discussed below – can push equilibrium content toward less differentiation than perceived underlying school preferences alone would imply.

To make this idea more concrete, we move from a one-dimensional representation of content c to representing content as a bundle of topics. Let R_g denote the set of topics required by governing body (in our case, a state) $g \in G$, and let B_g denote the set of topics banned or effectively disallowed by state g. A publisher chooses a baseline topic bundle A for a national edition and may then customize this baseline to produce a state-specific edition A_g. To be feasible in state g, a state-specific edition must satisfy:

$$R_g \subseteq A_g \qquad \text{and} \qquad A_g \cap B_g = \varnothing$$

Let customization carry a fixed cost $F > 0$ per state-specific edition. When F is large relative to the gains from customization, publishers economize by relying more heavily on a common baseline text. If a topic has high enough aggregate demand across states, or if including it once in the baseline edition is cheaper than repeatedly adding it through state-specific customizations, publishers have incentives to incorporate that topic directly into the national baseline. In the limiting case in which customization is sufficiently costly, the baseline bundle will contain the "safe union" of required topics:

$$A \supseteq \left(\cup_{g \in G} R_g\right) \setminus \left(\cup_{g \in G} B_g\right) \tag{1}$$

i.e., A contains topics that are required in at least one state and not banned elsewhere, making them especially suitable for inclusion in a national baseline edition. This "safe union" should be interpreted as a reduced-form implication of high customization costs and state standards.

This logic generates two predictions. First, textbook topics *required* in one state and *not banned* in other states can appear in *all* states' textbook editions even if only one state's standards explicitly require them (positive spillovers in topic coverage). Second, topics banned

by a sufficiently influential state become unattractive for the national baseline textbook, potentially reducing coverage even in states that do not ban them (a "chilling" effect on baseline content). Overall, our conceptual framework predicts that state standards, combined with high customization costs, can generate cross-state spillovers in textbook content. The absence of state standards, combined with differentiated preferences, will enable religious publishers to systematically differentiate their textbook content along dimensions that better align with the demographics, preferences, and beliefs of their target markets.

VI.C Empirical Evidence

The framework's predictions align with the main patterns documented in Section V: (i) Texas and California public-school textbooks are more similar than narratives of political polarization would suggest, while (ii) textbooks in the Religious collection differ systematically from public-school textbooks. We conduct two empirical checks on this framework. First, we test whether changes in one state's content standards spill over into textbooks adopted in the other state and discuss suggestive evidence of a related chilling mechanism for politically contentious topics. Second, we examine whether differences between public- and religious-school textbooks correspond to differences in beliefs and demographics of their target populations. These descriptive exercises help situate the observed content patterns.

VI.C.1 Spillovers in Curriculum Standards

In this section, we examine whether discrete changes in state content standards are reflected in textbooks and whether those changes spill over across states. Over our sample period, we observe multiple instances of a state altering curricular standards for all grades in a given subject.[23] For each subject, we restrict the sample to the longest period in which only one state changes its content standards and both states subsequently adopt new textbooks. For each subject and grade, we then construct a "new-standards" dictionary from the text of the revised standards. This dictionary consists of words that appear in the updated standards for the state changing its curriculum but do not appear in the prior standards of either state. We find approximately 250 new-standards words in our Social Studies standards change (such as *abolitionist, capitalism, pollution, Tuskegee,* and *urban*), 550 for Reading (such as *autobiography, brainstorm, monolingual, palindromes,* and *synthesize*), and 800 for Science (such as *flowering, diet, digital, reservoirs,* and *visual*).

We measure, for each textbook-year observation, the share of words belonging to this dictionary and track how this share evolves over time by state. Changes in the frequency of these new words by state and year are shown in Figure A.15. The dashed vertical line

[23] Changes in curricular standards occur less frequently than changes in textbook adoptions. In California, for example, we observe no changes in state standards, but we observe changes in textbook adoption.

marks the standards-change year and the solid vertical lines mark each state's subsequent textbook adoption cycle in that subject. The resulting patterns are consistent with the spillover mechanism predicted by Equation (1). In the state that revises its standards, the share of new-standards words rises sharply after the next adoption cycle, indicating a direct effect. In some cases, the state that does not revise its standards also exhibits an increase in new-standards words after its own subsequent adoption cycle, suggesting a spillover effect despite no change in its standards. To more precisely estimate the effects implied by these descriptive patterns, we estimate the following textbook-year level regression:

$$y_{tgjiy} = \beta^{D} DirectPost_{gjy} + \beta^{S} SpilloverPost_{gjy} + \alpha_{gji} + \varepsilon_{tgjiy} \tag{2}$$

where y_{tgjiy} is the share of words from the new-standards dictionary in textbook t from state g, subject j, grade i, in year y. $DirectPost_{gjy}$ is an indicator variable equal to one in years following a change in textbook adoption for subject j in the state that also had a change in curriculum standards. $SpilloverPost_{gjy}$ is an indicator variable equal to one in years following a change in textbook adoption for subject j in the state that had no change in curriculum standards. The omitted category consists of observations before either state adopts new textbooks, during which time standards change for one state. We include state-by-subject-by-grade fixed effects (α_{gji}) and cluster standard errors at the textbook level.

In the pooled specification, we find a direct increase of 16 new-standards words per 1,000 textbook words and a spillover effect of 2.3 new-standards words per 1,000 words. Disaggregating by subject, we find large, precisely estimated direct effects and noisier but generally positive spillovers for Reading and Social Studies (Table 2). In contrast, we do not detect spillovers between Texas and California Science textbooks. We hypothesize this reflects divergent Science standards. In 2013, California, along with many other states, adopted the Next Generation Science Standards, whereas Texas did not.[24] Figure A.8 illustrates this: "Next Generation Science Standards" appears as a topic unique to California and is absent from both the Texas and Religious collections. This divergence in standards likely contributes to the lack of evidence for spillover effects in Science textbooks and helps explain why Texas and California science textbooks have become less similar over time (Figure A.3).

Figure A.16 shows two robustness exercises: a placebo benchmark for the estimated effects and a leave-one-out test of sensitivity to individual words in the new-standards dictionaries. In the placebo exercise, we replace the true dictionaries with randomly drawn word lists from the standards-changing state's standards vocabulary outside the focal subject. The actual direct-effect estimates are meaningfully larger than the placebo estimates across the

[24]Since 2013, at least 39 states and the District of Columbia have adopted the Next Generation Science Standards (NGSS) or similar standards (NGSS Lead States, 2018).

pooled and subject-specific models, suggesting that the direct effects are not explained by vocabulary words present in standards for other subjects. The spillover estimates are also above the placebo benchmark in the pooled model and in Reading and Social Studies. The Science spillover estimate is less distinct from the placebo distribution, consistent with the interpretation that cross-state spillovers are weaker in Science. In the leave-one-out exercise, we drop one new-standards word at a time, recount the remaining dictionary words in textbooks, and re-estimate the model. The resulting estimates remain tightly centered around the full-dictionary estimates, indicating that results are not driven by any single word.

These analyses provide empirical evidence that changes in state standards translate into textbook content primarily at adoption-cycle boundaries, and that revisions in one adoption state can spill over into the other state's textbooks through publishers' incentives to update a common national baseline and then make only limited state-specific deviations ("minimum customization"). This spillover mechanism helps explain why California and Texas editions often move in parallel and why we observe substantial cross-state convergence in Section V.

VI.C.2 Chilling Effects in Curriculum

Identifying chilling effects in curriculum is more difficult than identifying spillover effects because topics are not always formally banned, even if they are politically contested or avoided in review/adoption processes. LGBTQIA+ representation provides a useful case study. In Texas, state law governing certain HIV/AIDS educational materials for minors requires those materials to state that "homosexual conduct is not an acceptable lifestyle and is a criminal offense," reflecting a policy environment restrictive toward LGBTQIA+ inclusion.[25] Consistent with this, we identify no passages in Texas state-adopted textbooks that discuss LGBTQIA+ topics, rights, or individuals. In California, we also found no similar passages before 2011. However, the 2011 California FAIR Education Act amended Education Code §51204.5 to require Social Studies instruction on the roles and contributions of lesbian, gay, bisexual, and transgender Americans.[26] In our corpus, the first substantive California discussion of LGBTQIA+ history appears only after this mandate, in a section of a third grade Social Studies textbook titled "California Heroes," discussed in Section V.D.[27] This pattern is consistent with the chilling effect suggested by our conceptual framework: even

[25]This language was added in 1991 and remains on the books even though the Supreme Court's 2003 decision in *Lawrence v. Texas* rendered the underlying criminal offense unconstitutional.

[26]The 2011 California FAIR Education Act also included persons with disabilities, other men, women, Native Americans, African Americans, Mexican Americans, Asian Americans, Pacific Islanders, European Americans, and other ethnic and cultural groups.

[27]This passage also contained Californians who supported other communities required to be discussed by the FAIR act including Charlotta Bass, an African American civil-rights activist; Rupert and Jeannette Costo, who founded a program at the University of California to study American Indians; and Annette Funicello, who raised awareness for people with disabilities.

though discussion of LGBTQIA+ history was not banned in California, it was not included in history textbooks until explicitly required because it was politically contested or disfavored in other states. Once LGBTQIA+ history was a required topic in California, publishers wrote a section that was clearly tailored to California and unlikely to be incorporated into the national baseline textbook.

VI.C.3 Product Differentiation by Religious Publishers

If parents select schools partly to reinforce beliefs, publishers may be incentivized to supply curricular materials that are consistent with the values of the populations they serve (Bisin and Verdier, 2001; Tabellini, 2008; Wheaton, 2020). Consistent with this idea, our framework in Section VI suggests that textbooks from religious publishers will differ from public-school textbooks along dimensions salient to consumers in religious private schools and homeschooling families. In this section, we present empirical evidence showing that several content differences between the religious and public-school textbook collections correspond with differences in the demographics and beliefs of the populations these textbooks target.

Demographics and Racial Representation. The religious-school textbooks contain a higher share of pictured characters and references to famous individuals who are White, and depict faces with lighter skin colors on average (Figure A.10). These patterns are consistent with differences in student demographics across school types. Figure 7 shows that students in religious private schools and homeschool settings are substantially more likely to be White than students in California and Texas public schools (e.g., the student population in religious private schools and among homeschoolers is approximately 70% White compared to 21% in California public schools and 26% in Texas public schools).[28] This demographic gap is large enough that even a publisher with no intrinsic preference over representation may rationally tailor examples and imagery toward the modal student audience in its target market.

Beliefs and Values. If religious publishers are tailoring materials to align with consumer values and demographics, we would expect to observe differences not only in the representation of individuals but also the emphasized norms and beliefs. To probe this mechanism, we use restricted geocoded General Social Survey (GSS) data to compare the beliefs of (i) religious respondents nationwide,[29] (ii) California respondents, and (iii) Texas respondents.[30]

Figure 8 shows significant differences in views on evolution and prayer in schools: our set

[28]Because data on the homeschooling population are inconsistently collected, current estimates likely understate the actual number of homeschooling students.

[29]We define "religious" as attending religious services at least once a week. Although many respondents who do not meet this criterion still consider themselves religious, our goal is to identify a group for whom religion is likely a central organizing identity.

[30]Note that the California and Texas samples include religious respondents (those from (i)) in those states.

of religious respondents are more likely to reject evolution and to oppose school-prayer bans than respondents in California and Texas. These belief gaps mirror the content differences we document in the Religious collection, where religious topics and values are much more prevalent (Figures 3 and A.7) and prayer is among the most prominent topics in reading materials (Figure A.4). This is consistent with alternative survey data that show that 84% of parents with children in private schools listed religious environment/instruction as a top desired school quality, relative to about 64% who listed better academics (Catt and Rhinesmith, 2016) and 59% of homeschooling parents listed religious instruction as a reason for homeschooling (relative to 73% who listed dissatisfaction with the academic instruction at other schools) (NCES, 2022). These same survey sources indicate that private-school and homeschooling families also list a preference for increased character/moral values education as a reason for choosing an alternative to public schools; approximately 81% for private-school parents and 75% for homeschooling families. This may help explain why we see a larger proportion of pages containing character values in our Religious collection of Reading textbooks (Figure 3a).

Religious respondents (89.1%) and Texas respondents (79.6%) are also more likely to report being "against" homosexuality, compared with 61.7% of California respondents (Figure 8d). This pattern aligns with textbook content: the Religious collection frames LGBTQIA+ rights as a threat to religious freedoms, while Texas contains no related content (Sections V.D and VI.C.2). Furthermore, religious respondents are more likely to endorse traditional gender norms (e.g., agreement that men are better suited for politics and that women belong in the home) (Figure 8). Consistent with this, the Religious collection exhibits lower female representation (Table 1) and systematically portrays females in contexts that are less dominant than males in both text and images relative to public-school textbooks (Figure 6).

A similar pattern appears for other salient beliefs. Figure 9a uses a long-running GSS module (1983–2018) asking respondents to choose the most important trait to teach children. Religious respondents place substantially more weight on teaching children to obey and less weight on teaching children to think for themselves than respondents in California and Texas; differences in "helping others" and "working hard" are comparatively smaller. Textbooks exhibit a qualitatively similar pattern (Figure 9b): Religious-collection textbooks are more likely to contain obedience-related values and much less likely to contain critical-thinking values than California and Texas textbooks. The direction of these differences aligns with the survey evidence, consistent with purposeful positioning toward a different demand segment. At the same time, Religious-collection textbooks place more emphasis on "caring" and "hard work" values than public-school collections.

These exercises suggest religious publishers differentiate textbooks along dimensions that their consumers value and that track observable differences in demographics and beliefs.

VII Conclusion

School curricula play a critical role in transmitting knowledge, shaping values, and constructing collective memory. Our study advances our understanding of educational materials' influence on cultural transmission by systematically analyzing textbook content across political and religious educational settings in the U.S. Our findings reveal that while public-school textbooks in politically distinct states such as Texas and California are more similar than public debate suggests, religious-school textbooks diverge meaningfully, emphasizing distinct sets of values, topics, and identities with some important overlap.

The similarities seen in public-school textbooks suggest that market and regulatory forces incentivize mainstream publishers to align public school textbook content towards a shared common baseline, whereas the regulatory flexibility of religious publishers allows them to differentiate their content to align more closely with the specific beliefs of their communities. This institutional distinction carries important implications for policy: expanding public funding for private education, without uniform curricular standards, may increase curricular and ideological divergence among student populations.

Given the similarities in curricula between politically distinct states, it is natural to ask what contributes to divergent political outcomes. An important consideration is that curricula can take differing forms: as written (in textbooks), as implemented (by teachers), and as discussed (among students) (Porter, 2012). Our analysis focuses on printed curriculum, but similarities in textbooks across different educational settings do not guarantee that children receive similar educational experiences in the classroom. Pedagogy plays a crucial role in shaping how material is taught, and differences in teacher training, beliefs, and classroom practices can lead to substantially different lessons even when the same textbook is used. While standardizing classroom instruction requires a highly prescriptive and often impractical implementation strategy, textbook adoption remains a more feasible lever for policy intervention. Prior research has shown that changes in curriculum standards can influence student outcomes, suggesting that curriculum matters, and the importance of analyzing textbook content remains despite potential variations in pedagogy. Our study contributes context to these prior findings, suggesting that regulated textbook markets may temper ideological divergence even when states' political environments differ. Indeed, state curricular standards may shape not only in-state content, they may also diffuse across markets.

Methodologically, our study introduces a scalable computational approach for content

analysis. By combining tools from computational social science with recent advances in artificial intelligence, we transform complex, high-dimensional objects such as text and images into structured, usable measures of topics, values, representation, and portrayal. This expands both the analytical toolkit available to researchers and the range of cultural, educational, and visual materials that can be studied systematically at scale. At the same time, these tools do not eliminate traditional measurement concerns; rather, they make validation a first-order issue. Because outputs from generative and multimodal AI models are probabilistic classifications rather than direct observations of ground truth, researchers must carefully benchmark, validate, and interpret these measures. In our study, we address this challenge through multiple validation exercises, including comparisons to human annotations and alternative measurement approaches, demonstrating both the promise of these tools and the importance of rigorous validation when using them for social science research.

These tools can be applied to an expansive range of content, including but not limited to: historical records; government and legal documents; news media; business and organizational documents; web and social media content; cultural and historical texts including letters, diaries, and speeches; health and clinical notes; and survey and feedback data such as open-ended responses or information from course or employee evaluations. Computer vision tools, in particular, open up a world of questions by enabling researchers to extract information from images, for example: everyday photographs; faces and identity documents such as IDs, yearbooks, or passports that can be linked to administrative information; medical and scientific images such as those from X-rays, MRIs, or microscopy of cells and tissues; remote sensing and geospatial images from satellites or aerial photographs; video footage from body-worn cameras and dashcams; artistic, cultural, and historical visuals; retail and commerce images such as advertisements; and environmental and ecological images including those captured by wildlife camera traps. As a result, the frontier of research will increasingly depend on researchers' creativity, curiosity, and imagination. This paper offers one example of that frontier by applying generative AI and other computational social science tools to examine forty years of values in textbook content used across U.S. schools.

Amid expanding educational gag orders and increased scrutiny of public-school content standards, our work shines a light on what messages children encounter across educational settings, which could have long-term implications for their worldviews. By documenting both inclusion in – and omission from – curricular content, we highlight how institutional structures, market incentives, and demographic composition help shape the content through which societies transmit values and construct collective memory.

References

Abdulkadiroğlu, Atila, Parag A Pathak, and Christopher R Walters. 2018. "Free to Choose: Can School Choice Reduce Student Achievement?" American Economic Journal: Applied Economics, 10(1): 175–206.

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.

Adukia, Anjali, Callista Christ, Anjali Das, and Ayush Raj. 2022a. "Portrayals of Race and Gender: Sentiment in 100 Years of Children's Literature." Proceedings of the 5th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies, 20–28.

Adukia, Anjali, Matthew Bonci, Paula Dastres, Emileigh Harrison, Jake Nicoll, and Teodora Szasz. 2026. "(How) Do We Teach Emotions?" Working paper.

Adukia, Anjali, Patricia Chiril, Callista Christ, Anjali Das, Alex Eble, Emileigh Harrison, and Hakizumwami Birali Runesha. 2022b. "Tales and Tropes: Gender Roles from Word Embeddings in a Century of Children's Books." Proceedings of the 29th International Conference on Computational Linguistics, 3086–3097.

Alesina, Alberto, Paola Giuliano, and Bryony Reich. 2021. "Nation-Building and Education." The Economic Journal, 131(638): 2273–2303.

Anderson, Simon P., André de Palma, and Jacques-François Thisse. 1992. Discrete Choice Theory of Product Differentiation. MIT Press.

Ansell, Ben, and Johannes Lindvall. 2013. "The Political Origins of Primary Education Systems: Ideology, Institutions, and Interdenominational Conflict in an Era of Nation-Building." American Political Science Review, 107(3): 505–522.

Arold, Benjamin W. 2024. "Evolution vs. Creationism in the Classroom: The Lasting Effects of Science Education." The Quarterly Journal of Economics, 139(4): 2331–2375.

Arold, Benjamin W., Ludger Woessmann, and Larissa Zierow. 2025. "Can Schools Change Religious Attitudes? Evidence from German State Reforms of Compulsory Religious Education." Journal of Human Resources. Article ID: 0723-13023R2.

Ash, Elliott, Ruben Durante, Maria Grebenschikova, and Carlo Schwarz. 2021. "Visual Representation and Stereotypes in News Media." Centre for Economic Policy Research CEPR Discussion Paper 16624.

Association of Christian Schools International v. Stearns. 2010. "362 F. App'x 640, United States Court of Appeals." 9th Circuit.

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 Pearson, and Patrick A Testa. 2026a. "The Confederate Diaspora." The Review of Economic Studies, rdag027.

Bazzi, Samuel, Masyhur Hilmy, and Benjamin Marx. 2026b. "Religion, Education, and the State." The Review of Economic Studies, 93(3): 1494–1535.

Bisin, Alberto, and Thierry Verdier. 2001. "The Economics of Cultural Transmission and the Dynamics of Preferences." Journal of Economic Theory, 97(2): 298–319.

California Department of Education. n.d.. "Character Education." https://www.cde.ca.gov/ls/yd/ce/, Accessed May 16, 2026.

California Legislature. 2026. "California Education Code Section 60061."

Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. 2017. "Semantics Derived Automatically from Language Corpora Contain Human-Like Biases." Science, 356(6334): 183–186.

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. 2016. "Why Parents Choose: A Survey of Private School and School Choice Parents in Indiana."

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).

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.

Crawfurd, Lee, Christelle Saintis-Miller, and Rory Todd. 2024. "Sexist Textbooks: Automated Analysis of Gender Bias in 1,255 Books from 34 Countries." PLOS ONE, 19(10): e0310366.

Deckman, Sherry L., Ellie Fitts Fulmer, Keely Kirby, Katharine Hoover, and Abena Subira Mackall. 2018. "Numbers Are Just Not Enough: A Critical Analysis of Race, Gender, and Sexuality in Elementary and Middle School Health Textbooks." Educational Studies, 54(3): 285–302.

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.

Doan, Sy, and Julia Kaufman. 2024. "What Role Do States Play in Selecting K–12 Textbooks?" State Education Standard, 24(1). Accessed 2026-04-13.

Downs, Anthony. 1957. An Economic Theory of Democracy. Harper.

EdChoice. 2023. "Fast Facts on School Choice." Accessed on March 9, 2024.

Fournier, Gaëtan, Karine Van der Straeten, and Jörgen Weibull. 2020. "Spatial Competition with Unit-Demand Functions." arXiv preprint arXiv:2001.11422.

Friedman, Jonathan, and James Tager. 2022. "Educational Gag Orders: Legislative Restrictions on the Freedom to Read, Learn, and Teach." PEN America.

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.

Goldstein, Dana. 2020. "Two States. Eight Textbooks. Two American Stories." The New York Times. Interactive article.

Hotelling, Harold. 1929. "Stability in Competition." The Economic Journal, 39(153): 41–57.

House Research Organization. 2002. "Fact or Faction: The SBOE's Role in Textbook Adoption." Texas House of Representatives 77-17. Accessed 2026-04-13.

Jamison, Peter, Laura Meckler, Prayag Gordy, C Ence Morse, and C Alcantara. 2023. "Home Schooling's Rise from Fringe to Fastest-growing Form of Education." Washington Post.

Kaleem, Jaweed. 2016. "Homeschooling Without God." The Atlantic. Accessed 2025-12-30.

Kamel, Donia, and L. Guillermo Woo-Mora. 2023. "Skin Tone Penalties: Quasi-Experimental Evidence on Colorism in Football." Working paper.

Katz, Phyllis A., and Jennifer A. Kofkin. 1997. "Race, Gender, and Young Children." In Developmental Psychopathology: Perspectives on Adjustment, Risk, and Disorder. , ed. Suniya S. Luthar, Jacob A. Burack, Dante Cicchetti and John R. Weisz, 51–74. Cambridge University Press.

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

Kessen, William, and Sally Weiskopf. 1976. "Color Vision and Hue Categorization in Young Human Infants." Journal of Experimental Psychology: Human Perception and Performance, 2(1): 115–129.

Khodak, Mikhail, Nikunj Saunshi, Yingyu Liang, Tengyu Ma, Brandon M. Stewart, and Sanjeev Arora. 2018. "A La Carte Embedding: Cheap but Effective Induction of Semantic Feature Vectors." Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 12–22.

Kinzler, Katherine D, and Elizabeth S Spelke. 2011. "Do Infants Show Social Preferences for People Differing in Race?" Cognition, 119(1): 1–9.

Klein, Rebecca. 2017. "Voucher Schools Championed by Betsy DeVos Can Teach Whatever They Want. Turns Out They Teach Lies." HuffPost.

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): 905–949.

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

Kunzman, Robert, and Milton Gaither. 2020. "Homeschooling: An Updated Comprehensive Survey of the Research." Journal of Educational Alternatives, 9(1): 253–336.

Lucy, Li, Camilla Griffiths, Claire Ying, JJ Kim-Ebio, Sabrina Baur, Sarah Levine, Jennifer L. Eberhardt, David Bamman, and Dorottya Demszky. 2025. "Racial and Ethnic Representation in Literature Taught in U.S. High Schools." Journal of Cultural Analytics, 10(1).

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 U.S. 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." Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 615–621.

Martinez, Joel E, and Alexander Todorov. 2023. "Mapping Varied Mental Representations: The Case of Representing Illegalized Immigrants." Social Cognition, 41(6): 507–536.

Michalopoulos, Stelios, and Christopher Rauh. 2024. "Movies." National Bureau of Economic Research Working Paper 32220.

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." Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 174–184.

Musaddiq, Tareena, Kevin Stange, Andrew Bacher-Hicks, and Joshua Goodman. 2022. "The Pandemic's Effect on Demand for Public Schools, Homeschooling, and Private Schools." Journal of Public Economics, 212.

National Center for Education Statistics. 2019. "School Choice in the United States: 2019." U.S. Department of Education, Institute of Education Sciences NCES 2019-106. Private school enrollment data from fall 2015.

NCES. 2022. "Homeschooled Children and Reasons for Homeschooling." U.S. Dept. of Education, National Center for Education Statistics. Accessed: 2026-04-20.

Newsom, Gavin, and Rob Bonta. 2023. "Governor Newsom and Attorney General Bonta Demand Answers from Temecula School Board." Accessed 2026-04-13.

NGSS Lead States. 2018. "Science Assessment Challenges." Achieve, Inc. Next Generation Science Standards.

Office of the Attorney General of Texas. 2025. "Attorney General Ken Paxton Sues Coppell ISD for Illegally Teaching Woke and Hateful Critical Race Theory Curriculum." Accessed 2026-04-13.

Paglayan, Agustina S. 2022. "Education or Indoctrination? The Violent Origins of Public School Systems in an Era of State-Building." American Political Science Review, 116(4): 1242–1257.

Porter, Andrew C. 2012. "Curriculum Assessment." In Handbook of Complementary Methods in Education Research. 141–159. Routledge.

Riley, Emma. 2024. "Role Models in Movies: The Impact of Queen of Katwe on Students' Educational Attainment." The Review of Economics and Statistics, 106(2): 334–351.

Rodriguez, Pedro, Arthur Spirling, and Brandon M. Stewart. 2023. "Embedding Regression: Models for Context-Specific Description and Inference." American Political Science Review, 117(4): 1255–1274.

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." The Heritage Foundation Backgrounder 3626.

Sallabaş, Muhammed Eyyüp. 2013. "Analysis of Narrative Texts in Secondary School Textbooks in Terms of Values Education." Educational Research and Reviews, 8(8): 361.

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.

Shaked, Avner, and John Sutton. 1982. "Relaxing Price Competition through Product Differentiation." The Review of Economic Studies, 49(1): 3–13.

Stevenson, Betsey, and Hanna Zlotnik. 2018. "Representations of Men and Women in Introductory Economics Textbooks." AEA Papers and Proceedings, 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, 462–471.

Tabellini, Guido. 2008. "Institutions and Culture." Journal of the European Economic Association, 6(2–3): 255–294.

Texas State Board of Education. 2020. "19 TAC Chap. 120, Subchap. A: Character Traits." https://tea.texas.gov/about-tea/laws-and-rules/texas-administrative-code/19-tac-chapter-120, May 16, 2026.

Voth, Hans-Joachim, and David Yanagizawa-Drott. 2024. "Image(s)." Centre for Economic Policy Research CEPR Discussion Paper 19219.

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.

Wheaton, Brian. 2020. "Laws, Beliefs, and Backlash." Working Paper.

Williamson, Danielle, and Jennifer Withrow. 2025. "Southern Academies: The Proliferation of All-White Private Schools After Brown and Their Legacy for Students."

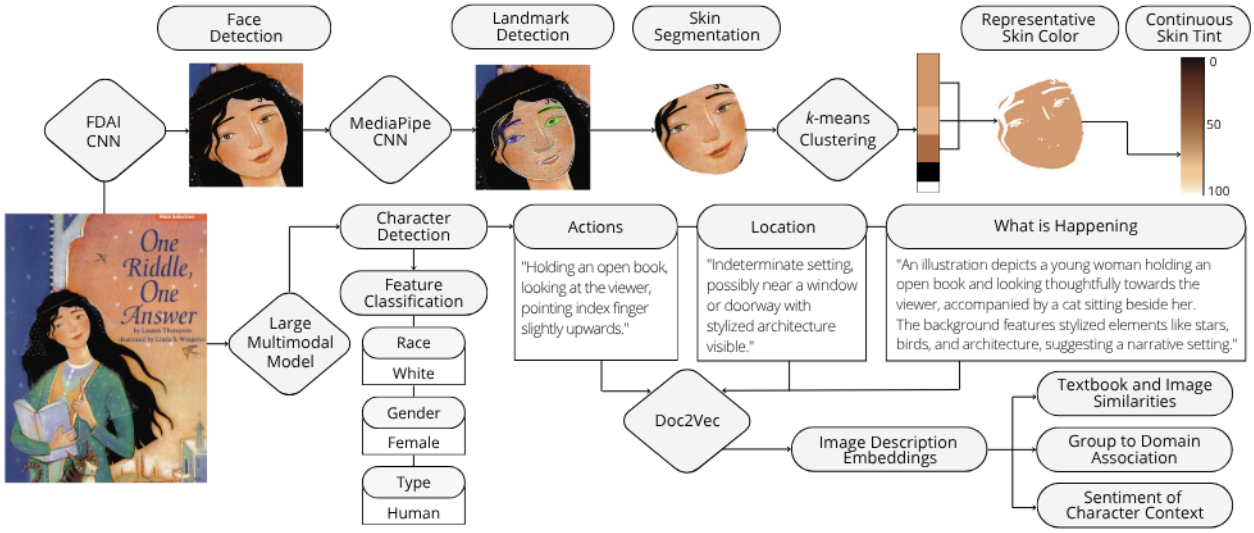
Yu, Amy Zhao, Shahar Ronen, Kevin Hu, Tiffany Lu, and César A. Hidalgo. 2016. "Pantheon 1.0, a Manually Verified Dataset of Globally Famous Biographies." Scientific Data, 2: 150075.

Ziems, Caleb, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. 2024. "Can Large Language Models Transform Computational Social Science?" Computational Linguistics, 50(1): 237–291.

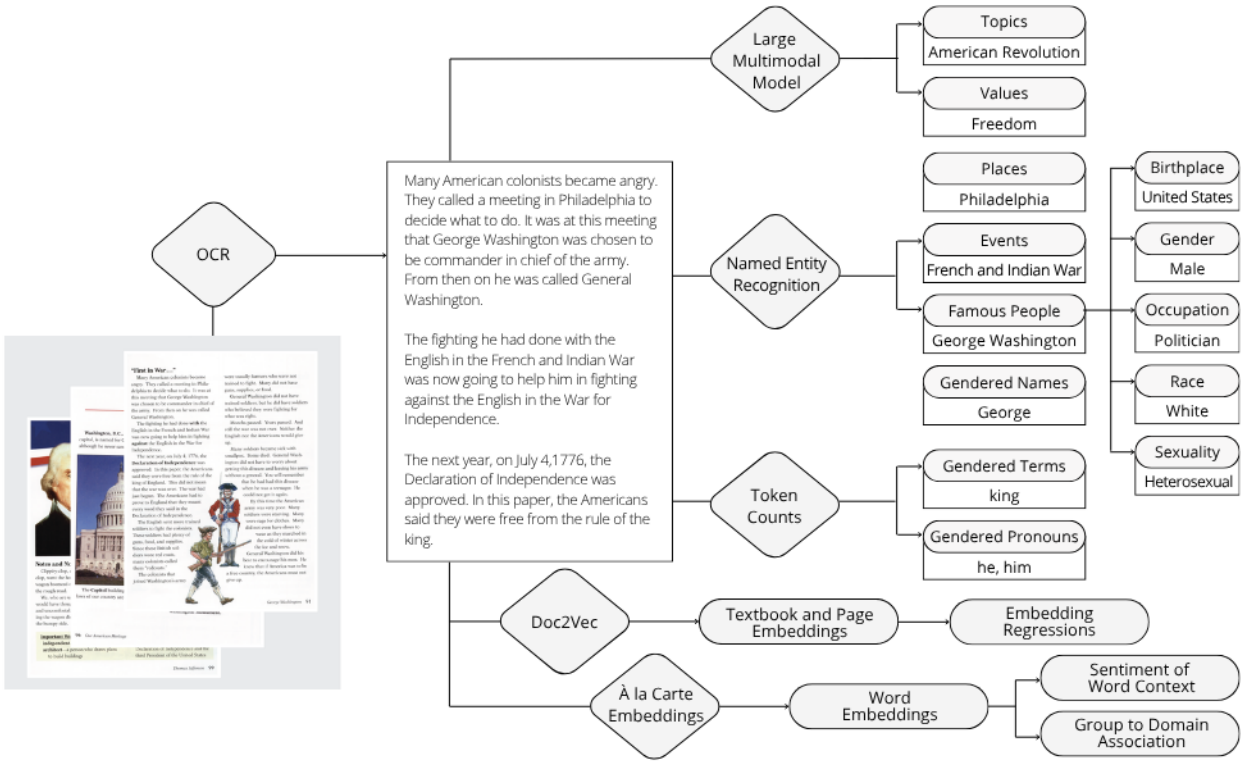
VIII Main Figures

Figure 1: Data Extraction Pipelines: Image and Text Analysis

(a) Image Analysis Pipeline



(b) Text Analysis Pipeline



Notes: 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.

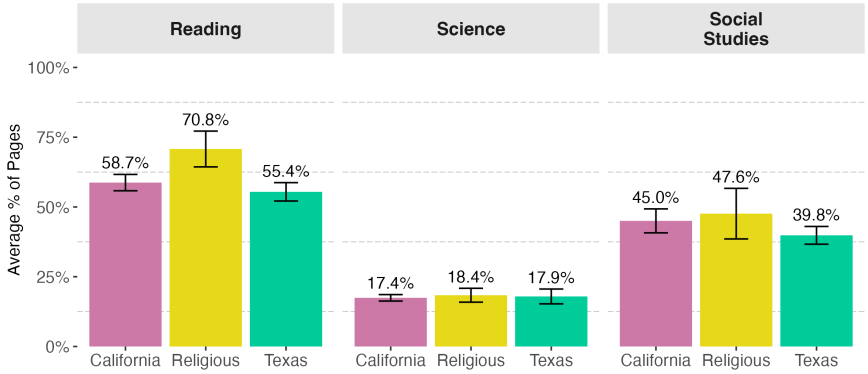
Figure 2: Textbook Embeddings



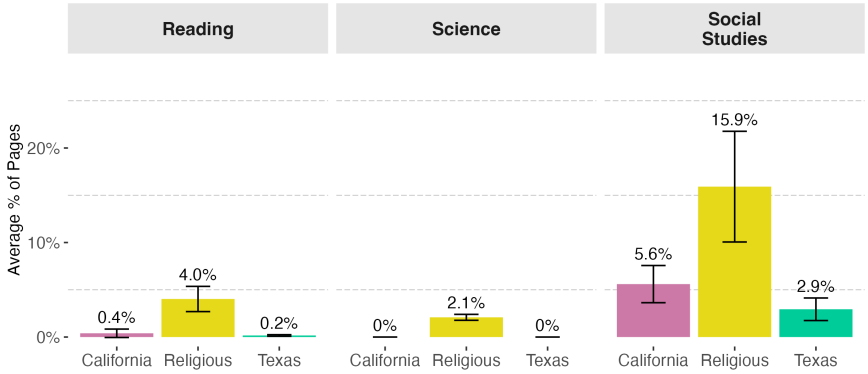
Notes: In panels (a) and (c), we show analyses using embeddings estimated from the text of each textbook. In panels (b) and (d), we show analyses using embeddings estimated from AI-generated image descriptions of a pictured character's setting, location, and actions as described in Section IV. In panels (a) and (b), each observation represents a book from a specific collection (color) and subject (shape). In panel (a), these are textbook-level embeddings estimated using text from the entire textbook; in panel (b), these are textbook-level embeddings constructed by averaging across image embeddings in a textbook. Their content, originally represented in a 300-dimensional space, is reduced to 2D via t-SNE while retaining meaningful information. Book representations should be interpreted in terms of their relative position to other observations. In (c) and (d), we report normed coefficients from page- or image-level embedding regressions, respectively, in the spirit of Rodriguez et al. (2023) as described in Appendix Section C. These normed coefficients measure the overall magnitude of semantic difference in embedding space between groups, aggregated across all embedding dimensions. Larger values indicate greater semantic differences between groups, while values near zero indicate little systematic separation in embedding space. We estimate two separate regressions; one where the Texas collection is the excluded group and one where the Religious collection is the excluded group. Statistical significance is assessed using permutation tests.

Figure 3: Values

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



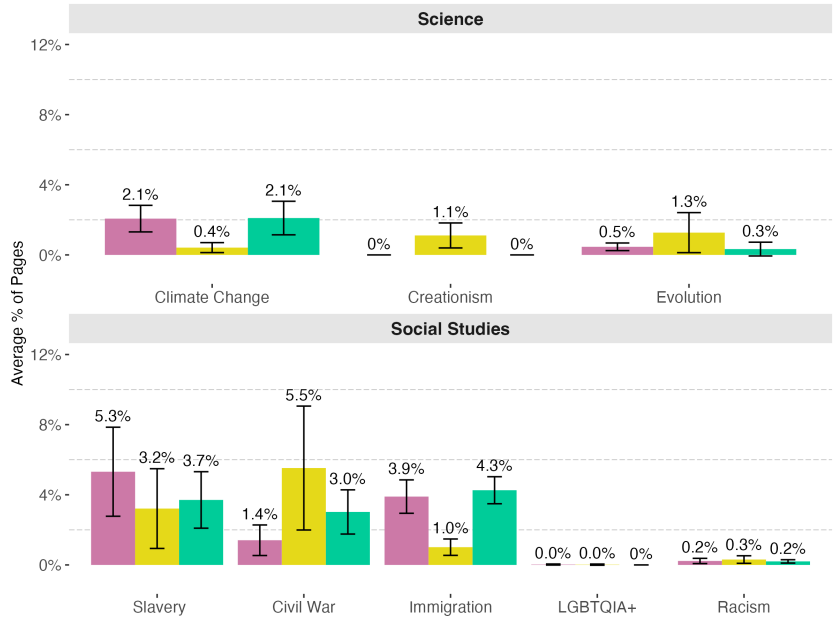
(b) Average Percent of Pages Containing Religious Values by Subject



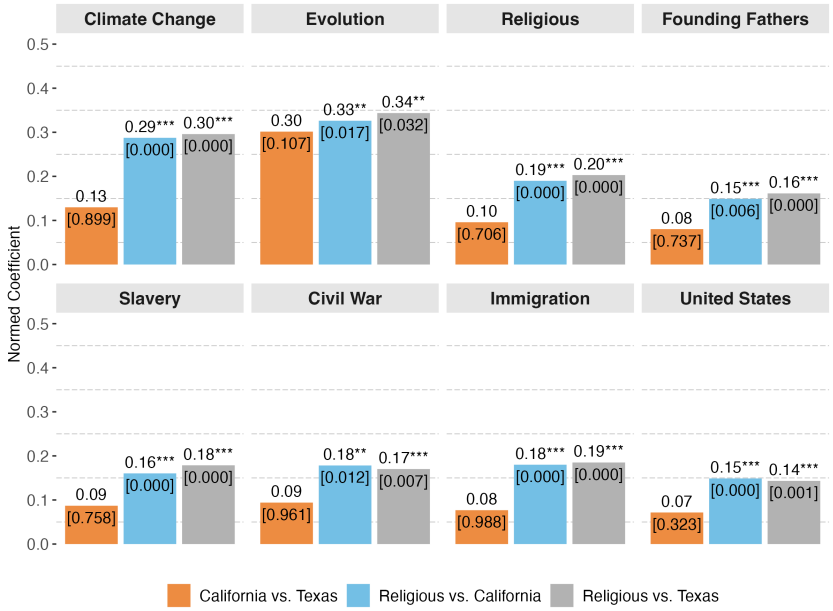
Notes: Panel (a) reports the average share of pages that contain character values by collection and subject and Panel (b) shows the average share of pages that contain religious values by collection and subject. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection and subject so that each grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals.

Figure 4: High-Profile Topics

(a) Average Percent of Pages Containing High-Profile Topics



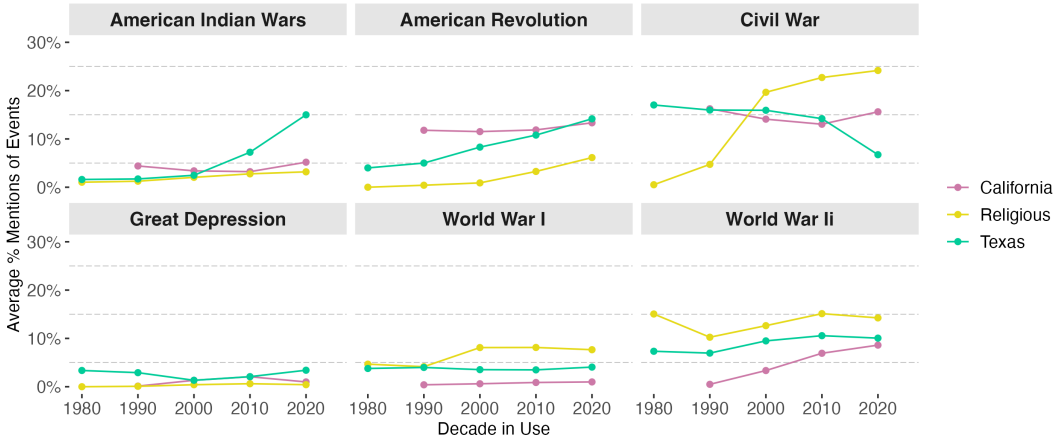
(b) Topic Embedding Regressions



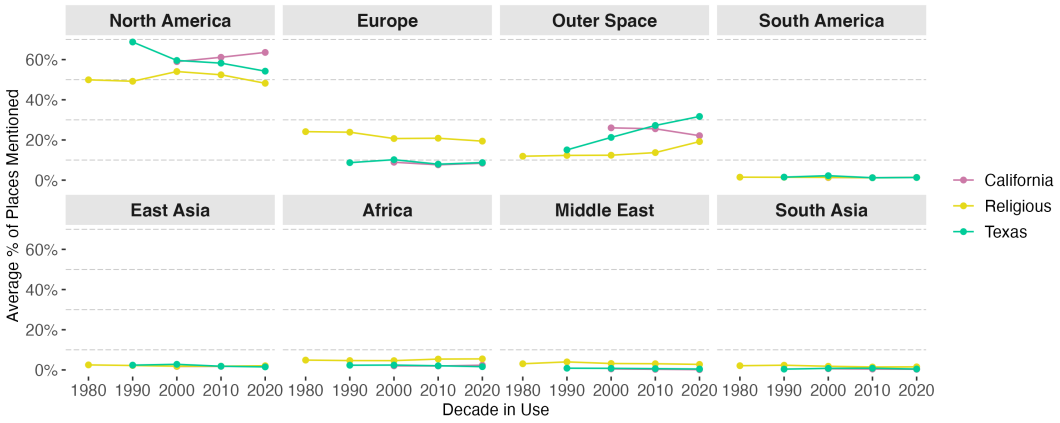
Notes: Panel (a) reports the average share of pages containing a specified topic in a given collection and subject. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection and subject so that each grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals. Panel (b) reports normed coefficients from a page-level embedding regression in the spirit of Rodriguez et al. (2023) as described in Appendix Section C except that for each regression we subset to pages discussing a given topic to test semantic differences in topics between collections. These normed coefficients measure the overall magnitude of semantic difference in embedding space between groups, aggregated across all embedding dimensions. Larger values indicate greater semantic differences between groups, while values near zero indicate little systematic separation in embedding space. We estimate two separate regressions; one where the Texas collection is the excluded group and one where the Religious collection is the excluded group. Statistical significance is assessed using permutation tests clustered at the textbook level.

Figure 5: Geography, History, and Birthplaces

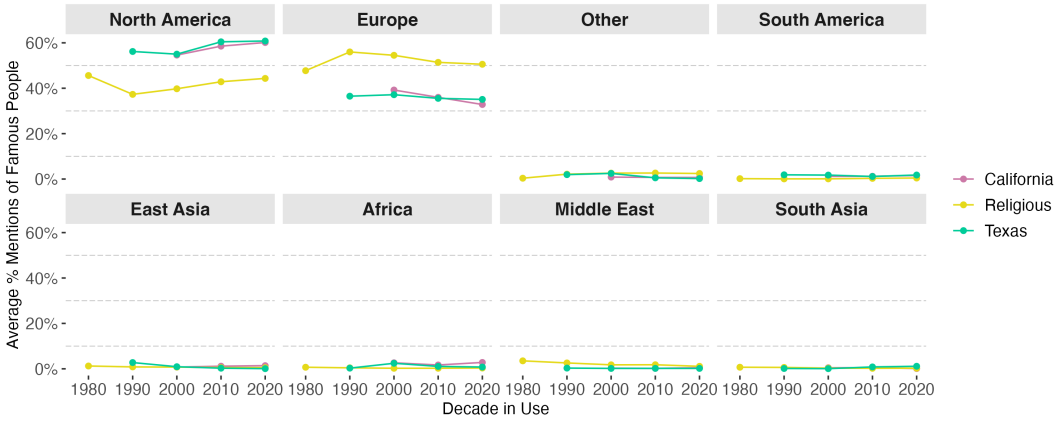
(a) Historical Events



(b) Geography



(c) Birthplaces of Famous Individuals

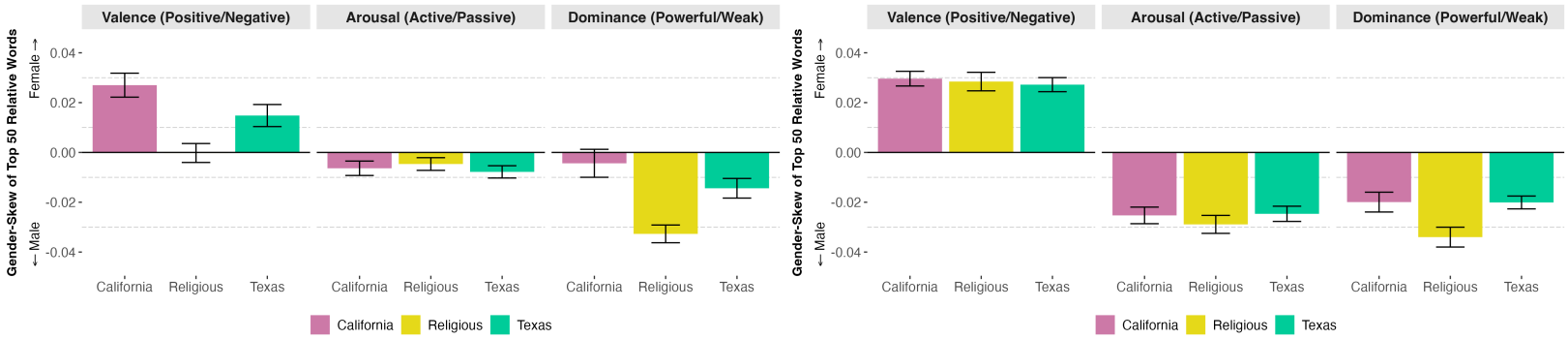


Notes: Panel (a) shows the average percentage of events mentioned accounted for by the most frequently mentioned events by collection over time. Panel (b) shows the average percentage of places mentioned, excluding an "Other" category. Panel (c) shows the average percentage of famous people mentioned by their place of birth for each collection over time. We determine mentions of places, events, and people using Named Entity Recognition. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection or collection and decade so that each subject-grade–year cell receives equal weight.

Figure 6: Gender Skew of Domains and Sentiment (VAD Scores)

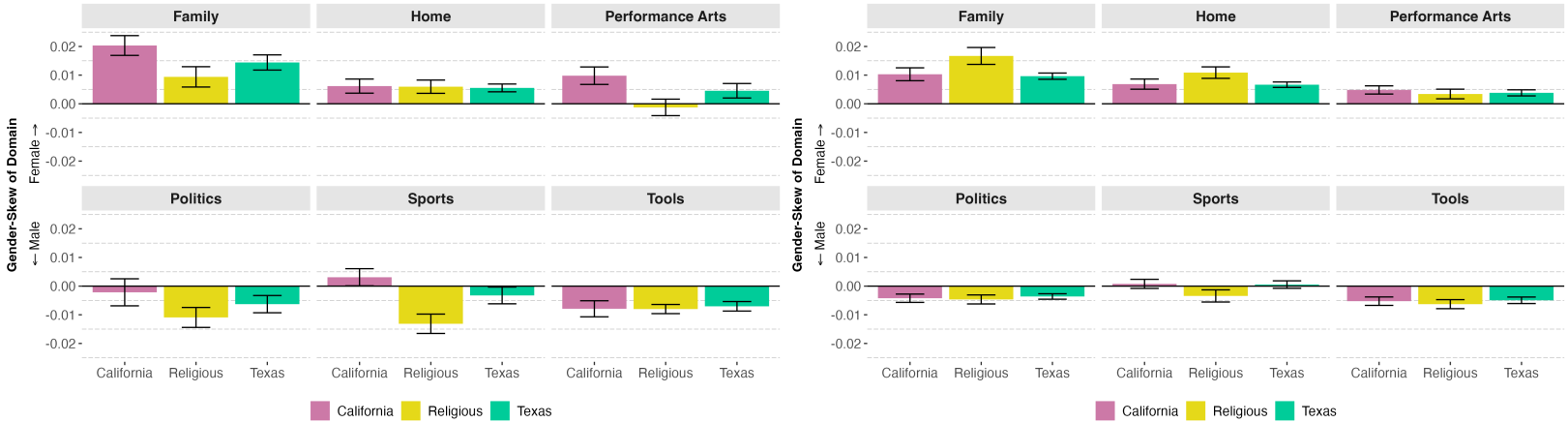
(a) Sentiment via VAD (Text)

(b) Sentiment via VAD (Images)



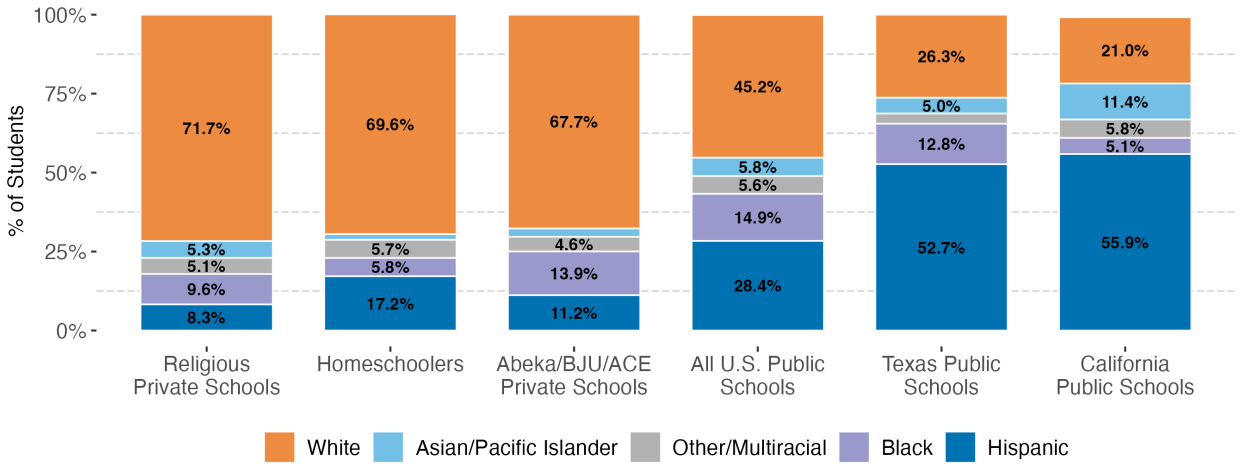
(c) Gender Skew by Domain (Text)

(d) Gender Skew by Domain (Images)



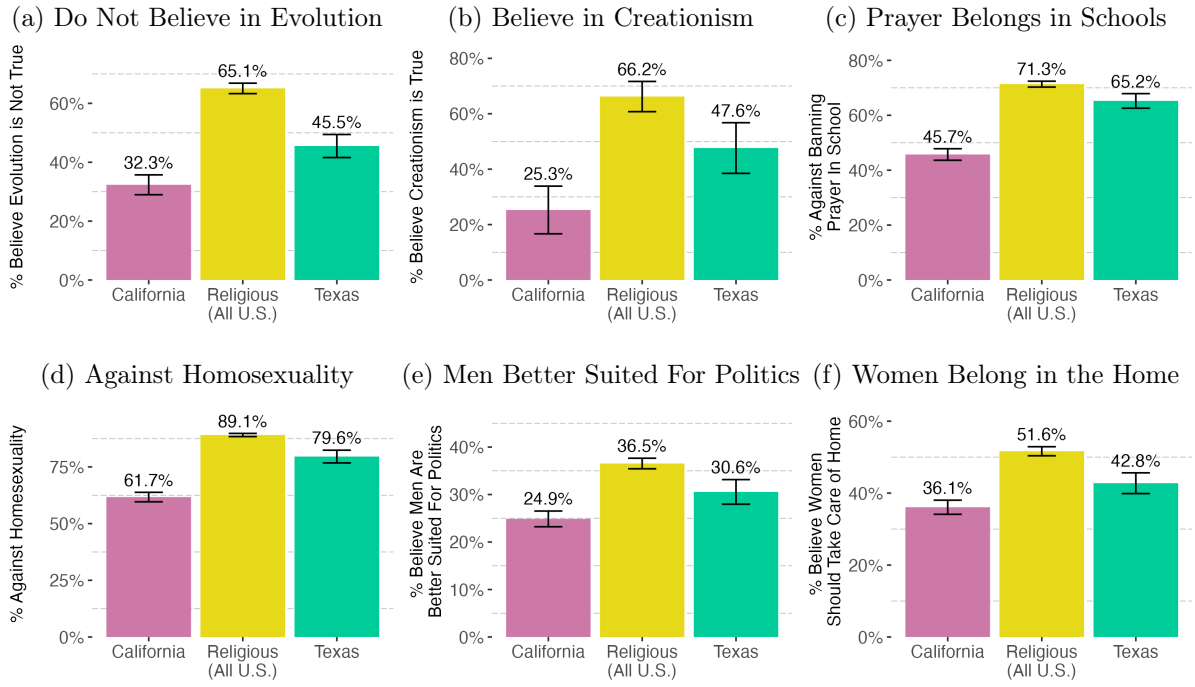
Notes: Bars report collection-specific coefficients estimating differences between males and females from embedding-based regressions; error bars denote 95% confidence intervals, with standard errors clustered at the textbook level. Panels (a) and (b) show gender differences in the average Valence, Arousal, and Dominance (VAD) scores of the top 50 relative words drawn from the NRC VAD lexicon. In panel (a), these relative words are identified from pronoun-instance text embeddings; in panel (b), they are identified from embeddings of model-generated image descriptions. Positive values indicate that female contexts are associated with words that are more positive, active, or powerful than male contexts within the same collection. Panels (c) and (d) show differences between males and females in average cosine similarity to pre-specified domain word sets (family, home, performance arts, politics, sports, and tools). Positive values indicate that a domain is more strongly associated with female than male contexts; negative values indicate the reverse. We call this "gender skew". More details available in Appendix C. All regressions control for subject, grade, and year in use.

Figure 7: Racial/Ethnic Composition of K–12 Students by School Type



Notes: Data compiled from several sources. Data on religious private-school students from NCES PSS 2021–22, Digest Table 205.30 (Other Religious, Non-Catholic); Data on public-school students from NCES CCD 2021–22, Digest Table 203.70; Data on homeschoolers from NCES PFI-NHES 2019, Digest Table 206.10. Homeschooler Asian/Pacific Islander category is back-calculated from reported totals and includes American Indian and Alaskan Native students. Data on students in Abeka/BJU/ACE schools are compiled from data on the curriculum of private schools receiving taxpayer dollars collected by Klein (2017) and merged to the NCES PSS 2021–22 school level data by researchers using school address.

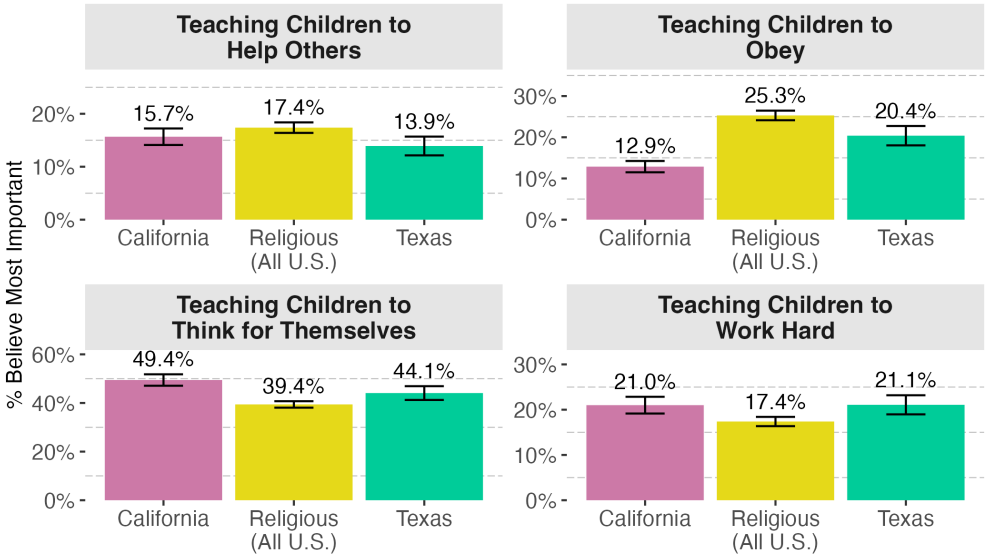
Figure 8: Beliefs by State and Religiosity



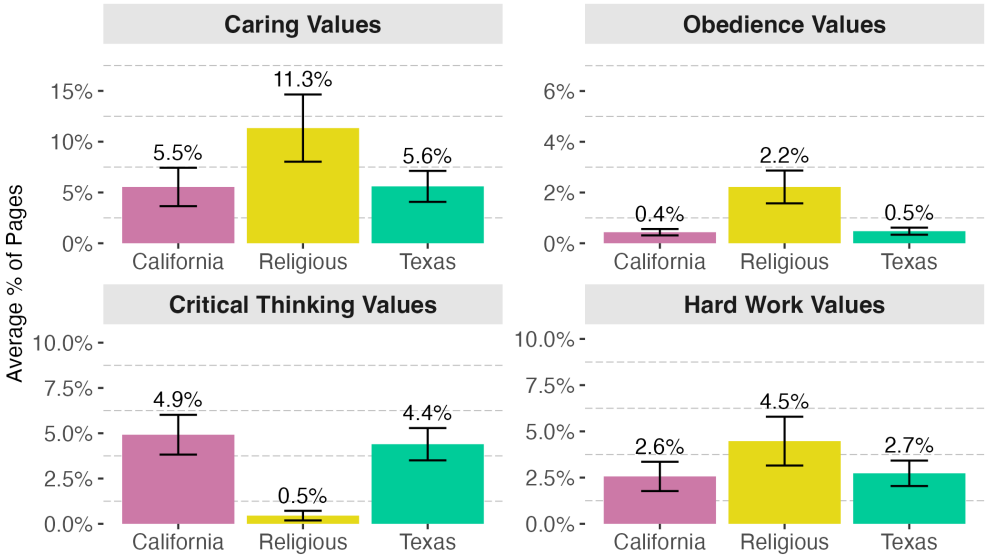
Notes: Bars report survey-weighted means from pooled General Social Survey cross-sections, using all years in which each question was asked; error bars denote 95% confidence intervals that account for the GSS survey design. The "Religious (All U.S.)" group consists of respondents nationwide who attend religious services weekly or more often. In panel (a), "Believe Evolution Is Not True" is a composite measure constructed from available GSS evolution/creationism items.

Figure 9: Character Values Categories

(a) GSS Survey on Most Important Things to Teach Children



(b) Average Percent of Pages Containing Values Related to GSS Survey



Notes: Panel (a) reports survey-weighted means from pooled General Social Survey (GSS) cross-sections, using all years in which this question was asked; error bars denote 95% confidence intervals that account for the GSS survey design. The "Religious (All U.S.)" group consists of respondents nationwide who attend religious services weekly or more often. Panel (b) shows the average share of pages that contain specific values related to the GSS question in Panel (a). Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection so that each subject-grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals.

IX Main Tables

Table 1: Differences in Representation by Collection

	Dependent variable:				
	Skin Tint		% Female Words	% Female Pictured Characters	% Human Pictured Characters
	(1)	(2)	(3)	(4)	(5)
Religious	6.191***	4.701***	−0.095***	−0.072***	−0.037
	(1.343)	(1.356)	(0.023)	(0.024)	(0.030)
Texas	−0.356	−0.815	0.024	0.007	−0.018
	(1.675)	(1.699)	(0.020)	(0.016)	(0.025)
Asian		0.718			
		(0.878)			
Black		−13.395***			
		(0.674)			
Latine		−7.866***			
		(0.738)			
Other		−2.388***			
		(0.609)			
Grade 5	−4.915***	−5.072***	−0.062***	−0.012	0.033
	(1.290)	(1.307)	(0.022)	(0.027)	(0.025)
Science	−9.127***	−8.742***	−0.114***	0.048**	−0.247***
	(2.120)	(2.186)	(0.020)	(0.020)	(0.034)
Social Studies	−2.687*	−1.884	−0.209***	−0.108***	0.204***
	(1.442)	(1.470)	(0.019)	(0.020)	(0.018)
Year in Use	−0.108*	−0.100*	0.0003	0.0005	0.0004
	(0.059)	(0.059)	(0.001)	(0.001)	(0.001)
Constant	64.022***	66.624***	0.517***	0.388***	0.683***
	(2.305)	(2.294)	(0.032)	(0.028)	(0.038)
Observations	347,294	227,000	2,521	2,521	2,521
Adjusted R^2	0.084	0.154	0.339	0.164	0.427

Notes: Column 1 shows a regression of skin tint of a face detected in a given textbook on indicator variables indicating which collection, subject and grade the textbook belongs to. We 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. The dependent variable in column 3 is the proportion of female words in a given textbook. The dependent variable in column 4 is the proportion of pictured female characters. The dependent variable in column 5 is the proportion of pictured human characters. Standard errors clustered at the textbook level. *p<0.1; **p<0.05; ***p<0.01

Table 2: Direct Effects and Cross-State Spillovers from State Standards Revisions

	Dependent Variable: % New-Standards Words			
	All Subjects	Reading	Science	Social Studies
	(1)	(2)	(3)	(4)
Direct Effect	0.0160***	0.0101***	0.0244***	0.0042**
	(0.0020)	(0.0018)	(0.0031)	(0.0017)
Spillover	0.0023***	0.0018*	0.0026	0.0027**
	(0.0008)	(0.0010)	(0.0021)	(0.0011)
Fixed-effects				
Collection-Grade-Subject	Yes			
Collection-Grade		Yes	Yes	Yes
Observations	735	285	271	179
Within R^2	0.49186	0.39748	0.67932	0.35463

Notes: This table reports estimates of Equation (2). The unit of observation is a textbook-year. For each subject, we restrict the sample to the longest period in which only one state changes its curriculum standards and both states subsequently adopt new textbooks. We construct a subject- and grade-specific "new-standards" dictionary consisting of words that appear in the revised standards of the state changing its standards but not in the prior standards of either state. The dependent variable is the share of textbook words from this dictionary. "Direct Effect" indicates the coefficient on post-adoption textbook-years in the standards-changing state, and "Spillover" indicates the coefficient on post-adoption textbook-years in the other state. The omitted category consists of textbook-year observations before either state adopts new textbooks. A positive and significant "Direct Effect" indicates that revised standards translate into textbook content at adoption-cycle boundaries, while a positive and significant "Spillover Effect" indicates that these changes diffuse across states through publishers' cross-market production decisions. In the pooled specification, the direct effect implies an increase of approximately 16 new-standards words per 1,000 textbook words, while the spillover effect implies an increase of approximately 2.3 new-standards words per 1,000 textbook words. Standard errors are clustered at the textbook level. *p<0.1; **p<0.05; ***p<0.01.

Online Appendix

Separation of Church and State Curricula? Public Standards, Private Values, and Textbook Content*

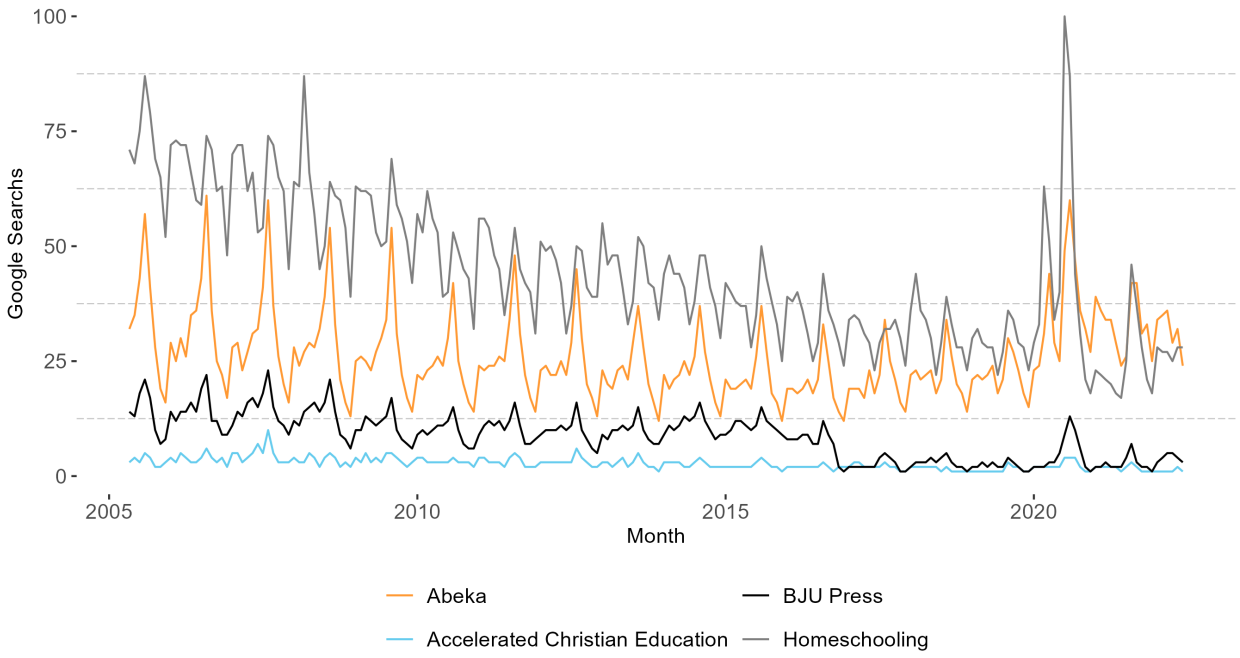
June 10, 2026

Anjali Adukia, University of Chicago, MiiE Lab, IZA, NBER
Emileigh Harrison, University of Texas at Austin, MiiE Lab, UChicago
Inclusive Economy Lab

*Contact: adukia@uchicago.edu, emileigh.harrison@austin.utexas.edu

A Appendix Figures

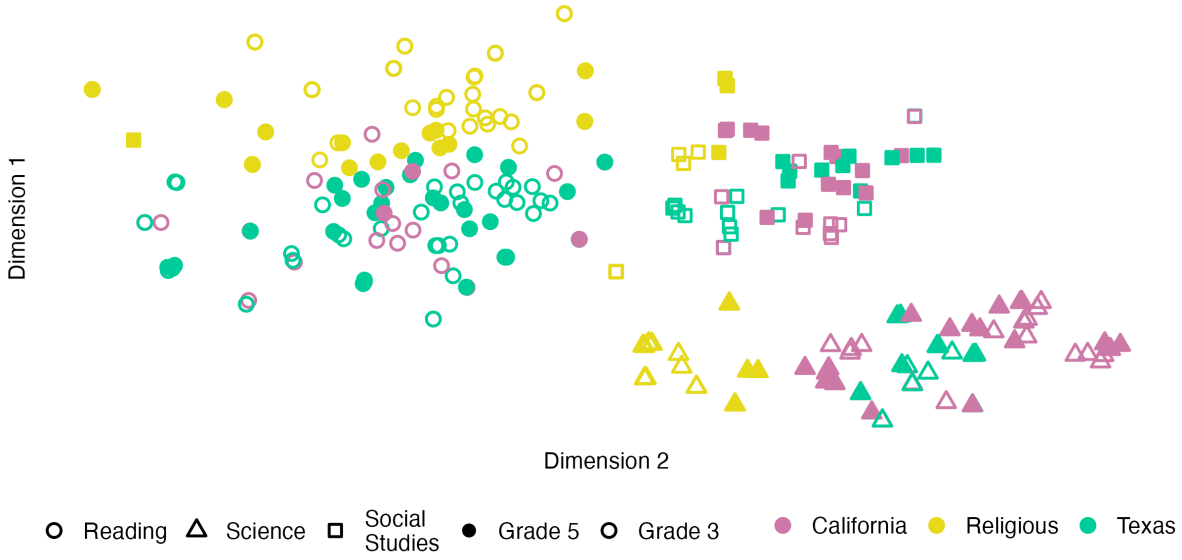
Figure A.1: Google Search Interest Over Time — Homeschooling and Religious Publishers



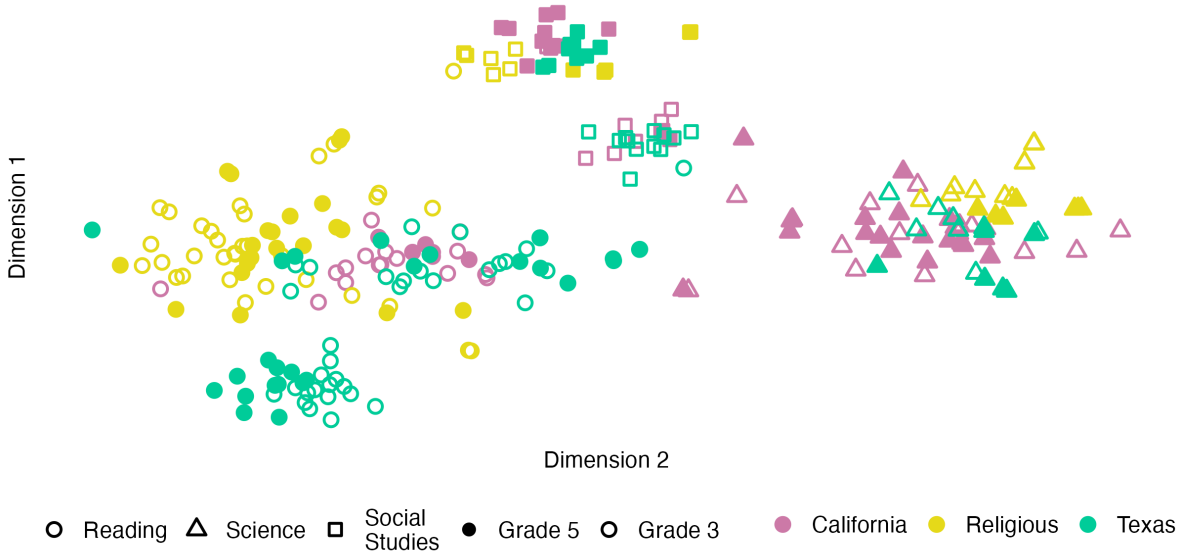
Notes: This figure shows Google search interest from Google Trends data for the search term homeschool and each of the education publishing companies in the religious school textbook collections; Abeka, Accelerated Christian Education (ACE), and BJU Press (BJU). Note that we use the company topics created 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 A.2: Textbook Embeddings by Grade, Subject, and Collection

(a) Text



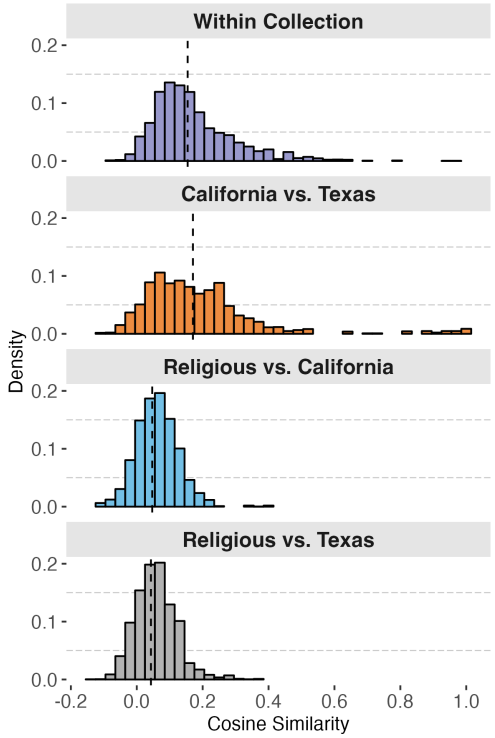
(b) Images



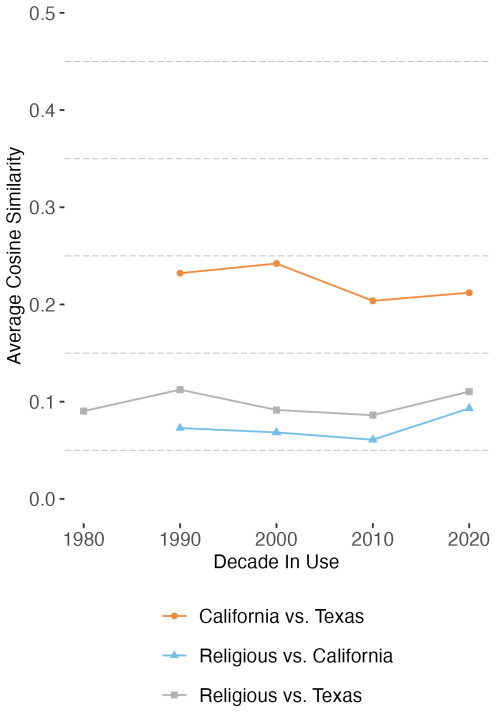
Notes: Each observation represents a book from a specific collection (color), subject (shape), and grade (outline). The filled shapes indicate 5th grade textbooks and hollow shapes indicate 3rd grade textbooks. Their content, originally represented in a 300-dimensional space, is here reduced to 2D via 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 circles (○) signify Reading, the triangles (△) signify Science, and the squares (□) signify Social Studies.

Figure A.3: Average Cosine Similarity Between Book Embeddings by Collection Pairs

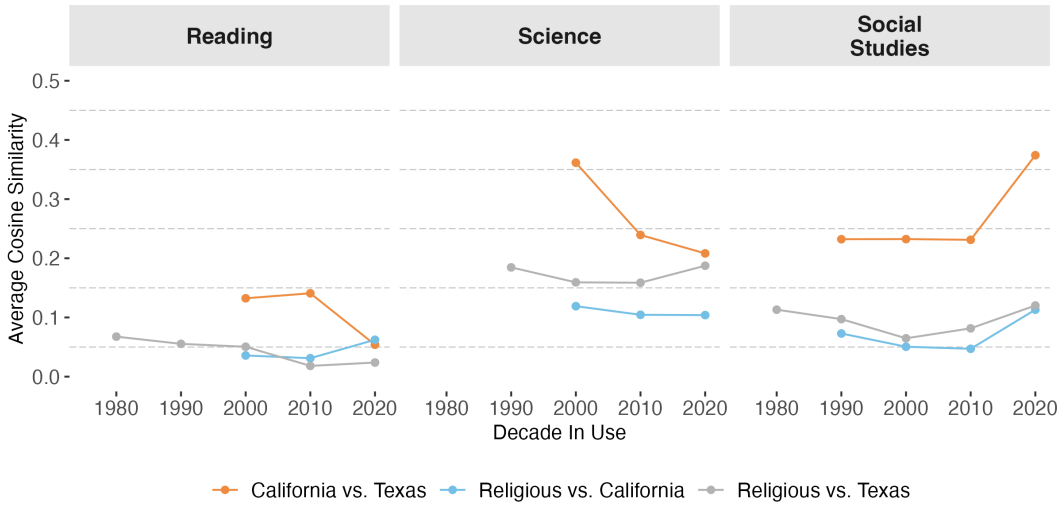
(a) Histogram of Similarities Between Book Embeddings by Collection Pairs



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



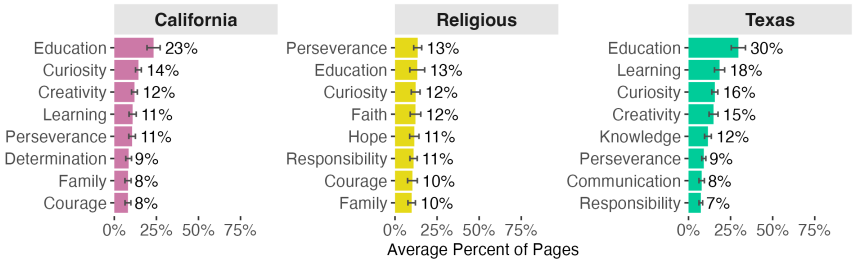
(c) Average Similarity Between Book Embeddings by Collection Pairs and Subject over Time



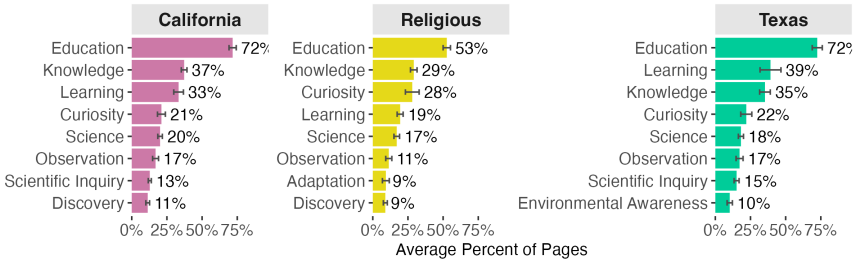
Notes: In panel (a), we plot a histogram of the cosine similarities between two books that share the same subject, grade, and year in use. We separate the densities by collection pairs (e.g. one California textbook and one Texas textbook) and plot within collection book pair similarities for comparison. Note that each book pair is repeated for every year those two books are both in use. In panel (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. panel (c) follows the same approach as panel (b) except that we calculate the averages separately by subject.

Figure A.4: Top Values and Topics by Collection

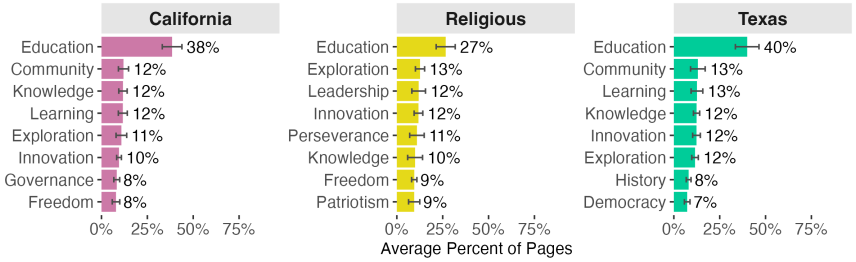
(a) Top Values in Reading



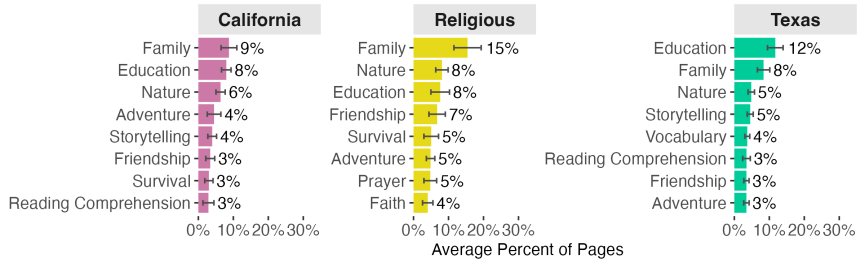
(b) Top Values in Science



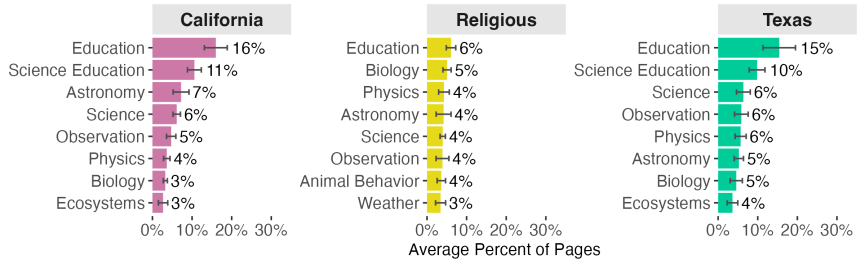
(c) Top Values in Social Studies



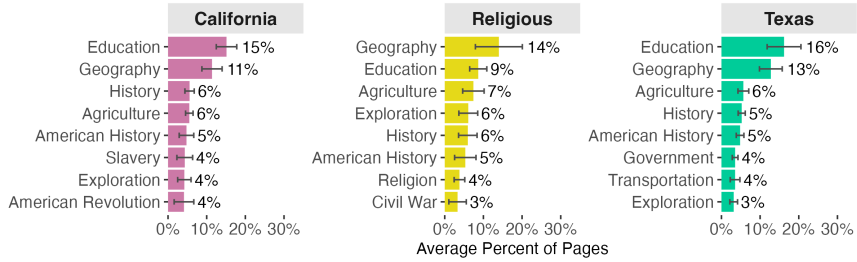
(d) Top Topics in Reading



(e) Top Topics in Science



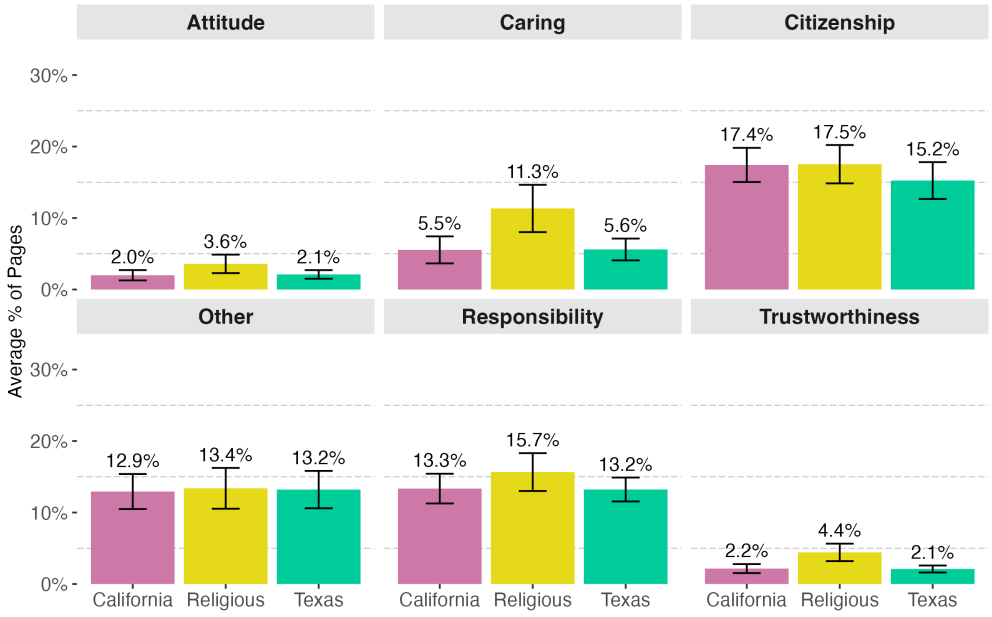
(f) Top Topics in Social Studies



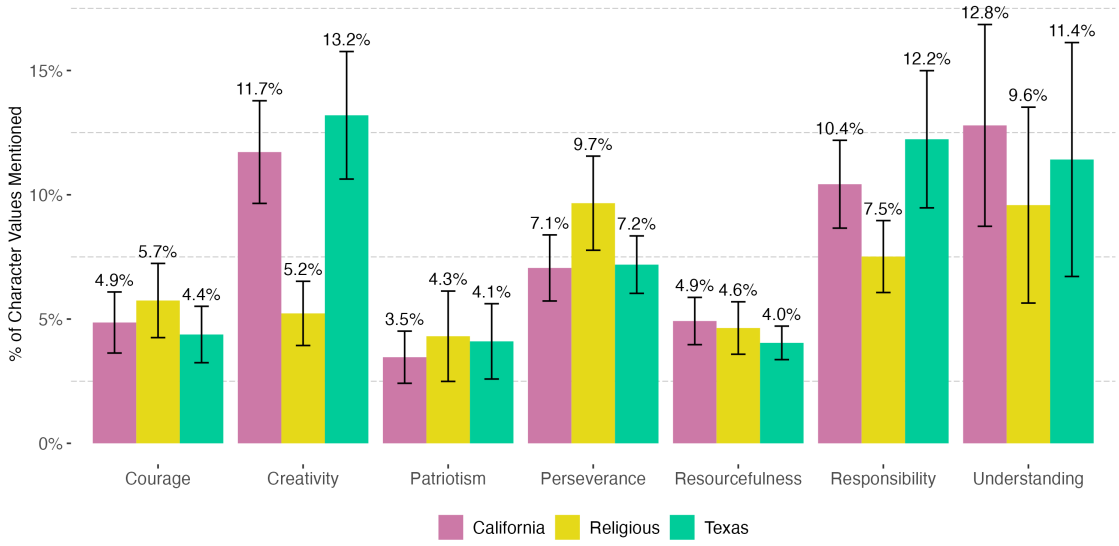
Notes: This figure shows the top values and topics (e.g. values and topics that appear on the highest average percentage of pages) within a collection. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection so that each subject-grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals.

Figure A.5: Character Values

(a) Average Percent of Reading Pages Containing Specific Character Values Categories

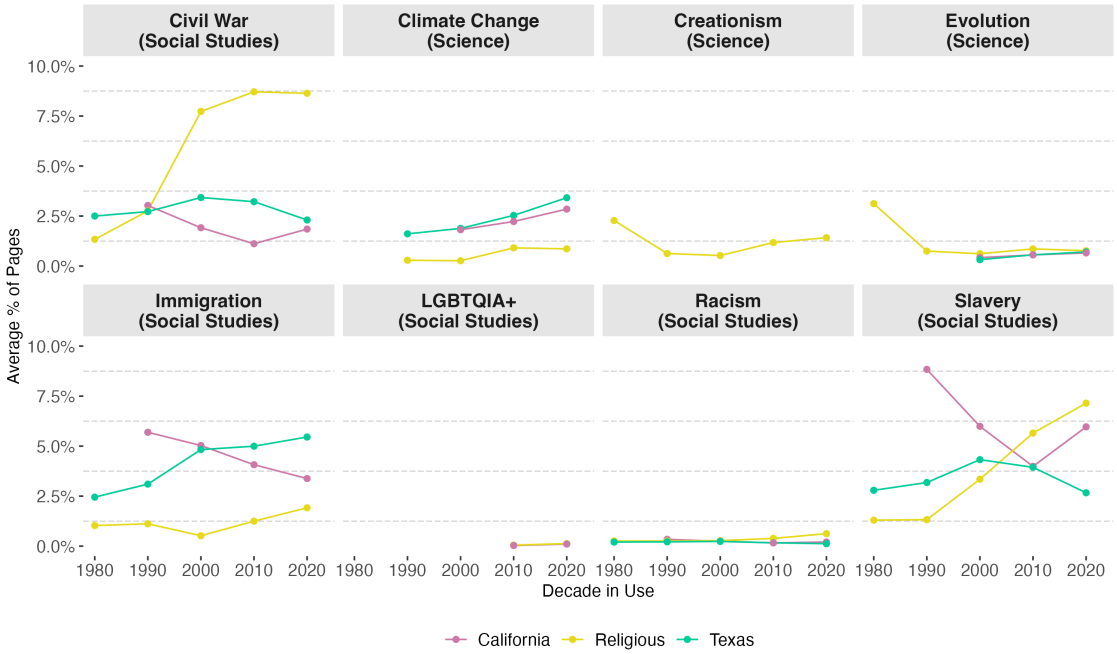


(b) Top Character Values by Collection



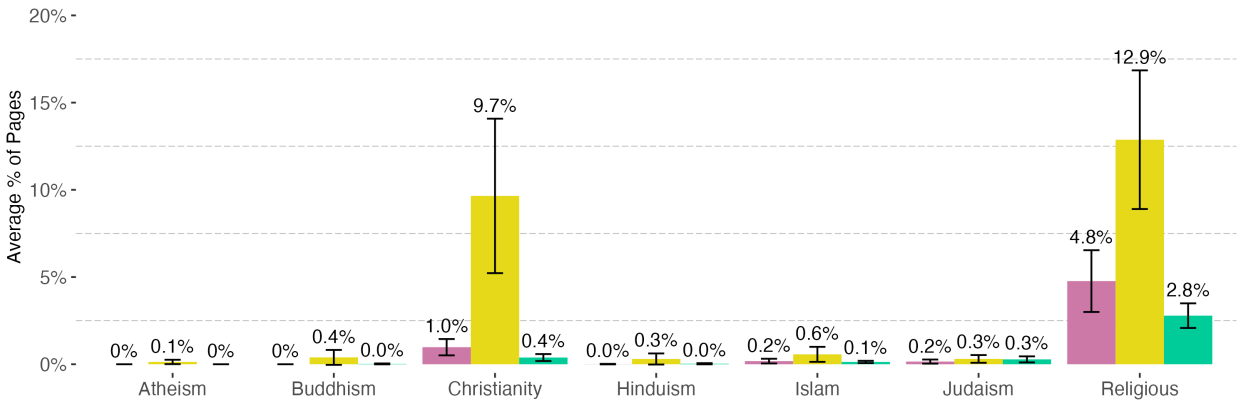
Notes: Panel (a) shows the average proportion of pages that contain character value groups from specific subcategories outlined in state standards as described in Appendix Table B.2. Panel (b) shows the average proportion of all character value mentions that is accounted for by the most frequently mentioned values in each collection by subject. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection so that each subject–grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals. Note that Panel (a) only examines individual character values while Panel (b) includes groups of character values (e.g. the Responsibility group contains values such as responsibility, responsibilities, citizen responsibility, taking responsibility, responsible behavior, etc).

Figure A.6: Average Percent of Pages Containing High-Profile Topics Over Time



Notes: This figure shows the average percentage of pages within a specific subject as denoted in the parentheses that contain a specific topic over time. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection and subject so that each grade–year cell receives equal weight.

Figure A.7: Average Percent of Pages in Social Studies Textbooks Containing Religious Topics or Topics Related to a Specific Religion



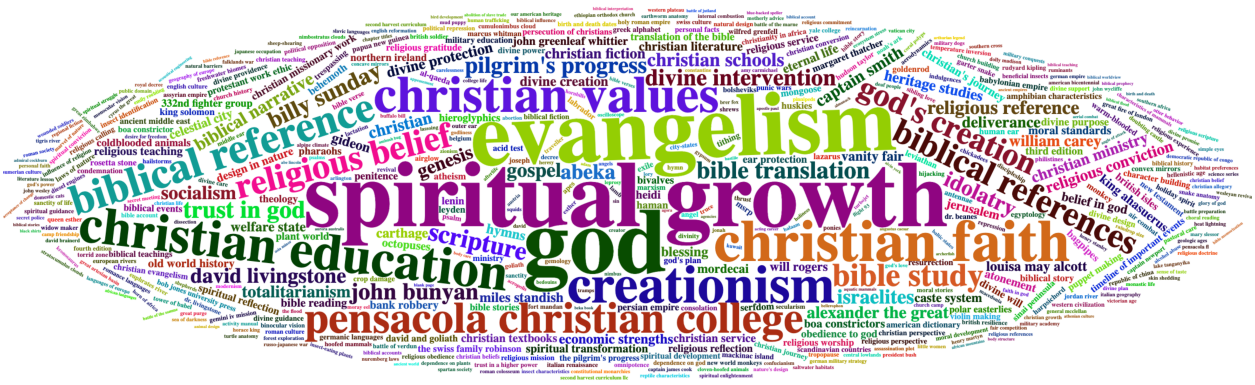
Notes: This figure shows the average percentage of pages within Social Studies textbooks that contain a religious topic or topics related to a specific religion. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection and subject so that each grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals.

Figure A.8: Topics and Values Exclusive to Each Collection: Word Clouds

(a) Topics and Values Exclusive to California Collection



(b) Topics and Values Exclusive to Religious Collection



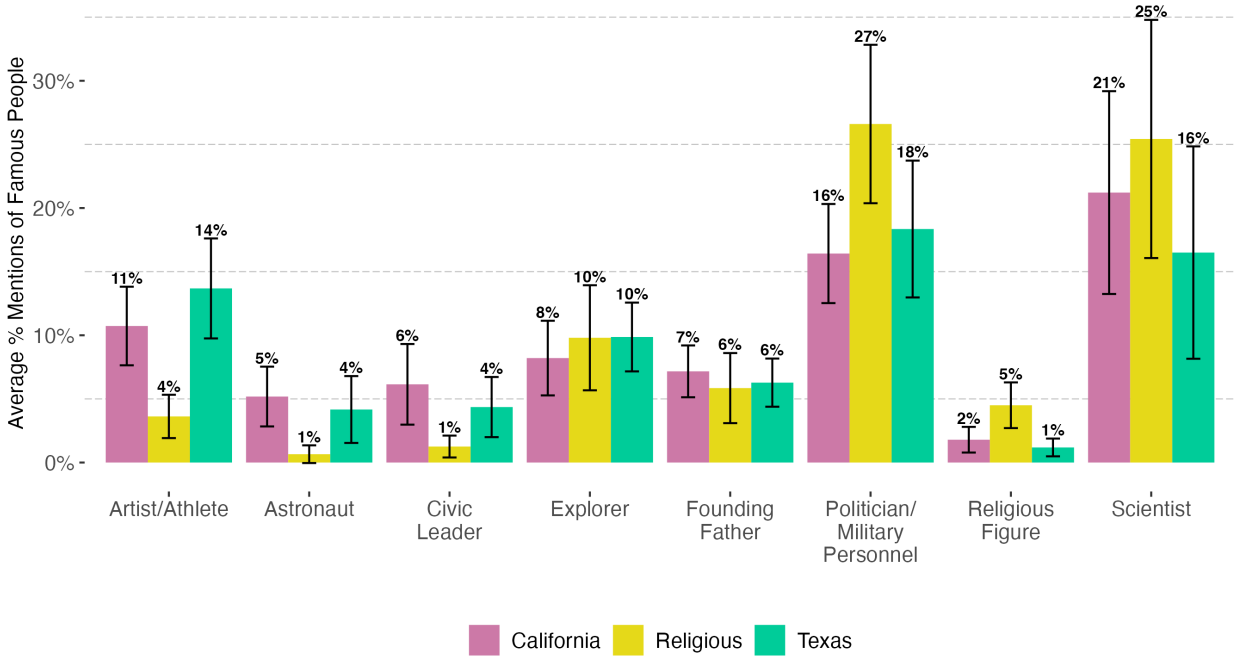
(c) Topics and Values Exclusive to Texas Collection



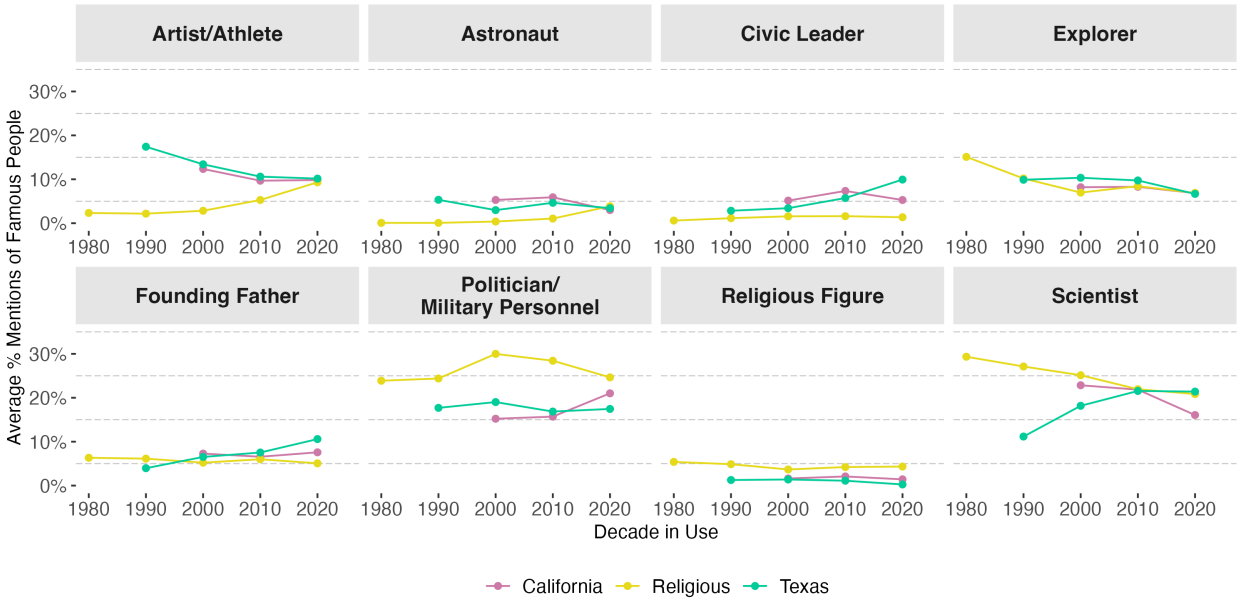
Notes: Each word cloud displays the topics and values metadata tags that appear exclusively within a single collection and not in either of the other two. Word size is proportional to frequency of occurrence across items in that collection. Publisher-specific tags (e.g., A Beka Book, BJU Press, Discovery Education) were excluded prior to generating the clouds.

Figure A.9: Historical Figures, by Occupation

(a) Overall



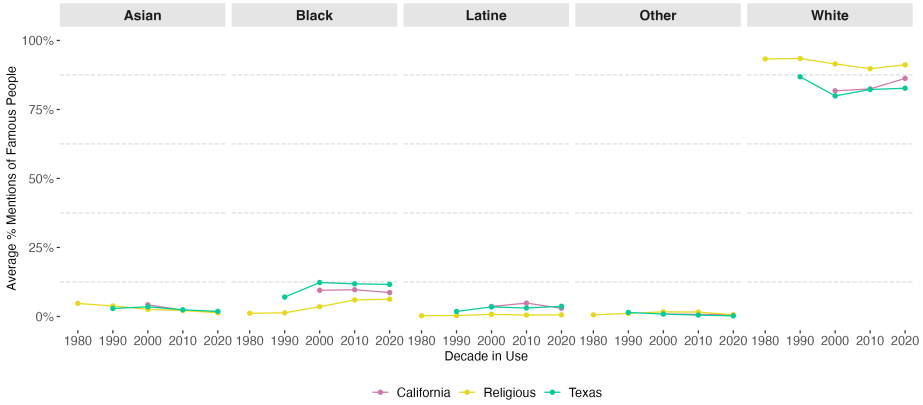
(b) Over Time



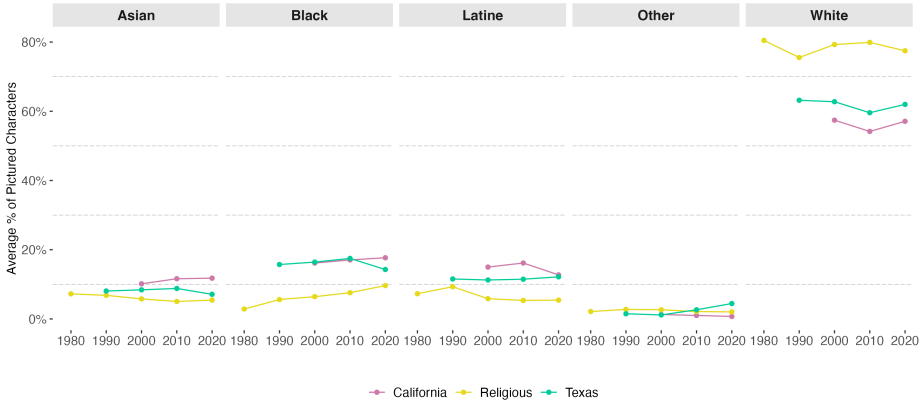
Notes: Panel (a) shows the average percentage of mentions of famous people by their occupation by collection and subject. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection and subject so that each grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals. Similarly, Panel (b) shows the average percentage of mentions of famous people by their occupation by collection and decade in use.

Figure A.10: Representation of Race and Skin Color in Textbooks

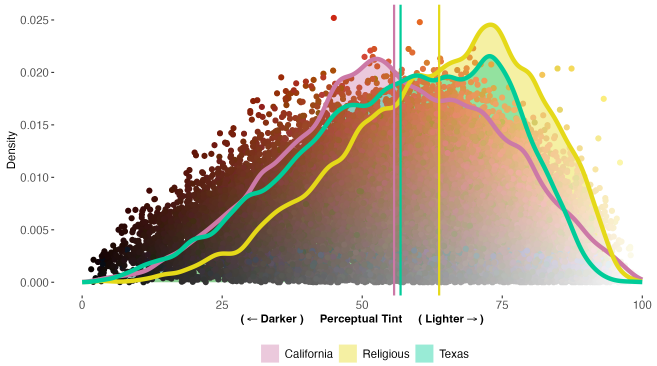
(a) Race (Text)



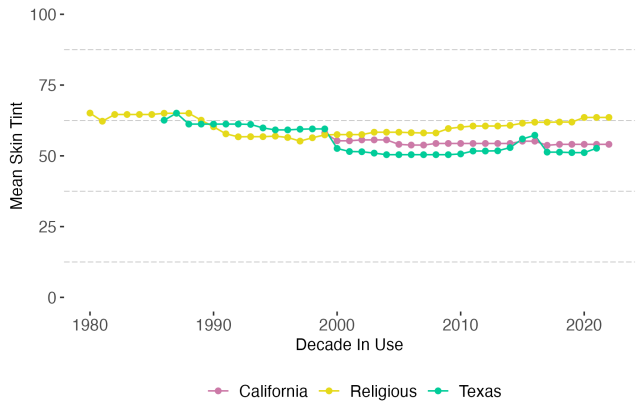
(b) Race (Images)



(c) Skin Tint Distributions



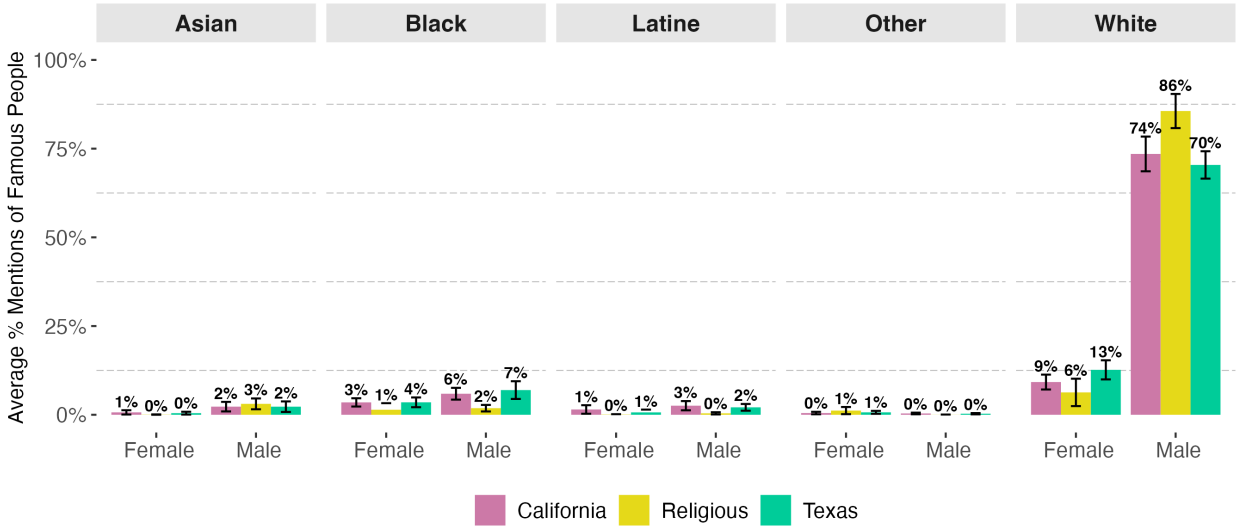
(d) Skin Tint Over Time



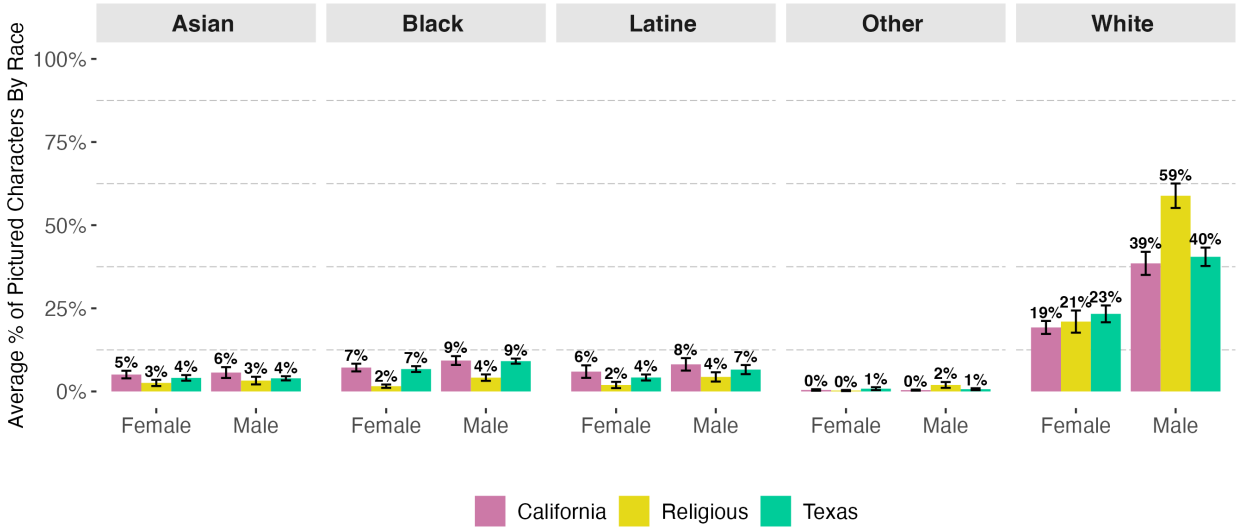
Notes: This figure shows the presence of predicted putative race and skin color. Panel (a) shows the average percent of mentions of famous people in the text that are of a specific race in each collection by decade in use. Panel (b) shows the average percent of pictured characters classified as a specific race in each collection by decade in use. Panels (c) and (d) show our analysis of the representative skin colors of detected faces in images that have human skin colors (polychromatic skin colors where $R \geq G \geq B$) or are from monochromatic images. Panel (c) shows the distribution of skin color tint for faces detected. The x-axis for both the density and scatter plot is skin tint, but the y-axis for the scatterplot is the vibrancy of each skin color. The mean for each distribution is denoted with a dashed line. Panel (d) shows the average skin tint over time. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection and decade so that each subject-grade–year cell receives equal weight.

Figure A.11: Race/Ethnicity and Gender Representation in Text and Images by Collection

(a) Text (Famous Individuals)



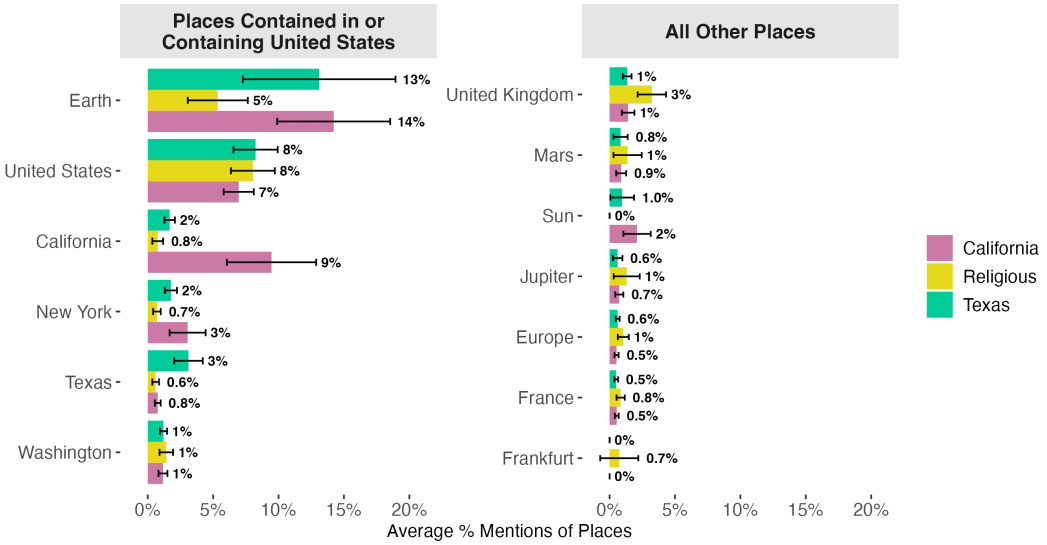
(b) Images



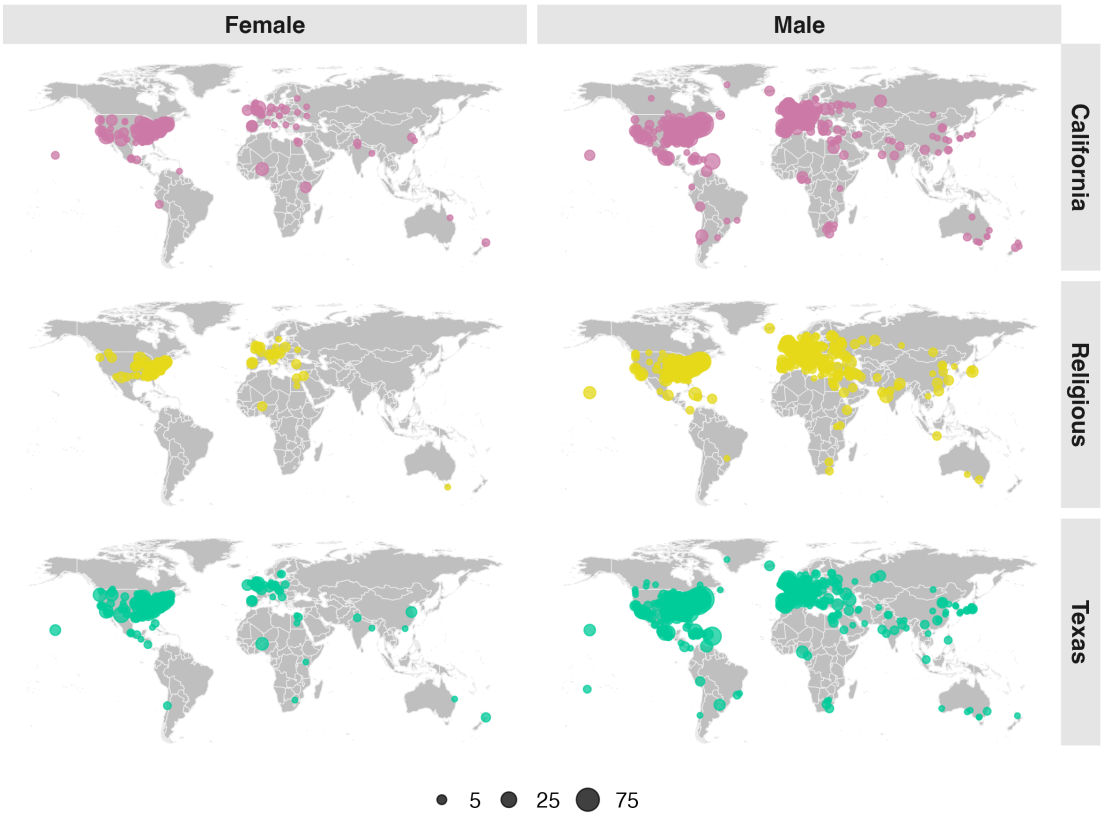
Notes: Panel (a) shows the average percentage of famous people identified as a specific race and gender out of all mentions of famous people by collection. Panel (b) shows the average percentage of pictured characters classified as a specific race and gender. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection so that each subject-grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals.

Figure A.12: Geographic Representation

(a) Top Geographies Mentioned



(b) Birthplaces of Historical Figures by Gender



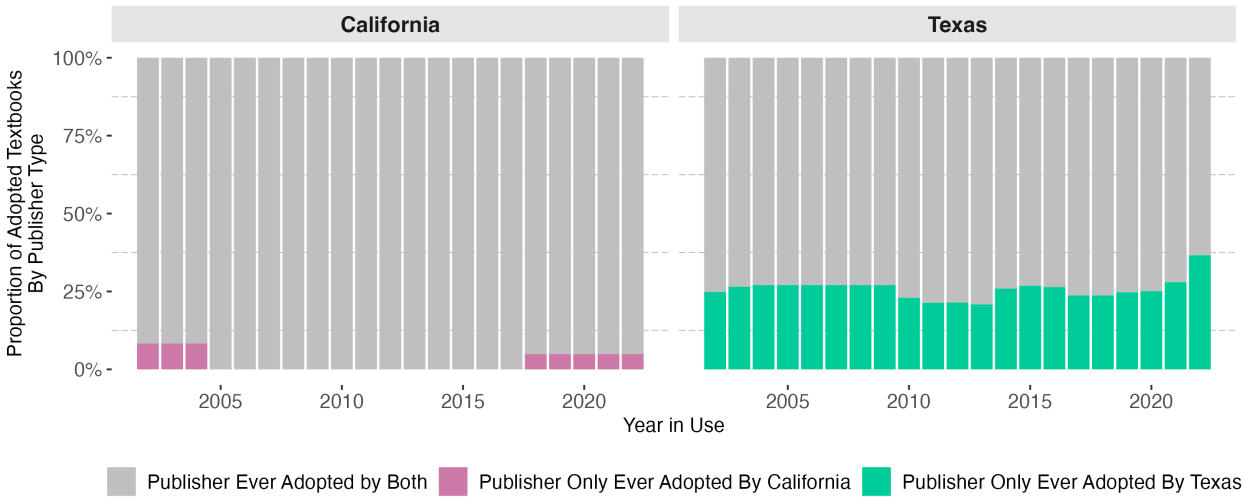
Notes: This figure shows additional measures of geographic representation. In Panel (a), we show the top places tagged as LOC or GPE by NER shown in textbooks by collection. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection so that each subject-grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals. In Panel (b), we show birthplaces of all famous people mentioned in a collection by gender. The size of the point represents the number of mentions.

Figure A.13: Female Representation



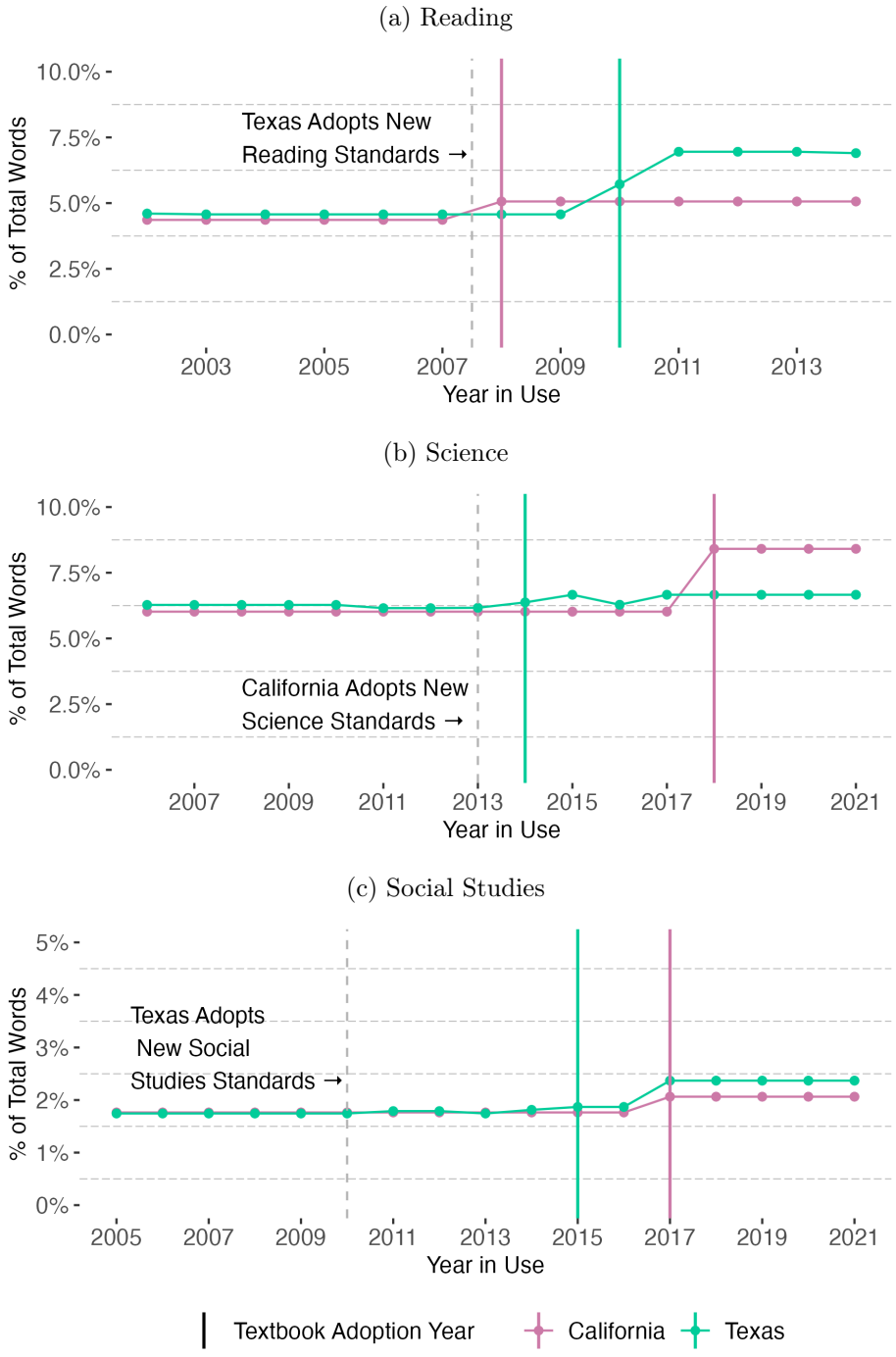
Notes: This figure shows the average percentage of female words and female pictured characters by collection over time (Panels (a) and (b)) and by collection and subject (Panels (c) and (d). Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection × decade cell (Panels (a) and (b)) or within each collection (Panels (c) and (d)) so that each subject–grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals.

Figure A.14: Publisher Presence in Texas and California Textbook Adoptions



Notes: This figure displays the proportion of textbooks in use each year by publisher market presence, separately for Texas and California. Publishers are classified as "Texas Only" if they appear exclusively in Texas adoptions, "California Only" if they appear exclusively in California adoptions, and "Both" if they have books adopted in both states at any point in the sample. Data are drawn from all state textbook adoption records for Reading, Science, and Social Studies subjects in grades 3 and 5 from 2002–2022. Note that we include all adopted textbooks, not just the textbooks digitized in our sample. The top publishers are McGraw-Hill, Scott Foresman, Harcourt, Houghton Mifflin, and Pearson. Examples of publishers with materials adopted by only Texas or only California include Nystrom, Studies Weekly, and Accelerate Learning.

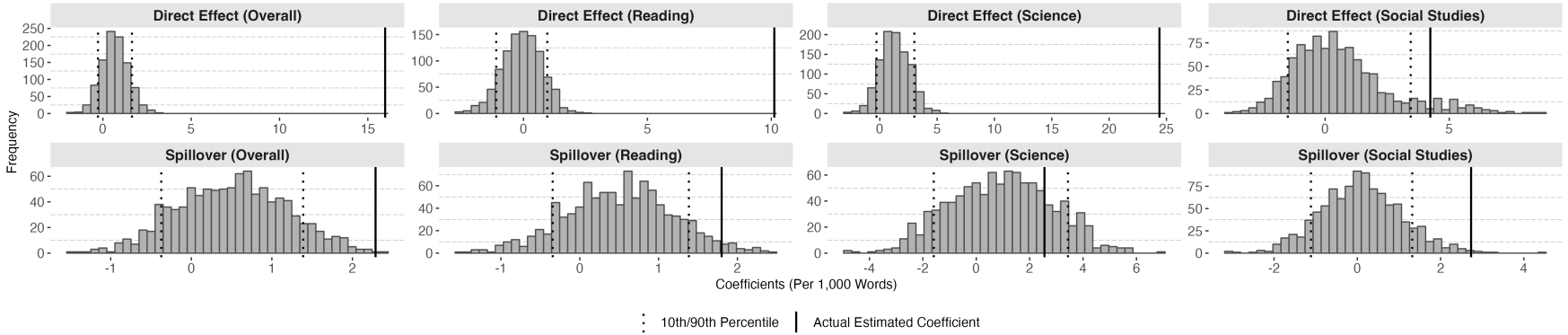
Figure A.15: Standards-Driven Vocabulary in Textbooks over Time, by Subject



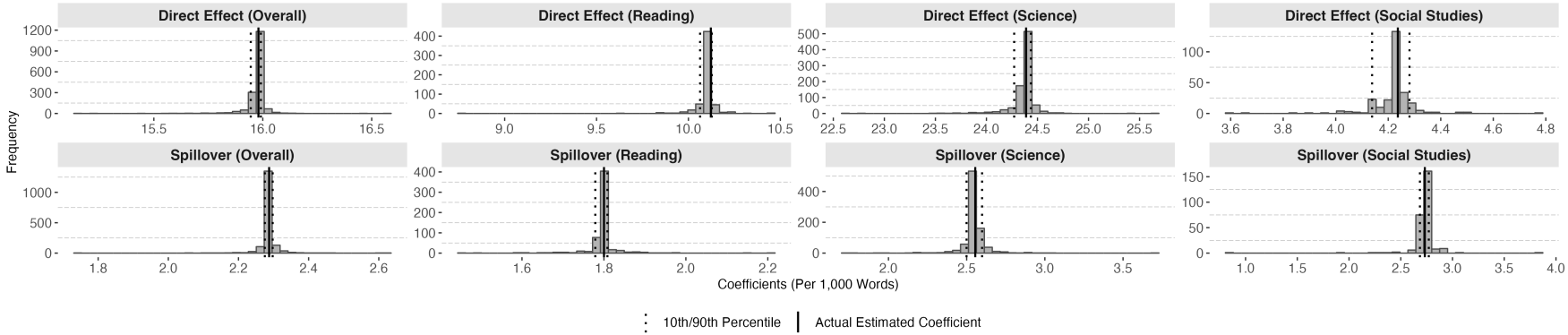
Notes: Each panel plots, by year in use, the share of words in California and Texas textbooks that belong to a dictionary of "new-standards" terms. For each subject and grade, we define this dictionary as new words that appear in the adopting state's revised standards but not in its prior standards and not in the other state's pre-period standards; we then count these terms in textbook text and average the resulting shares across grades by collection. The dashed vertical line marks the standards-revision year, and the solid vertical lines mark each state's subsequent textbook adoption cycle with different states denoted in different colors. The fact that these terms often rise in the non-adopting state after its adoption cycle is consistent with cross-state spillovers. Note that new terms are new to the state standards, not necessarily new to textbooks. For each subject, we subset to the longest time period for which we have one state standard change, and one textbook adoption cycle for each state after the standards change.

Figure A.16: Standards-Driven Vocabulary in Textbooks, Robustness Checks

(a) Placebo Distribution



(b) Leave-One-Out Distribution



Notes: Each panel reports a robustness exercise for the analysis in Table 2. Panel (a) benchmarks the estimated effects against placebo distributions estimated using words drawn from a placebo dictionary of words, matching the size of the true new-standards dictionary. We construct the placebo dictionary of words using vocabulary from the standards-changing state's standards outside the focal subject. Using this placebo dictionary, we randomly draw words and re-estimate the same regression specifications 1,000 times. The histograms show the resulting distribution of placebo coefficient estimates. Coefficients are scaled to represent changes in the number of dictionary words per 1,000 textbook words. Panel (b) assesses whether the estimates are sensitive to individual words in the true new-standards dictionary. In each iteration, we drop one word from the relevant dictionary, recount the remaining words in textbooks, and re-estimate the same specification. Within each panel, solid vertical lines mark the estimates obtained using the true full dictionary. Dotted vertical lines mark the 10th and 90th percentiles of the simulated distribution. In Panel (a), actual estimates that lie meaningfully outside the placebo distribution indicate that the estimated effects are unlikely to arise from random standards vocabulary words or dictionary size alone. In Panel (b), leave-one-out estimates that remain tightly clustered around the full-dictionary estimate indicate that the results are not fragile to dropping specific words.

B Appendix Tables

Table B.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 Studies 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
Total number of pictured characters	43,596	27,342	56,221
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	8	8
Average number of values per page	6	6	6
Weighted Average: Skin Color			
Average skin tint of faces	54	59	53
Weighted Average Percent: Putative Race of Pictured Characters			
Average percent of pictured characters classified as Asian	10%	6%	7%
Average percent of pictured characters classified as Black	15%	6%	15%
Average percent of pictured characters classified as Latine	14%	6%	11%
Average percent of pictured characters classified as White	51%	72%	57%
Average percent of pictured characters classified as other race	1%	2%	2%
Average percent of pictured characters classified as unsure race	9%	9%	8%
Weighted Average Percent: Putative Race of Famous People			
Average percent of famous people classified as Asian	3%	3%	3%
Average percent of famous people classified as Black	9%	3%	10%
Average percent of famous people classified as Indigenous	1%	1%	1%
Average percent of famous people classified as Latine	4%	0%	3%
Average percent of famous people classified as White	83%	92%	83%
Weighted Average Percent: Gender of Pictured Characters			
Average percent of pictured characters classified as female	25%	17%	26%
Average percent of pictured characters classified as male	43%	45%	44%
Average percent of pictured characters classified as unsure gender	32%	38%	30%
Weighted Average Percent: Gendered Words and Famous People			
Average percent of female gendered words	39%	26%	42%
Average percent of famous people classified as female	15%	9%	18%
Weighted Average Percent: Character Type			
Average percent of pictured characters classified as human	72%	63%	72%

Notes: Columns report the three U.S. textbook collections in our sample. The "Collection Totals" panel reports counts aggregated across unique textbooks in each collection. For all averages, we first compute the mean for textbooks within each collection × subject × grade × year-in-use cell and then average these cell means within each collection, so that each subject–grade–year cell receives equal weight. This procedure ensures that textbooks used over longer periods are given greater weight and that equal emphasis is placed on each subject and grade within a given year. Average number of topics and values per page are on conditional on having at least one topic/value identified on a page.

Table B.2: Character Trait Emphasis Across Curriculum

	Texas	California	BJU	Abeka	A.C.E.
Trustworthiness	loyalty, integrity, reliability, punctuality	trustworthiness	—	truthful	trustworthy, truthful, faithful, loyal, honest, sincere, integrity, steadfast, consistent, committed
Responsibility	accountability, perseverance, diligence, self-control	responsibility	—	dependable, diligent, orderly, self-control	dependable, diligent, responsible, perseverant, punctual, self-controlled, prepared, thorough
Caring	kindness, empathy, charity, generosity, patience, consideration, compassion	caring	—	attentive, unselfish, patient, thoughtful, kind, gentle, generous, helpful	affectionate, kind, gentle, loving, compassionate, generous, friendly, thoughtful, considerate, helpful, gracious, forgiving, merciful, affectionate
Citizenship	respect, courtesy, concern for the common good and community, fairness, justice, patriotism, respect for authority and law, gratitude	citizenship	—	thankful, courteous	grateful, patriotic, tolerant, understanding, courteous, flexible, appreciative, tactful
Respect	(in Citizenship)	respect	—	respectful, polite	respectful
Fairness	(in Citizenship)	fairness	—	—	fair, equitable, just
Obedience	—	—	—	obedient	obedient, submissive, meek, orderly
Attitude	—	—	—	cheerful	joyful, peaceful, happy, content, rejoiceful
Religious	—	—	—	—	godly, holy, righteous, reverent
Other	—	—	—	—	discerning, repentant, modest, pure, zealous, virtuous, wise, prudent, creative, resourceful, observant, courageous, thrifty, purposeful, decisive, efficient, industrious, secure, temperate, humble

Note: Character traits emphasized in Texas curriculum are from the Texas Administrative Code, Title 19, Part 2, Chapter 120, §120.2 "Character Traits." California's *Six Pillars of Character* reflect the state's Department of Education guidance on character education posted on its website, based on the CHARACTER COUNTS! framework. Abeka traits are taken from its K5 "Character Development Visuals" lessons. A.C.E. traits are from the "90 Character Traits of Christ" included in A.C.E. curriculum materials.

Table B.3: Top 20 Famous People by Collection

Collection	Name	Race	Gender	Occupation	Sexuality	Rank	Mentions	Textbooks
California	Benjamin Franklin	White	Male	Founding Father	Hetero	1	383	32
California	Thomas Jefferson	White	Male	Founding Father	Hetero	2	322	31
California	George Washington	White	Male	Founding Father	Hetero	3	417	30
California	Abraham Lincoln	White	Male	Politician	Hetero	4	197	25
California	Martin Luther King Jr.	Black	Male	Civic Leader	Hetero	5	133	24
California	John Adams	White	Male	Founding Father	Hetero	6	233	18
California	Abigail Adams	White	Female	Companion	Hetero	7	81	18
California	Frederick Douglass	Black	Male	Civic Leader	Hetero	8	95	17
California	Paul Revere	White	Male	Military	Hetero	9	194	16
California	James Madison	White	Male	Founding Father	Hetero	10	139	16
California	Anne Hutchinson	White	Female	Religious Figure	Hetero	11	88	16
California	Susan B. Anthony	White	Female	Civic Leader	Hetero	12	83	16
California	John Hancock	White	Male	Politician	Hetero	13	67	16
California	Harriet Tubman	Black	Female	Writer	Hetero	14	64	16
California	William Clark	White	Male	Explorer	Hetero	15	57	16
California	Theodore Roosevelt	White	Male	Politician	Hetero	16	32	16
California	Alexander Hamilton	White	Male	Founding Father	Hetero	17	217	15
California	Neil Armstrong	White	Male	Astronaut	Hetero	18	145	15
California	George III	White	Male	Politician	Hetero	19	124	15
California	Patrick Henry	White	Male	Politician	Hetero	20	112	15
Religious	George Washington	White	Male	Founding Father	Hetero	1	310	33
Religious	Abraham Lincoln	White	Male	Politician	Hetero	2	369	27
Religious	Benjamin Franklin	White	Male	Founding Father	Hetero	3	158	25
Religious	Paul Revere	White	Male	Military	Hetero	4	185	19
Religious	Thomas Jefferson	White	Male	Founding Father	Hetero	5	146	18
Religious	Noah Webster	White	Male	Linguist	Hetero	6	280	17
Religious	John F. Kennedy	White	Male	Politician	Hetero	7	102	17
Religious	James Madison	White	Male	Founding Father	Hetero	8	50	16
Religious	Robert E. Lee	White	Male	Military	Hetero	9	132	15
Religious	Daniel Boone	White	Male	Explorer	Hetero	10	103	14
Religious	David Livingstone	White	Male	Explorer	Hetero	11	59	13
Religious	Ulysses S. Grant	White	Male	Politician	Hetero	12	56	12
Religious	Charles the Bald	White	Male	Politician	Hetero	13	45	12
Religious	William Bradford	White	Male	Politician	Hetero	14	31	12
Religious	Rudyard Kipling	White	Male	Writer	Hetero	15	20	12
Religious	John Ball	White	Male	Religious Figure	Hetero	16	12	12
Religious	John Smith	White	Male	Explorer	Hetero	17	123	11
Religious	Woodrow Wilson	White	Male	Politician	Hetero	18	31	11
Religious	Francis Scott Key	White	Male	Writer	Hetero	19	20	11
Religious	Hans C. Andersen	White	Male	Writer	Hetero	20	17	11
Texas	George Washington	White	Male	Founding Father	Hetero	1	531	45
Texas	Benjamin Franklin	White	Male	Founding Father	Hetero	2	377	33
Texas	Abraham Lincoln	White	Male	Politician	Hetero	3	422	31
Texas	Thomas Jefferson	White	Male	Founding Father	Hetero	4	245	31
Texas	Alexander Hamilton	White	Male	Founding Father	Hetero	5	245	27
Texas	Paul Revere	White	Male	Military	Hetero	6	268	25
Texas	Martin Luther King Jr.	Black	Male	Civic Leader	Hetero	7	219	24
Texas	Neil Armstrong	White	Male	Astronaut	Hetero	8	198	24
Texas	Harriet Tubman	Black	Female	Writer	Hetero	9	127	22
Texas	John F. Kennedy	White	Male	Politician	Hetero	10	214	21
Texas	John Adams	White	Male	Founding Father	Hetero	11	175	21
Texas	Theodore Roosevelt	White	Male	Politician	Hetero	12	157	21
Texas	Thomas Edison	White	Male	Inventor	Hetero	13	101	21
Texas	William Clark	White	Male	Explorer	Hetero	14	63	21
Texas	Francis Scott Key	White	Male	Writer	Hetero	15	58	21
Texas	Franklin D. Roosevelt	White	Male	Politician	Hetero	16	454	20
Texas	Langston Hughes	Black	Male	Writer	Hetero	17	68	20
Texas	Meriwether Lewis	White	Male	Explorer	Hetero	18	59	19
Texas	Lyndon B. Johnson	White	Male	Politician	Hetero	19	58	19
Texas	Jane Addams	White	Female	Civic Leader	Lesbian	20	92	18

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

Table B.4: Top Relative Words by Gender

(a) Text

		Mean		
Top Words Associated with Female (vs. Male) Pronouns		V	A	D
California	mother, know, get, come, things, sure, way, couple, go, like	0.67	0.38	0.57
Religious	mother, mom, wife, get, know, daughter, crying, back, sure, tell	0.70	0.44	0.56
Texas	mother, get, know, come, things, sure, back, look, put, go	0.64	0.38	0.57
Top Words Associated with Male (vs. Female) Pronouns		V	A	D
California	cavalry, army, conquest, forces, territory, militia, troops, invasion, military, revolt	0.48	0.72	0.80
Religious	established, military, establish, army, forces, aim, major, notably, believed, history	0.61	0.56	0.75
Texas	army, forces, cavalry, military, territory, troops, armed, predecessor, war, militia	0.43	0.69	0.78

(b) Images

		Mean		
Top Words Associated with Female (vs. Male) Images		V	A	D
California	kids, classroom, children, gardening, decorating, healthy, coloring, homework, mom, sewing	0.76	0.41	0.47
Religious	mother, grandma, grandmother, aunt, mom, comforting, daughter, sister, nana, baby	0.84	0.32	0.52
Texas	mom, kids, children, grandma, homework, adult, kitchen, playroom, decorating, mother	0.80	0.43	0.52
Top Words Associated with Male (vs. Female) Images		V	A	D
California	cavalry, steamship, fort, steamboat, warship, regiment, garrison, lance, militia, republic	0.53	0.62	0.74
Religious	cavalry, regiment, army, militia, infantry, garrison, fort, troops, armored, lance	0.50	0.67	0.78
Texas	cavalry, militia, army, lance, fort, regiment, garrison, cannon, armored, war	0.47	0.68	0.77

Notes: This table shows the top relative words associated with a specific gender as well as the average Valence, Arousal, and Dominance (VAD) scores of those words. In panel (a), we use word embeddings estimated on the textbooks from each collection. In panel (b), we use document embeddings estimated on descriptions of pictured males and females in the textbooks. To find the top relative words associated with females vs. males, we find the words that are closest to females in word embedding space and farthest from the males. We do a similar exercise to find the top relative words associated with males vs. females. Note that we limit our analysis to words that are included in the VAD lexicon.

C Methods Appendix

In this appendix, we discuss in further detail the application of some methods. We note that each tool offers an informative yet partial perspective; since no single tool provides a complete view, employing multiple measures enables broader understanding.

C.A Gemini 2.5 Pro Preview: Character Detection and Feature Classification

We detect characters in our textbooks using a two-stage, fully automated pipeline implemented using Google Cloud Vertex AI and Gemini 2.5 Pro Preview (model version `gemini-2.5-pro-preview-03-25`). In step one, we identify pages containing visual content and separate them from pages containing only text. For each scanned page, the raw image bytes are submitted to the model with a binary classification task: determine whether the page is `TEXT_ONLY` or `CONTAINS_IMAGE`. We configure this stage with a low temperature parameter (0.1) to minimize stochastic variation in model outputs.

In the second stage, we prompt the model to output a JSON object capturing: (i) scene-level data (what is happening or a setting description) and (ii) character-level data (bounding-box coordinates, race, gender, human/non-human, actions, location). We recode the race output into categories (Black, Latine, Asian, White, Other, unsure, not applicable). Additional post-processing steps harmonize labels and remove non-informative outputs.

C.B Image Description Embeddings (Pictured Human Characters)

To compare visual portrayals at scale, we construct embeddings from the LMM-generated text describing each pictured human character. We subset to human characters with non-missing character identifiers and construct one document per pictured character by concatenating the model's descriptions of `actions`, `location`, and `what_is_happening`. We define a unique identifier `page_character_id` by combining page and character identifiers.

Before training embeddings, we standardize multiword expressions using a curated bigram dictionary. Specifically, we map common alias phrases to canonical bigram tokens using a dictionary-based lookup procedure and then reconstruct the cleaned text. We further clean text for embeddings by enforcing UTF-8/ASCII characters, lowercasing, removing isolated single-letter OCR artifacts, normalizing apostrophes, removing numbers, removing punctuation (while retaining apostrophes), and collapsing repeated whitespace.

We then train a 300-dimensional Doc2Vec model using the PV-DM architecture (distributed memory). We set `vector_size=300`, `window=2`, `min_count=30`, and train for `epochs=400` with `hs=1` and `negative=0`; we set `workers=1` and a fixed random seed to support reproducibility. We initialize the model's word vectors using pretrained 300-dimensional embeddings and freeze these pretrained word vectors during Doc2Vec training.[31] The resulting model produces one embedding per `page_character_id`, which we use for (i) book-level averages of character portrayals, (ii) top relative word and VAD analyses, and (iii) domain-association analyses.

[31] We use the Stanford GloVe `glove.2024.wikigiga.300d.zip` embeddings, trained on 2024 Wikipedia and Gigaword 5, available at https://github.com/stanfordnlp/glove.

C.C GPT-4 Preview: Topic and Value Classification

To extract the topics and values on a page, we use OpenAI's API which allows us to define a specific model, a system role, and a *temperature* for each of our prompts. We use the `gpt-4-1106-preview` model. The parameter *temperature* defines the freedom or creativity we allow the model to have for its responses, and is set to 1. Finally, we assign a role to the system to set its behavior. In our analysis, we instruct the model to assume the role of "a knowledgeable and unbiased judge". For each page, we extract (i) a list of topics conveyed in the textbook page, containing no more than 10 topics, and (ii) a list of values conveyed in the textbook page, containing no more than 10 values.

C.D Skin Color

To extract and classify skin color, we begin by identifying the facial area of interest in each detected face by "segmenting" the skin. To do this, we use a face landmarker model,[32] which uses convolutional neural networks to predict and detect facial landmarks and then to isolate the parts of the face which contain skin in order to remove hair, clothing, and other accessories (known as "skin segmentation"). We then extract dominant colors by grouping pixels of the segmented skin into clusters using k-means clustering. We remove the smallest clusters, which tend to contain irrelevant information such as shadows or highlights, and treat the remaining cluster centroids as dominant skin tones. Finally, we collapse segmented skin into a single "representative skin color" by taking a weighted average of dominant skin colors in CIE L*a*b* color space (where weights are determined by cluster size). To construct a continuous measure of how dark or light the skin color of a detected face is, we use the L* value of this representative skin color which ranges from 0 (darkest) to 100 (lightest).

C.E ALC (À La Carte) Word Embeddings (Text)

To characterize gendered context in our textbooks, we construct *à la carte* (ALC) embeddings for gendered pronouns using the `conText` framework, following Khodak et al. (2018) and Rodriguez et al. (2023). We begin with pre-trained 300-dimensional word vectors and estimate an ALC linear transformation matrix from our corpus of textbooks using a weighted word co-occurrence matrix (symmetric window size = 6, with weights that decay with distance).[33] We then apply this transformation to the local contexts of gendered pronouns (*he*, *him*, *his*, *himself*, *she*, *her*, *hers*, and *herself*) to obtain one embedding for each pronoun occurrence. Because these pronoun-instance embeddings are estimated in the same semantic space as the pretrained word vectors, we can use cosine similarity to compare pronoun contexts to individual words and to broader semantic domains.

C.F Top Relative Words

To characterize the contexts most strongly associated with female versus male pronoun or pictured character instances, we identify "top relative words" using the pronoun-instance ALC embeddings or image description embeddings described above. This approach follows a similar methodology outlined in Manzini et al. (2019) and Charlesworth et al. (2022), but in our implementation the analysis is conducted at the instance level rather than using only

[32] We use Google's MediaPipe Face Landmarker.

[33] We use the Stanford GloVe `glove.2024.wikigiga.300d.zip` embeddings, trained on 2024 Wikipedia and Gigaword 5, available at `https://github.com/stanfordnlp/glove` for our pretrained embeddings.

one average embedding per group.

Let v_{ic}^{g} denote the embedding for instance i of gender g in collection c, and let $\bar{v}_c^m$ and $\bar{v}_c^f$ denote the average male and female embeddings within collection c. For a female instance in collection c, we define the relative query vector as:

$$q_{ic}^{f} = \text{norm}\left(v_{ic}^{f} - \bar{v}_c^{m}\right), \tag{C.1}$$

and analogously, for a male instance:

$$q_{ic}^{m} = \text{norm}\left(v_{ic}^{m} - \bar{v}_c^{f}\right), \tag{C.2}$$

where norm$(\cdot)$ denotes length normalization. Intuitively, these relative query vectors capture the portion of an instance's contextual embedding that is distinctive relative to the opposite gender within the same collection.

For each instance, we identify the 50 nearest words to the relevant query vector using cosine similarity. To align this exercise with our sentiment analysis, we restrict candidate neighbors to words that appear in the NRC VAD lexicon (Mohammad, 2018). We then compute the mean valence, arousal, and dominance scores of these 50 nearest words and estimate collection-specific gender differences using regressions of the form

$$Y_{ibct}^{k} = \alpha + \sum_{c} \beta_c \left(Female_i \times \mathbf{1}\{Collection = c\}\right) + \lambda_c + \gamma_t + X_b'\delta + \varepsilon_{ibct} \tag{C.3}$$

where Y_{ibct}^{k} is the mean score for dimension $k \in \{V, A, D\}$ for pronoun or character instance i in book b from collection c used in year t. $Female_i$ is an indicator for female pronouns or pictured characters. λ_c and γ_t denote collection and year-in-use fixed effects, and X_b includes controls for subject and grade. Standard errors are clustered at the book level. The coefficients β_c capture the collection-specific female-minus-male difference in the average VAD score of top relative words. A more detailed description of the VAD lexicon used in this and other analyses is provided in the next section.

C.F.1 *Sentiment (VAD Lexicon)*

To characterize sentiment, we measure three attributes of words: their valence (V), arousal (A), and dominance (D). Valence (V) measures positivity (happy to unhappy), arousal (A) measures affective activation (active to passive), and dominance (D) reflects emotional control (powerful to submissive). To obtain these measures, we match the words in our corpora to the NRC VAD Lexicon, which provides valence, arousal, and dominance scores for over 20,000 English words (Mohammad, 2018).[34]

C.G Group-to-Domain Association

We also examine whether male and female contexts are more closely associated with specific semantic domains. In text, we look at instances of male and female gendered pronouns;

[34]Scores range from 0 to 1. For example, *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 high valence, but *leader* scores higher on arousal and dominance.

in images we look at instances of male and female pictured characters. This analysis is related to the intuition behind the Word Embedding Association Test (Caliskan et al., 2017), but rather than estimating a WEAT test statistic, we compute an instance-level measure of similarity between each embedding and pre-specified domain word lists.

Let W_D denote the set of words belonging to domain D (for example, family, sports, tools, performance arts, home, or politics). For each instance embedding i, we measure association with domain D as the average cosine similarity between the embedding and the words in that domain:

$$s_{iD} = \frac{1}{|W_D|} \sum_{w \in W_D} \cos(v_i, w). \tag{C.4}$$

We then estimate collection-specific gender differences in domain association by estimating the following regression specification:

$$s_{iD} = \alpha_D + \sum_c \lambda_{Dc} \mathbf{1}\{C_i = c\} + \sum_c \beta_{Dc} \left(\mathbf{1}\{C_i = c\} \times Female_i\right) + X_i' \gamma_D + \varepsilon_{iDb}, \tag{C.5}$$

where X_i again includes controls for subject, grade, and year in use, and standard errors are clustered at the book level. Positive values of β_{Dc} indicate that female contexts are more strongly associated than male contexts with domain D in collection c.

C.H Embedding Regressions

Following Rodriguez et al. (2023), we estimate $Y_i = X_i\beta + \varepsilon_i$ where Y_i is a 300-dimensional normalized document embedding for page or image i, and X_i consists of j indicator variables which denote the collection, subject, grade, and copyright year of the book the page or image is from. The estimated coefficient vector $\hat{\beta}_j$ captures the average difference in embedding space between the reference group and the indicated group for a given dimension, holding other covariates constant. We summarize these vector-valued differences across dimensions by their Euclidean norm, $\|\hat{\beta}_j\|$, which measures the overall magnitude of semantic difference in embedding space between groups, aggregated across all embedding dimensions. Larger values indicate greater semantic differences between groups, while values near zero indicate little systematic separation in embedding space. We estimate two separate regressions; one where the Texas collection is the excluded group and one where the Religious collection is the excluded group. We report the normed coefficients from these regressions. These show the estimated differences across collections. Statistical significance is assessed using permutation tests in which textbook characteristics associated with each page embedding are randomly reassigned across books (R = 1,000 permutations), preserving within-book correlation in the outcome. After permutation, the regression is re-estimated, and normed coefficients are recomputed. Two-sided p-values are calculated as the fraction of permuted normed coefficients that are greater than or equal to the observed norm.

C.I Survey Data

Our survey data come from the General Social Survey (GSS), a nationally representative repeated cross-sectional survey administered by NORC at the University of Chicago. Analyses use the restricted-use version of the GSS, which includes state FIPS codes back

to 1975, enabling state-level comparisons not possible with the public release. Because not all variables of interest are available in every survey year, the year range varies by analysis; the specific waves used for each outcome are noted where relevant. To allow weights to contribute equally across years regardless of within-year sample size variation, wtssall was normalized within each survey year prior to pooling. To facilitate comparison across three groups — weekly churchgoers nationally, all residents of Texas, and all residents of California — respondents were duplicated across groups where applicable, with estimates computed separately within each group. Survey weights were applied consistently across all three groups using the survey package in R. These weights correct for unequal probability of selection arising from household size and other design features, and post-stratify to national age, sex, and education distributions. While the GSS is designed for national inference and its post-stratification targets do not correct for demographic composition within individual states, applying consistent weights across all groups ensures methodological comparability and preserves the within-household selection correction that remains valid at the state level. Analyses treat Texas and California estimates as weighted descriptions of GSS respondents in those states rather than fully representative state-level population estimates, and state-level results should be interpreted accordingly.

D Methods Validation Appendix

In this appendix, we show the variety of ways in which we used manual validation to measure the effectiveness of our computational methods in measuring the people and topics present in our sample of educational curriculum.

D.A Feature Classification

In this section, we gauge the effectiveness of our feature classification model by comparing its performance with classifications completed by human annotators. The set of faces used in this validation procedure was randomly sampled from our textbooks and stratified by collection, subject and grade. Each randomly sampled face was randomly assigned three times to one of eight annotators. The annotators then determined whether the face was humanoid and, if so, labeled features of the face such as gender and race. Finally, the image was assigned the modal value out of the three labels, and any three-way ties were labeled as "unsure." The annotators also had the option to mark their answers as "unsure" along with their response. For some features, we incorporate this uncertainty in the modal aggregation. This aligns with the feature classification model's structure, which could also predict "unsure" labels for certain features.

There were 1,080 images chosen for this exercise, each of which were labeled 3 times, resulting in 3,240 image-by-labeler observations. After subsetting these observations to those labeled as containing a human face, and then keeping only the images with 3 labelers, our final validation set for race and gender consists of approximately 900 images.

We show measures of model accuracy and inter-labeler reliability shown in Table D.1. To calculate the accuracy, we take the modal human response for each image as the true value and subsequently find the percentage of model predictions equal to said value. To calculate the inter-labeler reliability, we use the Fleiss' kappa ($\hat{\kappa}$) measure commonly used to find the agreement between categorical labels given by three manual labelers. $\hat{\kappa} > 0$ means

labelers agree more often than would be expected by random chance and $\hat{\kappa}$ of 1 means perfect agreement. In particular, we apply the following formula:

$$\hat{\kappa} = \frac{I_o - I_e}{1 - I_e} \tag{D.1}$$

where I_o is the observed agreement and I_e is the expected value based on chance.

Table D.1 shows that human agreement is consistently higher for gender than for race: without "unsure" labels, $\hat{\kappa} = 0.88$ for gender versus 0.65 for race, and when "unsure" is included, $\hat{\kappa}$ declines to 0.72 for gender and more sharply to 0.45 for race. This indicates that race assignments are intrinsically more ambiguous and more sensitive to rater uncertainty in this labeling task. Model accuracy mirrors this pattern: both Gemini and FDAI achieve higher agreement with the modal human label on gender than on race, and both exhibit larger performance drops when "unsure" labels are retained, especially for race. For this reason, in our main analysis we drop Gemini-detected characters that have both an unsure race and an unsure gender prediction.

Table D.1: Accuracy and Inter-Labeler Reliability on Sample of Detected Faces

(a) Without Unsure Labels

	Fleiss Kappa		Gemini		FDAI	
	$\hat{\kappa}$	N	Accuracy	N	Accuracy	N
Human	0.81	1,080	0.95	722	0.89	1,080
Race	0.65	602	0.79	415	0.57	602
Gender	0.88	779	0.81	536	0.65	779

(b) With Unsure Labels

	Fleiss Kappa		Gemini		FDAI	
	$\hat{\kappa}$	N	Accuracy	N	Accuracy	N
Human	0.81	1,080	0.95	722	0.89	1,080
Race	0.45	899	0.68	612	0.48	899
Gender	0.72	900	0.79	611	0.63	900

Notes: Accuracy is calculated as the percentage correct between the modal value of human labels and the predicted value for a given feature. Inter-labeler reliability was calculated using Fleiss' kappa, a measure of agreement between categorical labels assigned by three manual labelers. $\hat{\kappa} > 0$ means labelers agree more often than would be expected by random chance and $\hat{\kappa}$ of 1 means perfect agreement. We repeat calculations with and without unsure labels. Race and gender labels were only provided for faces labeled as Human. Note that the Gemini and FDAI predictions were merged using a fuzzy bounding box matching. There is likely some error in this process, which would cause us to underestimate the accuracy of the Gemini predictions.

Overall, the results suggest that lower race accuracy reflects, in part, lower underlying human consensus rather than purely model error, while gender labeling is comparatively stable for both humans and models. Conditional on the set of characters that have a face detected, Gemini predictions consistently outperform the FDAI model predictions.

Note that for the Gemini model this is only the accuracy of characters in our sample that **also** had a face detected and not the overall accuracy of the model. We merged detected characters and detected faces using the largest overlap of bounding boxes, which may introduce some error. We were unable to match all detected faces from FDAI to detected characters using this strategy which is why the sample sizes are smaller for our validation exercise for the Gemini models. Our FDAI face detection model detected 59,769 faces and the Gemini model detected 137,718 characters (over twice as many). This is in part because Gemini finds more characters who do not have faces visible and more non-human characters. For this reason, the Gemini characters that do not also have a face detected are 3.5 times more likely to be non-human, 4.5 times more likely to have an unsure gender label, and 6 times more likely to have an unsure race label. To assess the accuracy of the Gemini predictions on the total sample of detected characters and not just the characters with a detected face, we manually labeled a random sample of Gemini-detected characters from our textbooks, stratified by collection, subject and grade. Each sampled character was randomly assigned to three annotators drawn from a pool of eight; we then measured inter-labeler reliability, modal labels, and accuracy of Gemini predictions relative to modal labels. The results are shown in Table D.2. The table shows that Gemini performs well on the full detected-character sample: excluding unsure labels (as we do in our analyses), accuracy is 97% for human classification, 89% for race, and 92% for gender, with lower but still substantial accuracy when unsure labels are retained.

Table D.2: Accuracy and Inter-Labeler Reliability On a Sample of Detected Characters

(a) Without Unsure Labels

	Fleiss Kappa		Gemini	
	$\hat{\kappa}$	N	Accuracy	N
Human	0.69	183	0.97	183
Race	0.58	70	0.89	70
Gender	0.63	107	0.92	107

(b) With Unsure Labels

	Fleiss Kappa		Gemini	
	$\hat{\kappa}$	N	Accuracy	N
Human	0.65	193	0.92	193
Race	0.53	104	0.73	104
Gender	0.72	193	0.86	193

Notes: Accuracy is calculated as the percentage correct between the modal value of human labels and the predicted value for a given feature. Inter-labeler reliability was calculated using Fleiss' kappa, a measure of agreement between categorical labels assigned by three manual labelers. $\hat{\kappa} > 0$ means labelers agree more often than would be expected by random chance. $\hat{\kappa}$ of 1 means perfect agreement. We repeat calculations with and without unsure labels. Race labels were only provided for faces labeled as Human.

D.B Image Descriptions

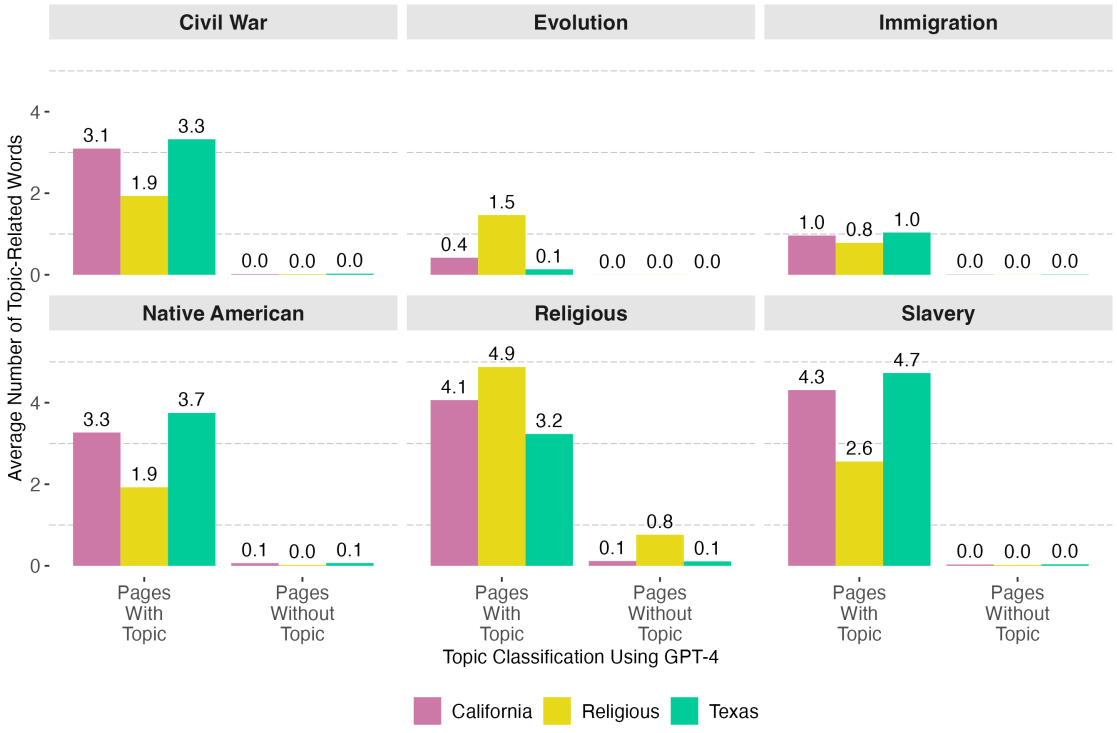
To assess the accuracy of the LMM-generated descriptions of character actions, locations, and setting, we use the same random sample of detected characters used to evaluate Gemini's race, gender, and human classification labels as described above. We asked three labelers per image to evaluate the accuracy of each image description type (e.g. action, location, and setting) on a five-point scale (1 = not accurate, 5 = accurate). Across all three description types, the vast majority of modal ratings per image were 5 (most accurate): 97% for action

descriptions, 97% for location descriptions, and 98% for what is happening descriptions. Fewer than 4% of modal ratings across any category fell below 5, and virtually no modal ratings fell below 4. We conclude from these results that the LMM-generated descriptions are sufficiently accurate for use in downstream embedding analyses.

D.C Topic Classification

First, we manually validate the topic classification performed by GPT using token counts of topic-related words. In Figure D.1, we show that on average, pages that were not classified as containing specific topic also do not have any mentions of those topic-related words on the page. For example, pages that were not classified as containing topics related to "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 yields examples such as a passage which discusses Martin Luther King, Junior and mentions that he was a pastor. On the corresponding page, there is a religious word, but religion is not a topic.[35]

Figure D.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 pages that GPT classified as containing the topic; "FALSE" indicates pages not classified as containing the topic.

[35]Note that lexicon-based approaches may miss relevant terms and separately may include terms with multiple meanings that do not reflect the domain of interest.

Next, we validate the LLM-generated topic and value classifications using manual human labels. Following the validation exercises above, we draw a random sample of pages stratified by subject, grade, and collection. Each page is independently reviewed by three manual labelers. The validation task is designed to assess both whether the LLM correctly identifies topics and values present on a page and whether it omits relevant ones. For each page, labelers are shown five candidate topics and five candidate values. Each set includes a random mixture of "true" items, defined as topics or values predicted by the LLM for that page, and "distractor" items, drawn without replacement from topics or values the LLM identified elsewhere in the same book but not on that page. The number of true items shown is drawn from a normal distribution with mean 3 and standard deviation 1, rounded to the nearest integer and truncated to fall between 0 and 5; the remaining slots are filled with distractors. To make responses comparable across labelers and to calculate inter-rater reliability, we fix the random draw at the page level: all three labelers assigned to a page see the same five topics and five values in the same randomized order. For each item, labelers indicate whether it is present on the page. We then compare the modal human label to the LLM prediction. Accuracy is defined as the share of items for which the LLM classification matches the modal human label. Table D.3 shows that the LLM classifications align closely with human judgments. For topics, the modal human label matches the LLM prediction in 89% of validation items, and inter-labeler agreement is high, with a Fleiss' kappa of 0.77. This suggests that topic classifications are both reliable across human labelers and accurate relative to the human benchmark. For values, the modal human label matches the LLM prediction 80% of the time and inter-labeler agreement is more moderate, with a Fleiss' kappa of 0.58. This lower agreement is consistent with the greater subjectivity of identifying values conveyed by a page, relative to identifying more concrete topics. Overall, the validation exercise indicates that the LLM performs sufficiently well in classifying both topics and values.

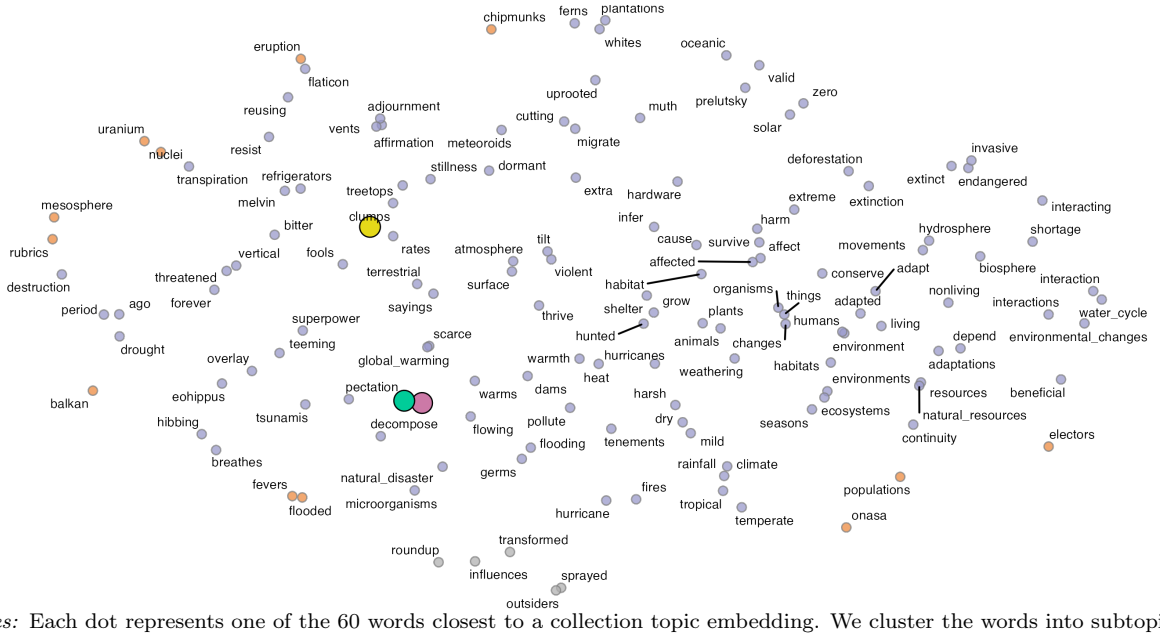
Table D.3: Accuracy and Inter-Labeler Reliability of Topics and Values on Textbook Pages

	Fleiss Kappa		GPT	
	$\hat{\kappa}$	N	Accuracy	N
Topics	0.77	540	0.89	540
Values	0.58	530	0.80	530

Notes: Inter-labeler reliability is measured using Fleiss' kappa ($\hat{\kappa}$), which captures agreement across the three labelers assigned to each page. $\hat{\kappa} > 0$ means that labelers agree more often than would be expected by random chance. $\hat{\kappa}$ of 1 means perfect agreement. Accuracy is the share of candidate topics or values for which the GPT prediction matches the modal human label.

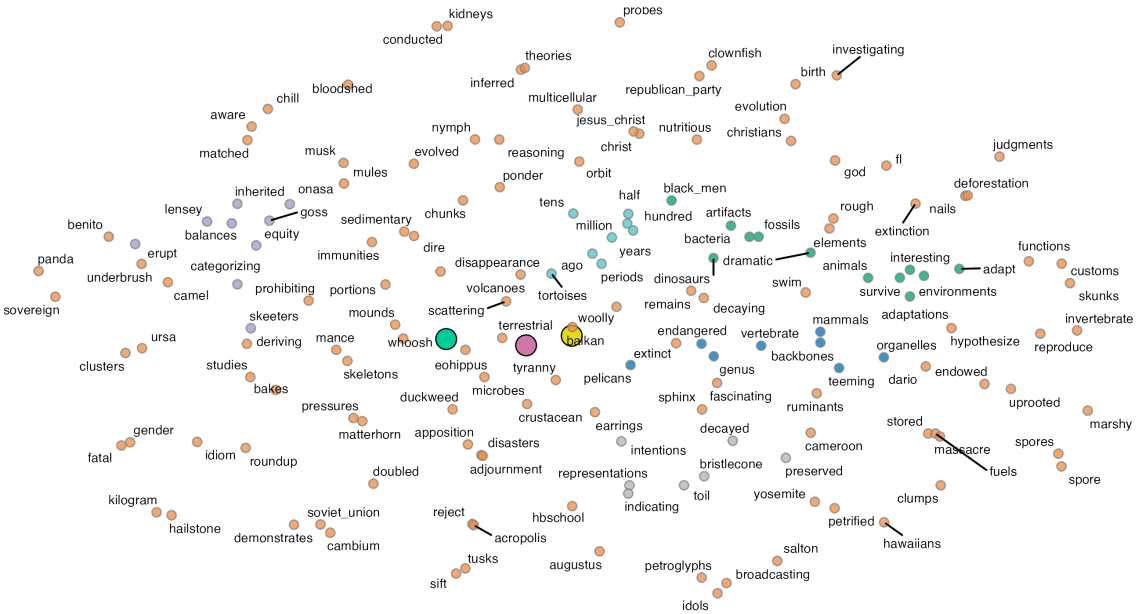
E Topic Embeddings

Figure E.1: Climate Change: Topic Embeddings (Top Words)



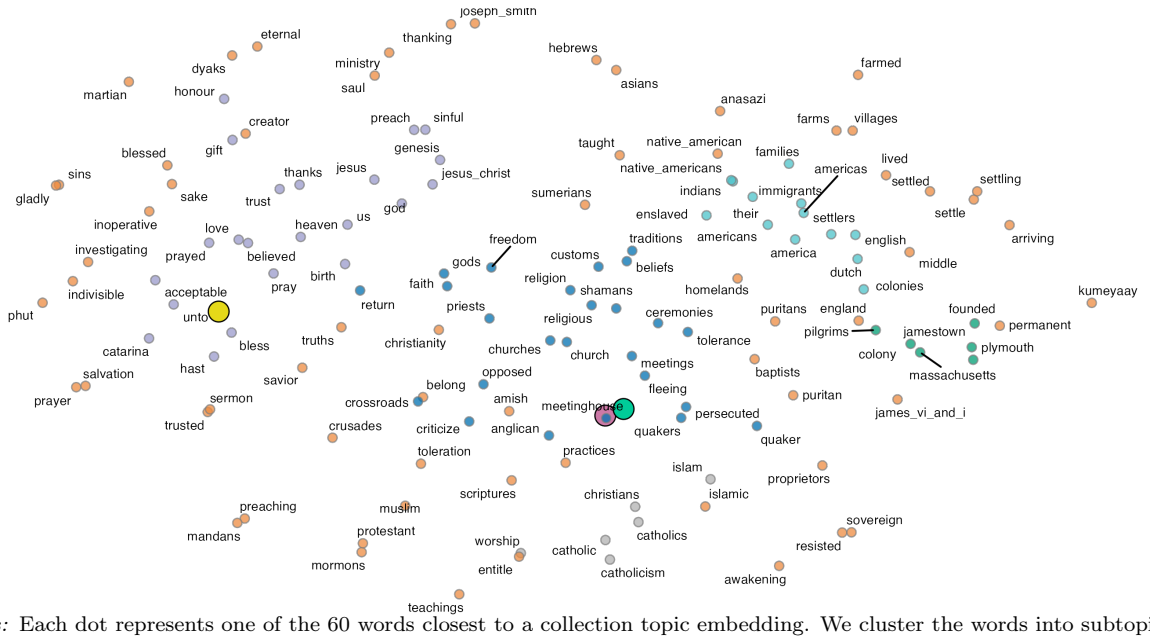
Notes: Each dot represents one of the 60 words closest to a collection topic embedding. We cluster the words into subtopics and denote these by color. Each large dot represents a collection topic embedding; pink for California, yellow for Religious, and green for Texas. Details on how these embeddings were estimated are discussed in Section IV. These embeddings are normalized and then reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and embedding observations should be interpreted in terms of their relative positions.

Figure E.2: Evolution: Topic Embeddings (Top Words)



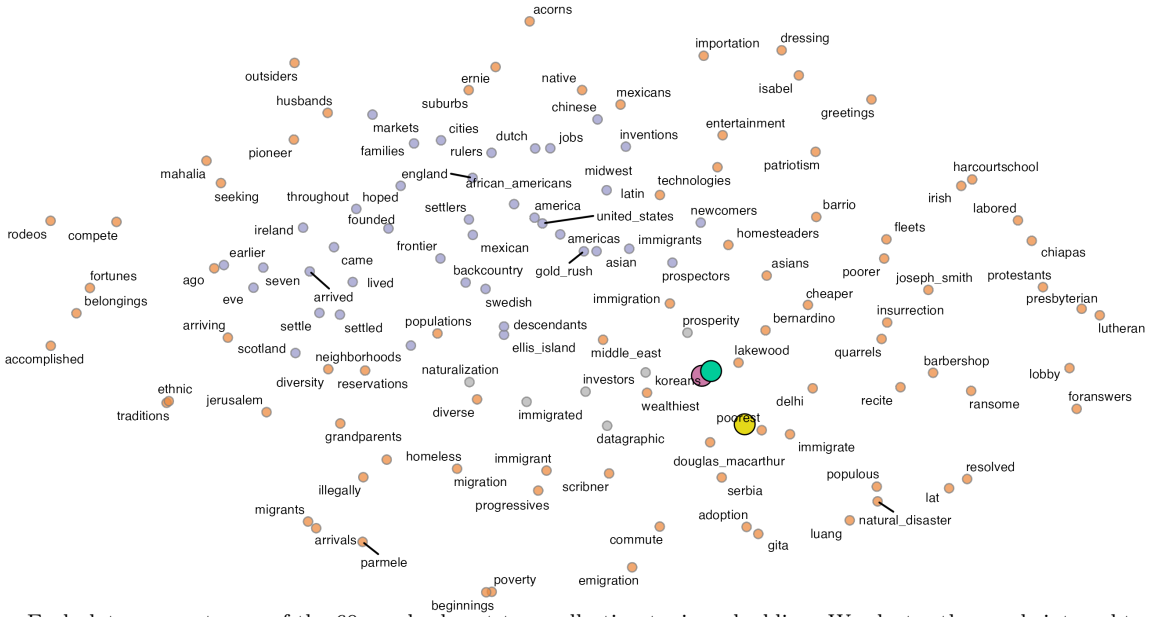
Notes: Each dot represents one of the 60 words closest to a collection topic embedding. We cluster the words into subtopics and denote these by color. Each large dot represents a collection topic embedding; pink for California, yellow for Religious, and green for Texas. Details on how these embeddings were estimated are discussed in Section IV. These embeddings are normalized and then reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and embedding observations should be interpreted in terms of their relative positions.

Figure E.3: Religious: Topic Embeddings (Top Words)



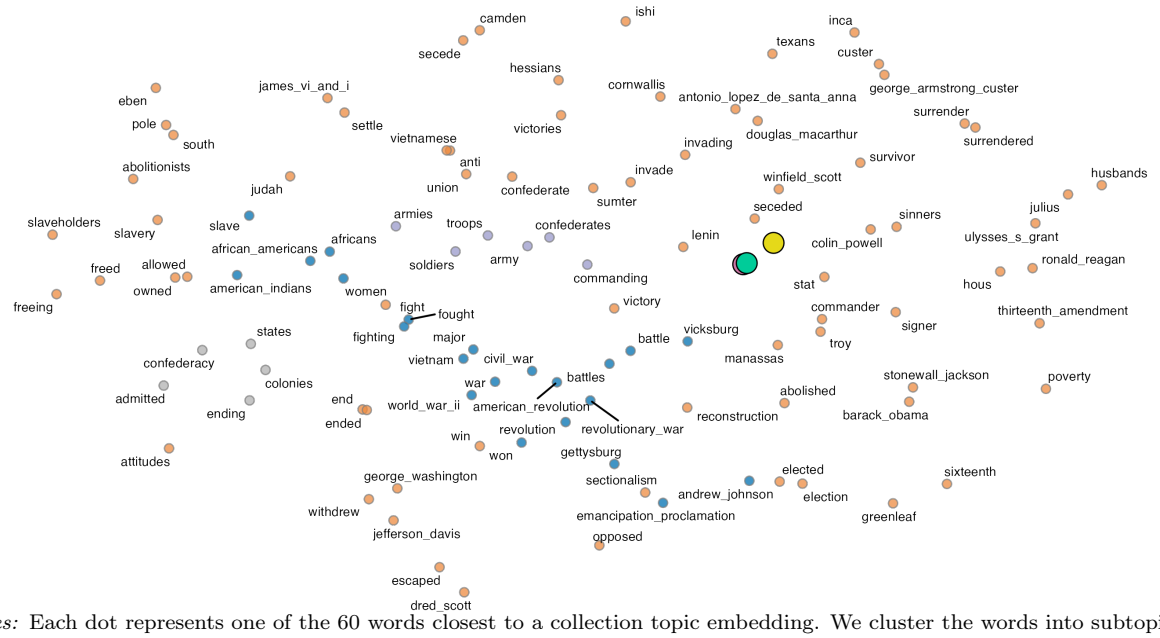
Notes: Each dot represents one of the 60 words closest to a collection topic embedding. We cluster the words into subtopics and denote these by color. Each large dot represents a collection topic embedding; pink for California, yellow for Religious, and green for Texas. Details on how these embeddings were estimated are discussed in Section IV. These embeddings are normalized and then reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and embedding observations should be interpreted in terms of their relative positions.

Figure E.4: Immigration: Topic Embeddings (Top Words)



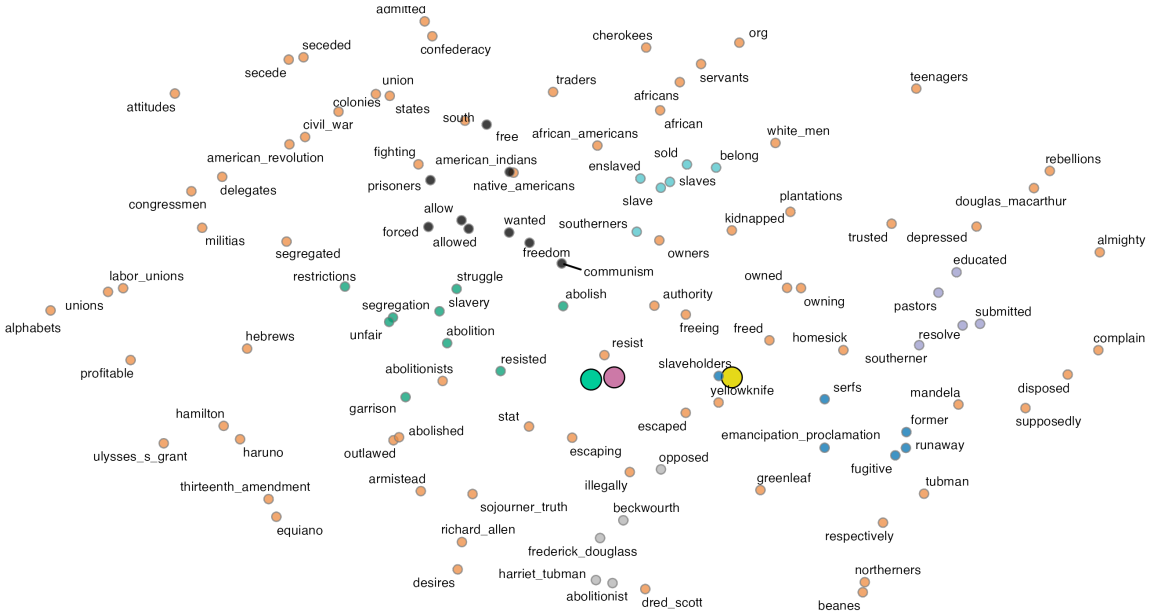
Notes: Each dot represents one of the 60 words closest to a collection topic embedding. We cluster the words into subtopics and denote these by color. Each large dot represents a collection topic embedding; pink for California, yellow for Religious, and green for Texas. Details on how these embeddings were estimated are discussed in Section IV. These embeddings are normalized and then reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and embedding observations should be interpreted in terms of their relative positions.

Figure E.5: Civil War: Topic Embeddings (Top Words)



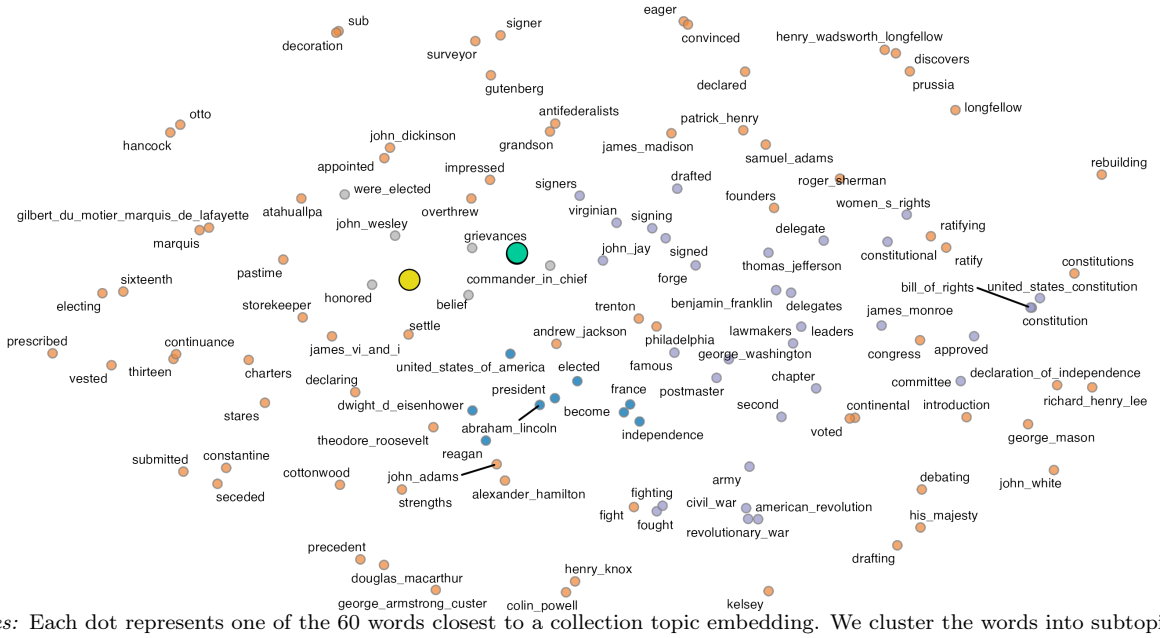
Notes: Each dot represents one of the 60 words closest to a collection topic embedding. We cluster the words into subtopics and denote these by color. Each large dot represents a collection topic embedding; pink for California, yellow for Religious, and green for Texas. Details on how these embeddings were estimated are discussed in Section IV. These embeddings are normalized and then reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and embedding observations should be interpreted in terms of their relative positions.

Figure E.6: Slavery: Topic Embeddings (Top Words)



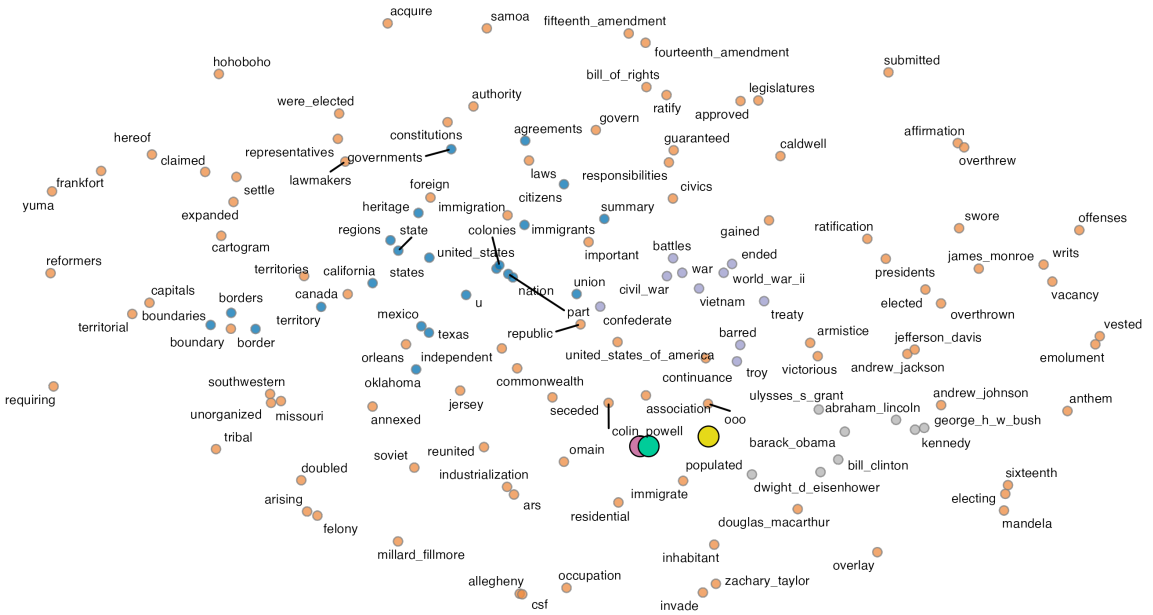
Notes: Each dot represents one of the 60 words closest to a collection topic embedding. We cluster the words into subtopics and denote these by color. Each large dot represents a collection topic embedding; pink for California, yellow for Religious, and green for Texas. Details on how these embeddings were estimated are discussed in Section IV. These embeddings are normalized and then reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and embedding observations should be interpreted in terms of their relative positions.

Figure E.7: Founding Fathers: Topic Embeddings (Top Words)



Notes: Each dot represents one of the 60 words closest to a collection topic embedding. We cluster the words into subtopics and denote these by color. Each large dot represents a collection topic embedding; pink for California, yellow for Religious, and green for Texas. Details on how these embeddings were estimated are discussed in Section IV. These embeddings are normalized and then reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and embedding observations should be interpreted in terms of their relative positions.

Figure E.8: United States: Topic Embeddings (Top Words)



Notes: Each dot represents one of the 60 words closest to a collection topic embedding. We cluster the words into subtopics and denote these by color. Each large dot represents a collection topic embedding; pink for California, yellow for Religious, and green for Texas. Details on how these embeddings were estimated are discussed in Section IV. These embeddings are normalized and then reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and embedding observations should be interpreted in terms of their relative positions.