(How) Do We Teach Emotions?*

Anjali Adukia Matthew Bonci Paula Dastres Gallardo
Emileigh Harrison Jake Nicoll Teodora Szasz

February 28, 2026
Preliminary Working Draft. Please do not cite or distribute.

Abstract

Children learn about emotional intelligence not only through interpersonal interactions but also through the norms and values transmitted in the media and educational materials they consume. These materials send messages about the emotional reactions that are and are not appropriate. This study applies computational methods combining natural language processing and computer vision to conduct the first large-scale, systematic examination of emotional representation across textual and visual modalities in two foundational sources of childhood education: public school textbooks and widely circulated children's literature. Our analysis reveals a stark mismatch between the emotions children read about and the ones they see in images. We find that textual context exposes children to a diverse emotional landscape, including happiness, sadness, anger, and calm, all with relatively balanced frequency. In contrast, pictured characters overwhelmingly display happiness and calmness, while emotions commonly classified as "negative" rarely appear. This pattern persists across time, genre categories, and demographic subgroups in our corpus. Within-page analysis demonstrates that regardless of which emotions are discussed in accompanying text, images on that same page disproportionately depict positive affective states, revealing a systematic text-image incongruence in emotional pedagogy. We further present evidence using individual-level book purchases and library-level inventory data suggesting that the overrepresentation of happiness and calmness in visual content reflects supply-side responses to consumer preferences.

Keywords: Culture, Emotions, Curriculum Analysis, Education Policy
JEL Category: I20, I21, I28, J15, Z13

*Affiliation: University of Chicago, University of Wisconsin-Madison, University of Texas at Austin. For helpful feedback, we thank Alex Eble, James Evans, Yuan Chang Leong, and seminar participants at Notre Dame, Zurich AI + Economics, Hong Kong AI Research in the Social Sciences and Humanities, Association for Education Finance and Policy, the Society for Research on Educational Effectiveness. For technical and research assistance, we thank Ping-Jung Liu. For financial support, we thank the 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. Researcher(s)' own analyses calculated (or derived) based in part on data from Market Track, LLC dba Numerator and marketing databases provided through the Numerator Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Numerator data are those of the researcher(s) and do not reflect the views of Numerator. Numerator is not responsible for and had no role in analyzing or preparing the results reported herein.

Introduction

Emotional experience is a universal human phenomenon; however, the manner in which individuals express emotions is highly variable and often context-dependent. The appropriateness of emotional expression is influenced by individual objectives as well as sociocultural expectations. Psychological research indicates that recognizing, understanding, and expressing a wide range of emotions fosters improved outcomes (Gross and John, 2003; Kelley et al., 2019), yet societal and labor market norms may discourage the open display of certain emotions, particularly those of negative valence, leading to the suppression of such feelings (Hochschild, 1983; Deming, 2017). Children acquire norms of emotional expression through multiple channels, including interactions with parents, peers, role models, various forms of media, and educational materials. These norms are transmitted both visually, through observation of others, and verbally, through spoken or written messages regarding what is considered appropriate in specific contexts. Given the psychological and economic implications, it is important to understand how societal norms and values regarding emotions are communicated to children through both formal and informal means.

In this study, we examine whether and how emotions are communicated through children's content presented in formal and informal settings: (i) three decades of public school textbooks from California and Texas and (ii) a century of award-winning children's books often found in homes, libraries, and schools. The textbooks have been used by children across two of the largest U.S. states and include books for third and fifth graders for the subjects of Reading, Science, and Social Studies. The award-winning children's books are highly influential; Newbery and Caldecott award-winning books are purchased five times more frequently and borrowed from libraries four times more often than other children's books (Smith, 2013; Cockcroft, 2018; Adukia et al., 2023). Using these two sets of influential children's media, we analyze temporal variation, character-level differences, and contextual patterns in emotional representation, exploiting computational tools' capacity for systematic measurement at scale compared to traditional manual content analysis (Neuendorf, 2016; Krippendorff, 2018).

To extract data on emotions from books, we apply artificial intelligence methods from natural language processing (applied to text) and computer vision (applied to images). We employ large multimodal models (LMMs) to classify eight primary emotions in the text and images, respectively: anger, calm, confusion, disgust, fear, happiness, sadness, and surprise.[1] To assess classification robustness, we compare outputs from our primary models

[1]This set of emotions was derived from the six basic emotions identified by Ekman, Sorenson and Friesen (1969) with two added emotions: calm and confusion, particularly relevant to children's media contexts.

with alternative approaches, including a lexicon-based emotion classification method applied to text, a RoBERTa model fine-tuned on an emotion-specific text data set, and a pre-trained convolutional neural network (CNN) used for feature-based emotion classification in images. We examine heterogeneity by context (e.g. corpora, state, grade, subject) and character identities (e.g. putative race, skin color, gender, and age).

Our analysis reveals a systematic divergence between textual and visual emotional content: written text exposes children to a broad emotional range including happiness, sadness, anger, and fear. By contrast, images almost exclusively depict emotions related to happiness and calm while rarely representing emotions traditionally categorized as "negative." These patterns persist across time periods, contexts, and demographic groups both in text and in images.

Moreover, we find a striking misalignment between the emotions portrayed in the text and images on the same page. Regardless of which emotions characters display in the text, images disproportionately depict characters as calm or happy. This demonstrates that the discrepancy between emotional representation in text and images does not simply reflect choices by illustrators to highlight certain positive components of their narratives; rather, the disconnect reflects a genuine incongruence between the emotions that are written about and the emotions that are shown. Almost uniformly, the books illustrate that, no matter what emotion a character experiences in the text, the appropriate response is to display a calm or happy outward appearance.

A natural question that arises is: why do we see these patterns? By merging information about the emotions in children's books with purchaser decisions – both of households and public libraries – we find that, controlling for book attributes in addition to award and year fixed effects, purchasers are less likely to choose children's books whose covers depict emotions traditionally considered "negative." This suggests that the lack of a range of emotions in images might be driven by consumer demand and reflects the societal value placed on displays of "positive" emotions. As a result, these emotions capture attention and make books look more appealing (Bloch, 1995; Underwood and Klein, 2002; Pieters and Wedel, 2004). Additionally, if consumers have limited bandwidth to accurately assess the quality of emotional representation in books and the messages they convey, then authors and illustrators could be incentivized to oversupply happy and calm images in equilibrium even relative to the underlying preference for these emotions (Ellison, 2006; Cusumano, Fabbri

Given that we are predicting these emotions in the text and images of children's media, we also allow for the model to predict "unsure" or "no emotion" to allow for cases that are not obvious or cases where a character is detected but their emotion is not visible.

and Pieroth, 2024).

Our findings contribute to multiple strands of literature. First, we expand prior research on emotional representation in children's books (Nikolajeva, 2014; Ku, 2019; Kardum, Dadic and Horvat, 2021), which has largely relied on manual content analysis and focused narrowly on text. Second, we add to the literature on cultural economics, which has recently rapidly expanded through the use of novel data sources to study naming conventions (Bazzi, Fiszbein and Gebresilasse, 2020), folklore (Michalopoulos and Xue, 2021), movies (Michalopoulos and Rauh, 2024), style (Yanagizawa-Drott and Voth, 2023), and now emotions. Third, we inform literature on non-cognitive skill development (Groves, 2005; Heckman, Stixrud and Urzua, 2006; Duckworth et al., 2007; Lundberg, 2017; Kautz et al., 2014), highlighting how curricular materials can shape early emotional intelligence. Fourth, we illustrate how books can effectively advertise themselves to consumers with limited attention to a book's content (Dröge and Darmon, 1987; Eliaz and Spiegler, 2011), in this case, through the emphasis of certain emotions. Finally, we advance computational social science methodology by demonstrating how natural language processing and computer vision tools can be leveraged for large-scale analyses of text and image data (Ash et al., 2021; Adukia et al., 2023; Ludwig and Mullainathan, 2024; Adukia and Harrison, 2025), specifically when examining emotional content in the text and images of children's media: public school textbooks from the last three decades and influential books from a century of children's literature.

The paper proceeds as follows: Section I discusses background related to the study and teaching of emotions. Section II details our data sources. Section III explains the processes used to transform the images and text into data, along with the tools employed for analysis. The main results are presented in Section IV. In Section IV.D, we further probe the apparent contrast between emotions in text and images by considering both media on the same page simultaneously. In Section V, we explore how consumer demand may contribute to the patterns we uncover. Finally, we offer a discussion in Section VI.

I Background

Emotions constitute a fundamental dimension of human experience. People experience a broad spectrum of emotions on a daily basis. In a smartphone-based survey that prompted participants to report their emotions multiple times per day over an average of 35 days, Trampe, Quoidbach and Taquet (2015) found that individuals commonly reported experiencing a range of emotions – including joy, anxiety, love, contentment, sadness, disgust, and anger – within a single day. This centrality of emotion to human experience has long been recognized across cultures and traditions. Philosophers and political thinkers

from across time and space such as Aristotle, Descartes, Hume, Adam Smith, Mozi, and Senghor have emphasized its importance. Emotions have also been the focus of sustained inquiry across a wide range of disciplines including psychology (Izard, 1977; Plutchik and Kellerman, 1980; Averill, 1983), political science (Marcus, 2000), sociology (Kemper, 1978; Hochschild, 1983), anthropology (Lutz and White, 1986), philosophy (Solomon, 1976; Rorty et al., 1980), and economics (Elster, 1998; Loewenstein, 2000)

Despite a range of emotions being a foundational part of being human, there is not consensus on how they are exhibited – or should be exhibited. On the one hand, emotional intelligence (the ability to recognize, process, and manage the range of our own emotions, as well as the emotions of others) has emerged as a significant predictor of well-being and a crucial factor for human flourishing (Mayer, Caruso and Salovey, 1999; Tomer, 2003). Research shows that individuals with higher emotional intelligence demonstrate superior physical health, self-perception, creativity, and problem solving skills, as well as increased prosocial behaviors (Mayer and Geher, 1996; Tomer, 2003; Lyubomirsky, King and Diener, 2005; Mayer, Roberts and Barsade, 2008; Martins, Ramalho and Morin, 2010). In educational settings, students with higher emotional intelligence exhibit improved stress management and stronger academic performance, contributing to higher retention rates and overall success (Parker et al., 2004; Austin, Saklofske and Mastoras, 2010). Moreover, classrooms characterized by higher levels of emotional intelligence are associated with increased student engagement, greater prosocial behaviors, and superior academic outcomes (Abdulkadiroglu, Pathak and Walters, 2015; Barrett, Lewis and Haviland-Jones, 2016). These findings contribute to accumulating evidence that non-cognitive skills constitute an important component of human capital and have a significant impact on labor market outcomes (Heckman and Rubinstein, 2001; Groves, 2005; Heckman, Stixrud and Urzua, 2006).

On the other hand, societal and labor market conventions often inhibit the overt expression of certain emotions – especially those characterized by "negative" valence – thereby fostering the suppression of these affective states (Hochschild, 1983; Deming, 2017). The high price of these skills in part reflects features of the modern labor market. Emotional intelligence contributes to one's capacity to work socially in teams and groups, a capacity that is essential in the modern economy (Deming, 2017; Bonhomme, 2021). Indeed, the value of these skills is growing even as the return to cognitive skills stagnates (Castex and Kogan Dechter, 2014; Edin et al., 2022). Regulating emotions and recognizing emotions in others can reduce frictions in teams and contribute to managerial effectiveness, which in turn improves earnings (Hansen et al., 2021; Hoffman and Tadelis, 2021). Furthermore, manipulating emotional displays to exhibit happiness and calmness is an essential feature

of what customer service workers provide (Hochschild, 1983). This emotional labor can be especially taxing when happiness and calmness are at odds with what the worker feels, especially in light of the fact that emotional suppression is a poor emotion regulation strategy that reduces well-being and increases experiences of negative emotions (Gross and John, 2003; Kelley et al., 2019). Emotional labor could thus produce a compensating differential that concretely connects emotional intelligence with increased earnings (Glomb, Kammeyer-Mueller and Rotundo, 2004), although this link has yet to be thoroughly explored.

In addition to these broad societal expectations about emotion management and restraint, different identities face a variety of stereotypes regarding the emotions they are disposed to express. For example, women are stereotyped as being more expressive of emotions but less logical than men, although these associations are heavily shaped by the distribution of social roles across genders (Eagly and Steffen, 1984). Additionally, the myth of the "angry Black woman" as an aggressive and ill-tempered entity produces deleterious consequences that lower people's perceptions of Black women's leadership capabilities (Ashley, 2014; Motro et al., 2022). These harmful stereotypes can instill beliefs in children that shape how they learn about emotions and which emotions they choose to display (Brody, 1997).

Educational institutions play an essential role in teaching kids about the exhibition of emotions, as well as the societal stereotypes attached to these expressions. In particular, these institutions foster emotional intelligence both through direct instruction (Brackett et al., 2012) as well as the materials they provide. Curricular materials – including textbooks and children's books – should reflect the richness and complexity of real-life emotional experiences, equipping children to navigate diverse future challenges. This could involve explicit modeling of emotional behavior as well as stories and images in textbooks or other educational materials that portray a broad and balanced range of emotions. At the same time, certain social situations and even occupations discourage the expression of so-called "negative" emotions and encourage maintaining a façade of positivity. To the extent that educational institutions socialize children to their roles in the labor market, this cultural tendency may, in turn, be reflected in educational materials that downplay or omit emotions such as sadness, anger, or fear (Bowles, Gintis et al., 1976; Entorf and Dohmen, 2025).

Children's books, while also influential as informal educational tools, are not explicitly designed for instruction. Consequently, their content is shaped not only by prevailing social norms but also by cultural values and preferences. A high cultural weight on the outward expression of positivity may further skew the emotional landscape these books portray. This has implications for emotional development. Indeed, exposure to emotionally rich storybooks and educational television programs improves children's emotional vocabulary,

perspective-taking skills, empathy, emotion recognition, and coping skills (Kumschick et al., 2014; Rasmussen et al., 2016; Foulds, 2023). As such, these books represent a dual window into both the messages about emotions being sent to children as well as the value placed by purchasers on emotional expression (or lack thereof).

In this study, we examine the emotional content of both textbooks and children's literature, and we compare patterns across the two corpora. We also incorporate data on both consumer and public library purchasing behavior to assess how preferences for particular emotional portrayals influence book acquisition decisions, which may in turn affect the supply of representation of emotions in children's media.

II Data Sources

We analyze two categories of children's media representing both formal and informal educational contexts: state-adopted public school textbooks and award-winning children's books. These materials constitute the primary printed content to which children are regularly exposed during critical developmental periods.

For the public school curricular materials, we analyze state-adopted textbooks from Texas and California in the subjects of Science, Social Studies, and Reading/Language Arts for 3rd and 5th grades. The selection of these states is justified by three key factors. First, they represent the largest public school populations in the United States, with California enrolling approximately 6.86 million students and Texas enrolling approximately 5.84 million. Second, both states employ advisory panels that customize textbook content to align with state-specific educational standards, unlike the majority of states that adopt materials without substantial modification. Third, these states occupy opposing positions on the political spectrum – Texas advisory panels typically reflect Republican perspectives, while California panels are more likely to reflect Democratic perspectives – providing a unique opportunity to examine how ideological orientation may influence the representation of emotions and identities in educational materials.

For children's literature, we focus on books recognized for being particularly distinguished or exemplary. Specifically, we include titles that have won awards featured by the Association for Library Service to Children. These awards honor books for their "exceptional literary" or "artistic" value, or for centering on the experiences of people with underrepresented identities. We analyze two sets of children's books: those that have won "Mainstream" awards, which emphasize literary quality and influence, and those recognized with "Diversity" awards, which specifically celebrate books featuring underrepresented identities. The corpus encompasses both fiction and nonfiction works, representing realistic and imaginary

settings, thereby capturing the full spectrum of narratives available to young readers in curated, high-quality literature.

III Methods

Emotions are communicated in multiple ways: through language – both spoken and written – as well as through non-verbal cues such as facial expressions, gestures, and body language. Previous research has established robust computational methods for emotion detection in textual data (Bollen, Mao and Pepe, 2011; Thelwall, Buckley and Paltoglou, 2011; Hasan, Rundensteiner and Agu, 2014; Demszky et al., 2020), and facial expressions have been extensively documented as reliable indicators of emotional states (Sullivan and Masters, 1988; Lucas, Diener and Larsen, 2003; Öhman, 2007, 2009). Building upon this foundation, we develop a multimodal framework to systematically quantify emotional content in both the textual and visual elements of children's books.

Our classification scheme encompasses eight discrete emotion categories: anger, calm, confusion, disgust, fear, happiness, sadness, and surprise. This taxonomy incorporates the six fundamental emotions originally delineated by Ekman, Sorenson and Friesen (1969) while extending the framework to include calm and confusion. To accommodate ambiguous cases, we incorporated two additional classifications: "unsure" for instances where multiple emotions appear equally plausible or emotional expression is ambiguous and "no emotion" for affectively neutral content, where a character's emotion is not obvious or visible. While the concept of universal basic emotions remains contested in the literature (Russell, 2003), empirical evidence supports the existence of discrete emotional categories (Cowen and Keltner, 2020). However, the precise set of discrete emotions is not agreed upon (Ekman, 2016), and our primary set aligns with emotion taxonomies validated across different contexts (Cowen and Keltner, 2017; Demszky et al., 2020).

III.A Text Analysis

We developed three complementary approaches to quantify emotional content in the text of children's printed content: a large multimodal model (LMM) for character-level emotion detection (depicted in Figure 1a), a lexicon-based frequency analysis, and a fine-tuned transformer model. This multi-method strategy allows exploration of the sensitivity of our findings and permits triangulation across distinct measurement paradigms.

Optical Character Recognition (OCR). We first obtain high-resolution scans of physical books, creating digital images of each page. These scans are then processed using optical character recognition (OCR), which converts each character identified in the scanned images into its corresponding ASCII encoding, thus transforming the visual representation of text

into machine-readable format.

Large Multimodal Model (LMM). We use a Large Multimodal Model (LMM) because it enables us to analyze and interpret complex textual information, such as the emotions exhibited by characters in a story, at scale. An LMM summarizes the information in the text by using patterns and relationships it learned from past data to predict relevant details in new data. Specifically, for emotion detection, we rely on the LMM's ability to recognize subtle cues and linguistic signals in the text to accurately extract and categorize the emotional states of the characters, making the analysis more efficient and objective than manual annotation.[2]

To prepare our data, we partition each book into page-level units and process each page through a single API call with a standardized prompt. In this prompt, we give the LMM various classification tasks and a standardized output format as a constraint. For each page, the prompt instructs the model to perform character identification and emotion classification simultaneously, linking each detected emotion to specific character attributes. For emotion categories, we specify a constrained choice set comprising eight discrete emotions: "anger," "calm," "confusion," "disgust," "fear," "happiness," "sadness," and "surprise."[3] We instructed the model to return "unsure" when confidence is insufficient or "no emotion" when characters appear without emotional valence. For each character, we additionally request predictions of character demographics (age, gender, race), and whether it is a human or non-human, in addition to whether the model is unsure. This allows us to connect each identified emotion to a character's demographic information when that information is available. This enables us to perform a variety of heterogeneity analyses and determine how, if at all, emotions in the text vary across demographic groups.

To ensure structured output alignment with specified requirements, we implement three quality control mechanisms within the prompt: (i) explicit variable delimitation, (ii) exemplar-based formatting instructions providing desired output structures, and (iii) guardrail statements constraining response space to mitigate hallucinations. We include the full prompt in Appendix C.B.1. Post-processing involves cleaning the output to ensure consistency within predefined choice sets, standardizing linguistic variations such as parts of

[2]Our primary textual analysis approach leverages OpenAI's GPT-4o-mini (version 2024-07-18) via API. We configured the model with temperature set to the default value of 1.0 and system prompt "You are a knowledgeable and unbiased judge," parameters chosen to balance response consistency with natural language understanding while minimizing hallucinations. The temperature controls the creativity of the model, with lower values allowing for more deterministic responses and decreasing the likelihood of hallucinations. Its default value is 1.0. The LMM plays the role that the system content provides, so this parameter structures the responses to mimic those of a knowledgeable and unbiased judge.

[3]Additionally, we ask for the six emotions identified by Ekman, Sorenson and Friesen (1969), which align with our choice set but exclude "calm" and "confusion", and we ask for the 28 emotions found in Demszky et al. (2020), which we refer to as the "GoEmotions" set.

speech (e.g., converting "angry" to "anger"), and filtering responses to retain only predefined categories. Hallucinations of emotions beyond simple morphological variants account for less than 1% of all observations, indicating high prompt adherence.

Alternative Approaches for Text Analysis. We implement two additional approaches separate from using an LMM to extract emotion information from text: a lexicon-based frequency analysis and a fine-tuned transformer model. The lexicon-based approach involves scanning the text for predetermined emotion-related words and counting their frequency, offering a straightforward and interpretable baseline. The fine-tuned transformer model, on the other hand, is specially trained on emotion-labeled data to enhance its sensitivity to emotional cues within the text. By comparing the outputs of all three methods, we can confirm the consistency and reliability of our emotional analysis, reducing the risk of bias or errors inherent in any single technique. We describe these further in Methods Appendix C.

Lexicon-Based Analysis. We construct comprehensive keyword dictionaries for each target emotion through three sources: WordNet semantic expansion, ChatGPT-assisted generation, and manual curation. Examples of keywords related to each emotion can be found in Appendix Table B.1.[4] For each sentence, we enumerate emotion-specific keyword occurrences and aggregate counts to our analytical units of interest (e.g., complete corpus, subject matter categories, demographic subgroups). We normalize these counts by total emotion word frequency within each aggregation level, yielding proportional emotion distributions. This frequency-based approach provides a transparent, replicable baseline measurement independent of neural network architectures.

Transformer Model Prediction. We apply a RoBERTa-base transformer model fine-tuned on the GoEmotions dataset created by Demszky et al. (2020).[5] This model predicts the emotion associated with each sentence based on the 28 emotion categories identified by those authors. We subsequently map these fined-grained emotion predictions to the eight emotion categories we use for our analysis. This intermediate-complexity approach bridges simple lexical matching and the more sophisticated LMM methodology, offering predictions grounded in a large-scale, supervised training regime while maintaining computational efficiency relative to generative models.[6]

[4]The full lexicon can be viewed at https://docs.google.com/spreadsheets/d/1xMSboViKOgjYI3ue-RMKROTWWhcN0WDQs2QREx1l3Lk/edit?usp=sharing.

[5]The model can be found at https://huggingface.co/SamLowe/roberta-base-go_emotions.

[6]We drop sentences labeled as "neutral" because we are only concerned with text that has emotional content. This parallels our lexicon method where we only consider emotion-related words. Moreover, the model is not equipped to identify calm, despite the fact that this emotion features prominently in images. The correspondence categories can be found at https://github.com/google-research/google-research/blob/master/goemotions/data/ekman_mapping.json.

III.B Image Analysis

Social scientists have increasingly utilized images as primary sources of information (Ash et al., 2021; Adukia et al., 2023; Yanagizawa-Drott and Voth, 2023; Adukia and Harrison, 2025). Building on this emerging methodological literature, we developed two complementary computational pipelines to systematically extract structured data on emotional expressions and character attributes from scanned pages of children's books. We use a large multimodal model and, separately, a two-step face detection model and feature classification model. Figure 1b provides an overview of the data extraction process applied to the images on book pages.

Large Multimodal Model (LMM) Classification. One way that we extract data from images is by using a Large Multimodal Model (LMM), which enables large-scale analysis of visual content, including the classification of emotions displayed by pictured characters. LMMs model information in images by applying patterns and relationships learned from extensive prior image datasets to predict salient features in novel visual material. Specifically, for emotion detection, we harness the LMM's capability to recognize facial expressions, body language, and other visual cues, enabling us to extract detailed and nuanced information about the emotions exhibited by pictured characters.

This strategy for image analysis involves the deployment of a two-stage automated pipeline implemented on Google Cloud's Vertex AI platform. This approach leverages Gemini 2.5 Pro Preview (model version gemini-2.5-pro-preview-03-25), a large multimodal model (LMM) capable of integrated reasoning across visual and textual inputs. The design of the two-stage architecture enhances processing efficiency and ensures consistency in feature extraction across our large-scale corpus.

Stage 1: Binary Page Classification Filter. The initial stage implements a computational filter to identify pages containing visual content and minimize computational expenses by excluding pages devoid of illustrations. For each scanned page, the raw image bytes are submitted to the LMM 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 and limit response length to 256 tokens to ensure concise, precise classification decisions. Pages classified as `TEXT_ONLY` are recorded and excluded from subsequent analysis stages, thereby concentrating computational resources on illustrated content requiring emotion analysis.

Stage 2: Structured Character and Emotion Extraction. Pages identified as containing illustrations proceed to a second analysis stage featuring expanded token allocation

(maximum 10,000 tokens) and schema-validated JSON output. We enforce a strict JSON Schema for the model's output through Vertex AI's structured output mechanism, which guarantees adherence to predefined fields, data types (e.g., `STRING`, `INTEGER`), and enumerated constraint sets for categorical variables. The extracted data is organized into two hierarchical components:

1. **Scene-level metadata** captures contextual information including image style classification (photograph versus illustration; realistic versus fantasy; polychromatic versus monochromatic), setting descriptions, depicted activities, atmospheric qualities, and image dimensions in pixels.

2. **Character array** contains one object per detected individual, each recording: (i) spatial localization via bounding-box coordinates (x_1, y_1, x_2, y_2) in pixel space; (ii) demographic attributes including race, gender (constrained to enumerated values: `female`, `male`, `unsure`), age category, and skin color (constrained values: `Dark`, `Medium`, `Light`); (iii) emotional expression classification across our target emotions using a hierarchical taxonomy spanning 6-class, 10-class, and 28-class granularity; and (iv) supplementary features such as character actions, occupations, perceptual-tint scores (0–100 scale), and human versus non-human classification.

This deterministic configuration (temperature = 0.0) ensures reproducible, schema-compliant outputs suitable for downstream statistical analysis. The pipeline processes files either sequentially or through parallel execution using multiple worker threads, with consolidated results periodically saved as checkpoints every 10 files to ensure data integrity during extended processing runs. Complete prompt specifications and schema definitions are provided in Appendix C.A.

Convolutional Neural Network-Based Face Detection and Feature Classification. For a second approach to classify emotions and other features in images, we developed a benchmark convolutional neural network (CNN) classifier using Google's Vertex AI platform. CNNs are well suited for image recognition tasks due to their capacity to learn hierarchical spatial representations in visual data, extracting low-level features (edges, textures, colors) in early layers and progressively combining these into higher-order semantic representations suitable for classification tasks.

First, we detect faces of pictured characters in the images using the face detection model (FDAI) trained on illustrations by Szasz et al. (2022). Second, we classify features using the feature classification model (FCA) from Szasz et al. (2022). We developed our model to perform multi-label image classification, enabling simultaneous prediction of mul-

tiple attributes per image. The model was trained to recognize 31 distinct labels spanning five categories: image properties (photo, illustration, monochromatic, polychromatic), human presence (human, nonhuman), demographic attributes (race, age, gender), emotional expressions (happy, sad, angry, surprised, disgusted, fear, calm, confused), and uncertainty indicators (unsure_race, unsure_gender, unsure_emotion).

We constructed a dataset of 981 manually annotated images, randomly partitioned into training ($n = 787,\ 80.2\%$), validation ($n = 99,\ 10.1\%$), and test ($n = 95,\ 9.7\%$) sets. These performance metrics establish a quantitative baseline for evaluating whether the LMM approach yields substantively different classifications compared to conventional computer-vision methods, thereby providing evidence regarding the reliability of our primary analytical framework.

IV Results

In this section, we present our findings on the emotional content expressed through the text and images of influential children's books.

IV.A Emotions in Text and Images Overall and Over Time

We first examine the overall patterns of emotions in text and in images, respectively. Figure 2 shows the distributions of detected emotions in text and images across all books in our sample. In the text, the most common emotions are happiness and sadness, which account for approximately 29% and 20% of characters' emotions, respectively (Figure 2a). With respect to the distribution of the other emotions (anger, calm, confusion, disgust, fear, and surprise), we observe a meaningful number of depictions of each in text (ranging between 4 to 12%), reflecting a balance between "positively" and "negatively" valenced emotions. In sharp contrast, there is a high degree of homogeneity in the emotions displayed in images. In images, calm and happiness together account for approximately 90% of visually depicted emotions (Figure 2b). Notably, negatively valenced emotions, which feature prominently in these books based on the text analysis, are largely absent from images. This pattern underscores a stark incongruence between the emotional narratives in the text and the limited emotional range portrayed in the images.

We further observe that these patterns persist over time. Figure 3 presents the proportions of emotions detected in text and images by year, separately for textbooks and children's literature.[7] We find that images consistently overrepresent calm and happiness across time in both textbooks and children's books, whereas a wider variety of emotions is

[7]Children's books are associated with their respective copyright years, whereas textbooks are associated with each year they are in use. Note that textbooks are in use for multiple years, whereas the children's books only appear one time in the time series panel.

displayed in text with relatively stable proportions over time. We see that the shares of emotions, in both images and text, are broadly similar for textbooks and children's books. Both collections of books consistently overrepresent happiness and calm in images relative to their respective proportions in text.

IV.B Emotions in Different Contexts

To better understand how emotions are represented across different types of books, we examine heterogeneity across a variety of different subcollections of our data. Overall, we find similar patterns across each context: a broader variety of emotions is conveyed in text, while images consistently emphasize calm and happy expressions.

Figure 4 presents the proportions of different emotions in the text of various subcollections. The distributions in children's literature and textbooks are broadly similar, with textbooks showing a slightly stronger overrepresentation of happiness compared to children's literature. As may be expected, children's literature displays a somewhat richer and more diverse range of emotions compared to textbooks, though both collections cover a variety of emotions in text.

Within the subcollections of these two categories, the patterns remain qualitatively similar. In particular, the Mainstream and Diversity subcollections of children's literature show comparable proportions of each emotion, with the Diversity subcollection exhibiting slightly greater heterogeneity of emotional representation. With respect to textbooks, we observe qualitatively similar patterns across those from California and Texas.

Although the broad patterns are consistent, we note some differences by grade and subject in textbooks. For example, happiness is represented more frequently in third-grade textbooks than in fifth-grade textbooks. Conversely, emotions such as anger, fear, and sadness appear more frequently in fifth-grade textbooks.

We also observe differences by subject. Science textbooks contain more confusion and happiness but less anger and sadness compared to reading and social studies textbooks. This may be consistent with the instructional goals of each subject: science textbooks may not require as rich an emotional narrative to convey concrete concepts, whereas reading and social studies naturally incorporate a broader range of emotions. Conversely, due to the importance of trials and experimentation in science, the lack of emotion representation in science textbooks could reflect a missed opportunity to illustrate emotions that may arise in the face of adversity. Even acknowledging small differences across subjects, we still observe that the representation of emotions in text remains broadly consistent across all subcollections.

Figure 5 displays the same heterogeneity analysis for images. We again find qualitatively similar patterns across subcollections. Calm and happiness are even more overrepresented in textbooks than in children's books. As in the textual analysis, this is consistent with a somewhat richer display of distinct emotions in children's literature compared to textbooks. This overrepresentation of happiness and calm is especially strong in science and social studies books. Reading textbooks still overrepresent happiness and calm, but the distribution of emotions in these books' images more closely resembles the distribution in children's literature than that in textbooks from the other subjects. This could reflect that these textbooks include age-appropriate stories, possibly designed to resemble the award-winning books in our sample.

We also observe that third-grade textbooks display more happiness and less calm than fifth-grade textbooks. The discrepancies nearly balance each other, so the total percentage in these two categories is similar across the two grade levels. This could reflect the fact that calmness and happiness may be difficult to disentangle in facial expressions. In any case, the overall pattern that calmness and happiness are massively overrepresented in images compared to other emotions holds across all subcollections.

IV.C Emotions Displayed by Different Identities

Next, we examine how displays of emotion vary for characters of different identities. As described in Section III, we obtain the predicted race and gender of characters, whether they are children or adults, and whether they are human or non-human. For both text and images, these attributes are predictions made by the respective models described in Section III.

Figure 6 displays this heterogeneity analysis in text. We see that females are slightly more likely than males to show sadness in text, whereas males are slightly more likely to show anger. Additionally, adults are more likely to express anger or calm, whereas children are more likely to express fear or confusion. Nonetheless, overall, we see qualitatively consistent patterns across identities.

Figure 7 presents the analogous heterogeneity analysis for images. We see that male faces are more likely to show calm than female faces. This is consistent with gender stereotypes that females are less calm than males. Reciprocally, female faces are more likely, by a similar amount, to show happiness than male faces. A similar pattern holds for the other identities; groups that are more frequently shown as calm are less frequently shown as happy and vice versa. As noted in Section IV.B, this could be due to the fact that calm and happiness are difficult to distinguish in images. In any case, all groups display calm or happy facial

expressions in almost all instances, and no group shows any other emotion with comparable frequency. This lack of heterogeneity in emotional expression across identities may be surprising given stereotypical associations between certain identities and certain emotional expressions.

IV.D Concurrent Text and Image Analysis

To further investigate our text and image data, we directly compare the emotions expressed through each modality on the same page. To do this, we construct eight binary variables separately for text and images which take on the value of 1 when a character on a given page expresses the corresponding emotion and zero otherwise. Ultimately, we obtain sixteen indicator variables that encode whether or not each emotion is found on the page in the text and images, respectively.

Figure 8 shows the co-occurrence of emotions for text and images on the same page, normalized by the text columns. In other words, each cell shows the proportion of times an image is assigned a specific emotion label, given a particular emotion label in the text of the same page. When comparing the presence of emotions in images and text on the same page, calm is the emotion most likely to align: when calm is mentioned in the text, it is also depicted in the image in 70% of cases. Following calm, when a character described in the text is happy, 27% of cases show a happy character in the corresponding image. In contrast, emotions such as anger, fear, and sadness, when mentioned in the text, are visually represented only 13%, 9%, and 10% of the time, respectively. When there is an emotion other than calm in the text, images on that page still depict a calm emotion between 58% and 67% of the time. This indicates that, not only are kids not exposed to a variety of emotional displays in images, but there is frequently a mismatch between what characters say, do, and feel in the text compared to what they convey in images. Moreover, this mismatch exhibits a consistent direction: negative emotions described in the text are accompanied by positive expressions in the visual cues. This incongruence not only complicates social emotional learning, but it reinforces an implicit message that suppressing negative emotions is desirable.

It could be the case that, despite broad differences between emotions in text and images, the images do faithfully convey the emotions of at least one of the characters in the text. To probe this possibility, we recompute the co-occurrence matrix for a sample in which we match characters in the text and images. To perform the matching, we identify cases in which there is a unique character of the same predicted race and gender in both the text and image of a particular page. Figure A.3 illustrates that the distributions of emotions in text and images in this matched sample are similar to the overall distributions displayed in

Figure 2, indicating that the matched sample is similar to the full sample of text and images.

Figure A.4 illustrates the misalignment of emotions in the text and images corresponding to these matched characters. The broad pattern of misalignment persists. We note that there is similar weight on the diagonal compared to Figure 8, reflecting the fact that emotions displayed in text and images frequently fail to align even when we restrict to considering the same character. Regardless of the emotions a character displays in the text, they are most likely to be depicted as calm in the accompanying image.

To complement these co-occurrence matrices, we also compute measures of the overall agreement of emotions in text and images. In particular, we combine the page-level indicators for the presence of different emotions into a single eight-dimensional vector separately for text and images. Each entry in these vectors corresponds to an emotion and takes a value of one if and only if the emotion is present on the page. After normalizing, we compute the total variation distance (TVD) between the page-level distributions of text and images as $\delta(p, q) = \frac{1}{2}\|p - q\|_1$.

Figure A.5 displays a histogram of the total variation distances for the distributions of emotions in text and images on individual pages. In a small proportion of cases, we see that the TVD is zero, indicating that the emotions in the text and images align perfectly. However, for nearly half of the pages, the TVD is one, indicating that there is no overlap between emotions featured in text and images on the same page.

Overall, these results demonstrate that the discrepancy between the distributions of emotions in text and images is not driven merely by illustrations that selectively emphasize specific positively valenced emotions in the text; rather, the illustrations convey happiness and calm even when these emotions are absent in the text. This mismatch produces a dilemma for young readers. They read about a range of emotions in text. At the same time, the illustrations consistently convey that the appropriate response to any feeling they encounter is to put on a calm façade.

IV.E Alternative Approaches to Measurement

To complement our text and image analyses, we apply two additional text models and one additional image model. The text models are described in Section III.A, and the image model is described in Section III.B. We note that the image model predicts emotions in each detected face. This is somewhat different from the LMM analysis of entire images and the characters therein. In a similar way, as discussed in Section III.A, our alternative textual analysis methods predict emotions globally or at the sentence level. By contrast, in our main analysis, we study textual displays of emotion at the level of the character when we

use the LMM in order to obtain results that are more comparable to our image results, which explicitly examine visually depicted characters. Therefore, we do not expect the results of these analyses to align perfectly with our main results. Nonetheless, the findings demonstrate the robustness of our results to an alternative system of classifying emotion (for example, at the sentence or subcollection level rather than at the character level). Moreover, the use of different models with varying complexity and transparency provides a source of validation for the findings we obtain leveraging generative AI.

Figure 9a shows the results of our lexicon-based approach. We see similar levels of anger, disgust, fear, and sadness in comparison to our main approach. In contrast, it appears that our main model identifies around 10% more happiness and around 10% less calm in our text than the lexicon method. This may be an artifact of the lexicon approach's sensitivity to keyword choice. Indeed, calm and happiness may be lexically difficult to differentiate. For example, consider the word "contentment" (which we use as one of the words related to the emotion of "calm"). Depending on the context of the surrounding text, this word could either indicate a state of calm or of happiness. We thus expect there to be some fluidity between these two categories with the lexicon model.

Figure 9b shows the results of the GoEmotions model. Here we note qualitatively similar patterns to Figure 2a in all but surprise, to which the GoEmotions model attributes 20% more of its classifications. As noted earlier, this may be due to the structure of our inputs. The GoEmotions model could, for example, classify sentences associated with discovery and learning as expressing surprise. (To see this, consider a history textbook chapter on the Age of Discovery.) This expression would not be counted by the LMM because there is no specific character who is exhibiting surprise. Nonetheless, the results from this method are arguably more comparable to those of the LMM than the Lexicon-based results since the GoEmotions model and LMM both use surrounding context to classify emotions.

Figure 9c shows the results from our alternative image model. This model predicts less calm and more happiness than the LMM for images, but the sum of emotions in these two categories is consistent across models. This further reflects the overlap between these two categories that. Overall, we find a qualitatively similar distribution to the one obtained using the LMM, which lends further confidence to our main results.

In summary, across a variety of different models ranging in complexity and interpretability, we obtain similar distributions of emotions in text and images to the ones we find using generative AI. This provides one way to validate the results from these more complex models. Additionally, while our main textual results only examine emotions in the text

displayed by characters in order to be more comparable to our image results, our robustness approaches allow for different levels of analysis. We observe that our findings are qualitatively consistent across levels of analysis. In particular, regardless of the models used, we find that happiness and calm are consistently overrepresented in images relative to text.

V Consumer Demand

A natural question that arises is: why do we see these patterns? In this section, we investigate economic factors which may contribute to the patterns of emotional representation in children's books that we report in Section IV. First, we discuss relevant prior theoretical and empirical research related to the economics of consumer demand and advertising to conceptually characterize a set of market forces which may shape the supply of emotions depicted. Second, we bring to bear data on two empirical exercises. We first show that both consumers and Seattle public libraries tend to choose books with happy and calm emotions. Here, we focus on emotions in the cover art because we expect the cover art to be highly salient and influential in purchase decisions, and we expect that consumers will not be attentive to the exact distributions of emotions in text and images *within* books. In order to connect our findings to purchasers' decisions about happy and calm emotions in general and not just on covers, we subsequently show that the representation of emotions on covers provides an informative signal of the emotional representation within the rest of the book.

V.A Related Literature on Consumer Demand

The apparent oversupply of happy and calm images could constitute a supply-side response to the preferences of consumers, who may favor positive expressions. In an analogous way, readers' demand has shaped news media in terms of its informativeness (Gentzkow, Glaeser and Goldin, 2006), slant (Gentzkow and Shapiro, 2010), and pessimism (Sacerdote, Sehgal and Cook, 2020). In the context of literature, narratives associated with "negative" emotions may be perceived as more demanding or less enjoyable, prompting readers under stress or seeking relaxation to favor covers promising pleasant experiences (Becker, 1996; Elster, 1998; Hogan, 2011). This tendency is further amplified during periods of collective uncertainty, when there is a marked increase in the demand for "feel-good" literature.

On the other hand, "positive" emotional signals – such as smiling or serene faces – on book covers tend to evoke an immediate and favorable affective response. Individuals may be drawn to stimuli that enhance their cognitive flexibility and mood (Isen, 1984). They may select books with happy or calm covers as part of the situation selection component of their emotion-regulation strategy (Gross, 1998). Beyond this type of intentional strategy, individuals may naturally gravitate toward "positive" affect and avoid "negative" emotional

cues, such as sadness or anger, which may be subconsciously interpreted as threatening or uncomfortable (Lang, Bradley and Cuthbert, 1990; Adams and Kleck, 2003). Furthermore, from an evolutionary standpoint, positive social signals such as smiles may reduce perceptions of threat (Ohman, 2002) and increase approachability, even in non-social contexts like retail browsing. Finally, as previously noted, these positive expressions may be demonstrations of an agreeableness that is valued in certain occupations and in the U.S. context more generally.

In studying consumer demand for these "positive" signals, we pay particular attention to emotional displays on covers since images on the covers of books exert disproportionate influence over consumers' purchasing behavior. Indeed, irrespective of the underlying driver of consumers' preferences for certain emotional displays, the displays on covers are particularly salient to consumers (Taylor and Thompson, 1982), and salient features play a particularly strong role in influencing purchase patterns (Ellison, 2006; Bordalo, Gennaioli and Shleifer, 2013). This is especially true for the online shoppers in our sample who do not even have the option to flip through the pages as they would in a traditional brick-and-mortar store. Furthermore, happy or calm character depictions can serve as visual cues that lower the cognitive threshold for consumer engagement (Bloch, 1995) and facilitate attention capture and transfer (Pieters and Wedel, 2004; Van Der Lans, Pieters and Wedel, 2008), which could increase the likelihood of purchase.

The crucial first impressions of cover art also serve to signal the quality of and content in the book (Nelson, 1970; Spence, 1973). The cover further enables a form of product differentiation referred to as "positioning" (Dröge and Darmon, 1987). Authors may emphasize certain emotions on covers as a signal to consumers about the emotions prevalent in the content. In this way, the art on the cover acts as a sort of advertisement in which depictions of happy or calm characters can signal an emotionally uplifting or reassuring story and attract consumers to the book (Underwood and Klein, 2002). Readers can be expected to use this information to screen books into or out of their consideration sets (Eliaz and Spiegler, 2011). If readers prefer "positive" expressions, then authors would be incentivized to position their books as featuring positive affect using happy and calm characters in order to capture consumers' attention. Moreover, even if readers would prefer books portraying a broader range of emotions under full attention, happy and calm characters may capture consumers' limited attention, generating an oversupply of these emotions (Gabaix and Laibson, 2006; Cusumano, Fabbri and Pieroth, 2024).

V.B Empirical Analysis of Economic Forces / Consumer Demand

In this section, we present empirical analyses probing the economic forces related to consumer demand. We study how readers value representations of different emotions

in the content and on the covers. We limit our analysis to the books in our children's literature collection because their consumption is more a measure of consumer demand than the consumption of textbooks, which are chosen at the school level.[8]

We study consumer demand using two separate data sources. First, we estimate a conditional logit model using micro-level book purchase data from the Numerator OmniPanel. These data contain information on purchases by individuals as well as detailed demographic information about those individuals, and they speak directly to consumers' revealed preferences. One challenge with using these data is that we cannot measure consumers' consideration sets. Although it is plausible that online purchasers – who comprise the majority of our sample – could select any book in the collection that we study, we still address this limitation by estimating a similar model using data on book stocks from the Seattle Public Library system. We expect that a public library would consider purchasing a much broader range of books than a typical individual consumer. Therefore, the assumption that every award-winning book in our sample is also in the library's consideration set becomes more plausible. Moreover, although public libraries may have different objective functions than individual consumers, we expect that they would generally stock more copies of books that readers prefer, so these stocks should indirectly reflect readers' choices.

In order to specifically study the seeming overrepresentation of happy and calm in images, we now classify emotions as either happy/calm or non-happy/calm, with the latter category indicating that some emotion besides happy or calm is displayed.[9] We estimate a conditional logit model using the Numerator data to explicitly study consumers' revealed preferences, and we subsequently use the public library data to regress the number of copies of a book held by the library against the same set of features. Our models include an indicator for whether there are exclusively happy/calm characters on the cover, an indicator for whether there is any character on the cover, and a set of controls including indicator variables for the award(s) the book won, the book's copyright year, whether the book is a picture book, whether it contains no pictured characters, the number of pages, and the average number of words per page.[10] We do not expect consumers to accurately perceive the proportions of emotions displayed in the text or images *within* the books. However, we still run two alternative specifications in which we control for the proportions of happy/calm emotions in both images and text by (1) directly including these proportions as controls and (2) by adding indicators for quintiles of these proportions within their respective dis-

[8]Of course, parents can select into different schools, but the content in textbooks is only one dimension of this multi-faceted choice.

[9]"No emotion" and "unsure" labels are not included in this category.

[10]These last two features serve as a proxy for the age of the target audience of the book.

tributions. These additional controls allow for consumers who attend more closely to the emotional content to respond to the emotional content of the books, and they can potentially proxy for unobserved confounders.

Table 1 displays the results from the Numerator data. We observe that the emotional representation on the cover plays a substantive role in shaping consumers' purchase decisions. In particular, across all specifications, we see that consumers place a significant premium on books depicting happy and calm characters on the cover compared to books depicting other emotions. Moreover, consumers significantly favor books with no characters on the cover to books with a character displaying an emotion which is neither happy nor calm. This suggests an explicit aversion to books that have these other emotions on the cover. In turn, this aversion could disincentivize book creators from illustrating non-happy/calm emotions, which would produce the patterns we uncover.

In Table 2, we display an analogous set of results using the Seattle Public Library system data. Overall, the results are qualitatively similar. As before, we find higher stocks of books with exclusively happy or calm cover characters or no character at all than books with non-happy/calm characters on the cover. This suggests that an inability to properly specify consumers' consideration sets does not drive our results in Table 1. Instead, these results jointly speak to consistent purchasing patterns of readers favoring books with depictions of happiness and calm over books with other emotional depictions.

Additionally, we find that the emotions displayed on the cover do provide an informative signal about the emotions in the content of the book. We estimate another conditional logit model to study the probability that a cover contains only happy/calm characters as a function of the proportion of happy/calm emotions in text and images in the content of the book. We also perform an analogous analysis to study the probability a cover contains only non-happy/calm characters. The results are displayed in Table 3. We find that having higher proportions of happy/calm emotions in text and images predicts having exclusively happy/calm characters on the cover. Likewise, lower proportions of happy/calm emotions in text and images predict having exclusively non-happy/calm characters on the cover.[11] These results indicate that covers do provide consumers with an informative signal about the emotional content within the book. Therefore, readers' tendency to favor books with happy/calm characters on the cover also indicates a tendency toward greater representation of happy/calm emotions in the content.

[11]The reason the latter result is not mechanical is because these two categories are not mutually exclusive. Covers could have a mix of happy/calm and non-happy/calm emotions, or they could display no characters at all.

In light of these results, the patterns we uncover in Section IV are consistent with a supply-side response to readers' observed purchasing patterns regarding certain emotions. Moreover, since images – especially those on covers – are particularly salient compared to text, authors would face a stronger disincentive to depict non-happy/calm emotions in the images compared to text. This is consistent with our finding that emotions are relatively diverse in text while images predominantly show characters who are either happy or calm. We also note that this market equilibrium is not necessarily innocuous; if consumers are boundedly rational but are attracted to depictions of positive emotions, then the resulting equilibrium may lead to an oversupply of happy/calm emotions relative to what is best for young readers.

VI Conclusion

Our analysis yields important insights into the role that educational materials—both textbooks and children's literature—play in shaping how emotions are represented and, potentially, how they are understood by students. These findings are particularly important in the context of emotional intelligence (EI) development, which is closely tied to students' social and academic success. By examining the portrayal of emotions in images and text, we uncover both strengths and limitations in how educational materials approach emotional representation.

One of the most striking results is the disproportionate emphasis on calm and happiness in images, which together account for about 90% of visually depicted emotions. While these emotions are crucial for fostering a positive and serene learning environment, the underrepresentation of other emotions—such as anger, sadness, or confusion—limits students' exposure to the full spectrum of emotional experiences. These images are particularly influential in shaping young readers' understanding of emotions, as they provide direct and intuitive cues that text alone cannot. Therefore, whether inadvertently or intentionally, this imbalance may signal to students that some emotions are more acceptable or valued than others, potentially affecting the way they recognize, understand, and navigate complex emotional landscapes. Moreover, the near absence of visual models of a broader range of emotions prevents these books from effectively teaching emotion recognition, which is an important component of emotional intelligence.

Our findings also reveal that this imbalance between depictions of "positive" and "negative" emotions in images persists across time, context, and character identity. Despite growing awareness of the importance of emotional intelligence in education, the representation of emotions in educational materials appears stagnant, with no clear trend toward broader inclusion of underrepresented emotions. Moreover, happiness and calm are overrepresented in

all subcollections and by characters of different identities. However, there are some notable sites of heterogeneity. For example, female characters are more often depicted as happy, while male characters are more often shown as calm, suggesting a subtle reinforcement of gendered emotional stereotypes.

While text offers a more diverse range of emotional representation compared to images, it often displays emotions that conflict with those depicted in images on the same page. Even within the same character on the same page, no matter which emotions the character displays in the text, they are disproportionately likely to be shown as either calm or happy. This incongruence could have important consequences for how children learn to interpret and respond to emotions. In the books they are most likely to be exposed to, they are implicitly shown that only calm and happy reactions are appropriate. When it comes to public school textbooks in particular, these patterns suggest that curriculum designers and publishers may not yet fully recognize the potential of educational materials to foster emotional literacy. Alternatively, these materials may intentionally emphasize the importance of positivity regardless of the situation.

In the case of children's literature, we uncover evidence that the disproportionate representation of happy and calm emotions is due to purchasers' revealed preferences. Both individual households and public libraries (whose inventories can be expected to reflect the interests of the public) are significantly less likely to purchase books that feature characters with "negative" emotions. As a result, authors and illustrators include images of happy and calm characters in an effort to attract consumers. These patterns may reflect the societal value placed on "positive" emotions. Importantly, we note that even if purchasers did want a greater diversity of emotions in their kids' books, they may be imperfectly attentive or unable to carefully peruse the emotional content of a book before purchase. In this case, smiling faces may capture their limited attention and carry appeal, and this could result in an inefficient overrepresentation of happy and calm emotions (Ellison, 2006; Gabaix and Laibson, 2006; Cusumano, Fabbri and Pieroth, 2024). We do not disentangle underlying preferences from actions taken under imperfect attention, but we simply illustrate that consumers' revealed preferences for happy and calm emotions could drive the patterns we uncover.

From an educational perspective, these patterns call for a more intentional approach to incorporating emotional diversity in textbooks and children's literature. Schools have a unique opportunity to foster emotional intelligence by exposing students to a wider range of emotions and helping them to recognize, understand, and label their own feelings as well as the feelings of others. Parents also have a deep-seated interest in providing their kids with appropriate books including those that portray a diversity of emotions so that kids can learn

strategies that allow them to successfully express and regulate emotions without suppressing them. Whether by parents or schools, the deliberate effort to include emotions like sadness, anger, and confusion in visual representations could provide children with the tools they need to effectively navigate real-world challenges and occupations without suppressing their feelings.

References

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

Adams, Reginald B Jr, and Robert E Kleck. 2003. "Perceived gaze direction and the processing of facial displays of emotion." Psychological Science, 14(6): 644–647.

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, and Emileigh Harrison. 2025. "Separation of church and state curricula? Examining public and religious private school textbooks." National Bureau of Economic Research.

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

Ashley, Wendy. 2014. "The angry black woman: The impact of pejorative stereotypes on psychotherapy with black women." Social work in public health, 29(1): 27–34.

Austin, Elizabeth J, Donald H Saklofske, and Sarah M Mastoras. 2010. "Emotional intelligence, coping and exam-related stress in Canadian undergraduate students." Australian Journal of Psychology, 62(1): 42–50.

Averill, James R. 1983. "Studies on anger and aggression: Implications for theories of emotion." American Psychologist, 38(11): 1145.

Barrett, Lisa Feldman, Michael Lewis, and Jeannette M Haviland-Jones. 2016. Handbook of emotions. Guilford Publications.

Bazzi, Samuel, Martin Fiszbein, and Mesay Gebresilasse. 2020. "Frontier culture: The roots and persistence of "rugged individualism" in the United States." Econometrica, 88(6): 2329–2368.

Becker, Gary S. 1996. Accounting for Tastes. Harvard University Press: Cambridge.

Bloch, Peter H. 1995. "Seeking the Ideal Form: Product Design and Consumer Response." Journal of Marketing, 59(3): 16–29.

Bollen, Johan, Huina Mao, and Alberto Pepe. 2011. "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena." Vol. 5, 450–453.

Bonhomme, Stéphane. 2021. "Teams: Heterogeneity, sorting, and complementarity." arXiv preprint arXiv:2102.01802.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2013. "Salience and consumer choice." Journal of Political Economy, 121(5): 803–843.

Bowles, Samuel, Herbert Gintis, et al. 1976. Schooling in capitalist America. Vol. 57, New York: Basic Books.

Brackett, Marc A, Susan E Rivers, Maria R Reyes, and Peter Salovey. 2012. "Enhancing academic performance and social and emotional competence with the RULER feeling words curriculum." Learning and Individual Differences, 22(2): 218–224.

Brody, Leslie R. 1997. "Gender and emotion: Beyond stereotypes." Journal of Social issues, 53(2): 369–393.

Castex, Gonzalo, and Evgenia Kogan Dechter. 2014. "The changing roles of education and ability in wage determination." Journal of Labor Economics, 32(4): 685–710.

Cockcroft, Marlaina. 2018. "Caldecott and Newbery Medal wins bring instant boost to book sales." School Library Journal, 64(2).

Cowen, Alan S, and Dacher Keltner. 2017. "Self-report captures 27 distinct categories of emotion bridged by continuous gradients." Proceedings of the national academy of sciences, 114(38): E7900–E7909.

Cowen, Alan S, and Dacher Keltner. 2020. "What the face displays: Mapping 28 emotions conveyed by naturalistic expression." American Psychologist, 75(3): 349.

Cusumano, Carlo M, Francesco Fabbri, and Ferdinand Pieroth. 2024. "Competing to commit: Markets with rational inattention." American Economic Review, 114(1): 285–306.

Deming, David J. 2017. "The growing importance of social skills in the labor market." The quarterly journal of economics, 132(4): 1593–1640.

Demszky, Dorottya, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. "GoEmotions: A dataset of fine-grained emotions." arXiv preprint arXiv:2005.00547.

Dröge, Cornelia, and Rene Y Darmon. 1987. "Associative positioning strategies through comparative advertising: Attribute versus overall similarity approaches." Journal of Marketing Research, 24(4): 377–388.

Duckworth, Angela L, Christopher Peterson, Michael D Matthews, and Dennis R Kelly. 2007. "Grit: perseverance and passion for long-term goals." Journal of personality and social psychology, 92(6): 1087.

Eagly, Alice H, and Valerie J Steffen. 1984. "Gender stereotypes stem from the distribution of women and men into social roles." Journal of personality and social psychology, 46(4): 735.

Edin, Per-Anders, Peter Fredriksson, Martin Nybom, and Björn Öckert. 2022. "The rising return to noncognitive skill." American Economic Journal: Applied Economics, 14(2): 78–100.

Ekman, Paul. 2016. "What scientists who study emotion agree about." Perspectives on psychological science, 11(1): 31–34.

Ekman, Paul, E Richard Sorenson, and Wallace V Friesen. 1969. "Pan-cultural elements in facial displays of emotion." Science, 164(3875): 86–88.

Eliaz, Kfir, and Ran Spiegler. 2011. "Consideration sets and competitive marketing." The Review of Economic Studies, 78(1): 235–262.

Ellison, Glenn. 2006. "Bounded rationality in industrial organization." Econometric Society Monographs, 42: 142.

Elster, Jon. 1998. "Emotions and economic theory." Journal of economic literature, 36(1): 47–74.

Entorf, Antonia K, and Thomas J Dohmen. 2025. "The effect of compulsory education on non-cognitive skills: Evidence from low-and middle-income countries." Economics of Education Review, 107: 102654.

Foulds, Kim. 2023. "Co-viewing mass media to support children and parents' emotional ABCs: An evaluation of Ahlan Simsim." Early Childhood Education Journal, 51(8): 1479–1488.

Gabaix, Xavier, and David Laibson. 2006. "Shrouded attributes, consumer myopia, and information suppression in competitive markets." The Quarterly Journal of Economics, 121(2): 505–540.

Gentzkow, Matthew, and Jesse M Shapiro. 2010. "What drives media slant? Evidence from US daily newspapers." Econometrica, 78(1): 35–71.

Gentzkow, Matthew, Edward L Glaeser, and Claudia Goldin. 2006. "The rise of the fourth estate. How newspapers became informative and why it mattered." In Corruption and reform: Lessons from America's economic history. 187–230. University of Chicago Press.

Glomb, Theresa M, John D Kammeyer-Mueller, and Maria Rotundo. 2004. "Emotional labor demands and compensating wage differentials." Journal of applied psychology, 89(4): 700.

Gross, James J. 1998. "The emerging field of emotion regulation: An integrative review." Review of General Psychology, 2(3): 271–299.

Gross, James J, and Oliver P John. 2003. "Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being." Journal of personality and social psychology, 85(2): 348.

Groves, Melissa Osborne. 2005. "How important is your personality? Labor market returns to personality for women in the US and UK." Journal of Economic Psychology, 26(6): 827–841.

Hansen, Stephen, Tejas Ramdas, Raffaella Sadun, and Joe Fuller. 2021. "The demand for executive skills." National Bureau of Economic Research.

Hasan, Maryam, Elke Rundensteiner, and Emmanuel Agu. 2014. "Emotex: Detecting emotions in twitter messages."

Heckman, James J, and Yona Rubinstein. 2001. "The importance of noncognitive skills: Lessons from the GED testing program." American economic review, 91(2): 145–149.

Heckman, James J, Jora Stixrud, and Sergio Urzua. 2006. "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." Journal of Labor Economics, 24(3): 411–482.

Hochschild, Arlie Russell. 1983. "The Managed Heart: Commercialization of Human Feeling."

Hoffman, Mitchell, and Steven Tadelis. 2021. "People management skills, employee attrition, and manager rewards: An empirical analysis." Journal of political economy, 129(1): 243–285.

Hogan, Patrick Colm. 2011. What Literature Teaches Us About Emotion. Cambridge University Press.

Isen, Alice M. 1984. "Toward understanding the role of affect in cognition." Handbook of social cognition, 3: 179–236.

Izard, Carroll E. 1977. "Differential emotions theory." In Human Emotions. 43–66. Springer.

Kardum, Rona Bušljeta, Katarina Dadic, and Martina Horvat. 2021. "Education for the cultivation of emotions through textbooks: the example of Croatian high school history textbooks." Educational process: international journal, 10(2): 28–41.

Kautz, Tim, James J Heckman, Ron Diris, Bas Ter Weel, and Lex Borghans. 2014. "Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success." , (Working Paper 20749).

Kelley, Nicholas J, James E Glazer, Narun Pornpattananangkul, and Robin Nusslock. 2019. "Reappraisal and suppression emotion-regulation tendencies differentially predict reward-responsivity and psychological well-being." Biological psychology, 140: 35–47.

Kemper, Theodore D. 1978. "Toward a sociology of emotions: Some problems and some solutions." The American Sociologist, 30–41.

Krippendorff, Klaus. 2018. Content Analysis: An Introduction to its Methodology. Sage publications.

Kumschick, Irina R, Luna Beck, Michael Eid, Georg Witte, Gisela Klann-Delius, Isabella Heuser, Rüdiger Steinlein, and Winfried Menninghaus. 2014. "Reading and feeling: The effects of a literature-based intervention designed to increase emotional competence in second and third graders." Frontiers in Psychology, 5: 1448.

Ku, Yunjung. 2019. "Collocations of English Emotion Expressions in High School Textbooks." , 20(4): 63–73.

Lang, Peter J, Margaret M Bradley, and Bruce N Cuthbert. 1990. "Emotion, attention, and the startle reflex." Psychological Review, 97(3): 377–395.

Loewenstein, George. 2000. "Emotions in economic theory and economic behavior." American economic review, 90(2): 426–432.

Lucas, Richard E, Ed Diener, and Randy J Larsen. 2003. "Measuring positive emotions." In Positive Psychological Assessment: A Handbook of Models and Measures. , ed. Shane J Lopez and Charles Richard Snyder. American Psychological Association.

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

Lundberg, Shelly. 2017. "Non-cognitive skills as human capital." Education, Skills, and Technical Change: Implications for Future US GDP Growth, 77.

Lutz, Catherine, and Geoffrey M White. 1986. "The anthropology of emotions." Annual Review of Anthropology, 405–436.

Lyubomirsky, Sonja, Laura King, and Ed Diener. 2005. "The benefits of frequent positive affect: Does happiness lead to success?" Psychological bulletin, 131(6): 803.

Marcus, George E. 2000. "Emotions in politics." Annual Review of Political Science, 3(1): 221–250.

Martins, Alexandra, Nelson Ramalho, and Estelle Morin. 2010. "A comprehensive meta-analysis of the relationship between emotional intelligence and health." Personality and individual differences, 49(6): 554–564.

Mayer, John D, and Glenn Geher. 1996. "Emotional intelligence and the identification of emotion." Intelligence, 22(2): 89–113.

Mayer, John D., David R. Caruso, and Peter Salovey. 1999. "Emotional intelligence meets traditional standards for an intelligence." Intelligence, 27(4): 267–298.

Mayer, John D, Richard D Roberts, and Sigal G Barsade. 2008. "Human abilities: Emotional intelligence." Annu. Rev. Psychol., 59(1): 507–536.

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

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

Motro, Daphna, Jonathan B Evans, Aleksander PJ Ellis, and Lehman Benson III. 2022. "Race and reactions to women's expressions of anger at work: Examining the effects of the "angry Black woman" stereotype." Journal of Applied Psychology, 107(1): 142.

Nelson, Phillip. 1970. "Information and Consumer Behavior." Journal of Political Economy, 78(2): 311–329.

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

Nikolajeva, Maria. 2014. "33 Emotion ekphrasis: representation of emotions in children's picturebooks." Visual communication, 711.

Ohman, Arne. 2002. "Automaticity and the amygdala: Nonconscious responses to emotional faces." Current Directions in Psychological Science, 11(2): 62–66.

Öhman, Arne. 2007. "Has evolution primed humans to "beware the beast"?" Proceedings of the National Academy of Sciences, 104(42): 16396–16397.

Öhman, Arne. 2009. "Of snakes and faces: An evolutionary perspective on the psychology of fear." Scandinavian journal of psychology, 50(6): 543–552.

Parker, James DA, Laura J Summerfeldt, Marjorie J Hogan, and Sarah A Majeski. 2004. "Emotional intelligence and academic success: Examining the transition from high school to university." Personality and individual differences, 36(1): 163–172.

Pieters, Rik, and Michel Wedel. 2004. "Attention capture and transfer in advertising: Brand, pictorial, and text-size effects." Journal of Marketing, 68(2): 36–50.

Plutchik, Robert, and Henry Kellerman. 1980. Emotion Theory, Research, and Experience. Academic press.

Rasmussen, Eric E, Autumn Shafer, Malinda J Colwell, Shawna White, Narissra Punyanunt-Carter, Rebecca L Densley, and Holly Wright. 2016. "Relation between active mediation, exposure to Daniel Tiger's Neighborhood, and US preschoolers' social and emotional development." Journal of Children and Media, 10(4): 443–461.

Rorty, Amélie Oksenberg, et al. 1980. Explaining Emotions. Vol. 22, University of California Press Berkeley.

Russell, James A. 2003. "Core affect and the psychological construction of emotion." Psychological review, 110(1): 145.

Sacerdote, Bruce, Ranjan Sehgal, and Molly Cook. 2020. "Why is all COVID-19 news bad news?" National Bureau of Economic Research.

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

Solomon, Robert C. 1976. The Passions: The Myth and Nature of Human Emotion. Anchor.

Spence, Michael. 1973. "Job Market Signaling." The Quarterly Journal of Economics, 87(3): 355–374.

Sullivan, Denis G, and Roger D Masters. 1988. ""Happy warriors"': Leaders' facial displays, viewers' emotions, and political support." American Journal of Political Science, 345–368.

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.

Taylor, Shelley E, and Suzanne C Thompson. 1982. "Stalking the elusive "vividness" effect." Psychological review, 89(2): 155.

Thelwall, Mike, Kevan Buckley, and Georgios Paltoglou. 2011. "Sentiment in Twitter events." Journal of the American Society for Information Science and Technology, 62(2): 406–418.

Tomer, John F. 2003. "Personal capital and emotional intelligence: an increasingly important intangible source of economic growth." Eastern Economic Journal, 29(3): 453–470.

Trampe, Debra, Jordi Quoidbach, and Maxime Taquet. 2015. "Emotions in everyday life." PloS One, 10(12): e0145450.

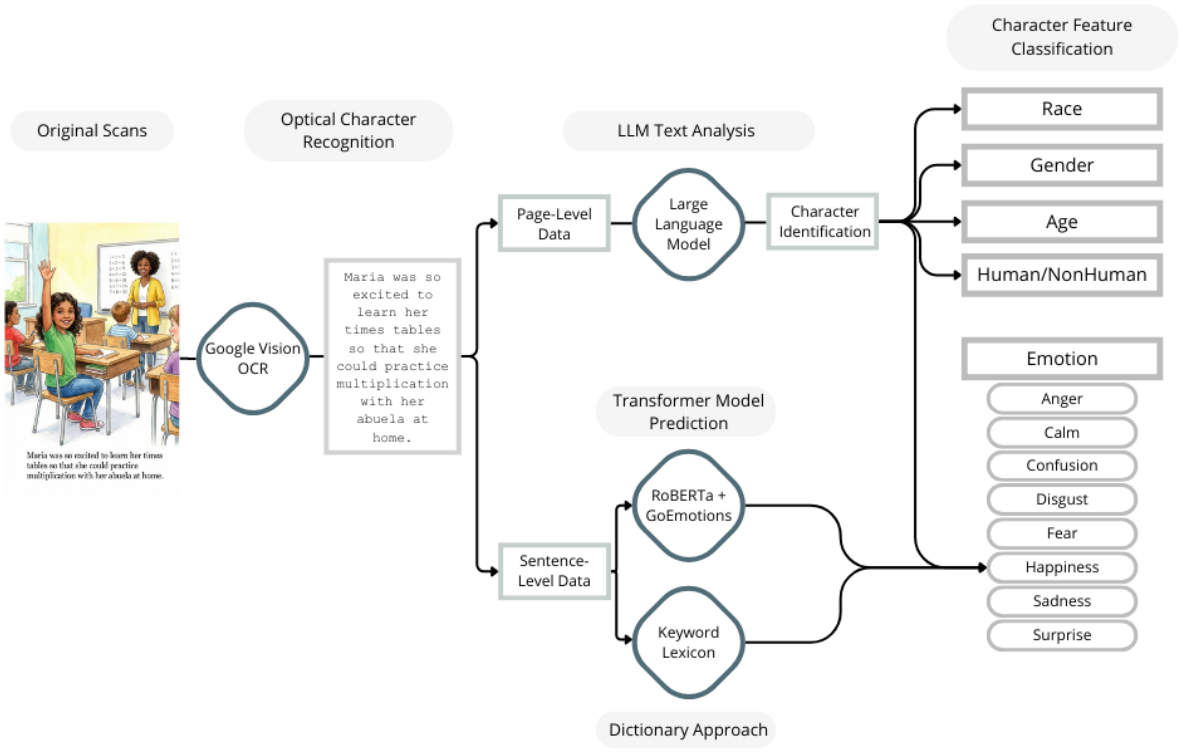
Underwood, Robert L, and Nicole M Klein. 2002. "Packaging as brand communication: Effects of product pictures on consumer responses to the package and brand." Journal of Marketing Theory and Practice, 10(4): 58–68.

Van Der Lans, Ralf, Rik Pieters, and Michel Wedel. 2008. "Research Note - Competitive Brand Salience." Marketing Science, 27(5): 922–931.

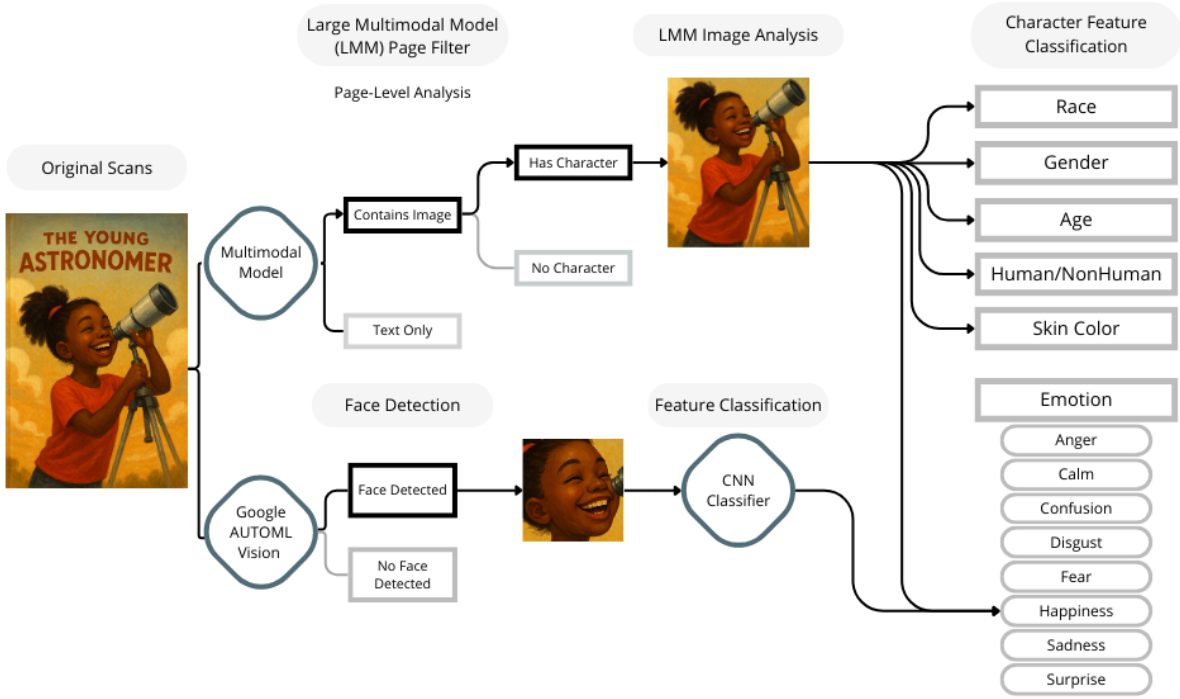
Yanagizawa-Drott, David H., and Hans-Joachim Voth. 2023. "Image(s)." Working Paper, University of Zurich.

FIGURE 1
Data Extraction Pipelines

(a) Text Analysis



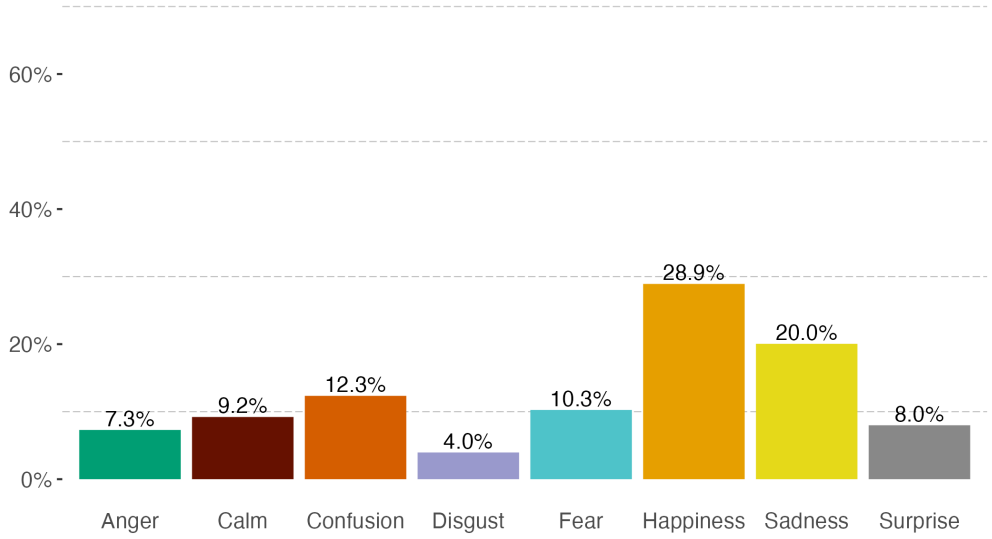
(b) Image Analysis



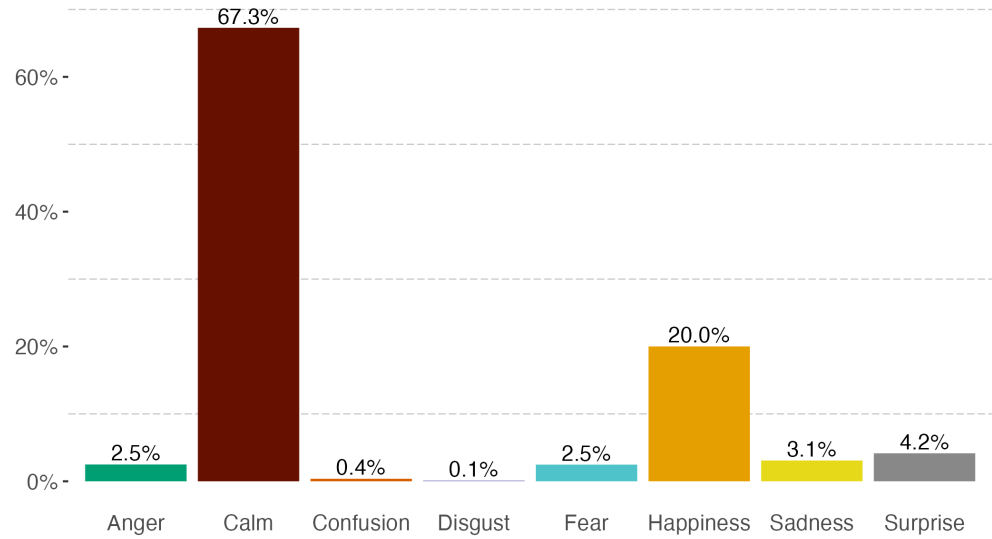
Notes: This figure illustrates the steps in our text analysis (top) and image analysis (bottom) pipelines beginning with the raw scan of each page in our sample of books.

FIGURE 2
Overall Representation of Emotions

(a) Text



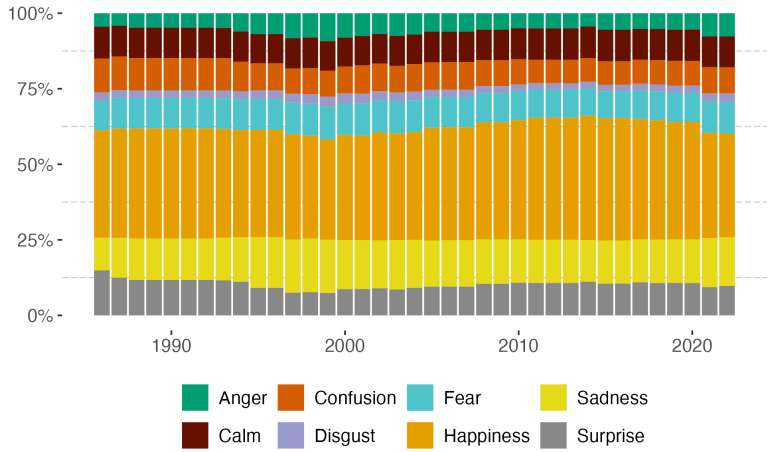
(b) Images



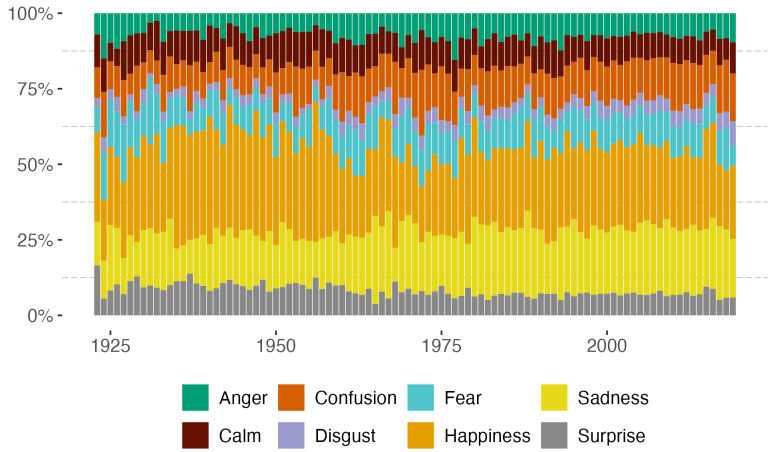
Notes: This figure shows the aggregate displays of emotions in text and images. The top panel shows the percentage of emotions conveyed by characters as a share of all emotions detected in the text of the books (literature and textbooks) using a large language model. The bottom panel shows the percentage of emotions conveyed by faces as a share of all emotions detected in the images of books (literature and textbooks) using a large multimodal model.

FIGURE 3
Representation of Emotions Over Time

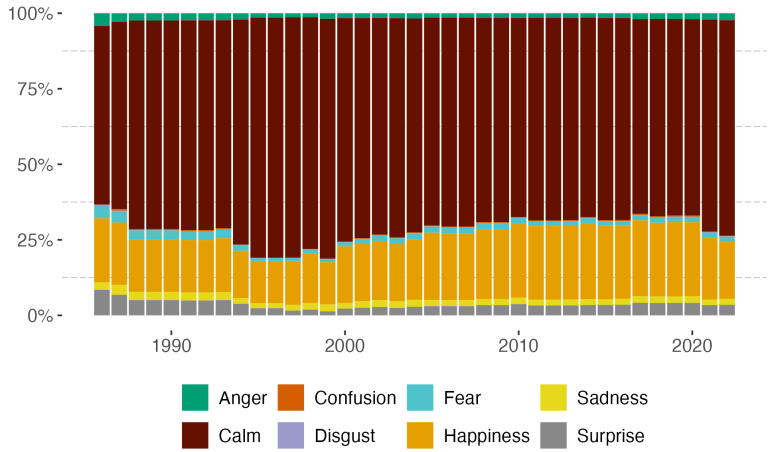
(a) Text: Textbooks



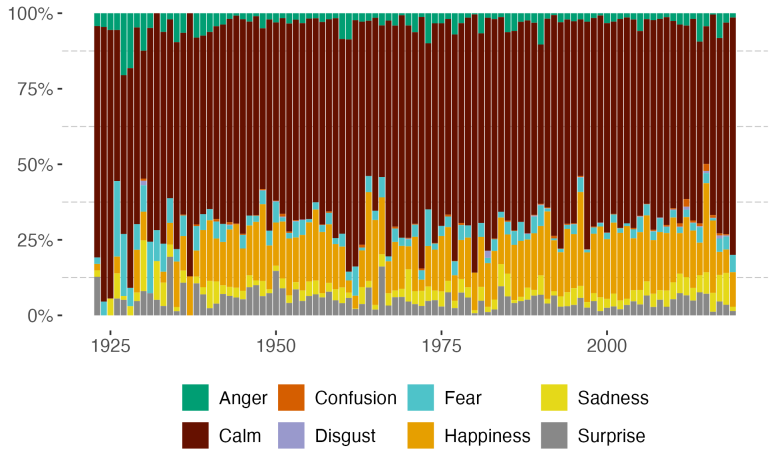
(b) Text: Children's Books



(c) Images: Textbooks



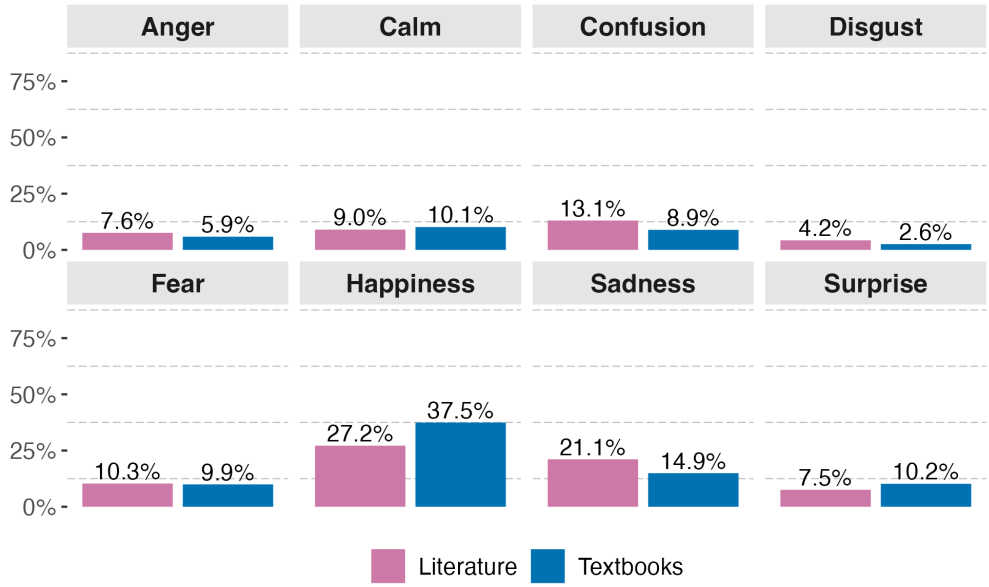
(d) Images: Children's Books



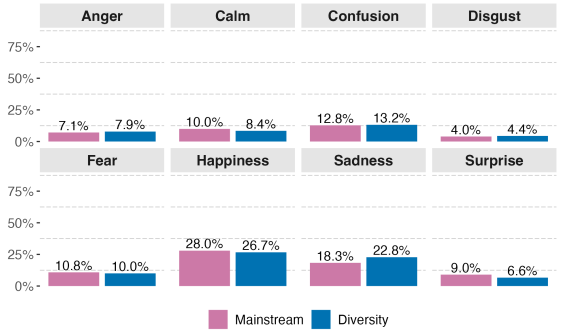
Notes: This figure shows the percentage of emotions detected over time. Panel (a) shows the share of all emotions detected in the sample of textbooks, and panel (b) shows the shares in literature. Emotions are detected using OpenAI. Emotions for each textbook are counted for each year the textbook is in use, and emotions for each children's book are counted for the book's copyright year. Panels (c) and (d) show the analogous results for emotions detected in faces in both corpora.

FIGURE 4
Representation of Emotions in the Text of Different Subcollections

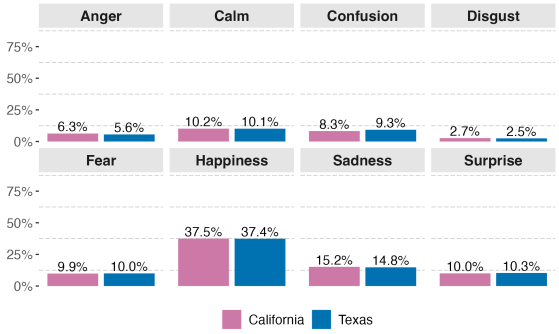
(a) Children's Literature vs. Textbooks



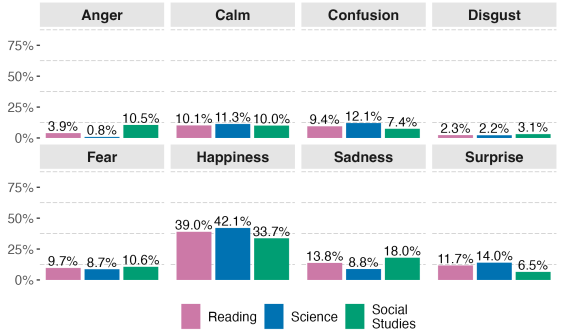
(b) Mainstream vs. Diversity



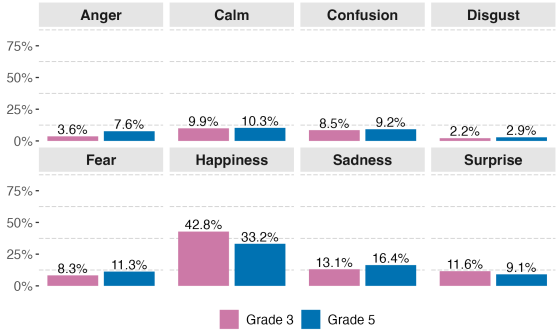
(c) California vs. Texas Textbooks



(d) By Textbook Subject



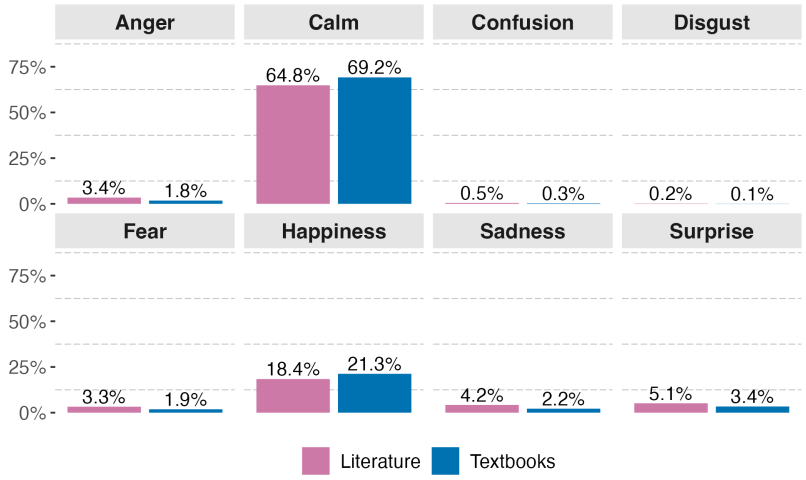
(e) By Textbook Grade



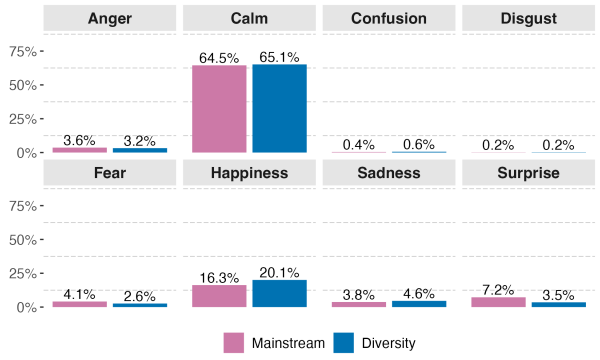
Notes: This figure shows the percentage of emotions conveyed by characters as a share of all emotions detected by OpenAI separately for different subcollections. Panel (a) compares textbooks and children's books, panel (b) compares California and Texas textbooks, panel (c) compares science, reading, and social studies textbooks, panel (d) compares grade three and grade five textbooks, and panel (e) compares the Mainstream and Diversity collections of children's books.

FIGURE 5
Representation of Emotions in the Images of Different Subcollections

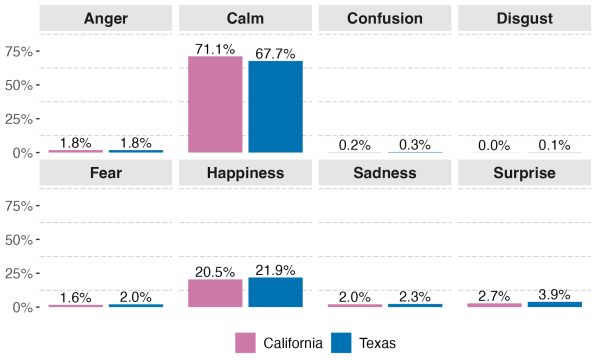
(a) Children's Literature vs. Textbooks



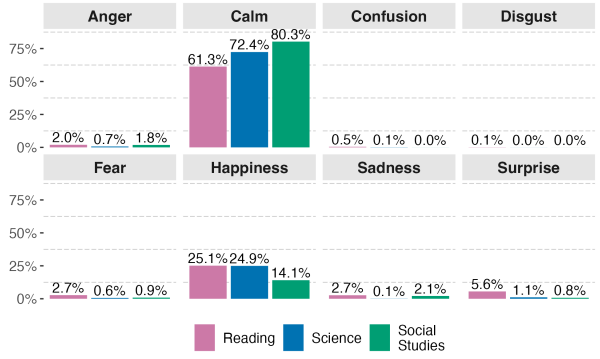
(b) Mainstream vs. Diversity



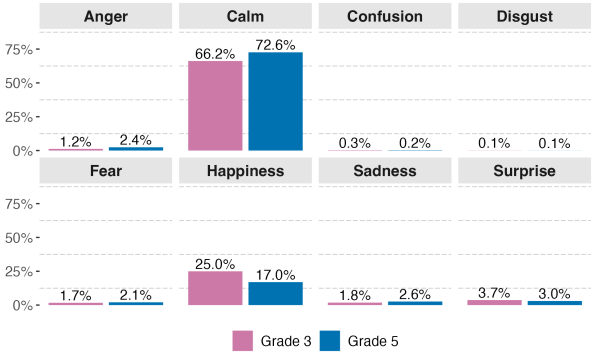
(c) California vs. Texas Textbooks



(d) By Textbook Subject



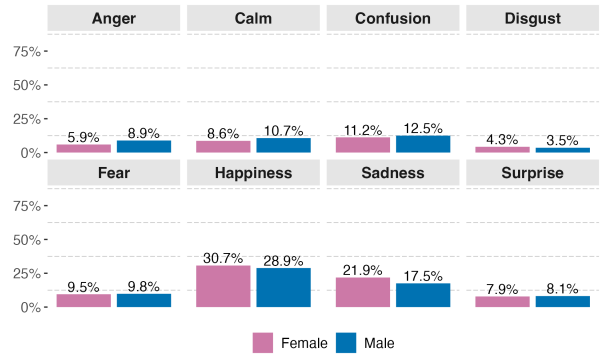
(e) By Textbook Grade



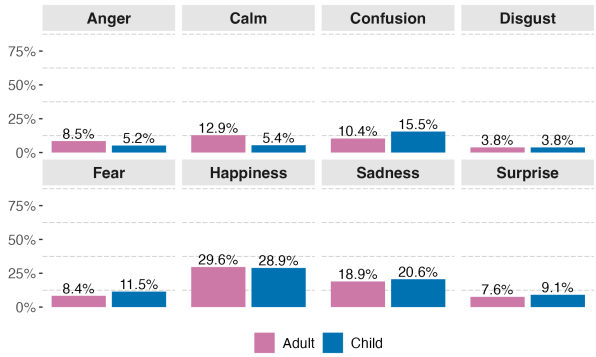
Notes: This figure shows the percentage of emotions conveyed by detected faces as a share of all emotions detected by an LMM separately for different subcollections. Panel (a) compares textbooks and children's books, panel (b) compares California and Texas textbooks, panel (c) compares science, reading, and social studies textbooks, panel (d) compares grade three and grade five textbooks, and panel (e) compares the Mainstream and Diversity collections of children's books.

FIGURE 6
Displays of Emotions in the Text for Characters with Different Identities

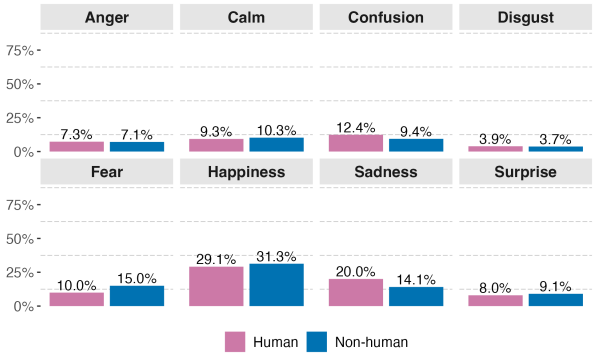
(a) Gender



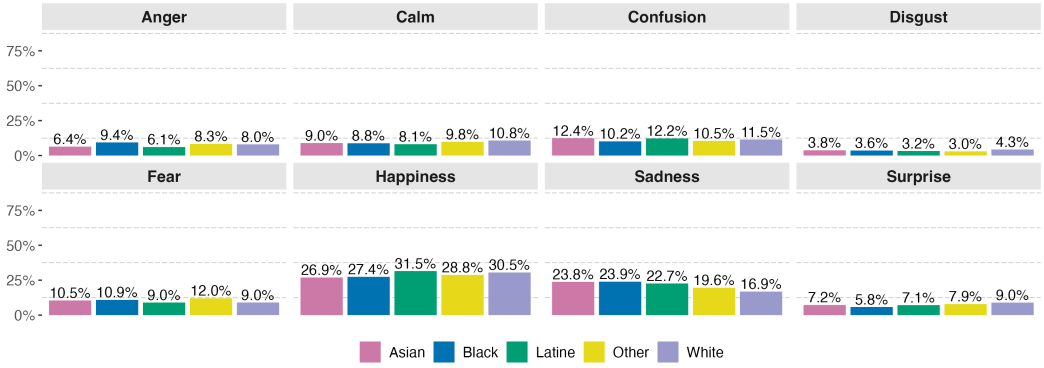
(b) Age



(c) Human vs. Non-Human



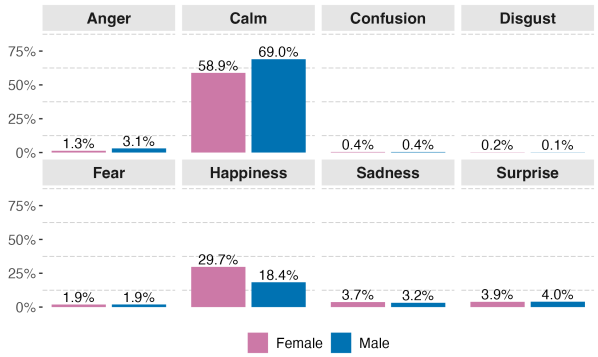
(d) Race



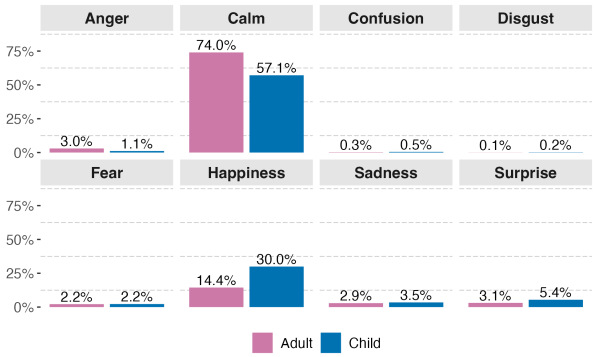
Notes: This figure shows the percentage of emotions conveyed in the text by characters of different identities. Panel (a) compares female and male characters, panel (b) compares human characters of different races, panel (c) compares characters of different ages, and panel (d) compares human and non-human characters.

FIGURE 7
Displays of Emotions in the Faces of Characters with Different Identities

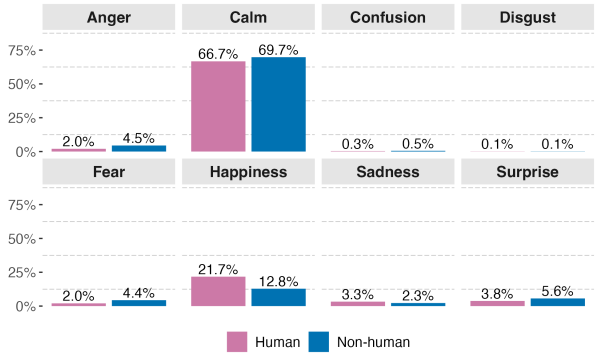
(a) Gender



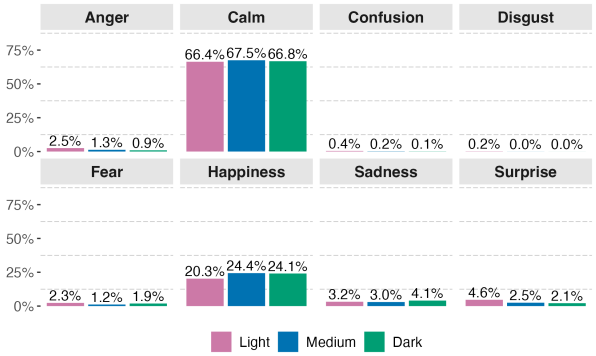
(b) Age



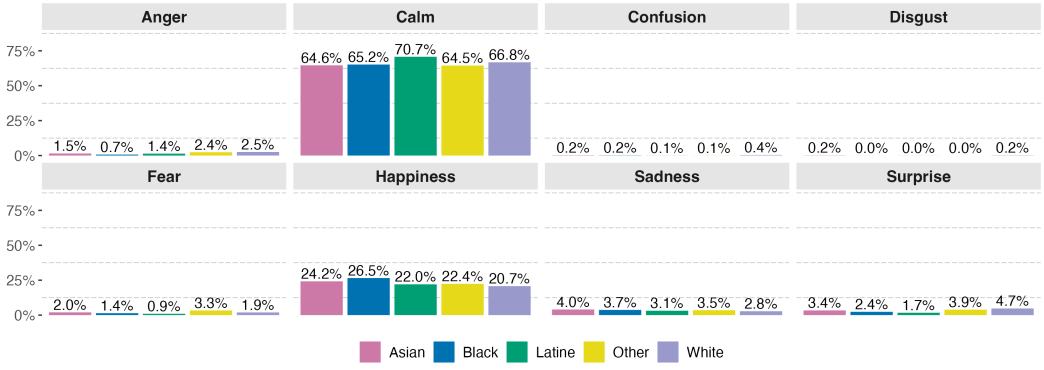
(c) Human vs. Non-Human



(d) Skin Color

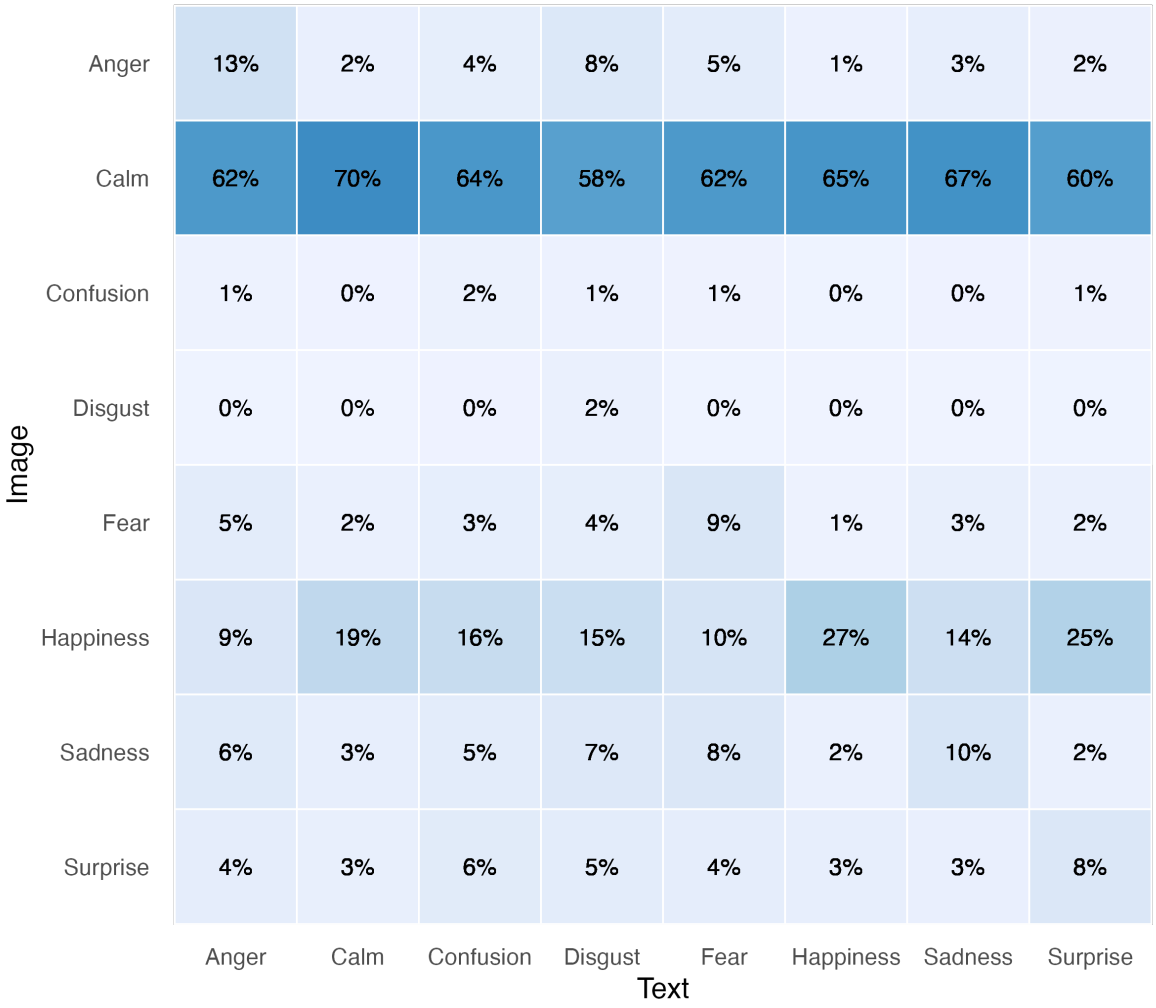


(e) Race



Notes: This figure shows the percentage of emotions conveyed in the detected faces of characters of different identities. Panel (a) compares female and male characters, panel (b) compares human characters of different races, panel (c) compares characters of different ages, panel (d) compares human and non-human characters, and panel (e) compares characters with different skin color.

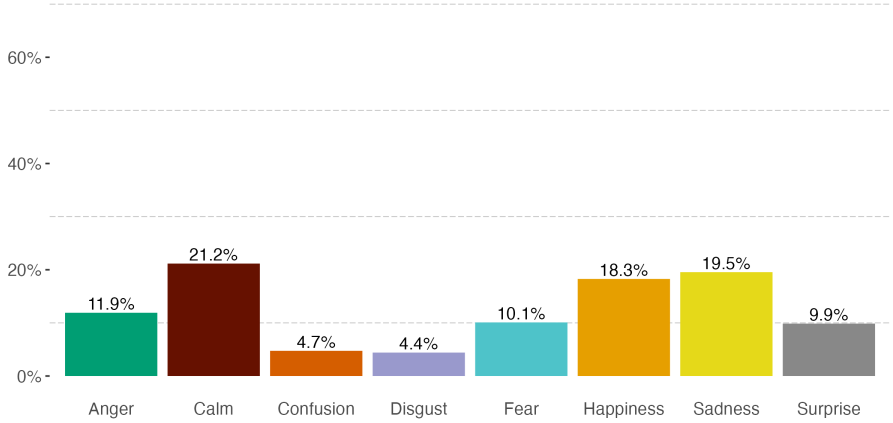
FIGURE 8
Normalized Co-occurrence Matrix



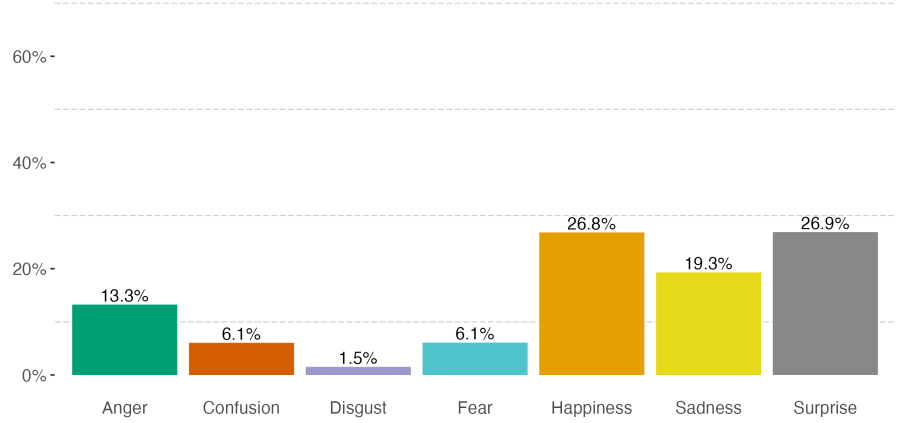
Note: This figure shows the co-occurrence of emotions for text and images on the same page, normalized by the text columns. In other words, each cell shows the proportion of times a specific image label occurred, given a particular text label. For example, the top left cell shows that when anger is found in text, an image on the same page will depict anger 22% of the time.

FIGURE 9
Robustness: Overall Representation of Emotions

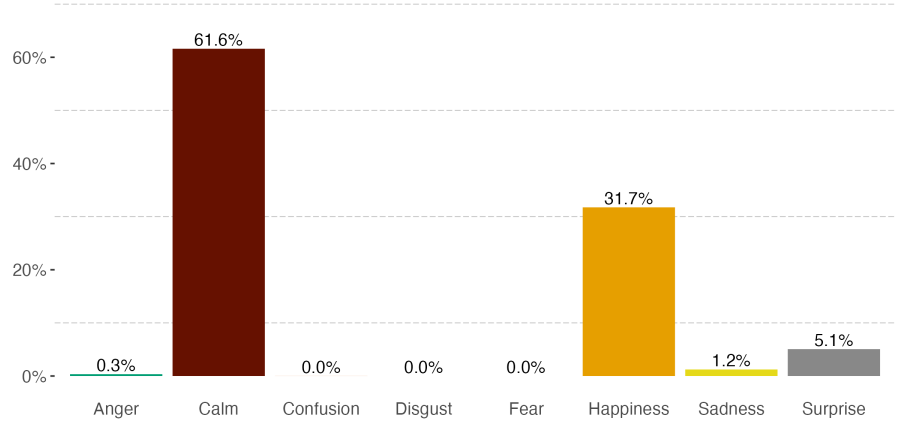
(a) Text: Lexicon Approach



(b) Text: GoEmotions AI Model



(c) Images: Feature Classification



Notes: This figure shows the aggregate displays of emotions in text and images using alternative approaches. Panel (a) shows the percentage of emotion-specific words as a share of all emotion words from a pre-specified lexicon. Panel (b) shows the percentage of emotion-specific sentences as a share of all emotion sentences as classified by the GoEmotions model. Panel (c) shows

TABLE 1
Book Purchases

	Dependent variable: Whether Purchased		
	(1)	(2)	(3)
Happy or Calm Only Cover	0.160***	0.554***	0.667***
	(0.025)	(0.031)	(0.031)
No Cover Character	1.186***	0.490***	0.542***
	(0.024)	(0.030)	(0.031)
Book Controls		X	X
Linear Emotion Controls		X	
Quintile Emotion Controls			X
Observations	28,521	28,521	28,521
AIC	351,716.6	296,521.2	295,216.7
BIC	351,745.1	298,218.3	296,999.3
RMSE	0.05	0.05	0.05

Notes: This table presents results from a logistic regression of the binary purchase outcome on cover emotion measures. The sample is restricted to literature books and characters that express a defined emotion. "Happy or Calm Only Cover" equals 1 when the cover's characters display only calmness or happiness; "No Cover Character" equals 1 when no characters are depicted on the cover; the omitted group is books that contain at least one character who portrays an emotion other than happy or calm. Book controls include a variety of book characteristics including indicator variables for the award(s) the book won, the book's copyright year, the number of pages, average words per page, whether the book is a picture book (characters with emotions only appear in images), and whether it contains no images (characters with emotions appear only in text). Observations refer to the number of times we observe a unique book being purchased by a household across all households. We cluster standard errors by household. In this table, ***, **, and * represent significance at the 0.01, 0.05, and 0.1 levels, respectively.

Researcher(s)' own analyses calculated (or derived) based in part on data from Market Track, LLC dba Numerator and marketing databases provided through the Numerator Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Numerator data are those of the researcher(s) and do not reflect the views of Numerator. Numerator is not responsible for and had no role in analyzing or preparing the results reported herein. LLMs and AI tools are never applied to the Numerator data.

TABLE 2
Library Inventory

	Dependent variable: Number of Books in Branch Inventory		
	(1)	(2)	(3)
Happy or Calm Only Cover	0.1334***	0.1174***	0.0760***
	(0.0191)	(0.0280)	(0.0270)
No Cover Character	0.1213***	0.1839***	0.1366***
	(0.0206)	(0.0420)	(0.0372)
Book Controls		X	X
Linear Emotion Controls		X	
Quintile Emotion Controls			X
Observations	6,944	6,944	6,944
Squared Correlation	0.00313	0.38340	0.38359
Pseudo R^2	0.00132	0.20337	0.20350
BIC	16,514.9	14,196.2	14,247.1

Notes: This table presents results from a linear regression of the library checkout outcome on cover and content emotion measures, controlling for the distribution of emotions in both text and images. The sample is restricted to literature books and characters that express a defined emotion. "Happy or Calm Only Cover" equals 1 when the cover's characters display only calmness or happiness; "No Cover Character" equals 1 when no characters are depicted on the cover; the omitted group is books that contain at least one character who portrays an emotion other than happy or calm. Book controls include a variety of book characteristics including indicator variables for the award(s) the book won, the copyright year of the book, the number of pages, average words per page, whether the book is a picture book (characters with emotions only appear in images), and whether it contains no images (characters with emotions appear only in text). We also add fixed effects for the specific library and cluster standard errors by specific library. In this table, ***, **, and * represent significance at the 0.01, 0.05, and 0.1 levels, respectively.

TABLE 3
Relationship Between the Book Cover and Content

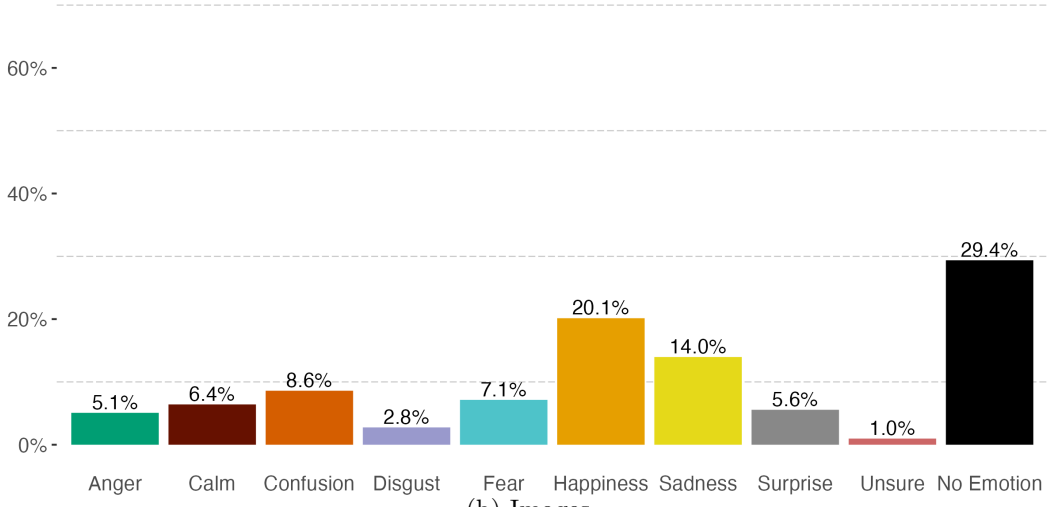
	Book Cover:		
Proportion Happy or Calm in Book Content:	Happy or Calm Cover (1 = yes) (1)	Mixed Cover (1 = yes) (2)	Not Happy or Calm Cover (1 = yes) (3)
Images	3.420***	-3.370***	-5.520***
	(0.471)	(0.867)	(0.982)
Text	2.202***	1.403	-3.510*
	(0.659)	(1.250)	(1.803)
Book Controls	X	X	X
Observations	1,113	1,113	1,113
AIC	1,310.5	557.3	461.4
BIC	1,917.3	1,164.1	1,068.2
Log Likelihood	-534.246	-157.658	-109.700
RMSE	0.40	0.20	0.17

Notes: This table presents results from a logistic regression where the outcome in column one indicates whether a book cover has only happy or calm characters, the outcome in column two indicates whether a book cover has a mix of happy or calm characters and characters that are neither happy nor calm, and the outcome in column three indicates whether a book cover has exclusively characters that are neither happy nor calm. The two variables shown are the proportion of happy or calm characters in images and text in the entire book. The book controls include the number of pages, average words per page, whether there are no characters in the text or images with an emotion, and indicator variables for the award(s) the book won and the copyright year. In this table, ***, **, and * represent significance at the 0.01, 0.05, and 0.1 levels, respectively.

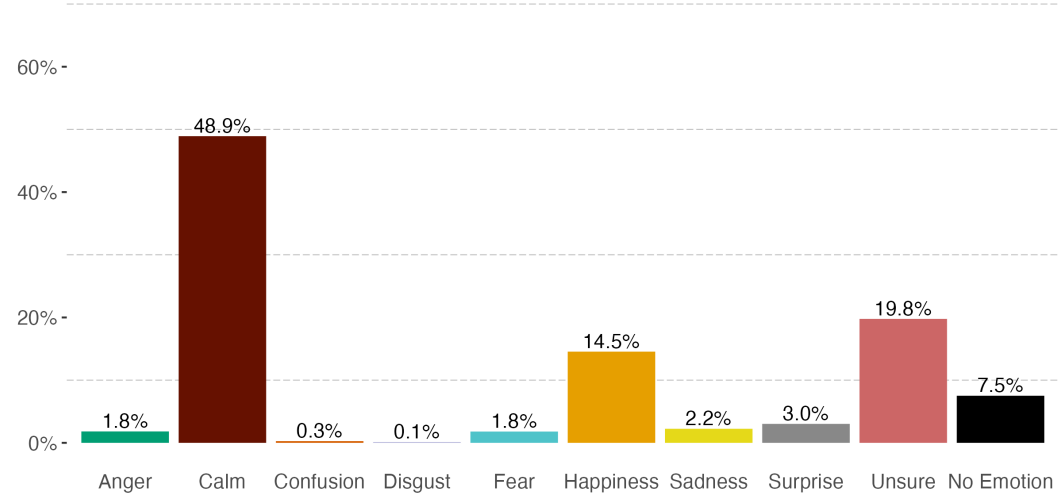
A Appendix Figures

FIGURE A.1
Overall Representation of Emotions (Including No Emotion and Unsure)

(a) Text



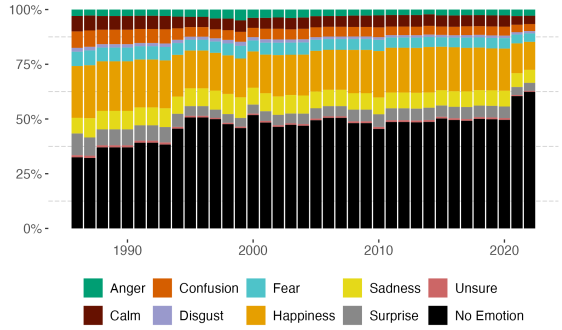
(b) Images



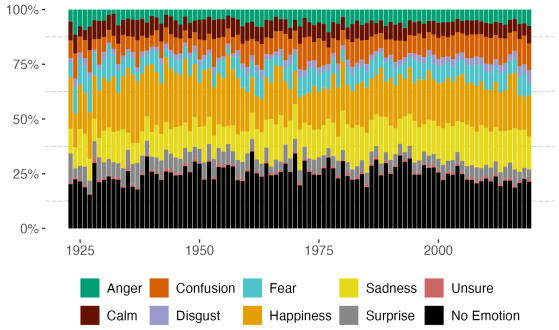
Notes: This figure replicates Figure 2 with additional categories for "No Emotion" and "Unsure." The top panel shows the percentage of emotions conveyed by characters as a share of all emotions detected in the sample of books (literature and textbooks). Emotions are detected using an LMM. The bottom panel shows the percentage of emotions conveyed by faces as a share of all emotions detected in the sample of books (literature and textbooks). Emotions are detected using an LMM.

FIGURE A.2
Representation of Emotions Over Time (including No Emotion and Unsure)

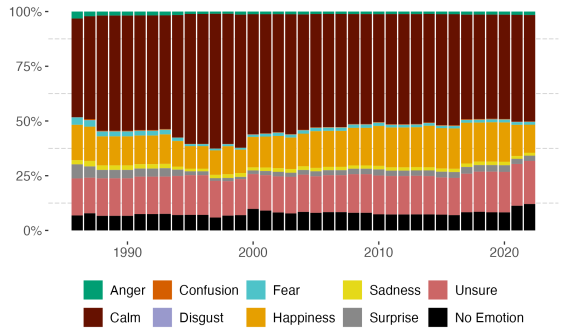
(a) Text: Textbooks



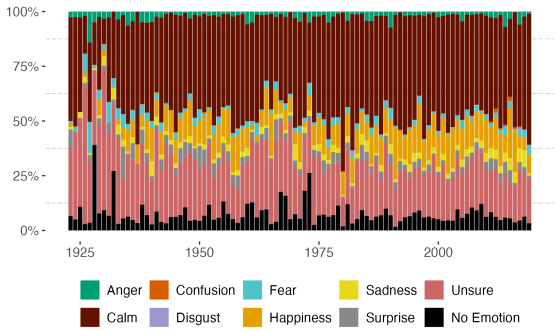
(b) Text: Children's Books



(c) Images: Textbooks



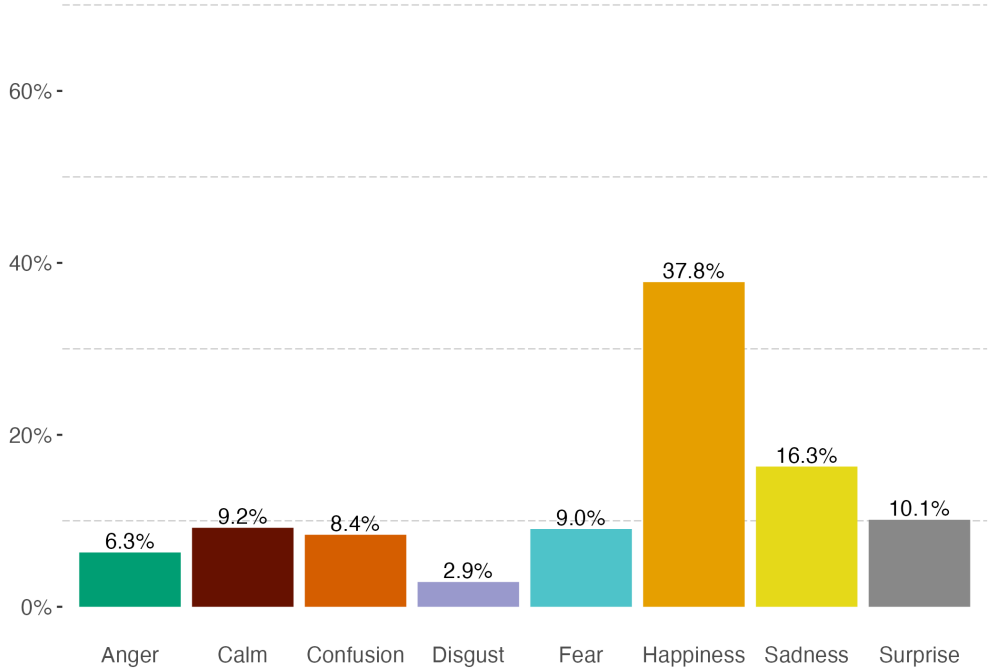
(d) Images: Children's Books



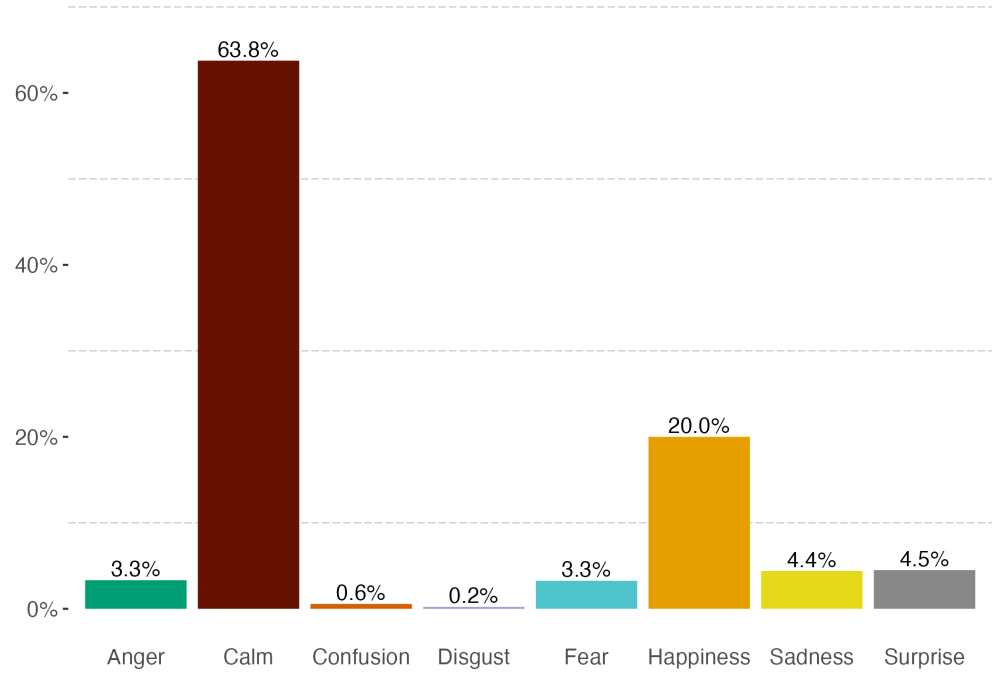
Notes: This figure replicates Figure 3 with additional categories for "No Emotion" and "Unsure." Panel (a) shows the share of all emotions detected in the sample of textbooks, and panel (b) shows the shares in literature. Emotions are detected using OpenAI. Emotions for each textbook are counted for each year the textbook is in use, and emotions for each children's book are counted for the book's copyright year. Panels (c) and (d) show the analogous results for emotions detected in faces in both corpora.

FIGURE A.3
Overall Representation of Emotions (Matched Sample)

(a) Text

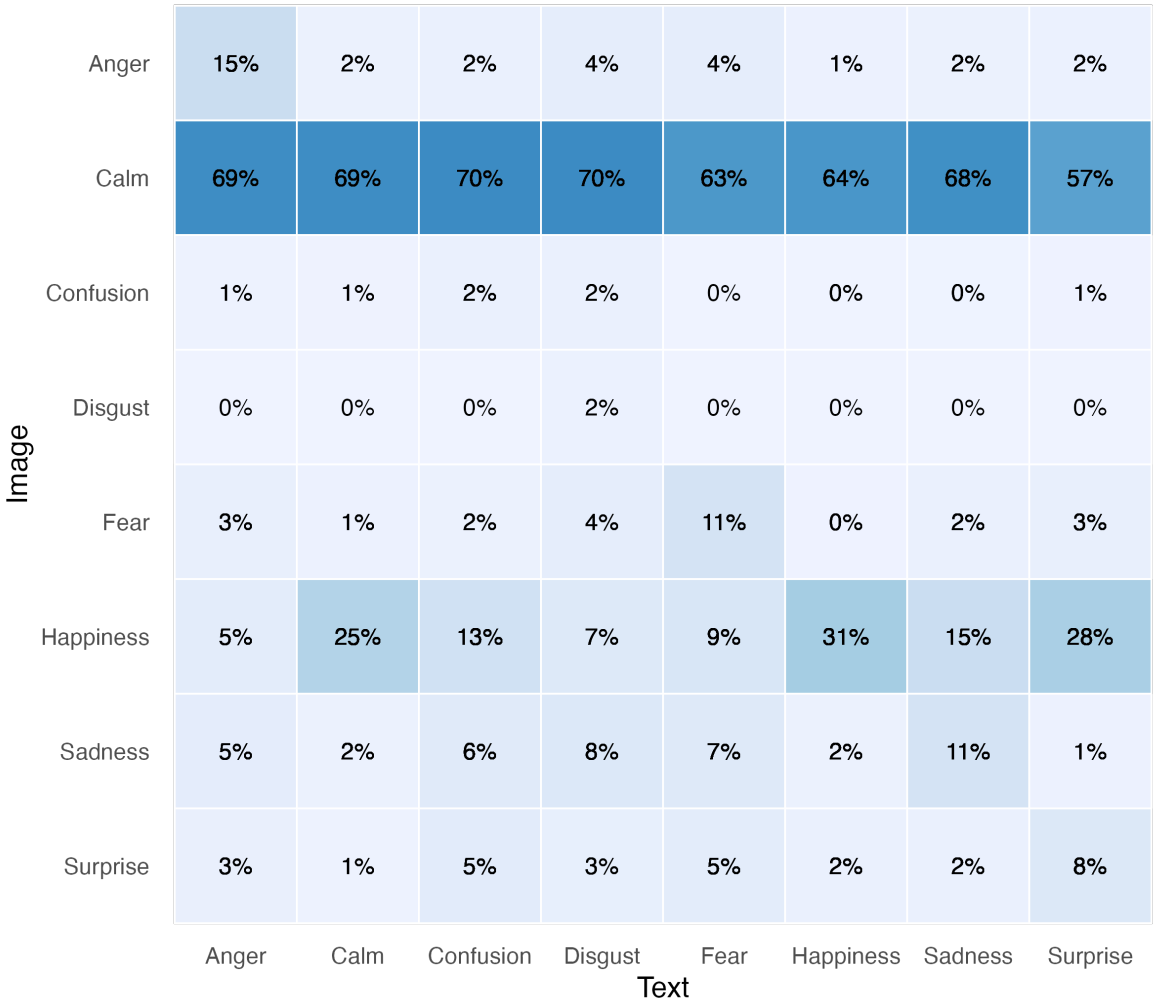


(b) Images



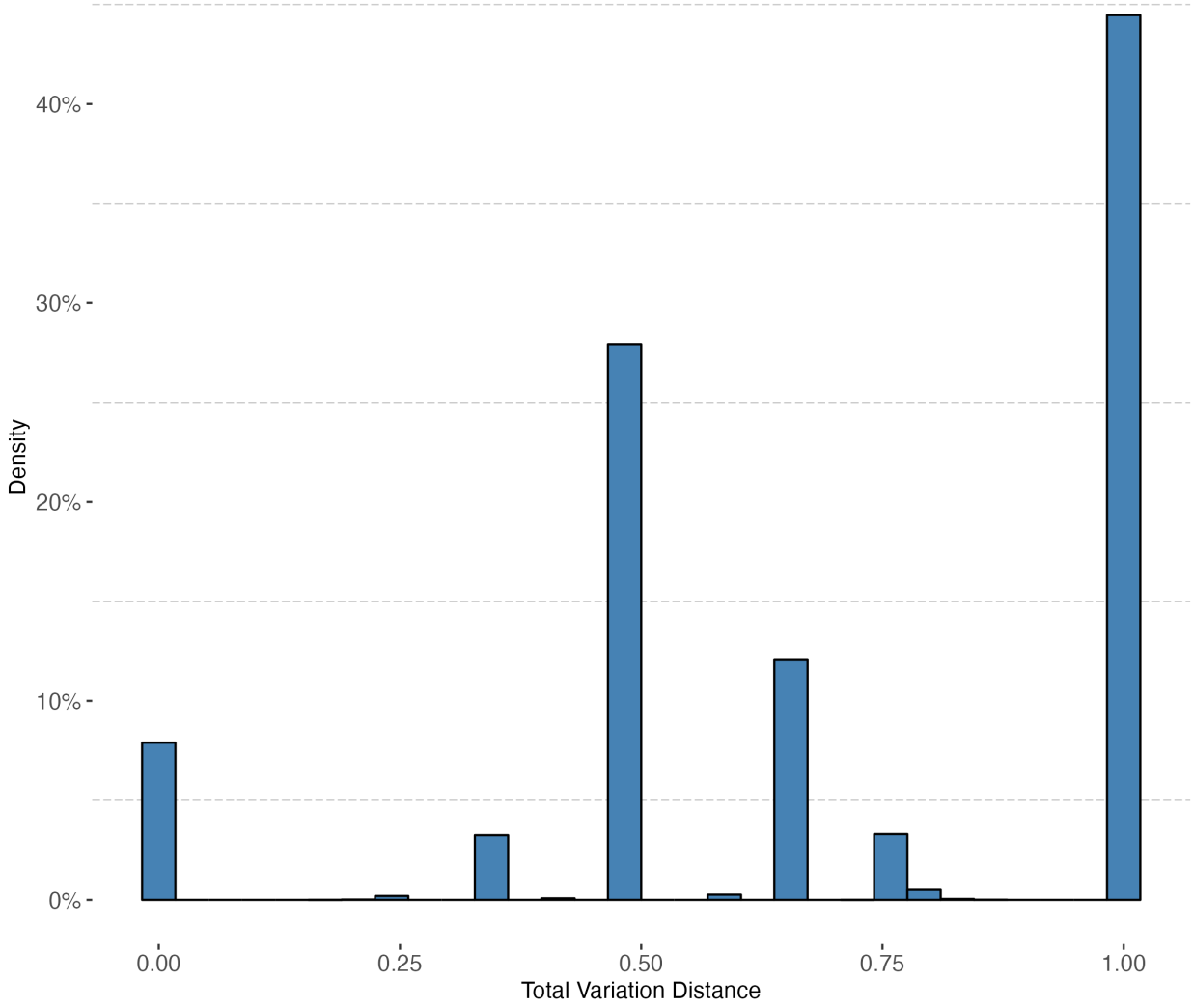
Notes: This figure shows the aggregate displays of emotions in text and images for the sample of matched pages. The top panel shows the percentage of emotions conveyed by characters as a share of all emotions detected in the sample of books (literature and textbooks); emotions are detected using an LMM. The bottom panel shows the percentage of emotions conveyed by faces as a share of all emotions detected in the sample of books (literature and textbooks); emotions are detected using an LMM.

FIGURE A.4
Normalized Co-occurrence Matrix for Characters Matched in Text and Images



Note: This figure shows the co-occurrence of emotions for text and images on the same page, where unique human characters are matched on their race and gender. Co-occurrence is normalized by the text columns. In other words, each cell shows the proportion of times a specific image label occurred, given a particular text label. For example, the top left cell shows that for pages where a character is shown as angry in text, that same character (i.e., a character with the same race and gender) will be depicted as angry in an image only 20% of the time.

FIGURE A.5
Total Variation Distance Between Text and Images



Note: This figures shows the total variation distance (tvd) between the emotional content of images and text on the same page. We calculate tvd by taking the l^1 distance between the empirical distributions of our 8 emotions in text and in images on a given page and multiply by $1/2$.

B Appendix Tables

TABLE B.1
Sample of Mapping from Keyword Lexicon to Emotions in Main Analysis

anger	**calm**	**confusion**	**disgust**	**fear**	**happiness**	**sadness**	**surprise**
acerb	abide	absentminde	abhor	affright	affection	abandon	aah
acerbic	amiable	absurd	abhorrence	afraid	amuse	abandonment	aflutter
acerbity	aplomb	absurdity	abhorrent	aghast	amusing	ache	ah
acrid	asleep	addle	abhorringly	alarm	banter	afflict	aha
acrimonious	assuage	baffle	abominable	alarum	beam	affliction	ahha
...	...	...	...	...	...	...	...

Note: This table presents a sample of our lexicon dictionary. Each of the eight emotion categories is associated with a mutually exclusive set of keywords. The full lexicon can be viewed here

TABLE B.2
Mapping from GoEmotions to Emotions in Main Analysis

anger	**confusion**	**disgust**	**fear**	**happiness**	**sadness**	**surprise**
anger	confusion	disgust	fear	amusement	embarrassment	curiosity
annoyance			nervousness	excitement	disappointment	realization
disapproval				joy	grief	surprise
				love	remorse	
				optimism	sadness	

Note: This table presents the mapping used in our main analysis to group the 28 fine-grained GoEmotions into the eight broader emotion categories used in our main analysis. We exclude the following fine-grained emotions from the joy category: admiration, approval, caring, desire, gratitude, pride, relief.

C Methods Appendix

C.A Image Analysis

We quantify emotional representation in illustrated pages of children's books using a two-stage, fully automated pipeline Figure 1b implemented using Google Cloud Vertex AI and Gemini 2.5 Pro Preview (model version `gemini-2.5-pro-preview-03-25`).

In Stage 1, each scanned page image (`.jpg/.png/.tif`), is processed using a lightweight Gemini classifier. This classifier assigns a binary label (either `TEXT_ONLY` or `CONTAINS_IMAGE`) using a low temperature setting of 0.1 and a response limit of 256 tokens to ensure consistent binary classification.[12] Pages identified as `TEXT_ONLY` are documented and excluded from further analysis, whereas pages labeled `CONTAINS_IMAGE` pages advance to Stage 2.

In Stage 2, involves a detailed analysis where the Gemini model is again utilized to produce a validated`application/json` output. This output includes structured data at both the scene and character levels (e.g., setting metadata, per-character bounding boxes, demographic attributes, and emotion labels). The schema is designed to enumerate categorical values and includes three distinct emotion codings (`emotion_6_options`, `emotion_10_options`, `emotion_28_options`). Additionally, we also request a perceptual lightness score (L*, 0–100) for skin tone. This stage allows for a maximum of10,000 tokens in this stage to accommodate images featuring numerous characters.. All outputs are serialized to a corpus-level file `{corpus}_consolidated_results.json` that we later aggregate for analysis and visualization. Full prompts, the exact JSON schema, and run-time arguments are provided below to support precise replication.

Stage 1: Automatic Page Filtering

The initial stage involves detecting whether a page contains illustrations to avoid unnecessary processing of text-only pages. For each scanned page, we instruct the model to respond with exactly one token: `text_only` or `contains_image`.

Model parameters:

- temperature = 0.1:
  Controls randomness in the model's output: lower values make the response more deterministic, reducing spurious labels.
- maxOutputTokens = 256:
  Sets an upper limit on the number of tokens (words or subwords) in the response, ensuring the model returns a concise label rather than a long explanation.

Pages labeled `text_only` are skipped; those labeled `contains_image` proceed to Stage 2.

Stage 2: Structured Emotion & Character Extraction

For pages flagged `contains_image`, the Gemini model outputs a strictly validated JSON object capturing:

[12] Implemented via a dedicated detection model with system instructions that restrict the output to exactly one of the two strings; see code for the detection prompt.

1. Scene-level context (`setting_description`): what is happening, the atmosphere, visual style, etc.
2. Character list (`characters`): one entry per detected figure, annotated with demographics and one of eight target emotions (happiness, sadness, calm, surprise, anger, confusion, disgust, fear).

JSON Schema

Embedding a JSON schema in the prompt enforces that every response contains exactly the prescribed fields (no extras, no omissions), uses the correct data types (string, integer, array), and respects enumerated options (e.g., gender $\in$ {male, female, unsure}.)

```
{
  "type":"OBJECT",
  "properties":{
    "setting_description":{
      "type":"OBJECT",
      "properties":{
        "what_is_happening":{"type":"STRING"},
        "how_are_activities_portrayed":{"type":"STRING"},
        "is_it_a_photograph_or_illustration":{
         "type":"STRING","enum":["Photograph","Illustration"]
        },
        "what_are_the_topics_portrayed":{
          "type":"ARRAY","items":{"type":"STRING"}
        },
        "what_are_the_values_portrayed":{
          "type":"ARRAY","items":{"type":"STRING"}
        },
        "is_it_realistic_or_fantasy":{
          "type":"STRING","enum":["Realistic","Fantasy"]
        },
        "is_it_present-day_or_historical":{
          "type":"STRING","enum":["Present-day","Historical"]
        },
        "what_is_the_atmosphere_of_the_setting":{"type":"STRING"},
        "is_it_a_polychromatic_or_monochromatic_image":{
          "type":"STRING","enum":["Polychromatic","Monochromatic"]
        },
        "image_size":{
          "type":"OBJECT",
          "properties":{"width":{"type":"INTEGER"},"height":{"type":"
             INTEGER"}},
          "required":["width","height"]
        }
      },
      "required":[
        "what_is_happening","how_are_activities_portrayed", "
           is_it_a_photograph_or_illustration","
           what_are_the_topics_portrayed", "what_are_the_values_portrayed"
           ,"is_it_realistic_or_fantasy",
        "is_it_present-day_or_historical", "
           what_is_the_atmosphere_of_the_setting", "
```

```
            is_it_a_polychromatic_or_monochromatic_image","image_size"
        ]
      },
      "characters":{
        "type":"ARRAY",
        "items":{
          "type":"OBJECT",
          "properties":{
            "character_id":{"type":"INTEGER"},
            "position_in_image":{
              "type":"ARRAY","items":{"type":"INTEGER"},
              "minItems":4,"maxItems":4,
              "description":"[x1,y1,x2,y2] in pixels"
            },
            "race":{"type":"STRING"},
            "gender":{"type":"STRING","enum":["male","female","unsure"]},
            "age":{
              "type":"STRING",
              "enum":["infant","child","teenager","adult","senior","unsure"]
            },
            "skin_color":{"type":"STRING","enum":["Light","Medium","Dark"]},
            "perceptual_tint":{"type":"NUMBER","minimum":0,"maximum":100},
            "assistive_devices":{"type":"STRING"},
            "hair_color":{"type":"STRING"},
            "eye_color":{"type":"STRING"},
            "actions":{"type":"STRING"},
            "emotions":{"type":"STRING"},
            "emotion_10_options":{
              "type":"STRING",
              "enum":[  "angry","calm","confused","disgusted","fear","happy"
                 ,"sad","surprised","unsure","no emotion"
              ]
            },
            "occupation":{"type":"STRING"},
            "is_human_or_non-human":{
              "type":"STRING","enum":["Human","Non-human"]
            },
            "location":{"type":"STRING"}
          },
          "required":[
            "character_id","position_in_image","race","gender","age","
               skin_color","perceptual_tint","assistive_devices","hair_color
               ","eye_color","actions","emotions","emotion_10_options","
               occupation","is_human_or-non-human",
            "location"
          ]
        }
      }
    },
    "required":["setting_description","characters"]
}
```

Model parameters:

- temperature = 0.0:
  Forces the model to be fully deterministic, which is crucial when strict adherence to the schema is required.

- maxOutputTokens = 10000:
  Allocates up to 10,000 tokens for the response, accommodating large JSON outputs for pages with many characters.

Information extracted
For each image presented in the books, we extract the following information:

1. Setting Description: This section provides context about the entire image. It includes:
   - `what_is_happening`: A general description of the activity in the image.
   - `how_are_activities_portrayed`: Whether the actions are calm, dynamic, aggressive, etc.
   - `is_it_a_photograph_or_illustration`: Visual style of the image ("Photograph" or "Illustration").
   - `what_are_the_topics_portrayed`: List of main themes or subjects (e.g., family, learning).
   - `what_are_the_values_portrayed`: List of abstract values (e.g., bravery, cooperation).
   - `is_it_realistic_or_fantasy`: Is the setting grounded in the real world?
   - `is_it_present-day_or_historical`: Temporal context (modern or past).
   - `what_is_the_atmosphere_of_the_setting`: Mood of the scene (e.g., tense, joyful).
   - `is_it_a_polychromatic_or_monochromatic_image`: Color style of the image.
   - `image_size`: Dimensions of the image in pixels (with `width` and `height`).
2. Characters: For each character, the model fills out:
   - `character_id`: Unique ID for each character.
   - `position_in_image`: Bounding box as `[x1, y1, x2, y2]` in pixels.
   - `race`, `gender`, `age`: Demographic labels.
   - `skin_color`: One of "Light", "Medium", or "Dark".
   - `perceptual_tint`: Score from 0–100 showing how tinted the skin appears.
   - `assistive_devices`: Any visible aids (e.g., glasses, cane).
   - `hair_color`, `eye_color`: Color of the eyes and hair.
   - `actions`: What the character is doing.

- `emotions`: Free-text description of their emotional state.
- `emotion_10_options`: Emotion classification within 'Anger', 'Calm', 'Confusion', 'Disgust', 'Fear', 'Happiness', 'Sadness', 'Surprise', 'None', 'Unsure'
- `occupation`: If a profession or role is visible (e.g., doctor).
- `is_human_or_non-human`: Human or non-human label.
- `location`: Where the character is in the scene (e.g., indoors, school).

Post-Processing and Recoding

After analyzing the scanned page using the Gemini-based model we process each JSON output file corresponding to each scanned files containing image, as follows:

- We parse it with a JSON library (e.g. `json.loads`); parse failures trigger up to two automatic retries.
- We recode `race` into categories (Black, Latine, Asian, White, Indigenous, Unsure, Other, Not applicable) using the exact keyword list presented below.
- We save both the raw JSON (for audit) and a flattened table (one row per character), including all schema fields plus recoded race and eight binary emotion flags.

Race Keywords List

The re-categorized variable for race (lower-cased and trimmed) was constructed according to the following classification of the Gemini model's output.

- Not applicable: cases and when human = "Non Human"
- Black: if race_original contains any of: black, african, afro, ashanti, mende
- Latine: if race_original contains any of: latin, cuban, rican, hispanic, mexica, mestizo, caboclo, mestiza
- Asian: if race_original contains any of: asian, afghan, chinese, hmong, indian, japanese, korean, mongol, tibetan, vietnamese, cambodian, filipino, kashmiri, kurdish, pashtun
- White: if race_original contains any of: white, greek, egyptian, european, german, russia, spanish, armenian, balkan, hungarian, italian, kosovan, mediterranean, norse, roman, celtic, jewish, ashkenazi
- Indigenous if race_original contains any of: indigenous, native, inuit, aboriginal, aben, aztec, guarani, hawaiian, inuk, inupiaq, kwakiutl, maya, navajo, olmec, penan, polynesian, saami, samoan, sioux, yup, zapotec, pacific islander, melanesian
- Unsure: if race_original contains any of: unclear, unknown, unspecified, unsure, ambiguous
- Other: if race_original contains any other classification

C.B Text Analysis

C.B.1 LMM Prompt Used for Emotion Identification

The following prompt was submitted to gpt-4o-mini (OpenAI, accessed April 2025) to generate structured character- and page-level emotion annotations:

```
Consider the following statement delimited by triple quotes:

<statement>: """STATEMENT"""

Your task is to find all emotions conveyed by each character in the text.
   For each character, determine the following items:

1. <character>: Provide the name of the character expressing emotions. If
   there is no character who expresses emotions, you can answer "not
   applicable". If you are unsure, you can answer "unsure". Otherwise,
   please answer with a string containing the name of the character
   expressing emotions.
2. <age>: Predict the age (in years) of <character>. If there is no
   character, return -999. If you are unsure, return -888. Otherwise,
   return the predicted age of the character in years as an integer.
3. <gender>: Predict the gender of <character>. Answer set: "male",
   "female", "other", "unsure", "not applicable".
4. <race>: Predict the race of <character>. Answer set: "white", "black",
   "asian", "latinx", "other", "unsure", "not applicable".
5. <human>: Predict whether <character> is human. If there is no
   character, return "not applicable". If you are unsure, return
   "unsure". Otherwise, return "yes" if <character> is a human and "no"
   otherwise.
6. <emotion_JSON>: Construct a JSON object that characterizes all
   emotions displayed by <character> in the format described below:

For each emotion displayed by <character>, determine the following items:
  - <emotion_ekman>: Classify the emotion conveyed. Answer set: "anger",
     "disgust", "fear", "happiness", "sadness", "surprise", "unsure", "no
     emotion".
  - <score_ekman>: Confidence score (float 0 1 ).
  - <emotion_miie>: Classify the emotion. Answer set: "anger", "calm",
     "confused", "disgusted", "fear", "happy", "sad", "surprised",
     "unsure", "no emotion".
  - <score_miie>: Confidence score (float 0 1 ).
  - <emotion_goemotions>: Classify the emotion. Answer set: "admiration",
     "approval", "annoyance", "gratitude", "disapproval", "amusement",
     "curiosity", "love", "optimism", "disappointment", "joy",
     "realization", "anger", "sadness", "confusion", "caring",
     "excitement", "surprise", "disgust", "desire", "fear", "remorse",
     "embarrassment", "nervousness", "pride", "relief", "grief",
     "neutral".
  - <score_goemotions>: Confidence score (float 0 1 ).

If a character appears but does not express any emotions, return:

{
  "1": {
```

```
    "emotion_ekman": "no emotion",
    "score_ekman": 1,
    "emotion_miie": "no emotion",
    "score_miie": 1,
    "emotion_goemotions": "neutral",
    "score_goemotions": 1
  }
}

For each character, organize results in JSON:

{
  "1": {
    "character": "character 1 in the text",
    "age": "...",
    ...,
    "emotion_JSON": {...}
  },
  "2": {
    "character": "character 2 in the text",
    "age": "...",
    ...,
    "emotion_JSON": {...}
  }
}

Important: do not misattribute emotions or assign characteristics not
   supported by the text.
```

Alternative Text Analysis Approach: Using a Lexicon-Based Approach The method employed to identify emotions representation in text is illustrated in Figure 1a

To create our lexicon, we primarily use the synsets of WordNet. A synset is a collection of cognitive synonyms that convey the same concept. Some words have a single synset, while others have multiple. For example, the word "anger" is part of the following synsets: 1) a feeling that is oriented toward some real or supposed grievance; 2) the state of being angry; and 3) belligerence aroused by a real or supposed wrong (personified as one of the deadly sins). Hence, with an initial emotion-related search term, we can access a comprehensive network of semantic relations.

We construct an initial set of WordNet inputs (see appendix), extract their synonyms, and manually review the results. Following this, we engineer a prompt to elicit more emotional synonyms from an LLM. Our final lexicon maps these synonyms to our 8 categorical emotions (confusion, anger, sadness, surprise, calm, happiness, fear, and disgust). In the process, we lemmatize all words using SpaCy and, through manual inspection, ensure that all categories are mutually exclusive and each keyword unique. Lemmatization is chosen in order to simplify the lexicon construction at the cost of less flexibility over word choice. Table B.1 shows a sample of the dictionary[13].

In total, our lexicon contains 1,706 entries. More specifically, breaking this down into

[13]The full lexicon can be viewed here

the number of synonyms for each emotion, we have 253 anger, 256 calm, 191 confusion, 224 disgust, 173 fear, 204 happiness, 246 sadness, and 159 surprise entries.

The sample of textbooks and children's literature is then processed by splitting it into sentences using SpaCy and then cleaned by removing certain punctuation and digits, by making it lowercase and by lemmatizing it. This gives us 3,361,503 sentences across 1,306 books. Subsequently, we apply a count vectorizer to each sentence, allowing us to identify if a sentence contains one or more keywords and the number of times it appears. We then sum these keywords to attain sentence-level counts for each emotion category. Afterwards, from the sentence level, we further aggregate the counts into our desired strata and in each case create a percentage distribution for analysis.

Alternative Approach: Using an Emotions AI-Based Approach This approach uses a RoBERTa-based model trained on the GoEmotions dataset created by Demszky et al. (2020). It is a multi-label classification model that takes a document as an input and returns 28 'probability' float outputs, one for each of the 28 emotion labels.

We use the same sample of 3,361,503 cleaned sentences as in the lexicon approach, but in this case do not lemmatize. We then feed these sentences into the model, receive the output, and assign each sentence the emotion with the highest predicted probability. Sentences labeled as "neutral" are dropped, as we are only concerned with text that has emotional content. This parallels our lexicon method, where our denominator consists only of emotion-related words. Next, we map each of the 27 emotions to 7 aggregate emotions guided by a correspondence released by Demszky et al. (2020). Ideally, we would like to have aligned the output to our MiiE Lab 8, but the model outputs did not allow us to easily identify "calm."

In comparison to the lexicon-based approach, this method takes into account the entire context of sentence when making a prediction. The base model was pretrained on a large volume of texts and learned the contextual representation of words based on their surroundings. Hence, it is able to reduce semantic ambiguity otherwise introduced by analyzing each word individually.

On the other hand, because these representations are learned by neural networks with many parameters, we lose insight into which components of a sentence may be contributing most to the classification of an sentence. Our lexicon-based method therefore allows us to approach the sentence-emotion classification problem in a more transparent way, providing insight into the words with which emotions are being conveyed.