Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 26-1584

(How) Do We Teach Emotions?

Anjali Adukia
University of Chicago

Matthew Bonci
University of Chicago

Paula Dastres Gallardo
University of Wisconsin-Madison

Emileigh Harrison
University of Texas at Austin

Jake Nicoll
University of Chicago

Teodora Szasz
University of Chicago

Emotional intelligence constitutes a key component of human capital, shaped partially by educational materials. Using machine learning and generative AI tools, we examine emotional content in public-school textbooks and children's literature. A stark mismatch emerges: text exposes children to a broad emotional range, but images overwhelmingly depict happiness and calm, regardless of emotions described in text on the same page. Nearly half of pages show zero overlap between the emotions described in text and those shown in images. This pattern persists across time, contexts, and identities. Household purchases and library inventories suggest content may be endogenously shaped by consumer demand favoring "positive" cover imagery, implying market forces narrow the emotional landscape children encounter.

VERSION: September 2026

Suggested citation: Adukia, Anjali, Matthew Bonci, Paula Dastres Gallardo, Emileigh Harrison, Jake Nicoll, and Teodora Szasz. (2026). (How) Do We Teach Emotions?. (EdWorkingPaper: 26-1584). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/hgn9-n385

(How) Do We Teach Emotions?*

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

September 2026

Abstract

Emotional intelligence constitutes a key component of human capital, shaped partially by educational materials. Using machine learning and generative AI tools, we examine emotional content in public-school textbooks and children's literature. A stark mismatch emerges: text exposes children to a broad emotional range, but images overwhelmingly depict happiness and calm, regardless of emotions described in text on the same page. Nearly half of pages show zero overlap between the emotions described in text and those shown in images. This pattern persists across time, contexts, and identities. Household purchases and library inventories suggest content may be endogenously shaped by consumer demand favoring "positive" cover imagery, implying market forces narrow the emotional landscape children encounter.

Keywords: culture, emotions, content analysis, education policy, curriculum, artificial intelligence tools, computational social science, natural language processing, large language models, computer vision

JEL Category: I20, I21, I24, L82, Z11, Z13

*Affiliation: University of Chicago, University of Wisconsin-Madison, University of Texas at Austin, NBER, IZA. Contact: adukia@uchicago.edu, mbonci@uchicago.edu, dastresgalla@wisc.edu, emileigh.harrison@austin.utexas.edu, jwnicoll@uchicago.edu, dorha.szasz@gmail.com. For helpful feedback, we thank Alex Eble, James Evans, Yuan Chang Leong, Lesley Turner, Isabeau Dasho, David Yanagizawa-Drott, and seminar participants at Notre Dame, Zurich AI + Economics, Hong Kong AI Research in the Social Sciences and Humanities, Duke, Association for Education Finance and Policy, the Society for Research on Educational Effectiveness, Stanford, The University of Chicago, and Yale. 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.

The centrality and universality of emotion to human experience has long been recognized across cultures, traditions, and disciplines (Izard, 1977; Lutz and White, 1986; Elster, 1998; Marcus, 2000; Gross, 2015; Denham, 2019). Beyond their psychological importance, emotional skills constitute a key component of human capital. Individuals with greater emotional intelligence exhibit stronger stress management, prosocial behavior, academic performance, and workplace functioning (Mayer and Geher, 1996; Parker et al., 2004; Austin et al., 2010; Brackett et al., 2012; Angelucci et al., 2026). This may be especially important in modern labor markets; as work has become more team-based and service-oriented, employers increasingly seek workers who can communicate effectively, respond to conflict, and manage interpersonal relations (Hochschild, 1983; Deming, 2017; Bonhomme, 2021; Hansen et al., 2021; Hoffman and Tadelis, 2021).

Despite growing evidence on the returns to social and emotional skills, less is known about the childhood inputs that develop these skills and the norms that reinforce their importance. Models of human-capital formation emphasize the importance of early investments, motivating attention to children's learning environments. Within these environments, textbooks and other children's media serve as one such potential input, providing models of emotional experience and expression through their text and images. As the primary instructional resources in classrooms, they not only transmit academic content, but they also convey to students – through both explicit content and implicit messaging – an expectation about which emotions are appropriate to display. We conceptualize these emotional norms as an input into emotional skill formation and therefore explore an upstream question: where do emotional norms come from and, specifically, what are children taught to recognize or display?

In this study, we examine emotions children encounter in educational materials across formal and informal settings. We also analyze the emotions presented across contexts and identities overall and over time using two sets of influential children's media: (i) three decades of public-school textbooks from California and Texas and (ii) a century of award-winning children's books commonly found in homes, libraries, and schools (Smith, 2013; Cockcroft, 2018; Adukia et al., 2023). These materials constitute two prominent sources of content to which children are regularly exposed during critical developmental periods. We then incorporate data on both individual and library purchasing behavior to understand the potential role of market forces in shaping the presence of particular emotional portrayals.

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). Specifically, we employ large multimodal models (LMMs) to classify eight primary emotions across both text and images: anger, calm, confusion, disgust, fear, happiness, sadness, and surprise.[1] We examine heterogeneity by context (e.g. corpora, state, grade, subject) and character identities (e.g. putative

[1] This set of emotions was derived from the six basic emotions identified by Ekman (1992) with two added emotions: calm and confusion, particularly relevant to children's media contexts. 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.

race, gender, age, skin color, and whether the character is human or non-human).

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 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, these books suggest that regardless of the emotions characters experience in the text, the appropriate outward response is to display a calm or happy appearance.

We check sensitivity of our results to the LMM approaches using alternative, non-generative approaches: a hand-curated emotion lexicon and a RoBERTa transformer model fine-tuned on the GoEmotions dataset for text, and a CNN-based face-detection and feature-classification pipeline for images. These approaches recover the same qualitative finding of a broad emotional range in text but a strong concentration of happiness and calm in images.

Since our primary emotion measures are model-generated, we further assess their reliability using human annotations. Following an approach inspired by Calderon et al. (2025), we compare model-human agreement with agreement among human annotators. Although performance varies across tasks and annotators, the model is never detectably worse than all human annotators. Using the validation sample to estimate the distribution of labels that humans would assign to the full corpus yields the same central result.

We next explore the role of consumer demand in the visual overrepresentation of calm and happiness in children's books. 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, 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 may be endogenously shaped by purchaser demand and reflects the societal value placed on displays of "positive" emotions. This may be because these emotions capture attention and make books look more appealing (Bloch, 1995; Underwood and Klein, 2002; Pieters and Wedel, 2004). Additionally, if purchasers 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 et al., 2024).

Our findings contribute to multiple strands of literature. First, they deepen our understanding of the production function of socioemotional skills by identifying one channel through which emotional norms may be produced through childhood exposure and transmitted across generations. In doing so, we inform work on non-cognitive skill development (Groves, 2005; Heckman et al., 2006; Duckworth et al., 2007; Kautz et al., 2014; Lundberg, 2017), highlighting how curricular materials can shape early emotional intelligence. This extends the literature beyond Elster's (1998) idea that emotions are inputs into economic behavior; emotions are not only inputs but are themselves shaped by institutions, markets, and cultural goods.

Second, we add to the literature on cultural economics, which has recently rapidly expanded through the use of novel data sources to study individualism through style (Yanagizawa-Drott and Voth, 2023) and naming conventions (Bazzi et al., 2020) as well as cultural norms through folklore (Michalopoulos and Xue, 2021), movies (Michalopoulos and Rauh, 2024), and children's books (Adukia et al., 2023). We extend this agenda by studying emotional norms as a core object: in children's media and curricular materials, a wide range of emotional displays are present in the text, but images primarily depict calm and happiness. This imbalance in the visual display of positively valenced emotions, regardless of the emotions described in the text, presents a discordance between the emotions that are felt and those that are shown, suggesting there are "emotional display rules" embedded in culture.

Third, we expand prior research on emotional representation in children's media (Nikolajeva, 2014; Ku, 2019; Kardum et al., 2021), which has largely relied on manual content analysis and focused narrowly on text. We extend this work to systematically examine emotional displays in images which may have greater influence than textual presentations of emotions on children's emotion identification and response capabilities (Arizpe and Styles, 2015; Nikolajeva, 2017). By examining elementary-school textbooks across multiple subjects alongside influential children's literature, our work shines a light on the socioemotional curriculum embedded in both formal and informal educational settings.

Fourth, we contribute to the economics of media by providing further evidence that producers tailor content in response to consumer demand (Gentzkow and Shapiro, 2010). We rationalize our findings as an equilibrium outcome in the market for children's media. Moreover, we illustrate how books can effectively advertise themselves to purchasers with limited attention to a book's content (Dröge and Darmon, 1987; Eliaz and Spiegler, 2011; Bordalo et al., 2013), in this case, through the emphasis of certain emotions on covers, which are salient, easy-to-process signals. Purchasers respond to this imagery, and this interaction characterizes an equilibrium with imbalanced visual emotional displays in children's books.

Finally, we contribute to computational social science by developing a multimodal approach to measuring emotional representation in children's media, and in particular documenting misalignment in emotional representation across modalities. We show that this misalignment is robust to a wide range of measurement choices, from simple token-based measures of text and CNN-based measures of images to a large multimodal model applied to both. We also offer a validation procedure adapted from Calderon et al. (2025), which demonstrates that these findings are not artifacts of poor model predictions but rather reflect a stylized fact about the underlying source material. Building on work that uses natural language processing and computer vision to study social representation, stereotypes, and cultural content (Ash et al., 2021; Adukia et al., 2023; Ludwig and Mullainathan, 2024; Adukia and Harrison, 2025), we examine the 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. Our work introduces new measures of emotional displays in text and images which can help to catalyze other research related to emotions and further expand our understanding of the relationship between emotions and societal outcomes.

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 modalities on the same page simultaneously. In Section V.A, we demonstrate robustness to alternative approaches to measurement, while Section V.B discusses our model validation strategy, which we hope offers guidance for others using LMMs for measurement in economics and in other disciplines. In Section VI, we explore how consumer demand may contribute to the patterns we uncover. Finally, we offer a discussion in Section VII.

I Background

Emotion constitutes a fundamental dimension of human experience, making it a natural subject of sustained inquiry across a wide range of disciplines. In psychology, foundational work has examined the nature, classification, and development of emotions (Izard, 1977; Plutchik and Kellerman, 1980; Averill, 1983), while more recent research has emphasized emotion regulation, emotional competence, preferences over affective states, the influence of emotions on economic and social behavior, and the role of early socialization in shaping how children learn to recognize, manage, and express emotions (Eisenberg et al., 1998; Gross and John, 2003; Tsai et al., 2006; Tsai, 2007; Côté et al., 2010; Genevsky and Knutson, 2015; Gross, 2015; Denham, 2019; Park et al., 2020; Perry et al., 2020). Sociology has highlighted how social roles, status relations, institutions, and labor markets shape emotional expression, including norms regarding which emotions are appropriate to display in different contexts (Kemper, 1978; Hochschild, 1983; Thoits, 1989). Anthropological research has emphasized that emotional experiences and displays reflect culturally shared meanings and expectations, including norms governing appropriate emotional responses across social settings (Lutz and White, 1986). Political science has examined the role of emotions in political judgment, participation, and collective behavior (Marcus, 2000). Philosophy has long debated the nature, moral significance, and rationality of emotions (Solomon, 1976; Rorty et al., 1980), while economics has examined how emotions influence preferences, decision-making, social interactions, and political attitudes (Elster, 1998; Loewenstein, 2000; Algan et al., 2026). Taken together, this work suggests that the capacity to recognize, express, and regulate emotions can be understood as an important dimension of human capital.

A large literature shows that non-cognitive skills shape educational attainment, employment, and earnings (Heckman and Rubinstein, 2001; Groves, 2005; Heckman et al., 2006). Emotional intelligence – the capacity to recognize, understand, and regulate emotions in oneself and others – is one such skill. For example, manipulating emotional displays to exhibit happiness and calmness is one feature of what customer service workers regularly provide (Hochschild, 1983). Additionally, regulating one's own emotions and recognizing emotions in others can reduce frictions in teams and contribute to managerial effectiveness, which in turn can improve earnings (Hansen et al., 2021; Hoffman and Tadelis, 2021). Importantly, the return to social-emotional skills has risen even as the return to some cognitive skills has stagnated (Castex and Kogan Dechter, 2014; Edin et al., 2022).[2]

Schools are a central institution through which children can acquire and practice emotional skills. In particular, these institutions foster emotional intelligence through direct instruction (Brackett et al., 2012), classroom interactions, and teacher expectations. Education materials and children's books are another important site of emotional display learning (Green and Sun, 2025). According to the constructionist hypothesis in developmental psychology, emotions are abstract concepts learned by children through an accumulation of encounters with emotions (Hoemann et al., 2019). Indeed, children learn to recognize emotions between the ages of two and eight, although the skill remains malleable after the age of eight (Widen and Russell, 2008; Ruffman et al., 2023). From this perspective, the images in children's books can provide a model for the accompanying emotion language that can help children calibrate both their emotion recognition skills and their understanding of how to express emotions during the age range that these skills are developing. This calibration could improve emotion recognition skills, which has been shown to have important social consequences: children who struggle to identify emotions such as sadness tend to exhibit lower prosocial behavior (Doherty, 2008; Dickerson and Quas, 2021; Brackett et al., 2012). Consistent with this hypothesis, research suggests that exposure to emotionally diverse content 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). In this way, children's materials may shape emotional skills not only by teaching children to recognize and regulate emotions, but also by transmitting norms about which emotions are appropriate to display.

Because emotional norms are relevant for later educational and labor market success, the way educational materials represent emotions is potentially important. Ex ante, however, it is not obvious which emotions these materials will emphasize. On the one hand, educational materials may aim to build emotional intelligence by exposing children to the complexity of real-life emotional

[2] Castex and Kogan Dechter (2014) compares the NLSY79 and NLSY97 cohorts and finds that measured cognitive skills (AFQT, mathematics, and verbal scores) had 30–50 percent larger effects on wages in the 1980s than in the 2000s. They suggest that the decline in returns to ability can be attributed to differences in the growth rate of technology between the 1980s and 2000s.

experience. Consistent with the constructionist hypothesis, this could involve stories and images that portray a broad range of emotions and model how characters recognize, express, and regulate them in healthy ways (Hoemann et al., 2019). On the other hand, educational materials may also reflect prevailing social and labor market norms about which emotions are appropriate to display. Many social situations and occupations discourage the expression of so-called "negative" emotions and instead reward calmness, compliance, or a façade of positivity (Hochschild, 1983; Deming, 2017). To the extent that schools socialize children into future social and labor market roles, these norms may be reflected in materials that downplay or omit emotions such as sadness, anger, or fear (Bowles and Gintis, 1976; Entorf and Dohmen, 2025).

This ambiguity creates an empirical question: do children's educational materials expose children to a broad emotional repertoire, or do they reproduce a narrower set of socially and economically rewarded emotional displays? Moreover, because children encounter these materials through both text and images, it is important to examine not only which emotions are described, but also which emotions are visually modeled. If images do not align with the emotions described in the accompanying text, materials may provide weaker or inconsistent cues for developing emotion recognition skills (Brady et al., 2024). Despite their potential importance in the formation of emotional skills, there is limited systematic evidence on how emotions are represented across both modalities.

We address this gap by examining the emotional content of both textbooks and children's literature, comparing patterns across formal and informal educational materials. We also incorporate data on purchaser and public library purchasing behavior to assess whether demand for particular emotional portrayals influences book acquisition decisions and, in turn, the emotional representations children are likely to encounter.

II Data Sources

We analyze two categories of influential children's media representing both formal and informal educational contexts: public-school textbooks and award-winning children's books.

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. These states are meaningful for several reasons. First, California and Texas are the two largest public-school systems in the United States, enrolling approximately 5.92 million and 5.53 million students, respectively, in fall 2023. Second, California and Texas both use state-level review processes in which appointed reviewers evaluate publisher materials for standards alignment, whereas most states leave textbook selection primarily to local districts (Doan and Kaufman, 2024). Third, the adoption decisions of these states influence adoption decisions of other states (House Research Organization, 2002; Doan and Kaufman, 2024). Finally, these states occupy opposing positions on the political spectrum with Texas advisory panels being more likely to reflect Republican perspectives and California panels being more likely to reflect Democratic perspectives. Differences across states could reflect the influence of ideology on the representation of emotions and identities in educational materials.

For children's literature, we focus on books likely to be particularly influential. Specifically, we include titles that have won awards featured by the American Library Association (ALA). These awards honor books for their "exceptional literary" or "artistic" value, or for centering 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.[3] The corpus encompasses both fiction and nonfiction works, representing realistic and imaginary settings, thereby capturing a spectrum of narratives available to young readers.

To study household purchasing behavior, we supplement the children's literature corpus with data from the Numerator OmniPanel, a large panel data set with information from over one billion shopping trips from over 44,000 retailers. The data cover purchases made between 2017 and 2020 through both online and brick-and-mortar retailers. We limit our analyses to purchases of titles in our award-winning children's literature corpus.

We complement the household purchase data with branch-level inventory records for 26 library branches from the Seattle Public Library system, obtained through Seattle's public data portal. The data report the number of copies of each title held by each library branch as of October 1, 2017. We match these records to the award-winning children's literature corpus and construct a book-by-branch sample in which the outcome is the number of copies of a title held in a branch's inventory. These data provide a useful complementary measure because a public library system is likely to consider a broader range of award-winning books than an individual purchaser. At the same time, library holdings reflect collection-development objectives, budgets, and other institutional constraints as well as anticipated or realized reader demand. We therefore interpret branch inventories as a measure of institutional acquisition that may indirectly reflect readers' preferences rather than as a direct measure of individual choice.

III Methods

Previous research has established robust computational methods for emotion detection in textual data (Bollen et al., 2011; Thelwall et al., 2011; Hasan et al., 2014; Demszky et al., 2020). However, emotions are communicated not only through language, but also through non-verbal cues such as gestures and body language. Facial expressions in particular have been extensively documented as reliable indicators of emotional states (Sullivan and Masters, 1988; Lucas et al., 2003;

[3] The Mainstream collection includes books that have received either the Newbery or Caldecott awards – the two longest-running children's book awards in the United States. We use the label "Mainstream" to reflect the substantial role these awards play in shaping which books children read (Smith, 2013; Adukia et al., 2023). The Diversity collection draws from awards highlighted by the ALA that foreground the voices and experiences of historically marginalized or underrepresented groups. Specifically, this collection encompasses titles honored by the following awards: American Indian Youth Literature, Américas, Arab American, Asian/Pacific American Award for Literature, Carter G. Woodson, Coretta Scott King, Dolly Gray, Ezra Jack Keats, Middle East, Notable Books for a Global Society, Pura Belpré, Rise Feminist (formerly the Amelia Bloomer Award), Schneider Family, Skipping Stones Honor, South Asia, Stonewall, and Tomás Rivera Mexican American.

Öhman, 2007, 2009), so only measuring emotional representation in text may miss a central channel through which children encounter and learn about emotions. We therefore develop a multimodal framework that systematically quantifies emotional content in both the textual and visual components of children's media.

We focus our analysis on eight emotion categories: anger, calm, confusion, disgust, fear, happiness, sadness, and surprise. These classifications incorporate the six basic emotions originally delineated by Ekman (1992) while extending the framework to include calm and confusion.[4] To accommodate ambiguous cases, we include 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 taxonomies accounting for displays of emotions across different contexts (Cowen and Keltner, 2017), including in faces (Cowen and Keltner, 2020).

III.A Text Analysis

We developed three complementary approaches to quantify emotions in the text of children's printed content. Our primary measure uses a large language model (LLM) for character-level emotion detection. We supplement this with a lexicon-based frequency analysis and a fine-tuned transformer model. This multi-method strategy allows us to test the sensitivity of our findings to different measurement strategies. Figure 1a provides an overview of the data extraction process applied to the text on book pages.

Large Language Model Classification. We use a large language model (LLM) (specifically GPT-4o-mini) to construct page-level measures of character emotions in text, motivated by recent evidence that LLMs can support scalable social-science annotation tasks, including emotion classification (Ziems et al., 2024).[5] For each page, the model identifies characters, assigns each character one of eight emotion labels – anger, calm, confusion, disgust, fear, happiness, sadness, or surprise – and records "unsure" or "no emotion" when appropriate. We also extract character attributes, including age, gender, race, and whether the character is human, allowing us to analyze heterogeneity in emotional representation across demographic groups. More details are found in Appendix C.A.

[4] Indeed, what constitutes an emotion and whether calmness is an emotion or a state of no emotion has been a subject of debate for quite some time, particularly starting when James (1884) first suggested that emotions to be considered are those with a "distinct bodily expression" followed by contention by Gurney (1884), who noted that even when emotion's "*mass*, and its character *quâ painful*, remain unaffected under these latter conditions, it is just as truly emotion as before, and emotion to which the physical signs...do not now even contribute." This debate has extended throughout time and disciplines (c.f. discussion by philosophers such as Zemach 2001 and neuroscientists such as Adolphs et al. 2019.)

[5] These models also have multimodal capacities and can be considered to be large multimodal models (LMMs). For clarity in the text, we primarily refer to these models as LLMs when we are referring to the text-based analysis and LMMs when we are referring to the image-based analysis.

Conceptually, GPT-4o-mini processes the raw text as input and generates character-emotion pairings by: (1) recognizing linguistic markers of emotional expression (e.g., descriptive verbs, dialogue, narrative commentary); (2) matching these markers to semantic patterns learned during model pre-training; and (3) mapping identified emotional content to our taxonomy. The model essentially performs linguistic pattern recognition at the character level, converting narrative descriptions of emotional states into discrete categorical labels. This differs from traditional lexicon-based approaches (which count emotion-related keywords and which we employ in subsequent analysis) by leveraging contextual understanding; the model interprets emotional expressions in relation to surrounding narrative, character relationships, and plot development.

Lexicon-Based Analysis. We also implement a lexicon-based approach which involves counting the frequency of predetermined emotion-related words, offering a straightforward and interpretable baseline. To do so, we construct comprehensive keyword dictionaries.[6] 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). This frequency-based approach provides a transparent, replicable baseline measurement independent of neural network architectures.

Transformer Model Prediction. As a second text-based robustness check, we apply a RoBERTa-base transformer model fine-tuned on the GoEmotions dataset created by Demszky et al. (2020).[7] The model returns a probability score for each of the 28 GoEmotions labels, which include 27 emotion labels and a neutral category. In our implementation, we assign each sentence to the label with the highest predicted probability and drop sentences classified as "neutral," since our goal is to characterize the distribution of emotional content.[8] We then map the remaining fine-grained GoEmotions labels to the corresponding aggregate emotion categories used in our analysis.[9] This supervised transformer-based approach provides an additional benchmark for our primary LLM-based text classification method.

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 images in children's books. We use an LMM and, separately, a two-step face detec-

[6] We construct these dictionaries using three sources: WordNet semantic expansion, ChatGPT-assisted generation, and manual curation; more details are found in Appendix C.B. Examples of keywords related to each emotion can be found in Appendix Table B.1. The full lexicon can be viewed at https://github.com/miielab/emotions_paper.

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

[8] Dropping neutral sentences parallels our lexicon-based approach, in which we only examine sentences containing emotion-related words.

[9] Because GoEmotions does not include a "calm" label, this approach maps the model outputs to the remaining aggregate emotion categories. We view this limitation as useful for assessing whether our conclusions are sensitive to the set of emotions available to the text classifier. Our mapping is adapted from the correspondence categories provided at https://github.com/google-research/google-research/blob/master/goemotions/data/ekman_mapping.json.

tion model and feature classification model. Figure 1b provides an overview of the data extraction process applied to the images on book pages. We note that these measures are capturing the outward emotional expression of pictured characters and may not reflect their true feelings, which may be more likely to be reflected in the text.

Large Multimodal Model Classification. We use a large multimodal model (LMM) (specifically Gemini 2.5 Pro Preview) to construct character-level measures of emotions and attributes from book images. We process scanned pages in two stages: first, we classify each page as having only text or containing an image; only pages with images proceed to the second stage. For each illustrated page, the model provides information for the illustration as a whole, such as a description of the setting as well as information on each depicted character including bounding-box coordinates, race, gender, age, skin color, human/non-human classification, and emotion. Mirroring our text analysis, our primary emotion taxonomy includes anger, calm, confusion, disgust, fear, happiness, sadness, and surprise, with "unsure" and "no emotion" allowed for ambiguous or neutral cases.[10]

Conceptually, Gemini 2.5 Pro Preview analyzes images by: (1) identifying human and non-human faces and figures; (2) extracting visual features associated with emotional expression (e.g. facial muscle configurations, body posture, spatial positioning); (3) mapping these visual patterns to learned semantic categories; and (4) simultaneously inferring demographic characteristics based on visual cues (e.g. phenotypical features). The model leverages pre-training on large datasets that include both text and images, allowing it to recognize associations between visual features and emotional states.[11] It is important to note that this process is fundamentally visual pattern recognition and should not be construed as interpretation. The model identifies statistical regularities between pixel configurations and emotion labels learned during pre-training; it does not "understand" emotions as subjective experiences.

Convolutional Neural Network-Based Face Detection and Feature Classification. As a second approach to measuring emotions and character attributes in images, we use a computer-vision pipeline that builds on the face-detection and feature-classification models developed in Szasz et al. (2022). We first apply FDAI, a face-detection model trained on illustrated faces from children's books, to identify faces in book images. We then use an extended version of our feature-classification model, originally developed for Szasz et al. (2022), to classify race, gender, age, and emotion for each detected face. We use these predicted labels as an alternative to our LMM-based classifications to ensure that our results are robust to different measurement choices.

IV Results

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

[10] More information is provided in Appendix C.B.

[11] For example, an upturned mouth with engaged eye muscles (as referred to by Ekman (1989) as Duchenne markers of smiling based on Duchenne de Boulogne (1990)) would be classified as "happiness," while a downturned mouth or furrowed brow would map to negative valences.

IV.A Emotions 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 approximately 4-12%), 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 characters' depicted emotions (Figure 2b).[12] Notably, negatively valenced emotions, which are meaningfully present in the text, are largely absent from images. Figure 3 shows that these findings are stable over time.

This captures our central finding: the emotional narratives conveyed through the text sharply contrast the limited spectrum of emotions depicted in the images. Importantly, as outlined in Section I, this could complicate the process of learning to decode emotions in faces for young readers and embed norms surrounding which emotional responses are socially appropriate. In the next two subsections that follow, we demonstrate that this finding persists across various subsets of the data and regardless of the predicted identity of the character depicted. We present a set of secondary findings exploring how emotional portrayals vary across contexts and identities. These patterns are less pronounced than the overall text-image discrepancy and should be interpreted as hypothesis-generating rather than definitive. In Section IV.D, we return to the aggregate text-image divergence and show that it does not arise from selective illustration of positive emotions; rather, our findings reflect a genuine incongruence of character emotions in text and images.

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: images remain concentrated in calm and happiness, whereas a wider variety of emotions is displayed in text. We discuss these findings further here.

We first examine the patterns across different book types. In Figure 4 we see that the distributions of emotions in the text of children's literature and textbooks are broadly similar, with textbooks showing slightly more representation of happiness compared to children's literature. Children's literature displays a somewhat more diverse range of emotions compared to textbooks, though both collections cover a variety of emotions in text.

[12] Note that these proportions reflect the proportion of all characters that have an emotion. We drop characters that are classified as having "No Emotion" or an "Unsure" Emotion. The proportion of pages that fit these categories are approximately 30% for characters in the text and 27% for characters in the images. Distributions including these classifications can be found in Figure A.1. Dropping these characters does not change our results in any substantive way and improves the interpretability of our analyses.

Within the subcollections of these two sources of children's text, the patterns remain qualitatively similar. In particular, the Mainstream and Diversity subcollections of children's literature show comparable proportions of each emotion. The primary difference is that the Diversity subcollection exhibits a slightly greater proportion of sadness (approximately 5 percentage points higher). With respect to textbooks, we observe similar patterns across those from California and Texas.

While the overall patterns hold across grade levels, we observe noteworthy variation. For example, happiness appears ten percentage points more frequently in third-grade textbooks than fifth-grade textbooks, while emotions such as anger, fear, and sadness appear slightly more frequently in fifth-grade textbooks. This may reflect curricular design priorities with earlier grades emphasizing reassurance and "positive" affect and upper elementary instruction introducing more complex narratives that require a broader emotional range.

Subject matter appears to slightly shape emotional representation. Science textbooks contain fewer depictions of anger and sadness compared to Reading and Social Studies textbooks. This likely reflects disciplinary differences in pedagogical approach for each subject: Science instruction emphasizes concept mastery through factual exposition and may not require as multifarious an emotional narrative, whereas Reading and Social Studies inherently rely on narrative and human experience, which naturally incorporate emotional complexity. Even acknowledging small differences across subjects, we still observe that the representation of emotions in text remains broadly consistent across all subcollections.

Image heterogeneity analysis corroborates the overall image analysis: calm and happiness dominate emotional depictions across all subcollections (Figure 5). There are a few suggestive patterns that emerge. First, Science and Social Studies textbooks show higher concentrations of calm compared to children's literature. Reading textbooks display more moderate patterns that more closely resemble children's literature. Second, grade-level differences in emotional composition emerge (grade 3 favors happiness; grade 5 favors calm). We note that these differences are compositional rather than substantive in valence: the total proportion of positive emotions remains stable across grades.

IV.C Emotions Displayed by Different Identities

Having established the aggregate text-image mismatch, we now examine whether emotional representation varies systematically by character identity. This analysis serves two purposes: it tests the sensitivity of our main findings across demographic subgroups, and it reveals whether educational materials encode different emotional norms for characters of different identities (e.g. genders, ages, races, and as well as whether the character is human). As described in Section III, all identity attributes are model-generated predictions derived from textual description for the text analysis and from visual features for the image analysis. These predictions should be read as the putative identity a reader would plausibly attribute to a character, not as ground truth about any character's identity.

Figure 6 presents the heterogeneity analysis in text. First, gendered differences emerge: female characters display sadness somewhat more frequently than male characters, while male characters more often display anger. This asymmetry reflects longstanding cultural associations linking women with vulnerability and men with assertiveness or aggression (Williams et al., 1977; Eagly and Steffen, 1984). Second, age-based differences appear: adult characters more frequently express anger and calm, whereas child characters more often express fear and confusion. This pattern may reflect children in narratives being more likely to be positioned as encountering unfamiliar or threatening situations, while adults are portrayed as composed authority figures or, alternatively, as sources of conflict. Importantly, these differences are modest in magnitude. Across all identity groups, textual content preserves substantial emotional diversity: every subgroup displays a meaningful range of emotions rather than a single dominant category.

Figure 7 presents the analogous heterogeneity analysis for images. The near-universal dominance of calm and happiness persists across every identity group, i.e. no subgroup displays negative emotions with meaningful frequency. This pattern applies uniformly to characters of all genders, ages, races, skin color, and human/non-human status. Within this constrained emotional space, however, subtle identity-based patterns emerge in how the two dominant emotions are allocated. Male faces are more likely than female faces to be depicted as calm, while female faces are correspondingly more likely to be depicted as happy: a difference of similar magnitude. This trade-off is consistent with a gendered emotional stereotype in which composure or emotional restraint is coded as masculine, while overt positive expressiveness (smiling) is coded as feminine. A parallel structure holds across other identity groups: subgroups depicted more frequently as calm are depicted less frequently as happy, and vice versa. This near-perfect substitution between calm and happiness underscores the central visual constraint: characters are almost always shown expressing one of these two positive emotions, and identity primarily shapes *which* positive emotion is displayed rather than whether the character is permitted to display a negative one.

IV.D Concurrent Text and Image Analysis

The preceding sections document a mismatch between the aggregate distributions of emotions in text and images. It is worth noting that images reflect the outward emotional expressions of pictured characters and may not necessarily correspond to their internal emotional states, which are more likely conveyed through the text. This distinction, far from being a limitation, is, in fact, central to our analysis: we are interested precisely in whether the emotions characters visibly display align with those that the text describes. That said, the discrepancies between emotions in text and images documented in the preceding sections need not imply a true misalignment between emotions in the two modalities; images may simply emphasize the positive moments of otherwise emotionally diverse narratives. We probe this by comparing emotions expressed through each modality on the same page and find systematic incongruence: images depict calm and happiness even when the accompanying text describes very different emotions.

Figure 8 presents the co-occurrence of emotions for text and images on the same page, normalized by the text columns. Each cell reports the share of a specific emotion label in images, given a particular emotion label in the text of the same page. The results reveal a striking asymmetry. Calm is the only emotion that consistently aligns across modalities: when calm appears in the text, 70% of image emotions on that page are also calm. Happiness follows calm, but it aligns more weakly: when the text describes happiness, only 27% of images on that page show happiness. For every other emotion, alignment collapses. When the text conveys anger, fear, sadness, or surprise, the shares of those same emotions in corresponding images are only 13%, 9%, 10%, and 8%, respectively.

Crucially, the misalignment is directional. When the text expresses an emotion other than calm, images on that page still depict calm between 58% and 67% of the time. In other words, negative textual emotions are systematically accompanied by positive visual expressions. Children reading these pages encounter a consistent implicit message: regardless of what characters say, do, or feel in the narrative, the appropriate outward display is calm or happy.

One concern is that page-level co-occurrence may conflate different characters. For example, the text might describe one character's anger while the image portrays a different, calm character. To probe this possibility, we recompute the co-occurrence matrix on a subsample of pages where we can identify 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 then shows the misalignment persists even for matched characters: the diagonal carries similar weight to Figure 8, meaning that a single character's textual and visual emotions frequently diverge. Regardless of the emotion a character displays in the text, that same character is most likely to be depicted as calm in the accompanying image.

We further quantify the overall text-image discordance by computing a page-level measure of the distance between the set of emotions in text and the set of emotions in images. In particular, for each page, we build an eight-dimensional vector separately for text and images, where each entry in these vectors corresponds to an emotion and equals one if and only if the corresponding emotion is present on the page. After normalizing, we compute the total variation distance (TVD) between the page-level text and image distributions, $\delta(p, q) = \frac{1}{2}\|p - q\|_1$. Figure A.5 displays the distribution of these total variation distances. In approximately 10% of cases, the TVD is zero, indicating perfect alignment between textual and visual emotions. However, for nearly half of the pages, the TVD equals one, indicating that there is no overlap between the emotions conveyed in text and those conveyed in images on the same page.

Together, these results establish that the text-image mismatch is not an artifact of aggregation or of illustrators selectively highlighting positive narrative moments. The discrepancy holds within pages and even within matched characters. Illustrations convey happiness and calm even when these emotions are entirely absent from the surrounding text. These patterns speak directly to the mechanisms discussed in Section I. Given the documented role of visual cues in children's emotion recognition, this systematic incongruence may narrow the range of emotional displays modeled to children, may confuse the learning signal that they receive where the emotional expression contradicts the narrative context, and may reinforce an implicit norm that negative emotions, however they are experienced, should be outwardly suppressed.

V Sensitivity Analyses and Validation

We check how robust our results are to the LMM approaches using alternative methodologies. To check sensitivity, we employ: (1) for text, a hand-curated lexicon-based approach counting emotion-related keywords; (2) for text, a fine-tuned transformer model (RoBERTa) trained on the GoEmotions dataset; and (3) for images, a two-stage CNN-based pipeline involving character detection and feature classification. The consistency of results across these methods substantiates our primary generative-AI-based findings and addresses concerns about model-specific artifacts. For validation, we additionally obtain ground truth labels from human annotators on a stratified sample and use these data to validate model prediction quality.

V.A Alternative Approaches to Measurement

Our primary results rely on large multimodal models. A natural concern is that our findings reflect idiosyncrasies of these particular models and decisions regarding measurement rather than genuine features of the underlying materials. To address this, we replicate our core analysis using three alternative classifiers: a lexicon-based word-count approach and a fine-tuned transformer model (RoBERTa/GoEmotions) for text, and a convolutional neural network (CNN) image classifier, described in Sections III.A and III.B. These methods span a wide range of complexity and transparency, from a fully interpretable keyword dictionary to less transparent neural network architectures, providing a ranging set of robustness exercises.

We note that these alternative approaches differ from our primary LMM approaches in their unit of analysis, so perfect agreement is neither expected nor desired. The CNN image classifier predicts emotions for each individual detected face, whereas the LMM analyzes full images and the characters within them. Similarly, the alternative text measures count emotion-related words (lexicon approach) or classify emotions at the sentence level (transformer method), while our primary text LLM produces character-level emotion labels. What matters is not whether the methods agree exactly, but whether they recover the same qualitative pattern: broad emotional diversity in text alongside visual concentration in calm and happiness. 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 A.6a reports results from the lexicon-based approach. The approach recovers levels of anger, disgust, fear, and sadness comparable to our primary model. The main divergence concerns the calm-happiness split: the LLM identifies roughly 10 percentage points more happiness and 10 percentage points less calm than the lexicon-based approach. This may be an artifact of the lexicon-based approach's sensitivity to keyword choice and the semantic proximity of the two emotions. Indeed, calm and happiness may be lexically difficult to differentiate. For example, consider the word "contentment." It appears in our "calm" dictionary, yet depending on the context of the surrounding text, it could signal either calm or happiness. We thus expect there to be some fluidity between these two categories with the lexicon model. Critically, the aggregate distinction that drives our results – emotional breadth in text – remains intact.

Figure A.6b shows the results of the transformer-based model. The distribution closely tracks our primary text results (Figure 2a) with one exception: the GoEmotions model assigns 19 percentage points more of its classifications to surprise. This may reflect the difference in the unit of analysis. Operating at the sentence level, the GoEmotions model may, for example, classify sentences associated with discovery and learning as expressing surprise even when no specific character experiences that emotion. This expression would not be counted by the LLM because there is no specific character who is exhibiting surprise. Nonetheless, the results from the transformer method are arguably more comparable to those of the LLM than the lexicon-based results since the transformer model and the LLM both incorporate surrounding context to classify emotions rather than relying on isolated keywords.

Figure A.6c shows the results from our alternative image classifier. Consistent with the text robustness checks, the CNN predicts somewhat less calm and more happiness than the LMM for images, but the sum of these two emotions is stable across models. This again reflects the difficulty of distinguishing calm from happiness, now in the visual domain, and reinforces our central image finding: whichever model is used, the two positive emotions jointly dominate visual representation, crowding out the full emotional range. 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 classifiers ranging widely in complexity and interpretability, we recover the same qualitative conclusions we obtain from generative AI: text exposes children to a broad emotional repertoire, while images concentrate overwhelmingly on happiness and calm. The specific magnitudes shift with the model and unit of analysis, and calm and happiness prove consistently difficult to separate, both textually and visually, but this ambiguity does not affect our central finding, which concerns the contrast between positive and negative emotions rather than the split within positive emotions. Moreover, because our robustness methods permit alternative units of analysis (word-, sentence-, and face-level), we can confirm that the text-image mismatch is not an artifact of our character-level framing. The convergence of results across independent methods provides meaningful validation of measurements derived from our large generative models.

V.B Manual Validation

In this section, we describe the validation procedures for our primary LMM-based emotion measures. Traditional metrics such as F1 scores or correlations between labels produced by models and humans exhibit several shortcomings. First, these measures do not admit an interpretable threshold above which the models are deemed acceptable. Second, some tasks will give rise to more disagreement between human annotators, and this uncertainty in the ground truth should be incorporated in an assessment of model performance. Third, even in settings where multiple enumerators label the same example, traditional comparisons of the correlations between model and human responses versus those between human and human responses are limited in that comparable correlations may be neither necessary nor sufficient for assessing a model's effectiveness for producing measurements used in downstream applications. Moreover, to the extent that human disagreement is the result of differing but legitimate interpretations of a given example, it may be admissible for a model to agree with *some* humans, even if it does not agree with all or even the modal annotator.

To address these concerns, we proceed in the spirit of Calderon et al. (2025) by examining how our models compare with human labelers in terms of their ability to agree with other human labelers. Our procedure directly tests the comparability of the model with each human annotator, allowing us to determine whether the model is statistically interchangeable with that annotator. We obtain a "pass rate" (value between zero and one), indicating the fraction of human annotators with whom the model is statistically interchangeable. The higher the pass rate, the more statistically indistinguishable the LMM's classification is from the human annotators' labels. The procedure is outlined in Figure A.7.[13]

More concretely, suppose we have a set of examples denoted $\mathcal{X}$ and human annotators $\mathcal{H}$. For a given example $x \in \mathcal{X}$, let $\mathcal{H}(x)$ denote the set of human annotators who labeled example x. Similarly, for $h \in \mathcal{H}$, let $\mathcal{X}(h)$ denote the set of examples labeled by annotator h. For annotator $h \in \mathcal{H}(x)$, we can also define $\mathcal{H}_{-h}(x)$ to be the set of annotators who labeled example x, excluding annotator h. For $h \in \mathcal{H}(x)$, let $h(x)$ denote the label that human annotator h assigns to example x. As in Calderon et al. (2025), let $f(x)$ denote the model's label assigned to example x. Then, for a given annotator $h_1 \in \mathcal{H}$ who labeled example $x \in \mathcal{X}$, we define the leave-one-out agreement score with annotator $h_2 \in \mathcal{H}$ excluded as

$$\mathrm{ACC}_{h_1}(x, h_2) = \frac{1}{\left|\mathcal{H}_{-h_2}(x)\right|} \sum_{h' \in \mathcal{H}_{-h_2}(x)} \mathbf{1}\{h_1(x) = h'(x)\}.$$

With this definition, we construct a measure of how the model performs compared to human h on example x in terms of relative agreement with all other human annotators who labeled x:

$$d(x, h) = \mathrm{ACC}(f, x, h) - \mathrm{ACC}(h, x, h).$$

Then, we compute

$$\mu(h) = \frac{1}{|\mathcal{X}(h)|} \sum_{x \in \mathcal{X}(h)} d(x, h).$$

[13]We also report traditional macro F1 scores in Table B.3.

We then perform a one-sided t-test of the null $H_0 : \mu(h) \geq 0$. A low (negative) value of $\mu(h)$ provides evidence that the model agrees less with other annotators than annotator h on the set of examples labeled by h.

We can perform the same test for every human annotator. Of course, there exists dependence across the test statistics. We apply the Benjamini-Yekutieli procedure to obtain adjusted p-values that control the false discovery rate under arbitrary dependence among the test statistics (Benjamini and Yekutieli, 2001). For a fixed level α, we determine the fraction W of adjusted p-values that are above α. This fraction represents the "pass rate," which intuitively is the fraction of human annotators with whom we cannot reject that the model is interchangeable.

We note that our procedure deviates from the alternative annotator test proposed by Calderon et al. (2025) in two important ways. First, they define an example-level win rate in place of our relative agreement. That is, they define

$$\tilde{d}(x, h) = \mathbf{1}\{\text{ACC}(f, x, h) - \text{ACC}(h, x, h) \geq 0\}.$$

This binarization discards information in the sense that the magnitude of the discrepancy between the model's agreement and h's agreement with other humans is not accounted for. The second departure we make from Calderon et al. (2025) is that we test for whether the model exhibits significantly lower agreement with other annotators rather than significantly higher agreement. Due to the fact that having equal agreement with humans is arguably the theoretical limit on how well a model can perform in a classification task, testing whether the model agrees with other human annotators significantly more than humans agree with each other sets an implausibly high bar. As a result, these authors introduce a cost parameter ε that biases the test in favor of the model. Given our definitions, the analog of their implementation would be to test $H_0 : d(x, h) \leq -\varepsilon$. The spirit of this cost parameter is to tolerate lower model performance since the model offers a cheaper source of measurement than human annotators. In practice, however, the choice of ε is somewhat arbitrary. We find it more natural and interpretable to test whether the model is statistically underperforming relative to human annotators and avoid the introduction of this cost parameter.

Table B.4 shows the results of this validation exercise. For each exercise, we show the adjusted p-values and pass rates at different significance levels. For several exercises, the LMM exceeds the 0.5 pass rate proposed by Calderon et al. (2025) as the threshold above which the model can justifiably replace human annotators. In many settings (and especially in text), demographic characteristics may simply be unavailable, so we are only assessing model performance on a selected set of examples. Nonetheless, in cases where demographic information can be ascertained, the model performs well.

At the same time, while the model performs comparably to some annotators in terms of emotion labeling, the 0.5 threshold is not always exceeded. The model's emotion labels in text are not statistically distinguishable from two out of six annotators. It performs better in image labeling, meeting or exceeding the 0.5 pass rate at each significance level. To the extent that humans can hold distinct but reasonable interpretations of an emotion, we view these pass rates favorably. However, after binning the emotions into two valence categories ($\{\text{happy, calm}\}$ and $\{\text{any other emotion}\}$, as well as $\{\text{no emotion}, \text{unsure}\}$), we see that the pass rate in image labeling declines. While the models may perform satisfactorily overall, these results motivate an analysis to verify that the types of mistakes the models make are not driving our primary results.

To do this, we use the validation sample to estimate the simulated distribution of human emotion labels we predict would be obtained if humans had labeled the entire dataset. We let $\mathcal{C}$ denote the set of emotion classifications. We observe that

$$\mathbb{P}(h(x) = c) = \sum_{c' \in \mathcal{C}} \mathbb{P}(h(x) = c | f(x) = c')\mathbb{P}(f(x) = c').$$

We obtain $\mathbb{P}(f(x) = c')$ from the model labels on the entire dataset, and we estimate $\mathbb{P}(h(x) = c | f(x) = c')$ using the validation sample, which was obtained by stratified random sampling from the full dataset (with LMM class labels as strata to ensure representation of each class). The results are shown in Figure A.8 with 95% confidence intervals obtained via stratified bootstrap resampling from the validation data. We observe that people ascribe calm to more examples in text, and, compared to the model, they label more emotions in faces as happy rather than calm. Nonetheless, we recover the same pattern from the predicted human labels that text displays both positively and negatively valenced emotions while images show happy and calm almost uniformly. This is further confirmed in Figure A.9, which shows the distribution of emotions predicted by humans and the model after binning into Happy/Calm and Not Happy/Calm.

In Figure A.10, we also show an additional robustness check for the LMM classifications in the text where we subset to emotion labels that the model has high confidence in (≥ 0.8 on a scale of 0 to 1).[14] This is a different subset of characters (likely characters where the emotion is clear). When we examine the overall representation of emotions in this subset, we see the largest changes in the proportion of happiness (10 percentage point increase) and confusion (9 percentage point decrease), but our main takeaways remain the same.

Overall, the results of the validation exercise lend confidence in the quality of the measurements produced by our LMMs. The model is never statistically worse than all annotators, and it exceeds the prescribed 0.5 threshold in several instances. To the extent that human disagreement reflects legitimate alternative interpretations, this means that the models are not out of line with the labels humans might propose. The model is not identical to human annotators, but we nevertheless recover the same qualitative conclusions after simulating the distribution of emotion labels that we predict humans would produce if they were to label the entire dataset.

[14] We note that the model-reported confidence score has no probabilistic interpretation, but model performance increases with this score. In particular, at this threshold, we obtain a pass rate of 5/6 for emotion labeling. We retain 282 validation examples after filtering on the confidence score, so this is not driven by a decrease in power of the test. We also see that the macro F1 score and LLM-human agreement relative to human-human agreement both increase with the confidence threshold.

VI Consumer Demand and Emotions

Our analysis was motivated by the fact that it is not ex ante obvious how children's materials will represent emotions. Because publishers operate in markets, the emotional content they supply may reflect the types of books purchasers are more likely to select. If purchasers (e.g. parents, libraries, schools) value materials that build emotional intelligence by exposing children to the richness and complexity of real-life emotional experience, then publishers would have incentives to produce books that portray a broad emotional repertoire across both text and images. If, instead, purchasers value materials that conform to prevailing social and labor market norms about which emotions are appropriate to display, then publishers would have incentives to emphasize calmness and happiness, especially in visible features such as illustrations and covers. A third possibility is that purchasers may prefer emotional complexity in principle but make choices based on salient, easily processed cues that do not fully capture those preferences. In each case, purchaser behavior can shape the emotional representations children are likely to encounter.

Having documented a systematic oversupply of happy and calm emotions in images relative to the text in two influential sources of children's media, we next ask whether these patterns are consistent with purchaser demand. If purchasers are more likely to select books whose visible features convey positive emotions, then authors, illustrators, and publishers may be incentivized to emphasize positive expressions, especially in images and on covers. This is consistent with other evidence that purchaser demand shapes media content (Gentzkow et al., 2006; Gentzkow and Shapiro, 2010; Sacerdote et al., 2020) and that salient product attributes influence choice when purchasers have limited attention (Ellison, 2006; Bordalo et al., 2013).

Book covers provide a natural setting to study this mechanism. Covers are highly visible at the point of purchase and can function both as attention-grabbing product attributes and as signals of a book's tone and content (Bloch, 1995; Pieters and Wedel, 2004; Underwood and Klein, 2002; Eliaz and Spiegler, 2011). This is especially relevant for online purchases included in our sample, where purchasers often cannot inspect interior pages before buying. If purchasers favor happy or calm images, this could create market incentives to supply books with these positive emotions on the cover, even when the text contains a broader range of emotions.

This discussion yields a simple empirical prediction: conditional on book characteristics, titles whose covers depict only happy or calm characters should be selected more often than titles whose covers include a character displaying another emotion. We test this prediction using two measures of choice in the children's literature sample: household purchases and library holdings.

VI.A Empirical Analysis of Purchaser Demand

We restrict the consumer demand exercise to the children's literature collection, whose consumption more directly reflects purchaser choice. Textbook adoption, by contrast, occurs at the school- or district-level and is largely insulated from individual purchaser choice.[15]

[15] Parents can select into schools, but textbook content is only one dimension of this multifaceted decision.

We study purchaser demand using two separate data sources. First, we estimate logit models using micro-level book purchase data from the Numerator OmniPanel, which records individual buying decisions and thus speaks directly to purchasers' revealed choices. A limitation of these data is that we cannot observe purchasers' consideration sets. In our analysis, we construct a common choice set consisting of all the children's book titles observed at least once in the Numerator transaction data. For a second measure, we also estimated demand using data on book inventories in the Seattle Public Library system. We expect that a public library would consider purchasing a much broader range of books than a typical individual purchaser, making the assumption that every award-winning book in our sample belongs to each library's consideration set more credible. Moreover, while public libraries may have different objective functions than individual purchasers, we expect that they would generally stock more copies of frequently borrowed books, in which case library holdings would indirectly reflect reader demand.

To study the disproportionate representation of happy and calm in images specifically, we collapse emotions into two valence categories: happy/calm or not happy/calm, with the latter category indicating that a character displays some emotion besides happiness or calm.[16] Using the Numerator data on purchasers' decisions, we estimate logit models. We then use the public-library inventory data to regress the number of copies of a book held by the library against the same set of covariates. In all specifications, we include an indicator for whether the cover depicts only happy or calm characters and an indicator for whether the cover depicts no character at all. We report a baseline specification without controls, and we also include specifications that control for the award(s) the book won, copyright year, whether the book contains no detected emotions in text or in images, and page count. For all analyses using library data, we include library branch fixed effects. We do not expect purchasers to accurately perceive the emotional composition of a book's interior. Nonetheless, we account for the proportions of happy/calm emotions in both images and text by (depending on the specification): (1) directly including these proportions as linear controls or (2) by adding indicators for quintiles of these proportions within their respective distributions. These controls account for purchasers who attend more closely to the interior emotional content and may additionally proxy for unobserved confounders.

Table 1 reports the household purchase results from the Numerator data. The emotional representation on the cover plays a substantive role in shaping purchasers' decisions. In particular, across all specifications, purchasers place a significant premium on books depicting happy or calm characters on the cover relative to books depicting other emotions. Notably, purchasers significantly favor books with *no* characters on the cover over books with a character displaying an emotion which is neither happy nor calm. This suggests an explicit aversion to negative emotional imagery: purchasers prefer even an emotionless cover to one signaling anger, fear, or sadness. Such an aversion would, in turn, discourage authors and illustrators from depicting these emotions, generating precisely the visual pattern we document.

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

In Table 2, we present the analogous results using the Seattle Public Library system data. The findings are qualitatively similar: the library stocks more copies of books with exclusively happy or calm cover characters, or with no characters at all, than of books with not happy/calm characters on the cover. Because the library's consideration set may be far broader than any household's, this suggests that our inability to observe individual purchasers' consideration sets does not drive the results in Table 1. Instead, these results jointly speak to a consistent pattern of readers favoring books with depictions of happiness and calm over books with other emotional depictions.

Additionally, in order to connect our findings to purchasers' decisions about happy and calm emotions in general and not just on covers, 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 a 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 not 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 featuring exclusively happy/calm characters on the cover. Likewise, lower proportions of happy/calm emotions in text and images predict featuring exclusively not happy/calm characters on the cover.[17] These results indicate that covers do provide purchasers with an informative signal about the emotional content within the book (e.g. they can indeed judge a book by its cover). Therefore, purchasers choosing books with more happy/calm characters pictured on the cover is also consistent with wanting these emotions displayed in the content of the book.

In light of these results, the patterns we document in Section IV are consistent with a supply-side response to purchasers' revealed choices regarding emotional displays in children's materials. The evidence does not allow us to distinguish whether purchasers genuinely prefer books that visually emphasize happy and calm emotions, whether they believe such displays are more appropriate for children, or whether they are imperfectly attentive to the broader emotional content of the book. However, all three interpretations help connect the demand results to our motivating empirical question. Rather than rewarding visual representations that expose children to the full range of emotional experience, purchaser behavior appears to reward books whose visible features align with socially valued emotional displays. Moreover, since images – especially those on covers – are particularly salient compared to text, authors would face a stronger disincentive to depict not happy/calm emotions in the images compared to text. This is consistent with our finding that text conveys a diversity of emotions while images predominantly show characters who are either happy or calm. We also note that this market equilibrium is not necessarily innocuous: if purchasers exhibit bounded rationality but prefer depictions of positive emotions, the resulting equilibrium may oversupply happy and calm imagery relative to a benchmark that places greater value on exposing

[17]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 not happy/calm emotions, or they could display no characters at all.

young readers to a broader emotional repertoire.[18]

VII Conclusion

This paper explores one channel through which emotional norms are embedded and transmitted into society. Using large multimodal models to classify emotions in public-school textbooks and influential children's literature, we document a systematic divergence between what children read and what they see. Text exposes children to a broad emotional repertoire including not only happiness and sadness, but also anger, fear, confusion, and surprise. Images do not. Calm and happiness together account for roughly 90% of visually depicted emotions, while emotions traditionally coded as "negative" are almost entirely absent, even on pages whose text describes them. This mismatch is not an artifact of aggregation: it holds within pages and even within individual matched characters. Nearly half of all pages show no overlap between the emotions conveyed in text and those conveyed in images.

These patterns prove remarkably stable. They persist across three decades of textbooks and a century of literature, across states occupying opposing sides of the political spectrum, across subjects and grade levels, and across every character identity we examine. Where identity shapes emotional representation, it does so within the boundaries of positive affect: for instance, female characters are more likely to be shown as happy while male characters are more likely to be shown as calm, subtly reinforcing gendered emotional norms rather than widening the range of emotions children encounter. We show that households and public libraries are significantly less likely to select books whose covers depict "negative" emotions, and covers reliably signal interior emotional content. This evidence suggests that not only do these cultural goods themselves shape emotional display norms, but they may also transmit them to the next generation.

Our findings speak to several literatures. First, they inform the production function of socioemotional skills. Standard models of human-capital formation emphasize early investments, and children's media constitute one such investment. Second, we contribute to the cultural economics literature by extending its focus from the identities, narratives, and styles represented in cultural goods to the behavioral norms those goods implicitly encode. Third, we expand prior work on emotional representation in children's books, which has relied largely on manual content analysis of text alone. Fourth, we introduce and validate multimodal measures of emotional expression, offering tools that can stimulate further research on emotions and deepen understanding of how they relate to broader societal outcomes. Finally, we contribute to the economics of media by showing that this pattern may be an equilibrium outcome rather than an exogenous feature of content. Consequently, authors and illustrators may face incentives to oversupply happy and calm imagery, especially in visual features, and these market forces would generate precisely the pattern we document.

[18] Whether this imbalance reduces welfare depends on assumptions about the objectives of children's media, how exposure to different emotional displays affects children's development, and how those effects should be weighed against purchaser preferences. Formalizing these trade-offs would require a model and welfare criterion beyond the scope of this paper.

Two features of this equilibrium warrant emphasis. First, we cannot disentangle three explanations for purchaser behavior: a genuine preference for positive imagery, a belief that such displays are more appropriate for children, or simple attraction to salient smiling faces under limited attention. All three fit our demand results but carry different welfare implications. If purchasers exhibit bounded rationality yet are drawn to positive expressions, the equilibrium may oversupply happy and calm imagery relative to a benchmark that values exposing children to a broader range of emotions. Second, the visual concentration in calm and happiness appears not only in children's books, whose acquisition reflects individual choice, but equally in public-school textbooks, which face different market forces and state-standard constraints. That the same display rule emerges in both corpora points to a diffuse cultural norm governing outward emotional expression, not demand alone. We therefore read consumer demand as one mechanism reinforcing this norm in the market for children's books, rather than the sole cause of the text-image divergence.

These conclusions come with limits. Our analysis documents emotional representation, not downstream effects; we do not measure how the text-image mismatch shapes children's emotional development, and establishing that link is an important direction for future work. Moreover, our emotion classifications, though validated against human annotators and alternative models, remain model-generated. Calm and happiness are difficult to separate, though this ambiguity does not affect the positive versus negative contrast that drives our results.

More broadly, our findings underscore the role of children's materials in the formation of emotional skills and norms. Emotional intelligence develops through repeated exposure to emotional language, visual cues, and models of expression. We show that influential educational materials offer children a rich emotional vocabulary in text while visually reinforcing a far narrower repertoire centered on happiness and calm. This pattern is consistent with social and market incentives that reward positive emotional displays, but it may also constrain the range of emotions children see modeled during the very years when emotion-recognition skills develop. From an educational standpoint, the results point to the value of more intentional emotional diversity across both text and images so that children encounter not only examples of emotional regulation, but visual models of the full range of emotions that shape everyday life.

References

Adolphs, Ralph, Leonard Mlodinow, and Lisa Feldman Barrett. 2019. "What is an emotion?" Current Biology, 29(20): R1060–R1064.

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.

Algan, Yann, Eva Davoine, Thomas Renault, and Stefanie Stantcheva. 2026. "Emotions and policy views."

Angelucci, Manuela, Raissa Fábregas, and Antonia Vazquez. 2026. "The well-being effects of digital mental health care."

Arizpe, Evelyn, and Morag Styles. 2015. Children reading picturebooks: Interpreting visual texts. Routledge.

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

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.

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.

Benjamini, Yoav, and Daniel Yekutieli. 2001. "The control of the false discovery rate in multiple testing under dependency." Annals of statistics, 1165–1188.

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."

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

Bowles, Samuel, and Herbert Gintis. 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.

Brady, Shannon M, Marissa Ogren, and Scott P Johnson. 2024. "Effects of conflicting emotional cues on toddlers' emotion perception." British Journal of Developmental Psychology, 42(3): 376–391.

Calderon, Nitay, Roi Reichart, and Rotem Dror. 2025. "The alternative annotator test for llm-as-a-judge: How to statistically justify replacing human annotators with llms." 16051–16081.

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).

Côté, Stéphane, Anett Gyurak, and Robert W Levenson. 2010. "The ability to regulate emotion is associated with greater well-being, income, and socioeconomic status." Emotion, 10(6): 923.

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."

Denham, Susanne A. 2019. "Emotional competence during childhood and adolescence." In Handbook of Emotional Development. 493–541. Springer.

Dickerson, Kelli L, and Jodi A Quas. 2021. "Emotional awareness, empathy, and generosity in high-risk youths." Journal of Experimental Child Psychology, 208: 105151.

Doan, Sy, and Julia Kaufman. 2024. "What role do states play in selecting K–12 textbooks?" State Education Standard, 24(1).

Doherty, Martin. 2008. Theory of Mind: How Children Understand Others' Thoughts and Feelings. Psychology Press.

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.

Duchenne de Boulogne, Guillaume-Benjamin. 1990. The Mechanism of Human Facial Expression (Original work published 1862). Cambridge University Press.

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): 735–754.

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.

Eisenberg, Nancy, Amanda Cumberland, and Tracy L Spinrad. 1998. "Parental socialization of emotion." Psychological Inquiry, 9(4): 241–273.

Ekman, Paul. 1989. "The argument and evidence about universals in facial expressions." Handbook of social psychophysiology, 143(164): 2.

Ekman, Paul. 1992. "An argument for basic emotions." Cognition & Emotion, 6(3-4): 169–200.

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

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.

Genevsky, Alexander, and Brian Knutson. 2015. "Neural affective mechanisms predict market-level microlending." Psychological Science, 26(9): 1411–1422.

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.

Green, Clarence, and He Sun. 2025. "Picturebooks increase the frequency and diversity of emotion vocabulary in children's language environments: Modeling potential benefits to emotional literacy, with pedagogical resources." Early Education and Development, 36(3): 568–586.

Gross, James J. 2015. "Emotion regulation: Current status and future prospects." Psychological Inquiry, 26(1): 1–26.

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.

Gurney, Edmund. 1884. "What is an emotion?" Mind, 9(35): 421–426.

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.

Hoemann, Katie, Fei Xu, and Lisa Feldman Barrett. 2019. "Emotion words, emotion concepts, and emotional development in children: A constructionist hypothesis." Developmental Psychology, 55(9): 1830.

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.

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

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

James, William. 1884. "What is emotion?" Readings in the History of Psychology (1948), pp. 290–303. Appleton-Century-Crofts.

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." NBER WP 20749.

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

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." Modern English Education, 20(4): 63–73.

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.

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.

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

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

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.

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

Nikolajeva, Maria. 2017. "Emotions in picturebooks." In The Routledge Companion to Picturebooks. 110–118. Routledge.

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.

Park, BoKyung, Alexander Genevsky, Brian Knutson, and Jeanne Tsai. 2020. "Culturally valued facial expressions enhance loan request success." Emotion, 20(7): 1137.

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.

Perry, Nicole B, Jessica M Dollar, Susan D Calkins, Susan P Keane, and Lilly Shanahan. 2020. "Maternal socialization of child emotion and adolescent adjustment: Indirect effects through emotion regulation." Developmental Psychology, 56(3): 541.

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.

Ruffman, Ted, Qiuyi Kong, Hui Mei Lim, Kangning Du, and Emilia Tiainen. 2023. "Recognition of facial emotions across the lifespan: 8-year-olds resemble older adults." British Journal of Developmental Psychology, 41(2): 128–139.

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.

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.

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.

Thoits, Peggy A. 1989. "The sociology of emotions." Annual Review of Sociology, 15(1): 317–342.

Tsai, Jeanne L. 2007. "Ideal affect: Cultural causes and behavioral consequences." Perspectives on Psychological Science, 2(3): 242–259.

Tsai, Jeanne L, Brian Knutson, and Helene H Fung. 2006. "Cultural variation in affect valuation." Journal of Personality and Social Psychology, 90(2): 288.

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.

Widen, Sherri C, and James A Russell. 2008. "Children acquire emotion categories gradually." Cognitive Development, 23(2): 291–312.

Williams, John E., Susan M. Bennett, and Deborah L. Best. 1977. "Sex-trait stereotypes in England, Ireland and the United States." British Journal of Social and Clinical Psychology, 16(4): 303–309.

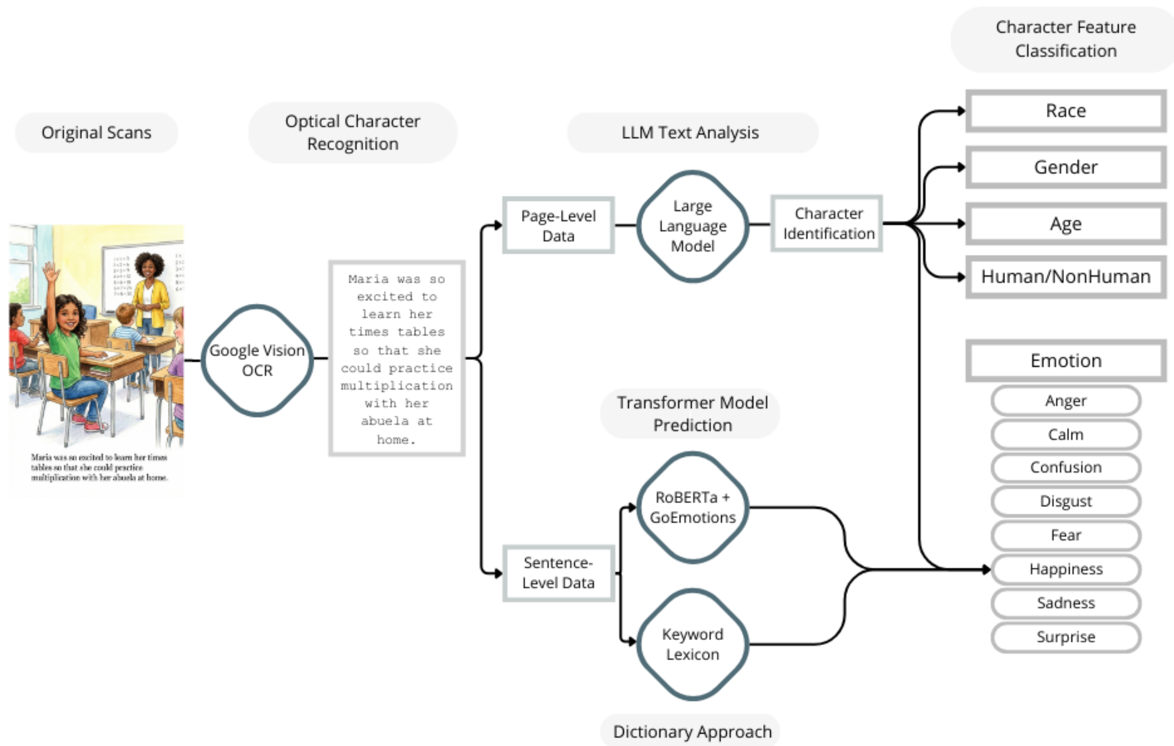
Yanagizawa-Drott, David H., and Hans-Joachim Voth. 2023. "Image(s)." Working Paper.

Zemach, Eddy M. 2001. "What is emotion?" American Philosophical Quarterly, 38(2): 197–207.

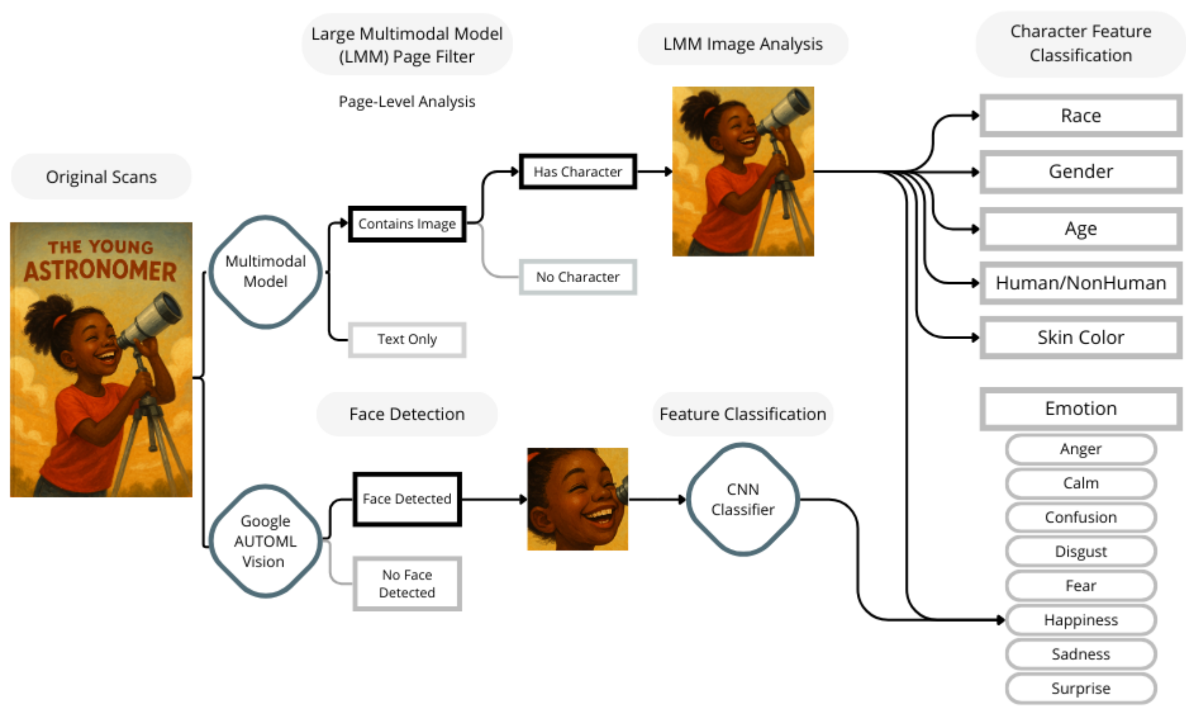
Ziems, Caleb, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. 2024. "Can large language models transform computational social science?" Computational Linguistics, 50(1): 237–291.

FIGURE 1
Data Extraction Pipelines

(a) Text Analysis



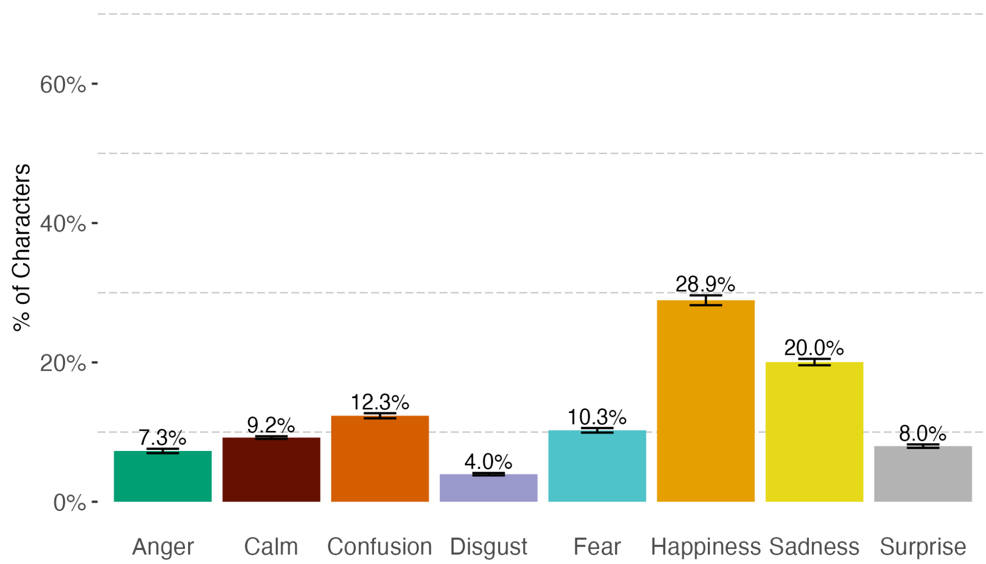
(b) Image Analysis



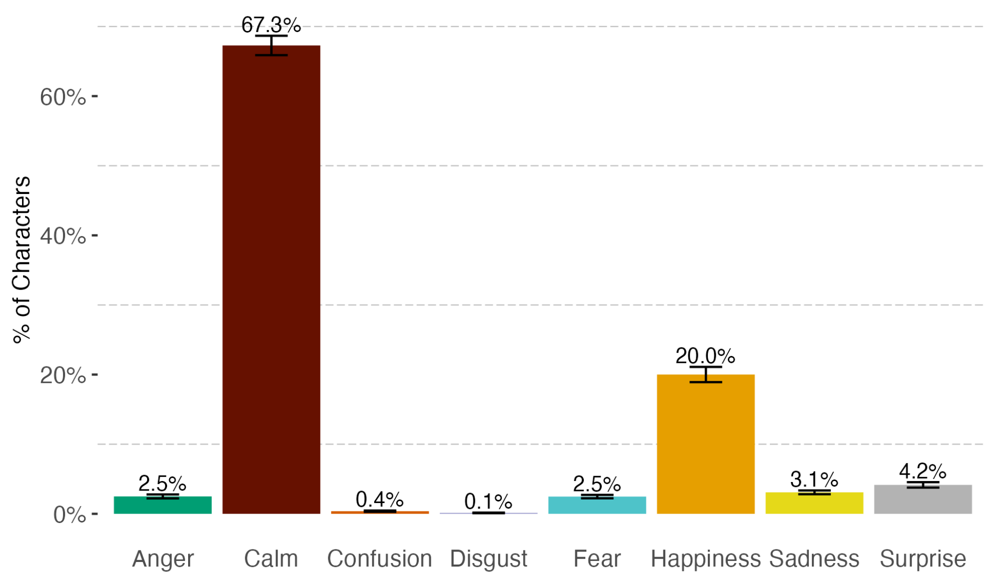
Notes: This figure illustrates the steps in our text analysis (top) and image analysis (bottom) pipelines beginning with the raw scan of each page in our sample of books.

FIGURE 2
Overall Representation of Emotions

(a) Text

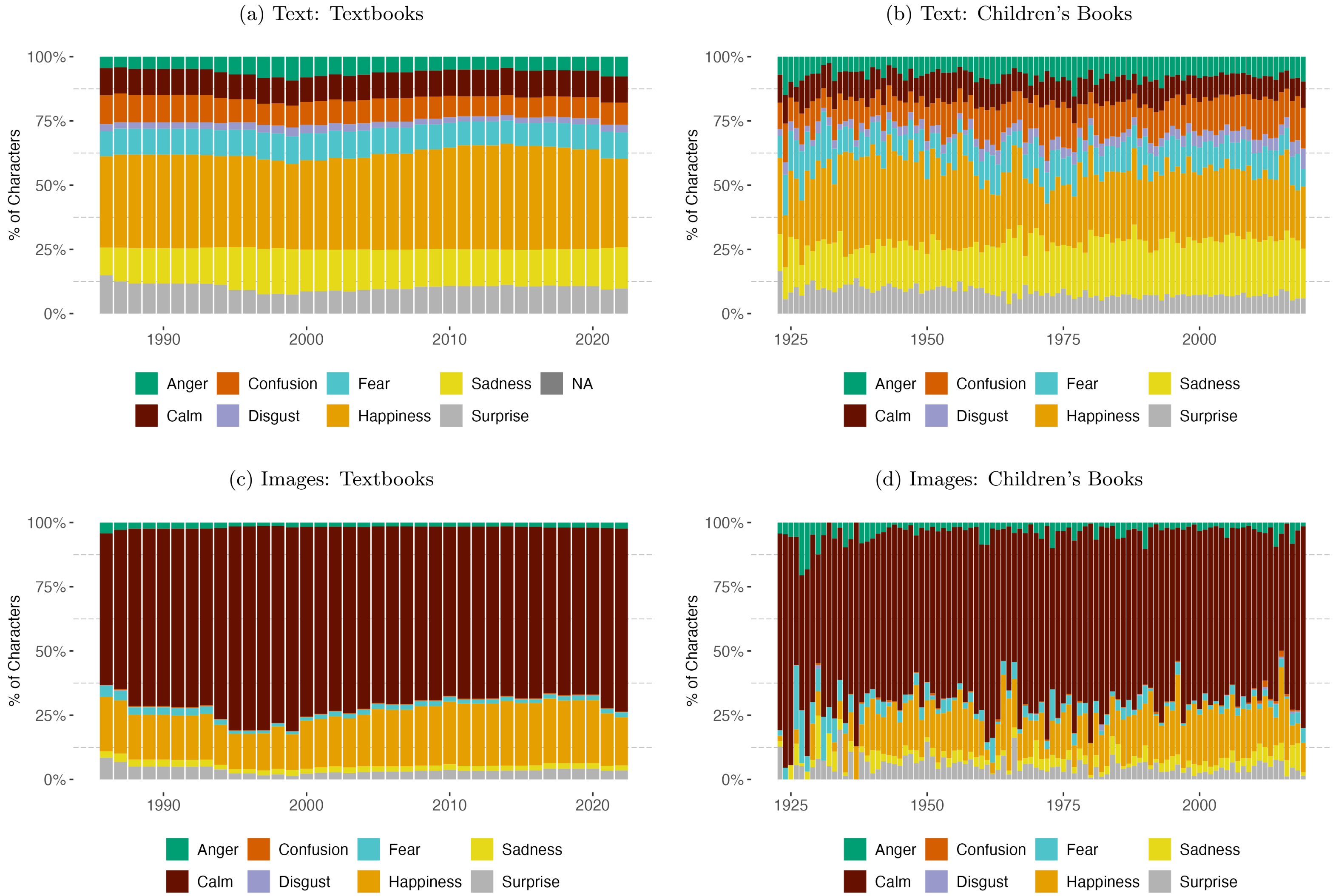


(b) Images



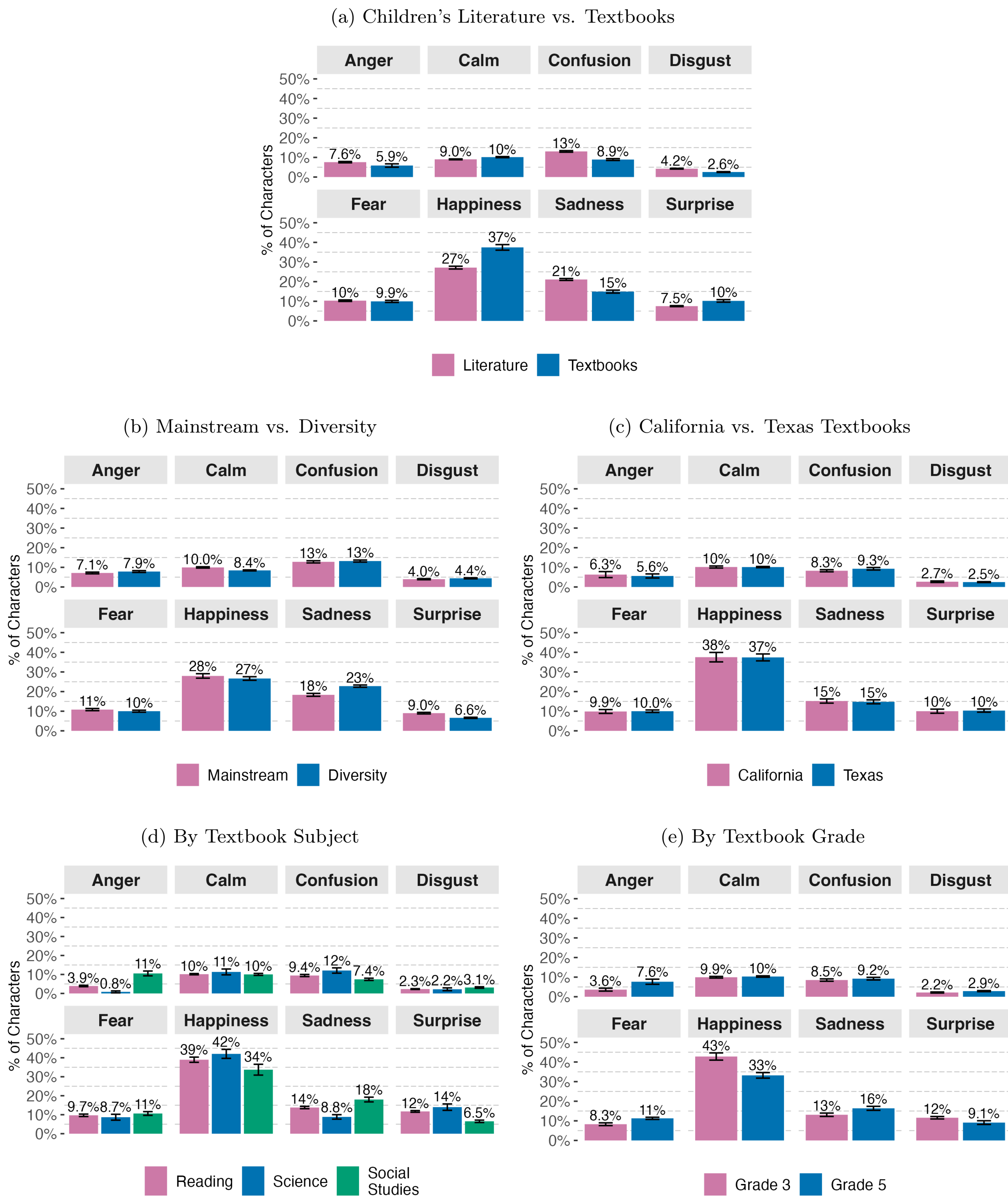
Notes: This figure shows the aggregate displays of emotions in text and images. The top panel shows the percentage of emotions conveyed by characters as a share of all emotions detected in the text of the books (literature and textbooks) using a large language model. The bottom panel shows the percentage 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. Error bars depict 95% confidence intervals constructed from standard errors that are clustered at the book level.

FIGURE 3
Representation of Emotions Over Time



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 a large multimodal model. 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. 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.

FIGURE 4
Representation of Emotions in the Text of Different Subcollections

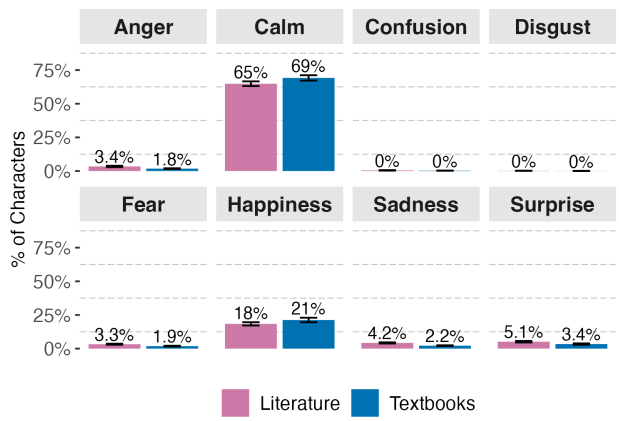


Notes: This figure shows the percentage of emotions conveyed by characters as a share of all emotions detected by a large multimodal model separately for different subcollections. Panel (a) compares textbooks and children's books, panel (b) compares California and Texas textbooks, panel (c) compares 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. Error bars depict 95% confidence intervals constructed from standard errors that are clustered at the book level.

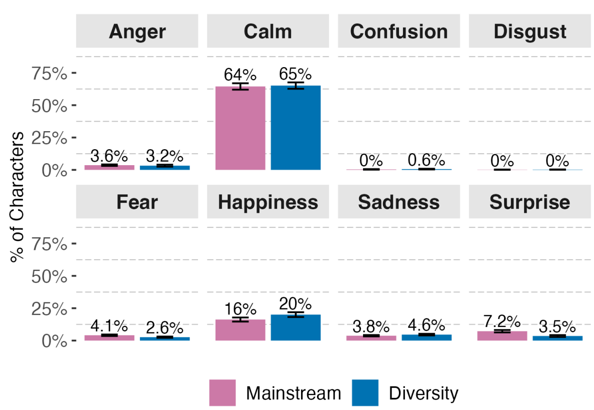
FIGURE 5

Representation of Emotions in the Images of Different Subcollections

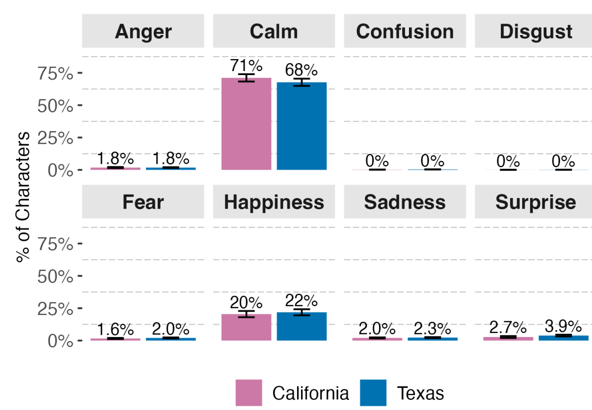
(a) Children's Literature vs. Textbooks



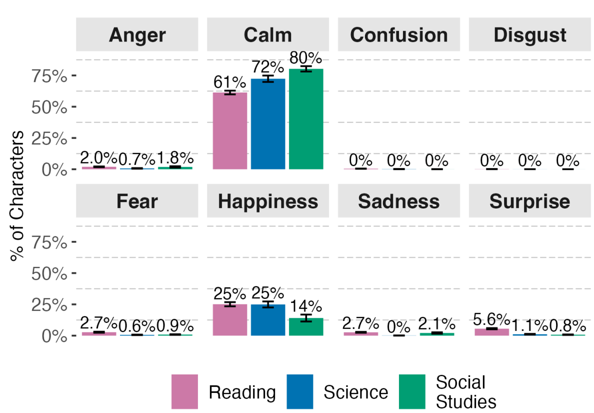
(b) Mainstream vs. Diversity



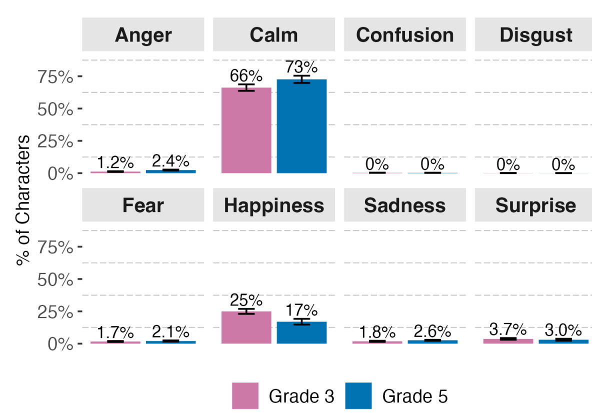
(c) California vs. Texas Textbooks



(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 a large multimodal model 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. Error bars depict 95% confidence intervals constructed from standard errors that are clustered at the book level.

FIGURE 6

Displays of Emotions in the Text for Characters with Different Identities

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. Error bars depict 95% confidence intervals constructed from standard errors that are clustered at the book level.

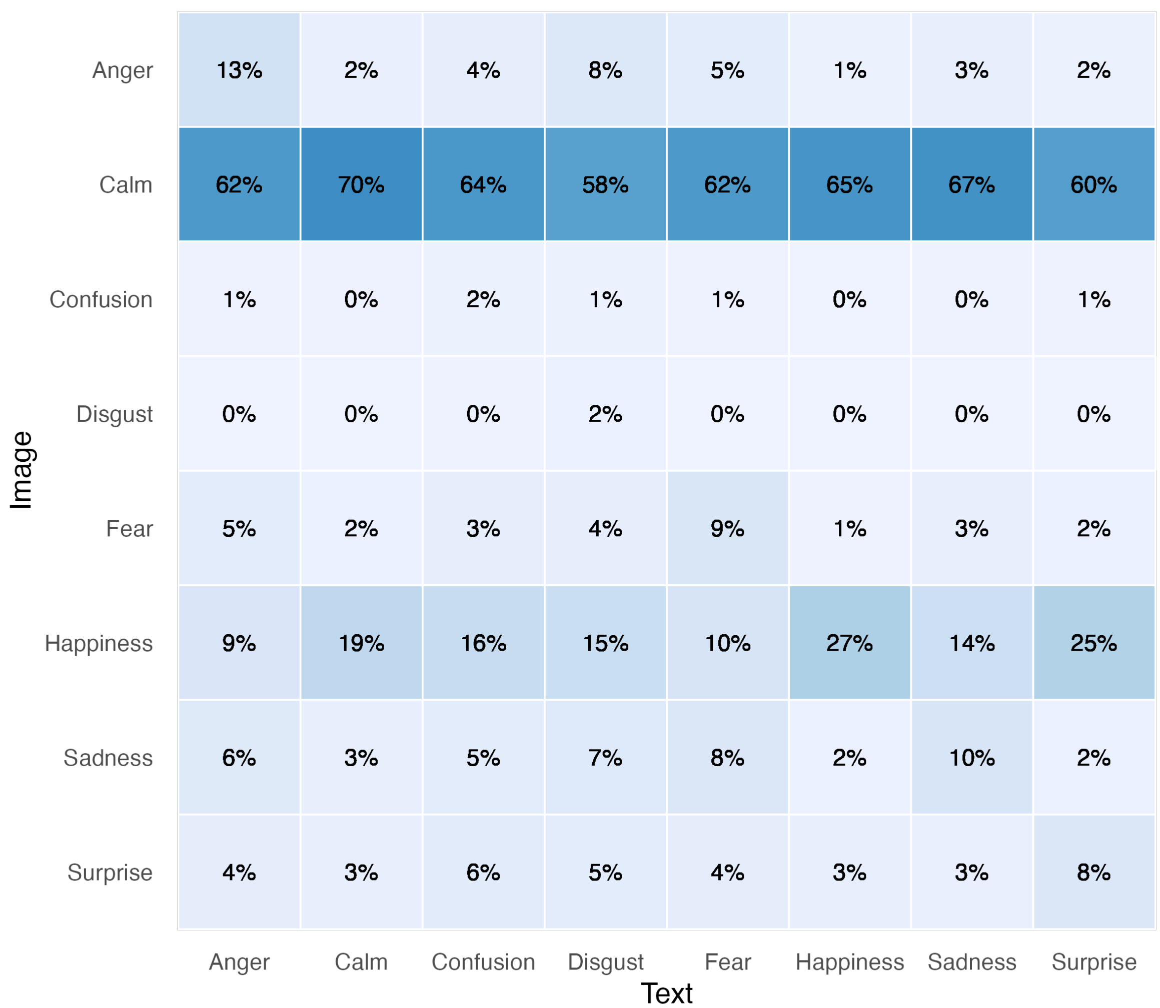
FIGURE 7

Displays of Emotions in the Faces of Characters with Different Identities

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. Error bars depict 95% confidence intervals constructed from standard errors that are clustered at the book level.

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, 13% of emotions detected on the same page show anger.

TABLE 1
Book Purchases

	Dependent variable:		
	Whether Purchased		
	(1)	(2)	(3)
Happy or Calm Only Cover	0.170***	0.537***	0.497***
	(0.025)	(0.036)	(0.037)
No Cover Character	1.186***	0.319***	0.407***
	(0.025)	(0.036)	(0.039)
Book Controls		X	X
Linear Emotion Controls		X	
Quintile Emotion Controls			X
Observations	26,501	26,501	26,501
AIC	926411.9	608962.4	606903.6
BIC	926440.9	609063.8	607092.0
RMSE	0.04	0.04	0.04

Notes: This table presents results from estimating logit models connecting the binary purchase outcome to preferences for cover emotion measures. The sample is restricted to children's books. "Happy or Calm Only Cover" equals 1 when the cover's characters display only calmness or happiness and 0 otherwise; "No Cover Character" equals 1 when no characters express emotions on the cover and 0 otherwise; the omitted group is books that contain at least one character who portrays an emotion other than happy or calm. Book controls include the number of pages and dummies for whether the book contains no emotions in images or in text. The models in columns (2) and (3) are conditional logit models stratified on the award(s) the book won as well as the book's copyright year. Linear Emotion Controls include the proportion of Happy or Calm emotions in text and images as linear controls, whereas Quintile Emotion Controls include indicators for the corresponding quintile bins as controls. When missing for text or images, the proportion of happy or calm emotions is imputed as the average over all books. Observations refer to the number of unique trips in the data. We cluster standard errors by household. In this table, *** and ** represent significance at the 0.01 and 0.05 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	0.1294***	0.0754***	0.0601**
	(0.0187)	(0.0243)	(0.0248)
No Cover Character	0.1213***	0.1670***	0.1584***
	(0.0206)	(0.0399)	(0.0398)
Book Controls		X	X
Linear Emotion Controls		X	
Quintile Emotion Controls			X
Observations	6,972	6,972	6,972
R^2	0.20983	0.39042	0.39361
Within R^2	0.00377	0.01661	0.02177

Notes: This table presents results from a linear regression of the library inventory stocks on cover and content emotion measures. The sample is restricted to children's books. "Happy or Calm Only Cover" equals 1 when the cover's characters display only calmness or happiness and 0 otherwise; "No Cover Character" equals 1 when no characters express emotions on the cover and 0 otherwise; the omitted group is books that contain at least one character who portrays an emotion other than happy or calm. Book controls include the number of pages, dummies for whether the book contains no emotions in images or in text, and fixed effects for the award(s) the book won as well as the book's copyright year. Linear Emotion Controls include the proportion of Happy or Calm emotions in text and images as linear controls, whereas Quintile Emotion Controls include indicators for the corresponding quintile bins as controls. When missing for text or images, the proportion of happy or calm emotions is imputed as the average over all books. All three specifications include library branch fixed effects, and we cluster standard errors by library branch. In this table, *** and ** represent significance at the 0.01 and 0.05 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	1.120***	-1.066	-1.567**
	(0.359)	(0.672)	(0.640)
Text	3.122***	-0.540	-4.225***
	(0.580)	(1.032)	(1.314)
Book Controls	X	X	X
Observations	971	971	971
AIC	1048.6	355.7	292.3
BIC	1063.2	370.4	306.9
RMSE	0.64	0.22	0.21

Notes: This table presents results from estimating conditional logit models 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 predictors shown are the proportion of happy or calm characters in images and text in the interior (non-cover) pages. Books with no detected emotions in the interior are dropped. The book controls include the number of pages and strata for the book collection (Mainstream, Diversity, or Both) and the copyright decade. In this table, *** and ** represent significance at the 0.01 and 0.05 levels, respectively.

Supplementary Appendix

(How) Do We Teach Emotions?*

September 2026

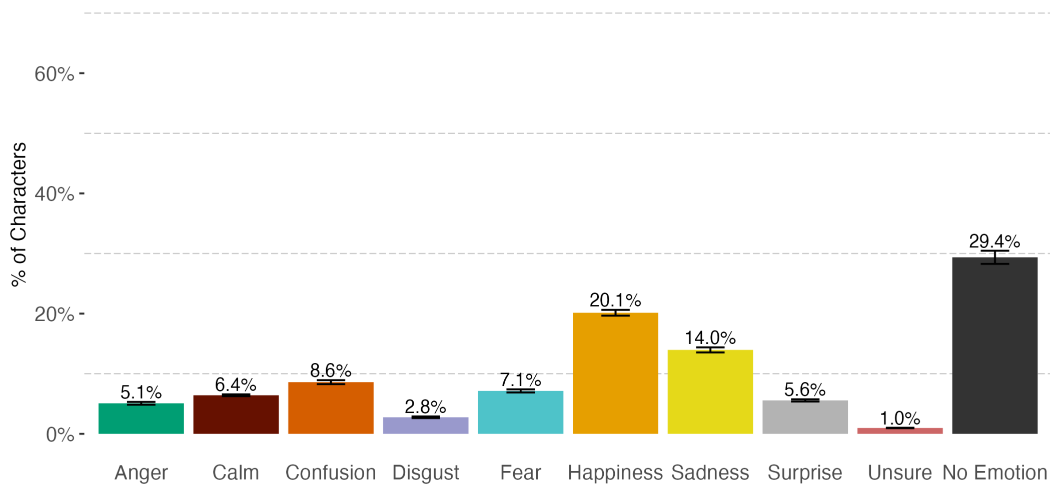
Anjali Adukia
Matthew Bonci
Paula Dastres
Emileigh Harrison
Jake Nicoll
Teodora Szasz

*Contact: adukia@uchicago.edu, mbonci@uchicago.edu, dastresgalla@wisc.edu, emileigh.harrison@austin.utexas.edu, jwnicoll@uchicago.edu, dorha.szasz@gmail.com

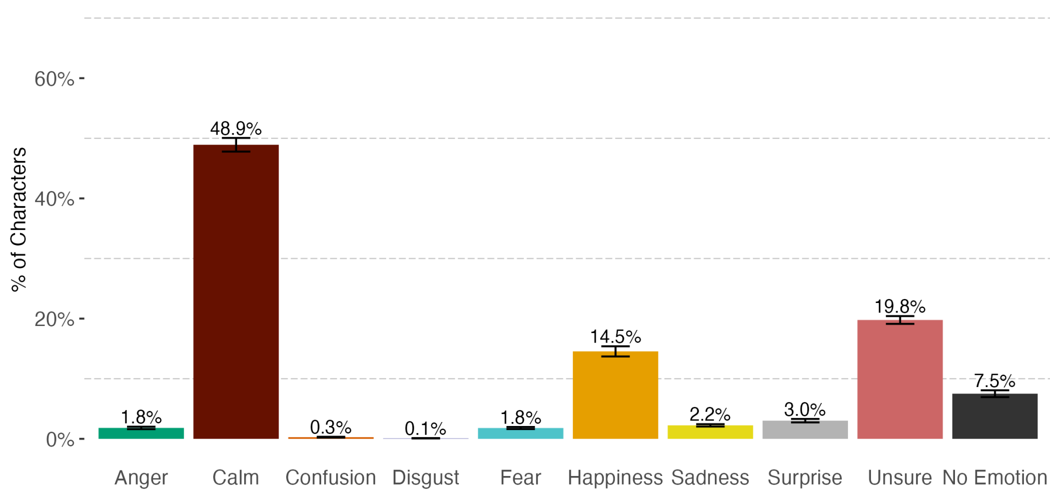
A Appendix Figures

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

(a) Text

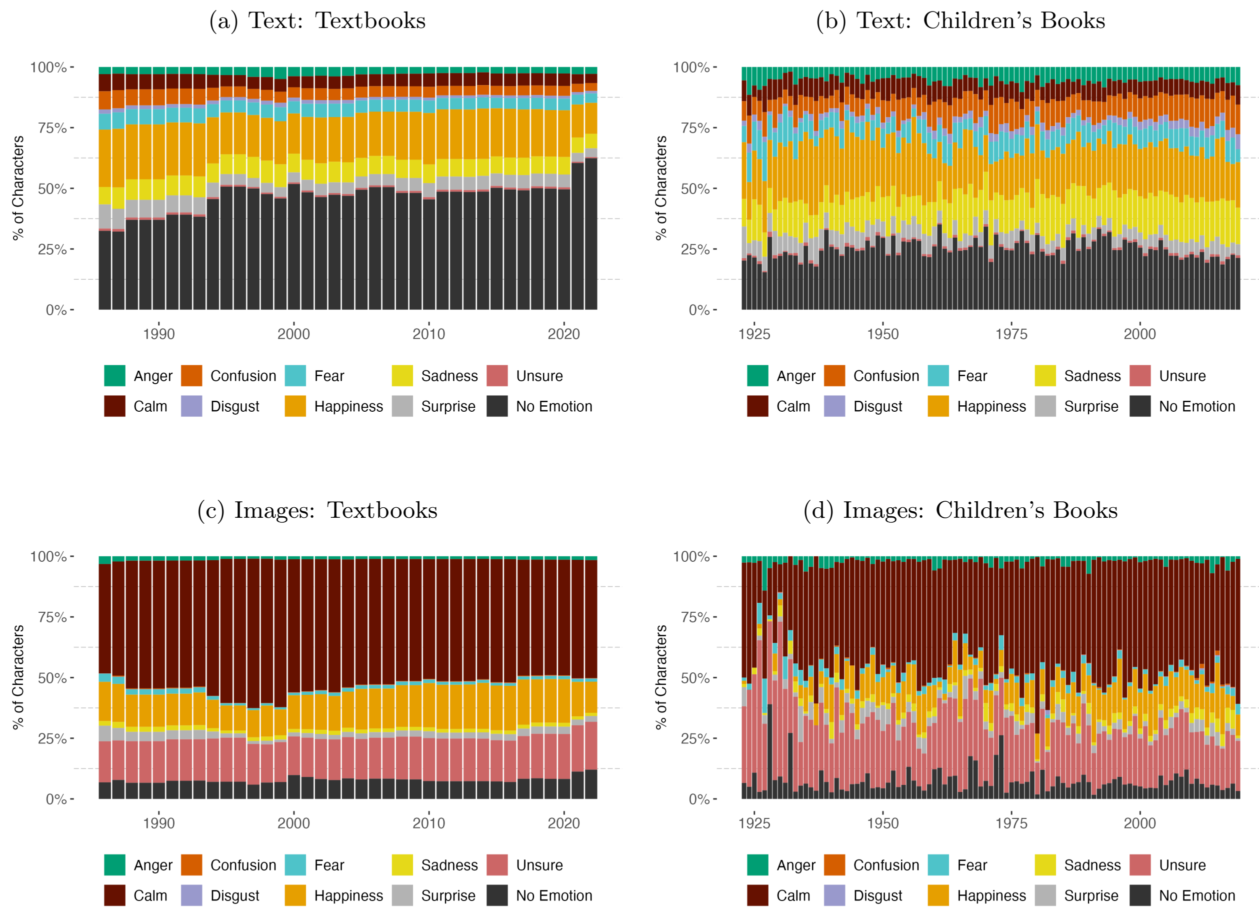


(b) Images



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

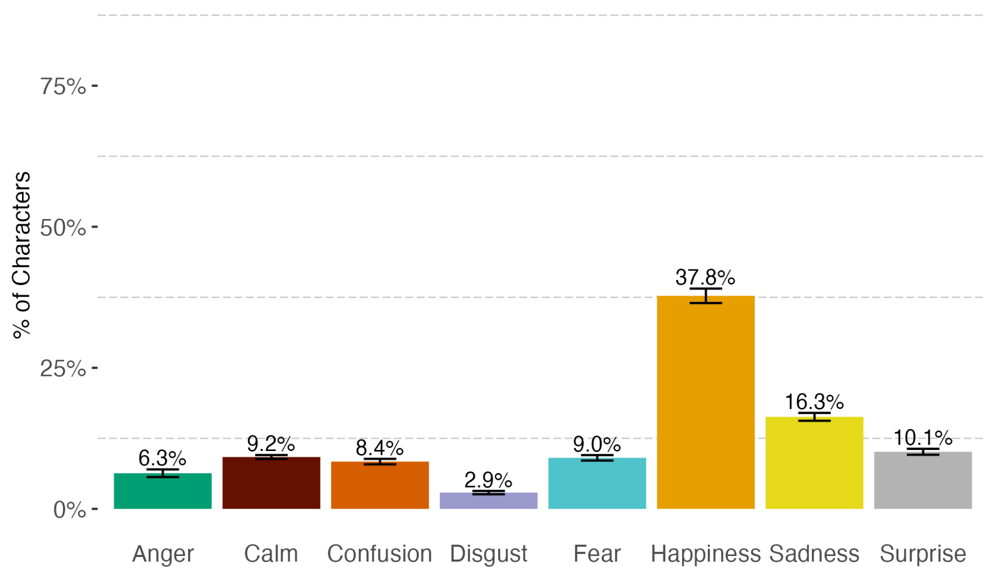
FIGURE A.2
Representation of Emotions Over Time (including No Emotion and Unsure)



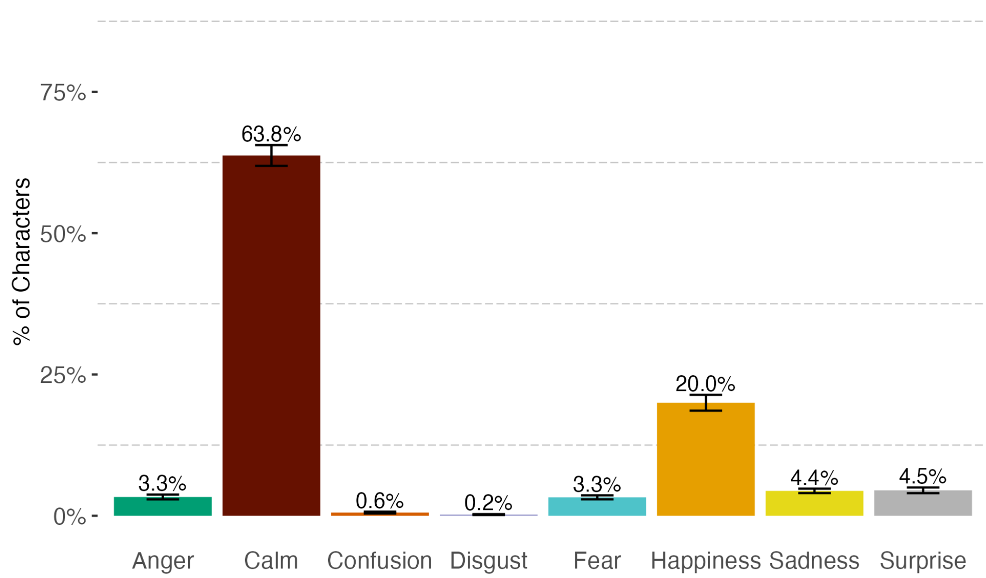
Notes: This figure replicates Figure 3 with additional categories for "No Emotion" and "Unsure." Panel (a) shows the share of all emotions detected in the sample of textbooks, and panel (b) shows the shares in literature. Emotions are detected using a large multimodal model. Emotions for each textbook 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. 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.

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

(a) Text

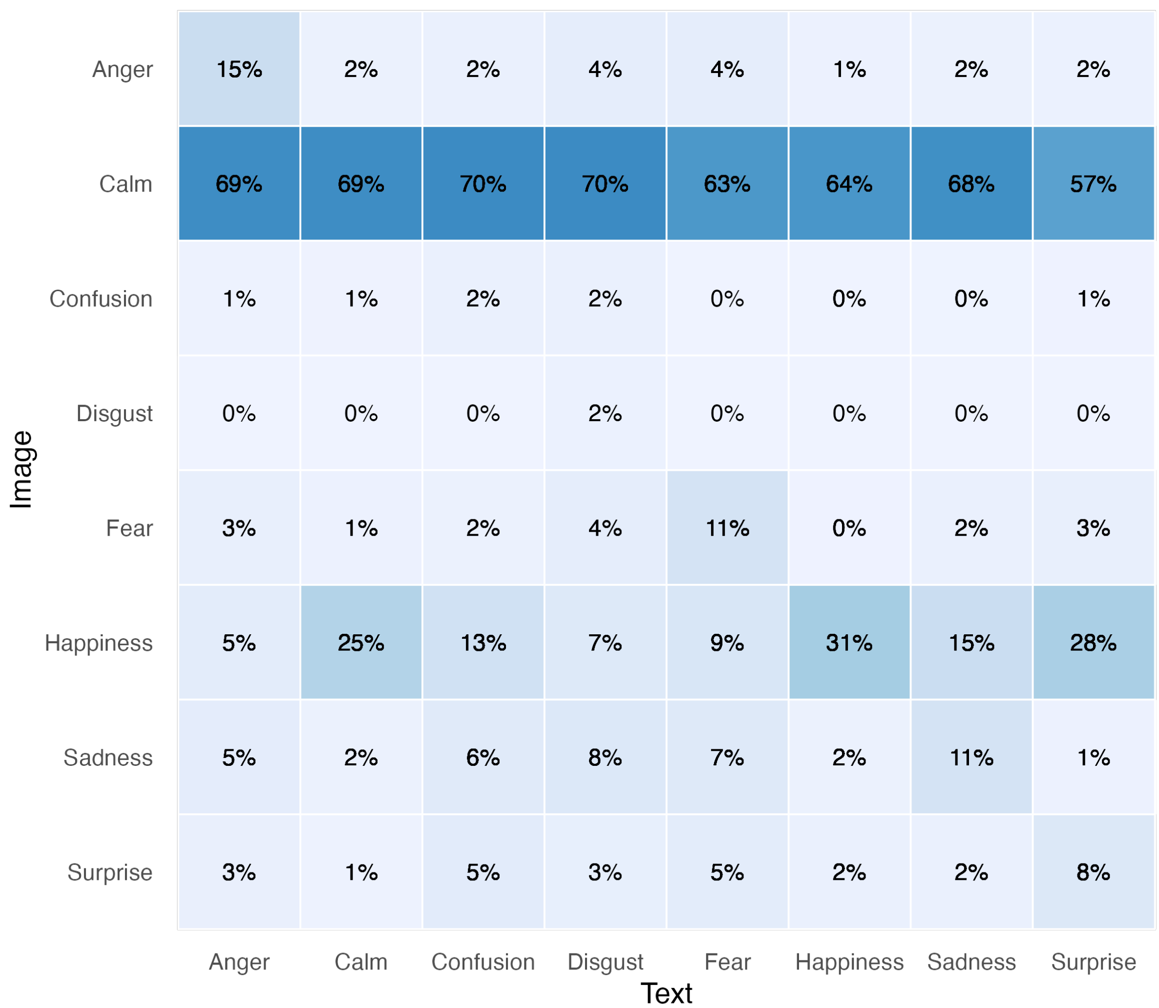


(b) Images



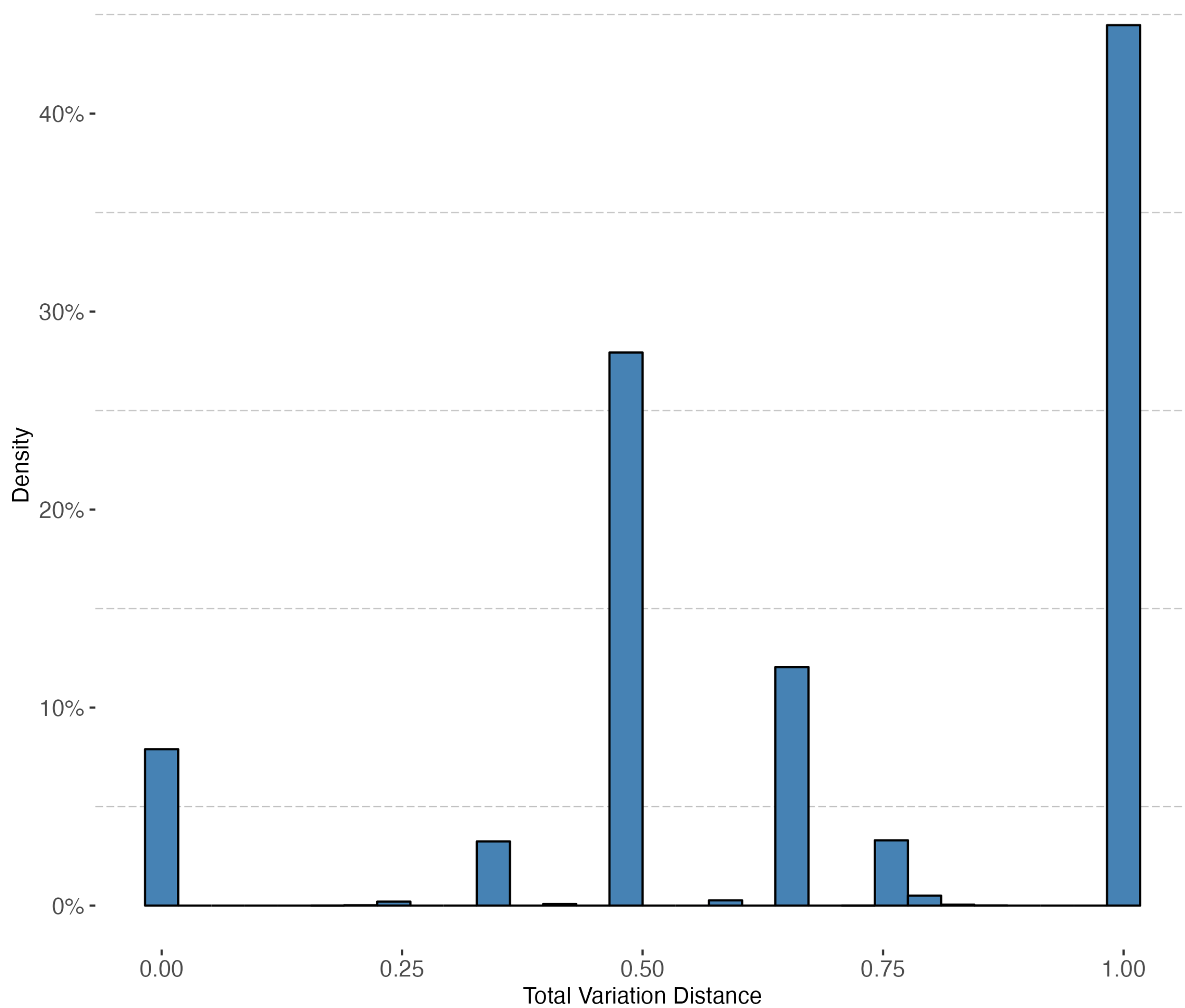
Notes: This figure shows the aggregate displays of emotions in text and images for the sample of matched pages. Emotions are detected using a large multimodal model. The top panel shows the percentage of emotions conveyed by characters as a share of all emotions detected in the text of the sample of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

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 there is one Black male character in the text and the image, when the character is shown as angry in text,the matched character will be depicted as angry in an image only 15% of the time.

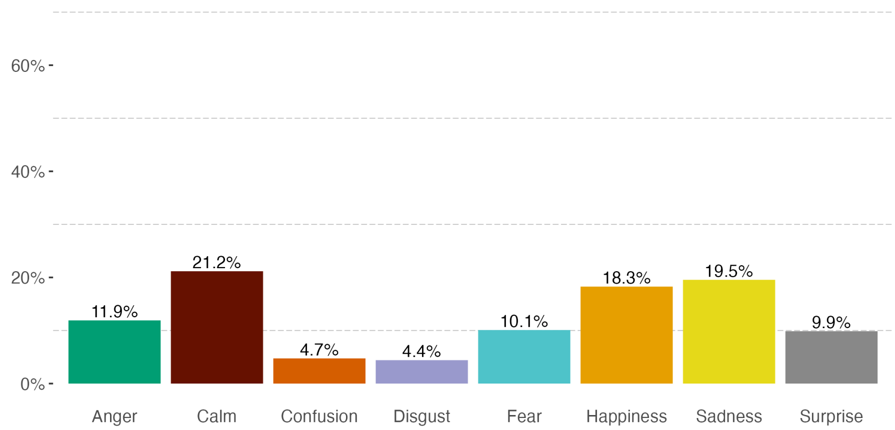
FIGURE A.5
Total Variation Distance Between Text and Images



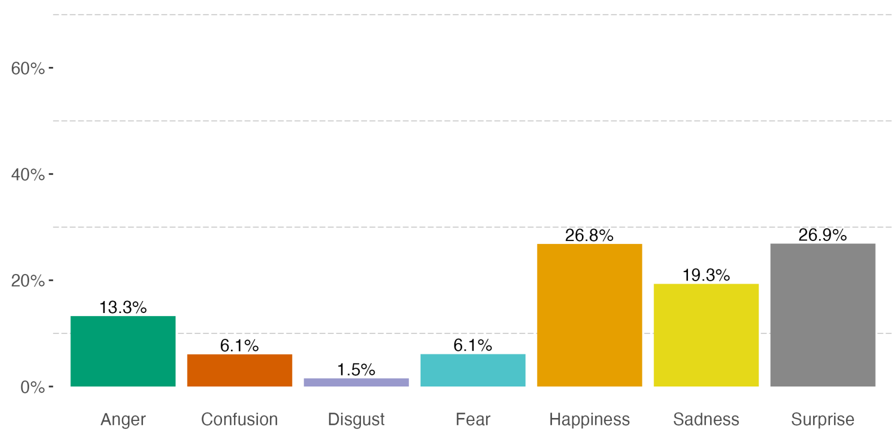
Note: This figures shows the total variation distance (tvd) between the emotional content of images and text on the same page. We calculate tvd by taking the l^1 distance between the empirical distributions of our 8 emotions in text and in images on a given page and multiply by $1/2$.

FIGURE A.6
Robustness: Overall Representation of Emotions

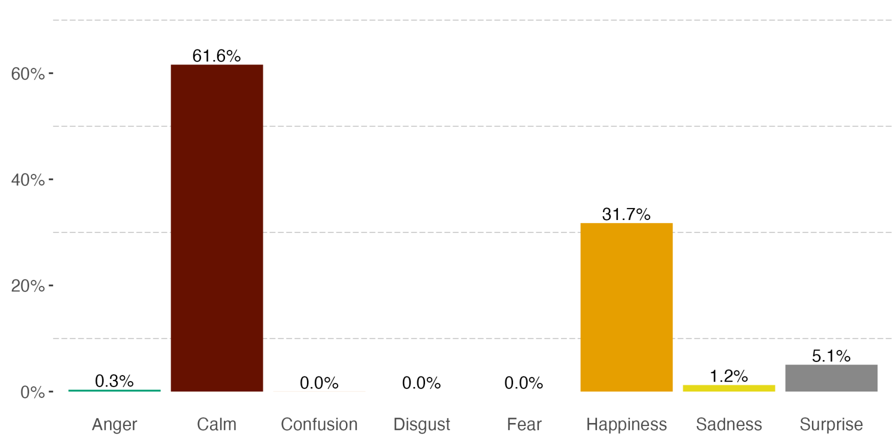
(a) Text: Lexicon Approach



(b) Text: Transformer/GoEmotions AI Model

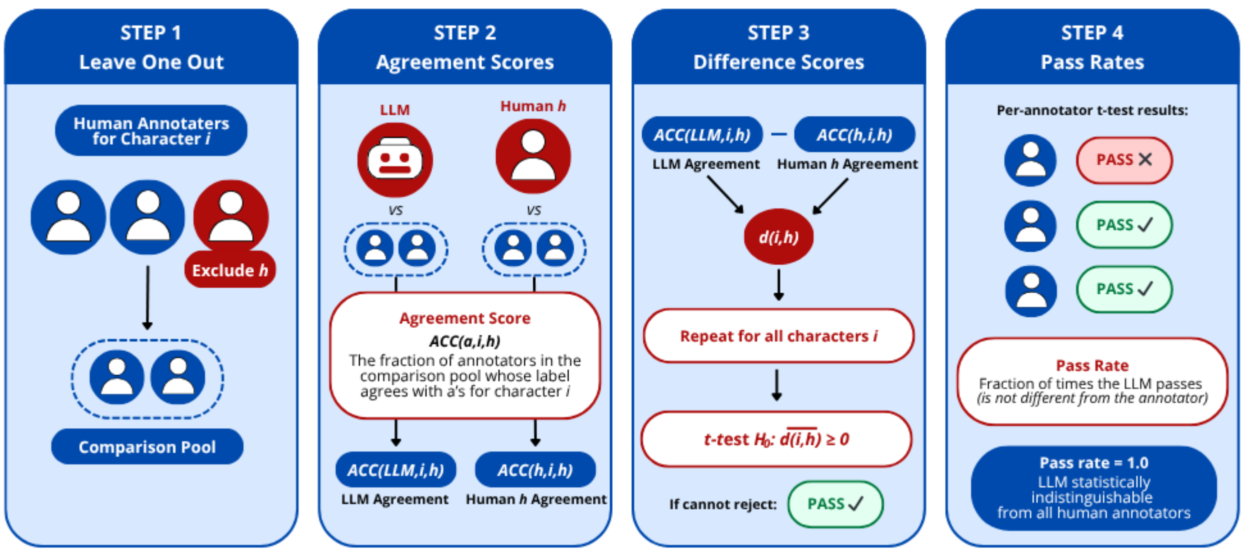


(c) Images: Feature Classification



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 transformer/GoEmotions model. Panel (c) shows the percentage of detected faces.

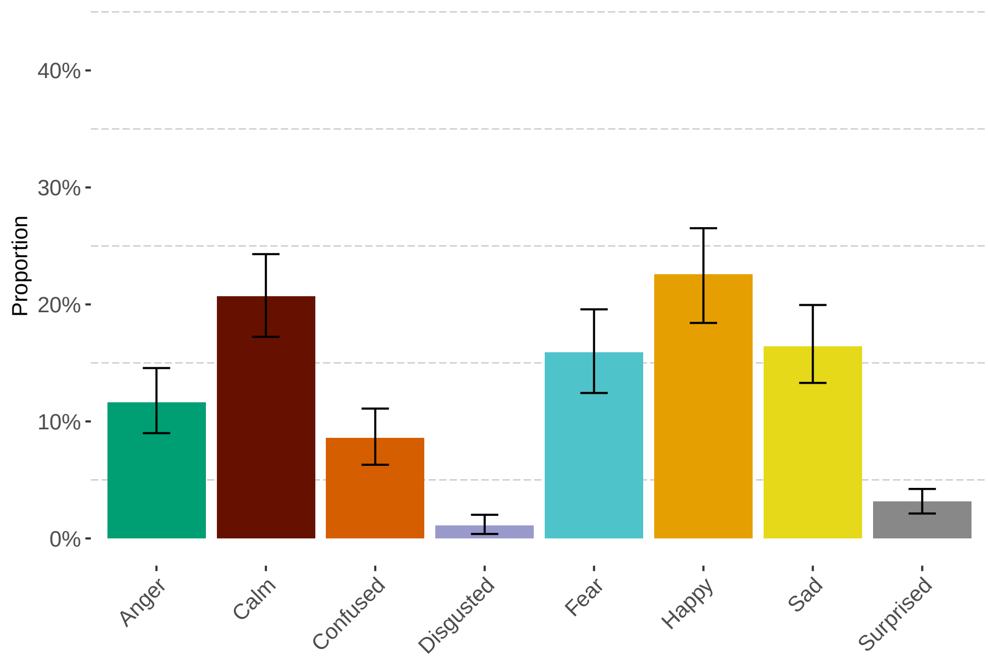
FIGURE A.7
Alternative Annotator Test Procedure



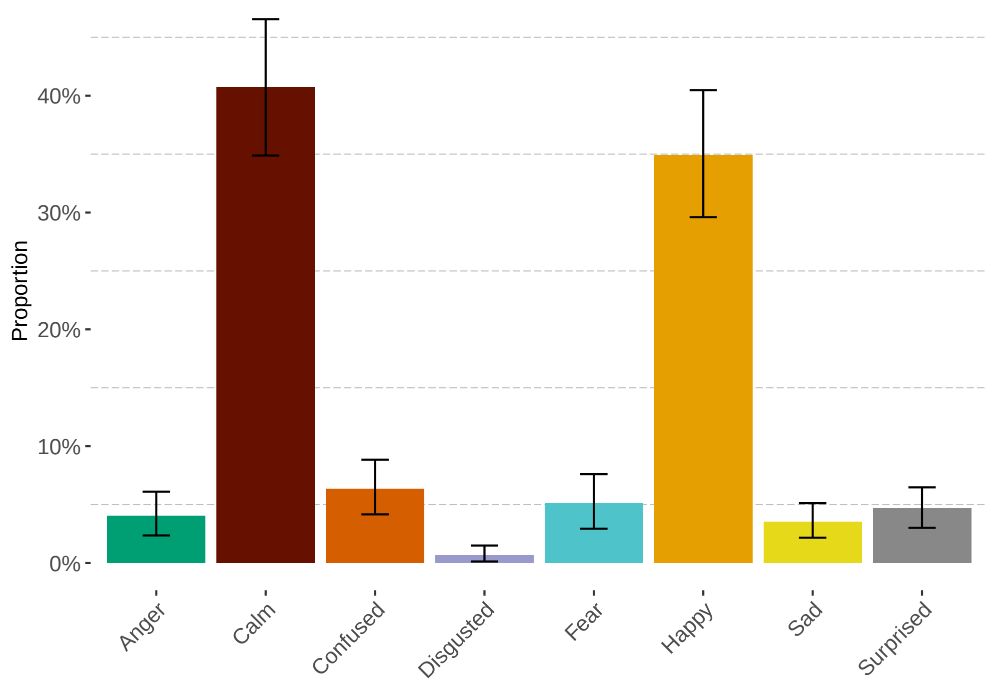
Note: This figures shows the procedure used to validate the output from the large multimodal models. The process involves assessing the model's reliability relative to human annotators in terms of its relative capacity to agree with other raters on the same set of examples.

FIGURE A.8
Simulated Distribution of Human Labels

(a) Text



(b) Images

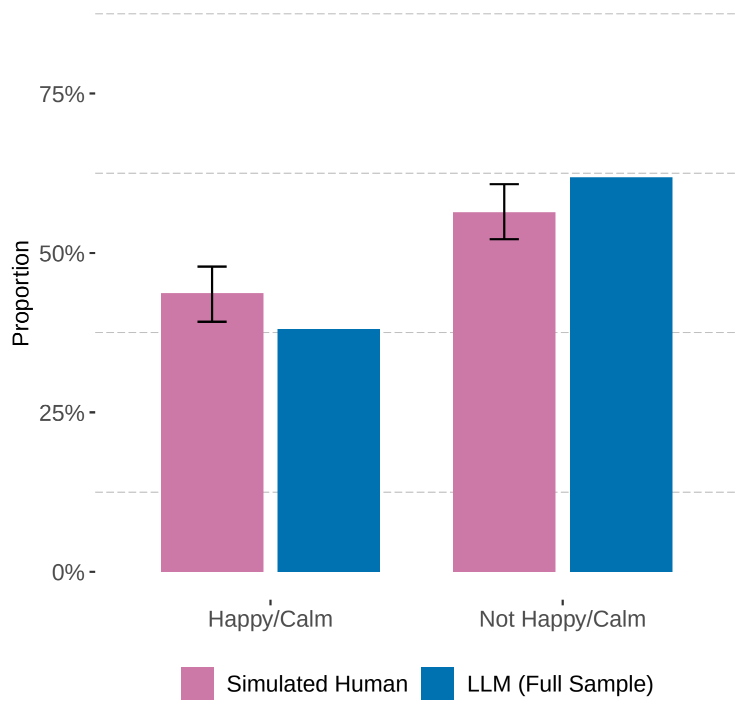


Notes: This figure shows the distribution of emotions predicted to be obtained by human labelers on the full sample. The simulation procedure is described in Section V.B. The top panel shows the distribution in text, and the bottom shows the distribution in images.

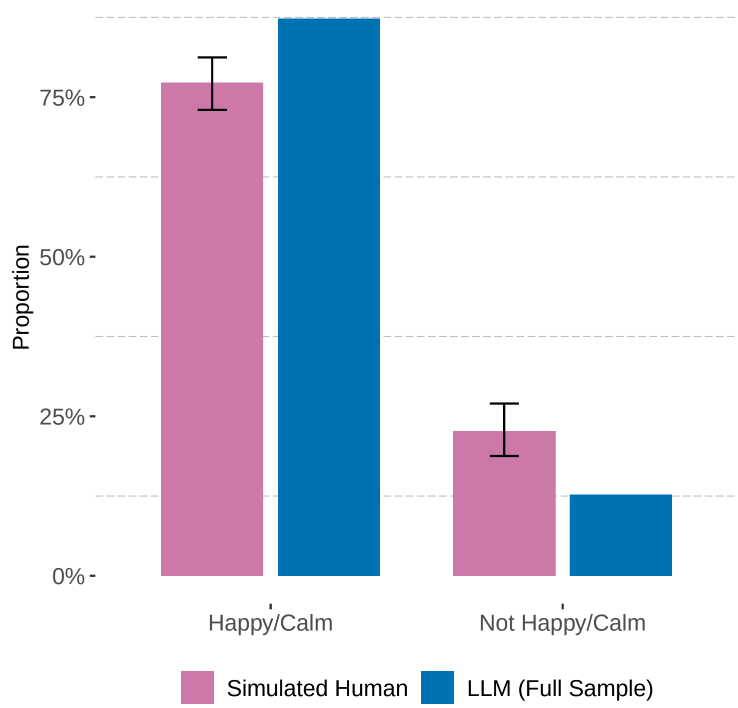
FIGURE A.9

Simulated Human Labels versus Models on Happy/Calm and Not Happy/Calm

(a) Text

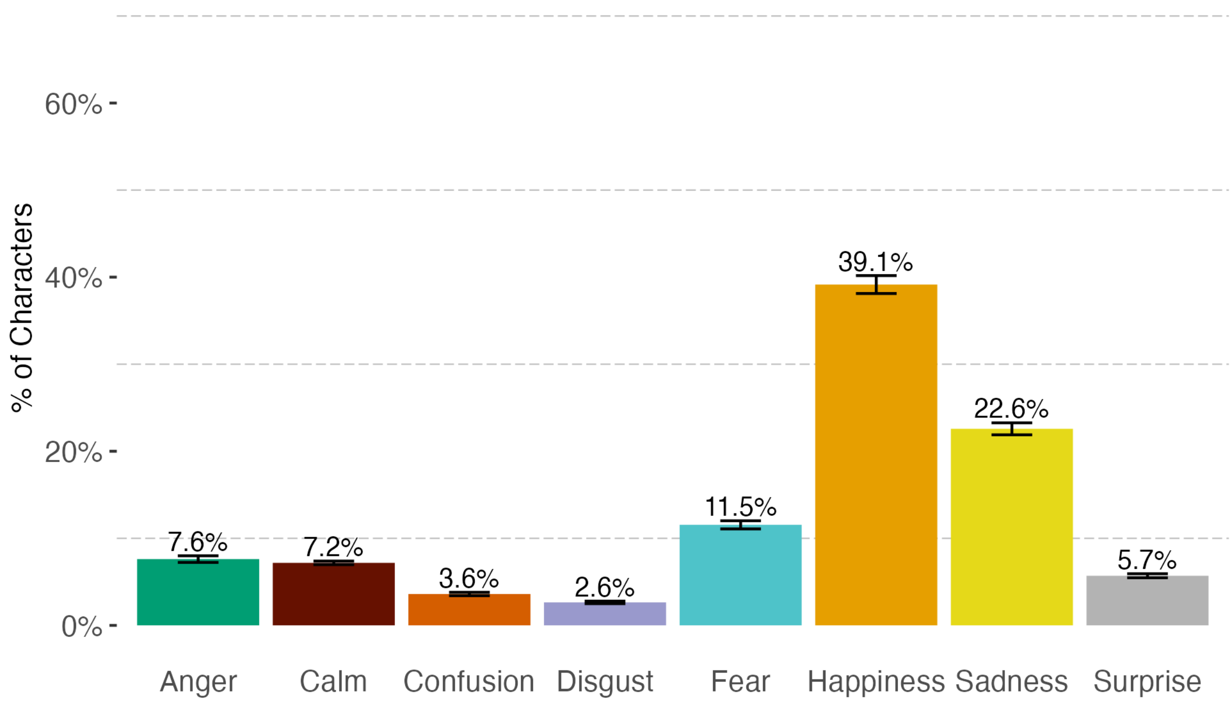


(b) Images



Notes: This figure compares the human and model predictions on the Happy/Calm and Not Happy/Calm split. The top panel shows the distribution in text, and the bottom shows the distribution in images.

FIGURE A.10
Overall Representation of High Confidence Emotions In Text



Note: This figure shows the aggregate displays of emotions in text when we subset to emotion labels that the model has high confidence in (≥ 0.8). Error bars depict 95% confidence intervals constructed from standard errors that are clustered at the book level.

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	absentminded	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.

TABLE B.3
Macro F1 Scores

Prediction	**Macro F1**	**Number of Examples**	**Exclusion Rate**
Panel A: Text			
Primary Emotion	0.289	588	35.0%
Emotion (Binarized)	0.467	588	11.7%
Is Human	0.862	564	0.7%
Gender	0.802	554	3.4%
Age	0.680	588	7.8%
Race	0.531	532	7.3%
Panel B: Image			
Primary Emotion	0.480	459	22.9%
Emotion (Binarized)	0.640	459	5.7%
Is Human	0.941	459	0.9%
Gender	0.821	459	2.6%
Age	0.782	368	3.8%
Race	0.685	305	12.5%
Skin Color	0.636	333	10.2%

Notes: This table reports macro F1 scores, which are obtained by computing the F1 score for each class and averaging over classes. The number of examples changes because raters may find demographic characteristics to be not applicable (i.e., the race of a non-human character). Examples are excluded if all three human raters disagree so that there is no modal rater. The high exclusion rate for text and image emotion labeling reflects the relative difficulty of this task.

TABLE B.4
Alternative Annotator Test

Prediction	**Avg Advantage**	**Pass Rate** ($q = 0.10$)	**Pass Rate** ($q = 0.05$)
Panel A: Text			
Primary Emotion	−9.3%	0.333	0.333
Emotion (Binarized)	−8.5%	0.333	0.333
Is Human	−0.8%	1.000	1.000
Gender	5.1%	1.000	1.000
Age	5.2%	1.000	1.000
Race	−11.3%	0.333	0.333
Panel B: Image			
Primary Emotion	−3.9%	0.500	0.833
Emotion (Binarized)	−6.3%	0.167	0.667
Is Human	2.3%	1.000	1.000
Gender	−1.1%	1.000	1.000
Age	−7.2%	0.333	0.500
Race	1.3%	1.000	1.000
Skin Color	−4.8%	0.500	0.500

Notes: This table reports the results from the alternative annotator test. The top panel shows the results for the text predictions, and the bottom panel shows the results for the image predictions. Binarized emotion is constructed from the primary emotion label by binning into Happy/Calm and Not Happy/Calm categories (along with a No Emotion/Unsure category).

C Methods Appendix

C.A Text Analysis

LMM Prompt Used for Emotion Identification Our primary text model is OpenAI's GPT-4o-mini, version 2024-07-18, accessed through the API in April 2025 with temperature set to 1.0 and the system prompt: "You are a knowledgeable and unbiased judge." For each page of OCR-extracted text, we submitted a standardized prompt which instructed the model to classify each character along with their age, gender, race, whether or not they were human, and any emotions expressed by that character. Emotion labels were recorded under three taxonomies: the six Ekman emotions, our eight-category MiiE taxonomy, and the 28-category GoEmotions taxonomy, each with a confidence score. The prompt restricted responses to predefined answer sets, allowed "unsure" and "not applicable" values for ambiguous cases, required "no emotion" for characters without expressed affect, and instructed the model not to assign emotions or attributes unsupported by the text. Hallucinations of emotions beyond simple morphological variants account for less than 1% of all observations, indicating high prompt adherence. The full prompt can be viewed at `https://github.com/miielab/emotions_paper`. This gives us a final dataset containing character-page level emotions.

Alternative Text Analysis Approach: Using a Lexicon-Based Approach 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 constructed an initial set of WordNet inputs, 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. The full lexicon can be viewed at `https://github.com/miielab/emotions_paper`.

In total, our lexicon contains 1,706 entries. More specifically, breaking this down into 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. This gives us a final dataset containing sentence level emotions.

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." This gives us a final dataset containing sentence level emotions.

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. However, we lose some of the transparency our lexicon-based method provides through insight into the words with which emotions are being conveyed.

C.B Image Analysis

We quantify emotional representation in illustrated pages of children's books using a two-stage, fully automated pipeline. We implement the pipeline on Google Cloud Vertex AI using Gemini 2.5 Pro Preview, model version `gemini-2.5-pro-preview-03-25`.

First, each scanned page 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. Pages identified as `TEXT_ONLY` are excluded from further analysis, whereas pages labeled `CONTAINS_IMAGE` advance to the next stage in the pipeline.

Next, for our illustrated pages, the model outputs character level data (e.g., character bounding boxes, demographic attributes, and emotion labels), using a temperature 0.0, a 10,000-token limit. Importantly, the character level output includes three distinct emotion codings (`emotion_6_options`, `emotion_10_options`, `emotion_28_options`).

Because the image prompt allowed for open-ended race predictions, we standardized the model output before using it in the analysis. In particular, the re-categorized variable for race (race_original) was constructed according to the following classification of the Gemini model's output.

- Not applicable: Occurs when the character is classified as 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