ORIGINAL ARTICLE

WILEY

Using structural topic modeling to understand ethnicity-related narratives

Rachel Jacobson[1] | **Dulce Wilkinson Westberg[2]** | **Edward Chou[3]** | **Moin Syed[3]** | **Sara J. Weston[1]**

[1]Department of Psychology, University of Oregon, Eugene, Oregon, USA

[2]Department of Psychology, University of California, Davis, Davis, California, USA

[3]Department of Psychology, University of Minnesota, Minneapolis, Minnesota, USA

Correspondence
Rachel Jacobson, Department of Psychology, University of Oregon, 1451 Onyx Street, Eugene, OR 97403, USA.
Email: rjacobs2@uoregon.edu

Abstract

Objective: Narrative identity is a promising approach for understanding the content of individuals' ethnic identities but can be limited by the time-intensive nature of human coding and the reliance on preestablished coding systems.

Background: The aim of our preregistered study is to elucidate the content of individuals' ethnicity-related experiences using a novel statistical approach.

Method: We applied structural topic modeling (STM), a natural language processing tool, to narratives written by an ethnically diverse sample of 1149 young adults about a moment they felt aware of their ethnicity.

Results: We identified 14 topics within ethnicity narratives and analyzed how each topic related to both the participant's ethnicity and the human-coded themes of agency and communion. For example, the topic *Gained perspective of structural inequality* was associated with greater agency, whereas *Peer dynamics* was associated with greater communion. *Ethnic/cultural celebration* was associated with both.

Conclusions: This study introduces STM as a useful tool for extracting topic content in narrative data and demonstrates how the multi-method assessment of ethnicity narratives provides greater insight into the content of ethnic experiences. These findings contribute to our understanding of contextualized aspects of personality, including the innovative ways we might examine them.

KEYWORDS

agency, communion, ethnic identity, narrative identity, STM

1 | INTRODUCTION

Research on identity development in general, and ethnic[1] identity in particular, has long emphasized the *processes* involved in developing a sense of identity, or the affective, behavioral, and cognitive strategies individuals use to better understand who they are and how they fit in the world (Galliher et al., 2017). In contrast, the *content* of ethnic identity—the actual experiences of what it means to hold a particular identity—has received less attention, even though variations in such contents are important for understanding the processes (Syed, 2015).

One challenge of understanding ethnic identity content is that this content is highly contextualized and not subject to any existing taxonomies and thus cannot be reliably assessed using rating-scale instruments. A potential solution is to use narratives of people's ethnicity-related experiences, through which respondents provide contextualized details about their lives while recounting their personal past (Syed, 2015). These narratives are almost

always human-coded, either using an existing protocol or inductively based on the data (Adler et al., 2017). This approach presents two challenges. Pragmatically, human-coding narratives is a difficult and time-consuming process. More substantively, the subjectivities associated with human coding may lead to overemphasizing some aspects of narratives or overlooking others. Recent advances in natural language processing have opened new doors for narrative researchers to efficiently capture topics within narratives using a bottom-up approach.

In the current study, we applied structural topic modeling (STM; Roberts et al., 2014) to narratives of ethnicity-related experiences to better understand ethnic identity content and its potential relevance for ethnic identity. Specifically, we examined how topics identified through STM varied based on self-identified ethnicity to understand diversity in participants' ethnicity-related experiences. We then examined the association between the prevalence of the identified topics and human-coded themes of agency and communion within the same narratives. Agency and communion are two motivational themes implicated in identity development (McAdams et al., 1996) and are considered relevant for navigating ethnicity-related experiences and expectations (Syed & McLean, 2022). Through this study, we demonstrate STM to be a useful tool for examining narrative content. Moreover, integrating human-coded themes with topic modeling can provide more nuanced description of ethnicity-related experiences and why they matter.

1.1 | Narrative identity and ethnicity-related experiences

Researchers interested in personal narratives related to ethnicity have used the *narrative identity approach* (Lilgendahl et al., 2018; Syed & Azmitia, 2010). Narrative identity is a person's internalized and evolving life story, integrating the reconstructed past and imagined future to provide a degree of unity and purpose (McAdams, 2001; McLean et al., 2020). This approach has been used to study personalities in general (e.g., McLean et al., 2020) but has also been applied to better understand individuals from marginalized groups, such as those defined by sexuality (Weststrate & McLean, 2023), gender (McLean et al., 2017), and disabilities (Adler, 2018). Narratives are believed to provide insights into aspects of personality that are not well-measured by or even absent from self-report surveys of personality (Dunlop et al., 2020; McAdams, 1995). Narratives are particularly well suited for highlighting the experiences of those who have been overlooked, ignored, or actively silenced (Fish & Syed, 2021).

Indeed, contextualized, lived experiences are not mere descriptions of peoples' lives; rather, varying domains of narrative identity content have been shown to be related to ethnic identity developmental processes (Syed, 2015). Individuals develop their ethnic identities through a complex interplay of personal experience and cultural context, and—in the context of life histories—participants can frame their life stories through their ethnic backgrounds. In other words, just as ethnic identity is constructed, so too is the telling of life stories, reflecting the complexity of ethnicity in the individuals' life. Thus, deciphering the content of ethnicity-related experiences is important insofar as it can provide real-life insights into the cultural experiences faced by people of diverse backgrounds.

Despite the promise of examining ethnicity-related narratives for content related to identity, the approach also comes with some challenges. Most narrative identity research involves human coders, who use a codebook outlining various themes and categories to guide the assessment of particular narrative themes of interest (e.g., Adler & Poulin, 2009; McAdams et al., 2004; Singer, 2004; see Syed & Nelson, 2015 for details). Human coding offers advantages, such as the flexibility to choose themes and the ability to compare studies using standard manuals for theoretical consistency (McLean et al., 2020). However, human coding can be labor-intensive, time-consuming, expensive, and limited to the constructs that are chosen for assessment. Specifically, when coding hundreds of narratives, time and resources may be saved by analyzing narrative content with a more automatic approach, one that captures granular details within qualitative responses. Indeed, new opportunities for text data have emerged with the advent of open-access approaches like topic modeling.

1.2 | Topic modeling as a useful tool for narrative identity research

In recent years, methods for analyzing text have significantly advanced, offering new tools for psychological research. One such method is topic modeling, which can be used to efficiently summarize and code a set of narratives. Topic modeling works by identifying clusters of co-occurring words that represent underlying topics (Wallach, 2006). Essentially, this method helps researchers understand the themes present in a large set of texts and how these themes relate to each other and to other variables of interest. Topic modeling can be compared to factor analysis, a technique more familiar to psychologists. Just as factor analysis identifies latent factors that explain the patterns in a set of observed variables, topic modeling identifies latent topics that explain

the patterns in a set of words across multiple documents. Each topic is characterized by a group of words that frequently appear together. The researcher then interprets these groups of words to assign meaningful labels to the topics. Topic modeling provides two key insights: (1) it identifies a coherent set of topics that can summarize the entire collection of narratives and (2) it estimates the extent to which each topic is represented in any given narrative. This method allows for a nuanced analysis of textual data, making it possible to explore the relationship between identified topics and other variables, such as psychological constructs or demographic characteristics. By incorporating topic modeling into their toolkit, psychologists can gain deeper insights from qualitative data, enhancing the richness and depth of their research findings.[2]

Topic modeling is considered an open-vocabulary analysis, meaning that topics are estimated from the narratives themselves rather than an external, preexisting dictionary (e.g., Linguistic Inquiry and Word Count [LIWC]; Tausczik & Pennebaker, 2010). In short, dictionary methods define latent topics a priori by assigning words to topic lists (e.g., a family list might contain words like "mom," "brother," and "cousin"). Topic modeling, by contrast, estimates topics via the observed clustering of words. A benefit is that, in contrast to pre-existing dictionaries, the open-vocabulary approach honors the words of the participants themselves; in other words, the open-vocabulary approach centers the language and experiences of marginalized groups.

Among available topic modeling algorithms, STM (Roberts et al., 2014) uniquely allows for the integration of covariates into the estimation of the topics and thus may be of particular interest to personality psychologists. Within narrative psychology, STM can be used to connect human-coding and topic modeling methods. For example, a researcher may code a set of narratives for previously validated themes—such as positive tone—and integrate these scores as covariates in an STM analysis. This allows the researcher to identify common content topics that arise from a specific narrative prompt and to determine whether some topics are more likely to include positive tone. An additional feature of STM analysis is the opportunity to examine the relationship of covariates to topic vocabulary. For example, a study of decision-making found that when discussing the concept of intuition, men were more likely to use the words "interest," "gamble," and "certain," while women were more likely to use "god," "middle," and "doubt" (Roberts et al., 2014). STM, like many forms of topic modeling, uses Bayesian estimation to derive its solution.

Overall, topic modeling is ideal for questions that involve extracting content and gaining a descriptive understanding of the data at hand. Topic modeling is well suited to summarize both broad (such as "Gene," "Evolution," and "Computer," for journal topics in *Science*; Blei, 2012) and narrow topics (such as "landing gear issues," "engine, oil pressure," etc., in aviation incident reports; Kuhn, 2018) and to quantify the relationship between narrator and narrative factors to these topics (Roberts et al., 2014). Topic modeling has been used to describe documents as short as tweets (Sterling et al., 2019) and as long as *Pitchfork* reviews (Light & Odden, 2017), making them suitable for narratives of varying lengths. Experts have not reached a consensus on the necessary number or length of documents for topic modeling. In general, we may assume that more narratives are needed when short in length. One study (using a different topic modeling algorithm known as Latent Dirichlet Allocation) found that the smallest number of essays needed to accurately recover the model parameters was 55, given a small set of topics (2–3) and essays of varying lengths (200, 250, and 500 words; Wiberg et al., 2021).

1.3 | The intersection of STM and narrative identity

Structural topic modeling analysis can help us understand the nuances of identity narratives, including identity narratives focused on ethnicity. Ethnicity is a salient social category in the United States (U.S.; Chae et al., 2021; Williams et al., 2020); therefore, capturing ethnicity-related experiences can reveal individuals' subjective beliefs about their place in the larger society. Past work—using human coders—identified content themes of connection to culture, awareness of cultural differences, discrimination, and underrepresentation in ethnic identity narratives (Syed & Azmitia, 2008, 2010). Moreover, narrative content varies based on ethnic group membership (Westberg, 2022; Yip et al., 2021). For example, Syed and Azmitia (2008) found that multiracial participants were more likely than monoracial participants to tell narratives about awareness of differences and connection to culture. Relatedly, Mitchell et al. (2018) found Latinx participants were more likely than other ethnicities to describe spirituality as a reason for nontypical ethnic experiences. As evident from these examples, the content of ethnic experiences may vary based on ethnic group membership and represent a distinct aspect of ethnic identity. Structural topic modeling may reveal additional features that further illuminate the content of ethnic identity. Thus, the first aim of the current study was to identify the topics that are extracted by applying STM to narratives of ethnicity-related experiences.

Even when a topic is equally present in narratives across groups, the words used to describe that topic may

differ. This is especially important when considering how participants come from different ethnic and cultural backgrounds, as words reflect the larger semantic systems of a narrator's culture (Thompson et al., 2020). Such differences in *topic vocabulary* provide additional nuance to the understanding of narratives. For example, when using an online peer-support platform for mental health, Asian American participants used fewer negative words compared to White, Black/African American, and Latinx participants (Loveys et al., 2018). To date, we are aware of little research that explicitly compares vocabulary within identity narratives and across ethnic groups. However—given prior work demonstrating the importance of even pronoun choice for revealing social standing (e.g., Kacewicz et al., 2014) and linguistic code-switching as a means of asserting identity (Loyd et al., 2023)—topic vocabulary is likely to hold important clues about ethnic identity. Thus, the second aim of our study was to examine how word choice differs across members of various ethnic groups.

Independent of ethnic background, a narrator's psychological profile may be gleaned from their narrative, primarily through coding for motivational themes. Two motivational themes, communion and agency, capture the need for connection to a community and establishing one's purpose, respectively (McAdams et al., 1996). These themes play a crucial role in shaping one's sense of self because they act as overarching motivational frameworks for structuring ethnic experiences in the context of balancing group expectations and personal aspirations (Bakan, 1966; Syed & McLean, 2022). For example, a person's sense of who they are can be based on their group memberships, if they actively reflect on their social groups (Tajfel & Turner, 1979). In this case, communion-related constructs—ethnic group belonging and cultural connectedness—positively affect psychosocial outcomes by buffering experiences of discrimination (Brondolo et al., 2009) and improving well-being (Webber & Waru-Benson, 2022). Agency, on the other hand, enables individuals to adhere to their ethnic group's norms and values and to redefine what their ethnicity means to them on a personal level. For example, students who are more individuated from, yet value attachment to, their family had higher college adjustment compared to those who did not (Kalsner & Pistole, 2003). These goals seem on the surface oppositional, yet they are interdependent, particularly within the context of people's experiences involving ethnicity (i.e., people can be both agentic and communal; McCabe & Dinh, 2016).

Agency and communion can co-occur within the same ethnicity-related narrative (McLean & Syed, 2015).

For example, an experience can be communal and agentic when the participant describes feeling cared for and accepted by other ethnic group members. Identifying ethnicity-related instances that empower and connect an individual is important for highlighting positive experiences to be fostered by communities. In contrast, other research suggests that accomplishment (agency-related) is achieved at the expense of communion (McCabe & Dinh, 2016). Understanding the relationship between ethnic identity content with agency and communion can elucidate in which contexts people feel valued, included, or ostracized (Yip et al., 2006) and the extent to which those contexts vary as a function of the narrator's ethnicity. Our third aim was therefore to understand the relationship of agency and communion to topic prevalence within ethnic identity narratives.

1.4 | The present study

In the present preregistered study, we integrate STM with the traditional narrative identity approach to provide further insight into the content of ethnic narratives and how they correspond with major motivational anchors of ethnicity-related experiences (i.e., agency and communion). In so doing, our methodological contribution is to introduce an additional, efficient way of analyzing content in narrative data, especially in larger quantities. In contrast, our theoretical contribution is to further elucidate the kinds of ethnicity-related experiences individuals navigate. In this study, we first use STM to unearth topics representing the content of ethnicity narratives (Aim 1). Next, to determine the extent to which topics derived from STM are overlapping or different across ethnic groups, we examine differences in topics based on self-identified ethnicity (Aim 2). Last, we explore relations between topics and human-coded themes of agency and communion (Aim 3).

2 | METHOD

In this study, we made use of existing data collected by the fourth author between 2005 and 2012.[3] Structural topic modeling and qualitative analyses were conducted by the first three authors and the last author who had not analyzed any portions of the data prior to preregistration. The methods, materials, exclusion criteria, and analysis plan of this study were preregistered before analyzing the data (https://osf.io/tbmd3). Ethical approval was obtained from the Institutional Review Board (IRB) at each data collection institution.

2.1 │ Participants and procedure

Participants were 1184 college students recruited from psychology department participant pools at two large universities in California and one large university in the Midwest. Participants completed a 50–60-min survey in a research lab on campus in exchange for course credit. From this sample, only participants who responded to the relevant narrative prompt were retained for analysis, resulting in a final sample size of 1149 participants (M_{age} = 21.14, range = 18–57; 64% women, 35% men, 0.5% nonbinary, and 0.5% unspecified).

For ethnic identification, participants reported their self-identified ethnic group membership using both close-ended and open-ended formats. The close-ended responses were of seven broad ethnic categories including Asian American, Black/African American, Hispanic/Latinx, White, Native American/Indigenous, Multiracial,[4] and Other/Noncategorizable. Participants who did not select a close-ended ethnic category were labeled "noncategorizable."[5] Within our sample, 42% participants identified as White, 31% as Asian American, 8% Latinx, 5% Multiracial, 4% Black, 1.6% as Native American/Indigenous, and 6.5% as Other/Noncategorizable. In recognizing the inherent challenges in all classification systems, including ours, we adopted a categorization strategy for participant ethnicity/race when their close-ended was left blank and their open-ended was not to ensure inclusivity and to better reflect the diversity of identities who are part of these pan-ethnic groups. See the supplemental materials (Table S1) for a table with self-identified ethnic group membership by pan-ethnic group.

2.2 │ Measures

2.2.1 │ Ethnicity-related narrative prompt

Participants were asked to respond to the following narrative prompt:

> Please describe a particular time in your life, either positive or negative, that you felt aware of your race/ethnicity when you were hanging out with one or more of your close friends. Tell us how old you were when this happened, where you were, whom you were with, what happened, how you reacted. Include details that would help us see and feel as you did.

Participants were asked to consider memories that included a close friend or family member to elicit salient experiences that involved other people. Narratives were 145 words on average (range = 4–628 words). One narrative (0.09%) was shorter than 5 words, 3 (0.26%) were shorter than 10 words, and 20 (1.7%) were shorter than 25 words. Given the comprehensive nature of our dataset, this singular occurrence of a 4-word narrative does not significantly affect the validity or the insights derived from our model. Therefore, we have opted not to adjust the STM model for this anomaly, as we believe it maintains the integrity and relevance of our findings.

2.2.2 │ Human-coded agency and communion

Narratives were coded for themes of agency and communion by three teams of two trained undergraduate research assistants (six coders total). Members of each team coded all narratives randomly assigned to that team; in other words, each team coded one third of the narratives from the total sample, resulting in approximately 383 narratives assigned to each coder (1184 narratives total). Coders were trained to rate themes of agency and communion using coding systems established by past researchers (McLean et al., 2019; https://osf.io/d2xts). First, coders familiarized themselves with the narrative identity approach by reading McAdams et al. (1996), which outlined themes of agency and communion. Then, coders rated ethnicity narratives for agency (e.g., autonomy, self-mastery, control) and communion (e.g., connection, unity, togetherness) on two separate 5-point scales ranging from 0 to 3 (0 = low agency/communion, 1.5 = does not contain agency/communion, 3 = high agency/communion; this was incremented as 0, 1, 1.5, 2, and 3 for each theme). Since not all participants followed the instructions to describe an experience with a close friend or family member, mention of a close friend or family member was not necessarily indicative of higher scores for communion.

Coding for each team was overseen by the first, second, and third authors, respectively. In line with conventions established in past research (Adler et al., 2017; Syed & Nelson, 2015), coders met weekly with the author leading their team, during which time they were given the chance to change or maintain discrepant codes. Weekly meetings fostered reliability and validity, allowing coders to share their thoughts, and come to an agreed-upon understanding of how agency and communion manifested within narratives. In addition, the first three authors convened in a biweekly meeting separate from the coders to discuss common problems and ensure the coding process was standardized across teams. Using benchmarks suggested by Koo and Li (2016), we found moderate interrater agreement after computing the intraclass correlation coefficient across the three

pairs of raters for agency (ICC = 0.71, 0.74, 0.90) and good interrater agreement for communion (ICC = 0.82, 0.75, 0.87) using two-way random effects models to assess absolute agreement (one model for each of three sets of raters and two themes).

2.3 | Analysis plan

2.3.1 | Aim 1: Identify ethnicity-related topics and label them using top words and example narratives

Topics were extracted using STM, a technique that estimates topics based on the distribution of words in a corpus of text (Roberts et al., 2019). Analyses were conducted in R (version 4.1.2; R Core Team, 2021) using the packages *stm* (Roberts et al., 2019) and *tm* (Feinerer et al., 2008). Topic modeling determines the likelihood of each topic in each document (narrative) on a probabilistic scale from 0 to 1. We included agency and communion as prevalence covariates, meaning the prevalence of topics is freed to vary across narratives as a function of these human-coded themes. Estimates of these relationships—between topic prevalence and human-coded themes—can be easily extracted and statistically tested. We also included ethnicity as a content covariate, meaning that the rate of word use within a topic is freed to vary across ethnic groups. While prevalence covariates are well suited for research questions like "Is topic 4 highly associated with emotional tone?", content covariates meaningfully examine questions like "Which words in topic 4 are exclusive to members of positive versus negative tone?"

We preprocessed the data by removing stop words (e.g., "the," "and"), numbers, special characters (e.g., "!"), and low prevalence words (appears once; Denny & Spirling, 2017).[6] Next, we determined the number of topics to extract by comparing models with varying numbers of topics (e.g., 2, 3, 4, etc.) on their model fit. We used four key metrics to compare model fit: (1) exclusivity,[7] (2) variational lower bound, (3) residuals, and (4) semantic coherence. These are recommended from prior literature (Fu et al., 2021; Weston et al., 2023). Exclusive words are those that are uniquely associated with a single topic; the exclusivity metric indexes the degree to which the words strongly associated with topics are exclusive to their topics. High exclusivity indicates each topic is clear and distinct from one another. Variational lower-bound indexes convergence. The estimation procedure of STM iterates until the variational lower bound is below a set threshold. All else being equal, lower values of variational lower bound indicate better fit. Residuals in this instance specifically refer to the dispersion of residuals. Recommendations for

the use of this metric vary—some suggest looking for local minima, while others claim that residual dispersion lower than 1 is sufficient to indicate a good fit. Last, semantic coherence indices are the degree to which high-probability words within a topic co-occur. This index has been found to positively correlate highly with human ratings of the coherence of topics, although there is some question as to the validity of this inference (Koltcov et al., 2019). Each of these metrics is limited in this capacity to identify a good-fitting and interpretable model. However, taken together, these four metrics can be used to identify a tractable number of candidate models for deeper investigation.

Frequently, these four metrics indicate that more than one model may provide sufficient fit to the data.[8] In those cases, we then compare estimated models in terms of topic overlap (i.e., how redundant the models are) and topic prevalence. Prevalence refers to the proportion of words within a single narrative that comes from the topic. Average topic prevalence across all narratives provides an index of the proportion of words in the entire corpus that come from the topic. Topics that are especially rare (low prevalence), have low utility, as they appear in a few narratives and may not generalize beyond a few key examples (Roberts et al., 2019). Overall, this comparison of multiple potential topic models allows us to determine the best set of models with little overlap and high utility. We calculated the four metrics across STM models from 3 to 20 topics. Once the narrowed number of topics was agreed upon, we estimated, labeled, and compared the candidate models.

We reported one optimal topic model in the final manuscript. Focusing on one model allowed for a more in-depth interpretation and optimization of the model's parameters and results. This approach facilitated a deeper understanding of the underlying themes, making it easier to draw meaningful insights into the relationship between ethnic experiences and agency and communion across ethnicities. The choice between a parsimonious model and a more descriptive one hinges on the research objectives. A simpler model with few topics is advantageous when the researcher aims to replicate the model to various datasets; it distills information into broader themes that are less likely to be data set-specific. On the contrary, a model with more topics may be ideal when the goal is to fully characterize a particular data set. This decision is much like the choice of the number of factors in a factor analysis. Broad factors are more likely to replicate across datasets but may lack predictive power compared to narrow factors or items and may overly aggregate similar but not identical constructs.

To identify robust topics found across the models and determine which topics are unique, we examined the overlap in topics across models with a correlation matrix

between the two and computed prevalence (the mean of each topic across all narratives). For robust topics, we looked for topics highly associated with one another (i.e., Topic 13 in the 14-topic model and Topic 1 in the 5-topic model). For unique topics, we looked for topics of high prevalence that did not overlap with topics across models (i.e., Topic 5 in the 14-topic model). This approach added conceptual value in that models with more topics have the benefit of capturing more granular and specific details but the limitation of including more topics that only describe a small set of narratives.

Among the STM output are two matrices. The first, the *beta matrix*, contains the term-topic probabilities. These represent the likelihood that a given term is produced by a given topic. The second, the *gamma matrix*, contains topic-document probabilities, which represent the share of a given document (narrative) that contains each topic. Together, these matrices can be used to evaluate the content of topics (from the beta matrix), the prevalence of topics (from the gamma matrix), and even the association of topic prevalence with narrator- and narrative-specific covariates (by connecting the gamma matrix with other sources of data). From these matrices, we extracted 10 topic-related keywords and five topic-related narratives. To label topics, co-authors independently looked at the 10 topic-related keywords and five key narratives from the data set most likely generated from each topic. Then, a meeting took place during which the authors came to a consensus on topic labels using narrative content and keywords. Strategies for labeling topics are thoroughly described in Weston et al. (2023). We would like to acknowledge that our interpretation of the topics may have been influenced by our backgrounds, experiences, and identities. The authors represent two graduate students, one postdoctoral researcher, and two university faculty. Regarding ethnicity, members of the research team identify as Hispanic/Latina ($n=1$), Multiracial ($n=2$), White ($n=1$), and Asian ($n=1$).

2.3.2 | Aim 2: Determine whether the content of the STM topics changes as a function of participants' ethnicity

To assess whether the content of STM topics changed as a function of participants' ethnicity, we examined self-identified ethnicity as a content covariate. Notably, there is no significance test for these associations; thus, the evaluation of Aim 2 was qualitative in nature. Instead, this covariate indicates which words for a given topic were most frequent within and exclusive to each group. For this aim, we provided highly frequent and exclusive words for each topic and each group.

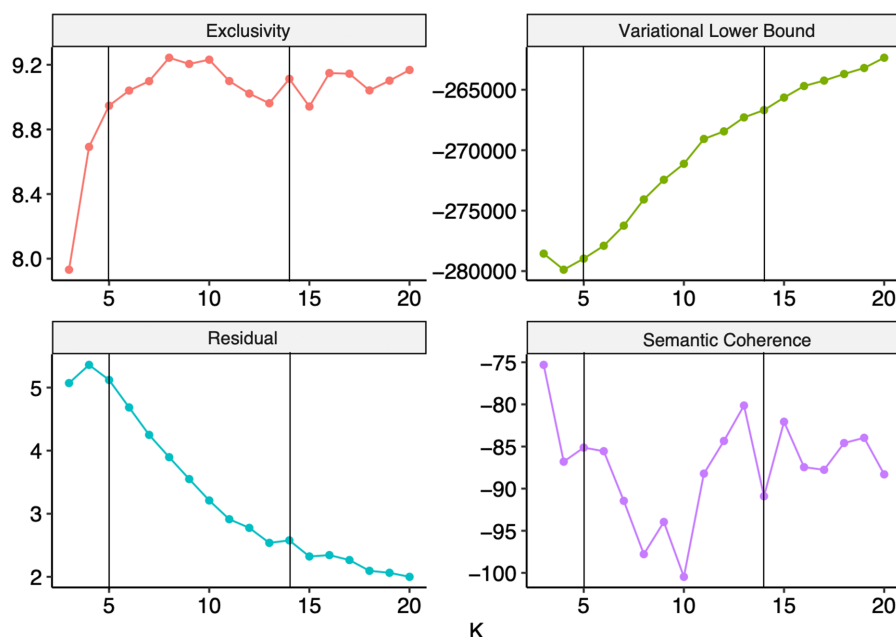## 2.3.3 | Aim 3: Examine the relationship between extracted topics and themes of agency and communion

To address Aim 3, we examined the parameter estimates associated with the human-coded agency and communion covariates of topic prevalence. This allowed us to detect which topics reflect themes of agency and communion. Model output includes coefficient estimates (and standard errors) that communicate the association of topic prevalence to human-coded agency, communion, and their interaction.
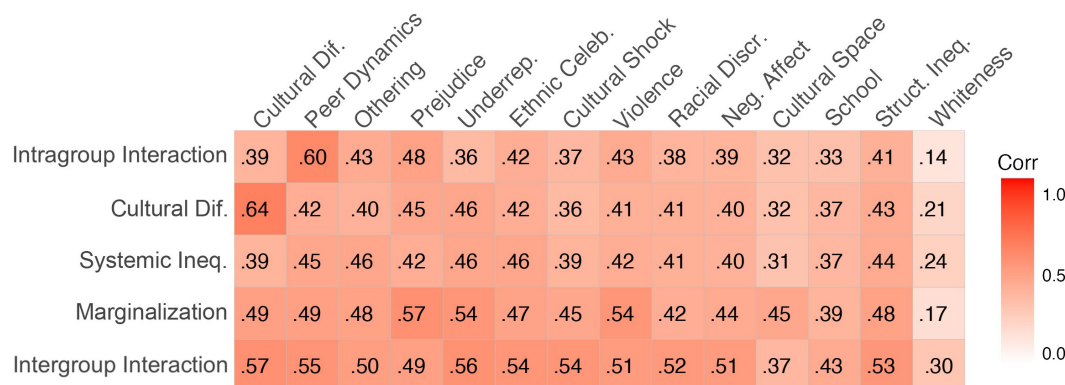
3 | RESULTS

3.1 | Aim 1: Identify/label STM topics

To determine the ideal models, we looked for models with high exclusivity, low residual value, high semantic coherence and high variational lower bound. Based on these standardized metrics, we identified two topic models: a 5-topic model and a 14-topic model (see Figure 1). The 14-topic model had a lower residual, higher exclusivity, and higher variational lower bound than the 5-topic model, indicating higher overall topic quality. As described in the Methods section, each author of this paper independently reviewed and labeled the 5- and 14-topic models based on frequent and exclusive words for each topic, words associated with each topic by ethnicity, and five highly associated narratives, which are strongly linked to the specific topic, as determined by the topic assignment probabilities and the thematic content of the documents themselves. After labeling the topics, we found that the 14-topic model was more nuanced than the 5-topic model in that the topics covered a broader range of concepts and with more specificity. (see supplemental materials for more detail about the 5-topic model; Table S2).

After labeling the STM topics, we examined correlations between topics across the two models to determine thematic overlap (see Figure 2). Correlations between topics were generally high, an average of $r=0.43$, indicating substantial overlap between them. Importantly, no correlations were above 0.70, indicating that no topic was replicated across the two solutions. Instead, we take this pattern of results to suggest that topics in the 14-topic model likely represent subtopics or more specific instances of topics in the 5-topic model. We observed that the topics we labeled *Awareness of cultural difference* in both solutions had a positive association of $r=0.64$. Another example, the topic we labeled *Peer dynamics* in the 14-topic model had a positive association of $r=0.60$ with the labeled topic *Intra-intergroup interaction/tension*

FIGURE 1 The four standardized metrics used to determine candidate models with the 5-topic and 14-topic models highlighted as vertical lines and the number of topics on the *X*-axis.

FIGURE 2 Correlations between the 14-topic model and the 5-topic model.

in the 5-topic model. Overall, the 5-topic model exhibited a strong association with the 14-topic model. Given that both models exhibited a significant degree of topic overlap, we report results from only one model to avoid redundancy in interpreting the similar underlying themes they capture. For the rest of the manuscript, we describe the 14-topic model for its higher overall topic quality and the degree of overlap across the two models.

Table 1 illustrates the topics in the 14-topic model by prevalence, along with key narratives. After determining the 14-topic model, we explored the ethnicity-related topics. STM topics varied in prevalence, which is the proportion of the word documents associated with each topic. *Awareness of cultural differences* and *Peer dynamics* were the most prevalent across the narratives (see Table 1). *Othering* and *Reactance to prejudice* were the third and fourth most prevalent across narratives. The least prevalent STM topics were *Whiteness*, *Gained perspective of structural inequality*, and *Ethnic stereotypes at school*.

We examined bivariate correlations between topics in the 14-topic model to determine whether any co-occurred and found there to be little to no association between topics within the model (the range of correlations from −0.13 to 0.10; see supplemental materials (Figure S2) for the correlation matrix). In what follows, we detail each topic by reporting the keywords and describing highly associated narratives. See the supplemental materials (Figure S1) for the high probability words to occur in each topic. The topics are ordered by highest to lowest prevalence from 1 to 14.

3.1.1 | Topic 1: Awareness of cultural difference

Topic 1, labeled *Awareness of cultural difference*, describes the interaction of cultures either intra or interpersonally, where the individual learns the significance of their ethnicity. Compared with Topic 3 (*Othering*), many of

TABLE 1 Labeled structural topic modeling topics in the 14-topic model with exemplary narratives and prevalence represented as the mean (and median) gamma probability of documents.

Number	Label	Sample narrative excerpt	Topic prevalence
1	Awareness of cultural difference	When I was with other nonnative Chinese/Taiwanese teenagers, I realized that my peers were more characterized by who they actually were and not by their ethnicity, unlike in America, where I was—I am assuming—subconsciously labeled as "Asian"	0.14 (0.02)
2	Peer dynamics	I felt very conscious of having brown skin where nobody else did and it kind of made me wonder if the kids were trying to make me feel bad or if they were just being annoying	0.14 (0.02)
3	Othering	I knew I was different from them but I just never noticed it because they were so friendly and nice to me	0.09 (0.01)
4	Reactance to prejudice	I turned around, and even though I was not really yelling, I told them off. I said that how could they say that we were considered as stuck up when they too had a table full of ONLY Hispanic girls at their table?	0.09 (0.01)
5	Awareness of underrepresentation	It is a predominately white school with about a quarter of Natives going there like us, two or three black people, one Asian kid, so you see, there really is not diversity among the student body	0.08 (0.01)
6	Ethnic/cultural celebration	The conference was spread over 3 days and included workshops pertaining to Filipino culture, an entertaining variety show, and a club afterset	0.08 (0.01)
7	Experience of cultural shock	Living back home was a huge culture shock. I felt myself so caught up to the American lifestyle, however, when I was living back home to where I belonged I felt that a huge part of my life was missing	0.08 (0.01)
8	Violence	It turned out that the crazy black girl assaulted the Indian lady very harshly right before me for no reason, too. There were several white guys who saw the whole situation around us, but they did not take any action	0.08 (0.01)
9	Contact with racial discrimination	When we were walking to our car, two African American men were standing close by it. My two friends panicked and made us stall for a while until the two men left	0.06 (0.01)
10	Negative affect	I could not do anything as an individual because the whole community was scorned upon by a lot of people. Of course, there were people who understood the situation, but there were those who absolutely despised us too	0.05 (0.01)
11	Lack of belonging in cultural space	I brought up the topic of LGBTA people to my mom and she said to me, only Americans have that, we Hmong people do not have LGBTA people, we dont even have a word for it. That really broke my heart. I felt really alone	0.04 (0.01)
12	Ethnic stereotypes at school	From everyones' perception, Indians are never good at soccer and so no-one really considered putting me on the team	0.03 (0.00)
13	Gained perspective of structural inequality	Of course, the education system was lacking, focus was primarily on reading/writing in school while trying to keep up with the other counties and their education systems. In her situation, there was nothing restraining her education	0.02 (0.00)
14	Whiteness	I proposed the idea that class is more important than race in the city music scene, and my guitarist countered by suggesting that the reason that I booked the shows in our band was due to my race, being that he was black and I was white	0.02 (0.01)

Note: Sample narratives have not been edited by the authors.

the narratives in this topic were neutrally valenced. The highly associated narratives depicted contrasts between Chinese/Taiwanese and American cultures. The top words were "else," "integrated," "blended," "lecture," "others," and "odd."

3.1.2 | Topic 2: Peer dynamics

Topic 2, labeled *Peer dynamics*, describes exposure to ethnic/cultural norms from peers. The keywords were "latino," "mattered," "sixth," "saturday," "defensive," "cook,"

and "friends." The highly associated narratives for Topic 2 included both positive and negative experiences. The example in Table 1 presents a negative experience where the individual, surrounded by other kids, felt aware of having brown skin.

3.1.3 | Topic 3: Othering

Topic 3, which is labeled *Othering*, features narratives focused on feelings of confusion or frustration that stemmed from experiences of oppression, violence, or alienation. The highly associated narratives include racial profiling (being pulled over by the police) and rejection of one's heritage from a perceived in-group. The top keywords were "called", "slurs", "hit", "point", and "throw". While the representative narratives for Topic 3 reflected feeling othered, not all were due to explicit acts of discrimination or prejudice and were instead sometimes characterized by awareness of difference (see Table 1).

3.1.4 | Topic 4: Reactance to prejudice

Topic 4, labeled *Reactance to prejudice*, features the following top words: "story," "disappointed," "furious," and "roommate." The key narratives described experiences of prejudice and related reactions, including disappointment and frustration, by an out-group member. For example, in Table 1, the narrative described prejudice from an out-group member whom the participant deemed hypocritical.

3.1.5 | Topic 5: Awareness of underrepresentation

Topic 5, labeled *Awareness of underrepresentation*, reflected individual feelings of underrepresentation, either numerically and/or by society. The top words were the following: "fought," "people," "verbal," "gifted," and "tastes." In Table 1, the representative narrative includes an individual of Native American heritage who described their school with a nondiverse student body.

3.1.6 | Topic 6: Ethnic/cultural celebration

Topic 6, labeled *Ethnic/cultural celebration*, involved organized celebrations of one's culture. There was often a feeling of connection, and the highly associated

narratives took place in organized cultural events. The top keywords were "performers," "prominent," "dances," "relatives," and "annual." The highly associated narratives took place at a conference held by the Association of Filipino-Americans, a Hispanic Scholarship Consortium, a Vietnamese new year community event, a Japanese community center, and a Filipino debutante ball.

3.1.7 | Topic 7: Experience of cultural shock

Topic 7, labeled *Experience of cultural shock*, reflected unfamiliarity in navigating particular cultural contexts. The representative narratives comprise positive and negative experiences (i.e., "I was very open to meeting and interacting with other ethnicities. I wasn't sure how other ethnicities would respond to me, but I was hoping to make friends" and the excerpt in Table 1, where the person returning to Egypt describes a culture shock in a positive way). The keywords were "american," "roommate," "egypt," "world," "begin," and "complete."

3.1.8 | Topic 8: Violence

Topic 8, labeled *Violence*, included experiences of anger and frustration (such as aggressive language and physical altercations). The highly associated words were "street," "walking," "guys," "innocent," and "died." In Table 1, we present a representative narrative, where the participant and friend were physically assaulted by a woman.

3.1.9 | Topic 9: Contact with racial discrimination

Topic 9 was labeled *Contact with racial discrimination* and involved learning about or experiencing racial inequality. This includes narratives that contained mention of either witnessing or receiving discrimination. The highly associated narratives described discrimination against Black/African American participants and were primarily from the perspective of White individuals. The top keywords were "african," "americans," "classroom," "trash," "crossing," and "slaves."

3.1.10 | Topic 10: Negative affect

Topic 10, characterized by its globally negative language, is labeled *Negative affect*. The highly associated words were

"unkind", "upset[i]," "again," "helping," and "glendale." The keywords were more reflective of negative affect than the highly associated narratives. However, an example narrative included a participant of Asian American descent who felt "scorned" and "despised" by members of the White community.

3.1.11 | Topic 11: Lack of belonging in cultural space

Topic 11, labeled *Lack of belonging in cultural space*, involved the individual being in an out-group's environment (e.g., home of a family friend, friend's house, unfamiliar neighborhood) and feeling hurt, sad, or isolated. This was distinct from Topic 3 (*Othering*) in that narratives containing this topic mentioned the physical space of another person (i.e., a friend's neighborhood and house), whereas Topic 3 narratives contained diffuse mention of marginalization. However, the top keywords were noninterpretable and included: "interesting," "holds," "debate," "culture," and "employee."

3.1.12 | Topic 12: Ethnic stereotypes at school

Topic 12, which is labeled *ethnic stereotypes at school*, features experiences of stereotyping that occur in a school environment. The highly associated words were "studies," "teammates," "gathering," "soccer," and "merely." The key representative narratives for this topic reflected a heightened awareness of stereotypes. The highly associated narrative describes a high schooler who was stereotyped on the basis of their ethnicity as a bad soccer player and defied expectations by scoring a goal. Another narrative was about an elementary school student of Latinx descent who was placed into a "special" bilingual studies class despite speaking fluent English.

3.1.13 | Topic 13: Gained perspective of structural inequality

Topic 13, labeled *Gained perspective of structural inequality*, included encounters with inequality at the structural level. Highly associated narratives involved critical reflection, a form of critical consciousness that captures the recognition of systems of inequality and desire to advocate for social change (Watts et al., 2011). The keywords were "perspective," "writing," "analyzing," "lately," "ruin," and "immigrants."

3.1.14 | Topic 14: Whiteness

Topic 14, labeled *Whiteness*, revolves around eye-opening experiences of White-conforming norms and privilege. The associated keywords were "class," "opener," "race," "visit," "minority," "ancestry," and "white." The representative narratives elucidate instances of white privilege.

3.2 | Aim 2: Topic vocabulary by participants′ ethnicity

Table 2 contains the group-specific vocabulary words by ethnicity for all topics. These refer to words that are uniquely used by a specific group when discussing a topic. In some cases, vocabulary words simply reflect the names of the ethnic groups themselves (such as "Irish") but in other cases may relate to disparate experiences. For example, Latinx participants uniquely used the word "illegal" when discussing prejudice, likely reflecting politically-influenced stereotypes about race and immigration. Importantly, group-specific vocabulary words are distinct from high-probability words, which are likely to appear for a specific topic regardless of the narrator's ethnicity and are central to defining the essence of a topic based purely on word frequencies within that topic. As an example of group-specific words, for *Ethnic/cultural celebration*, the unique vocabulary of participants of Asian American descent included the words: "vietnamese," "persian," "ceremony," "keynote," and "skits." This is in contrast to Black/African American participants, the unique words were "michelle," "rally," "adult," "invite," and "protested." This finding suggests that types of organized events (i.e., ceremony versus rally) differed based on self-identified ethnicity.

To further illustrate the differences in topic content by ethnicity, Figure 3 has the highest probability words for *Awareness of underrepresentation* across Black/African American and Latinx participants. For Latinx participants, high probability words revolved around specific reactions (i.e., complained, prejudice, assumptions, insulting) and ethnicities/races (i.e., Mexican and Asian-American). For Black/African American participants, the high probability words were related to physical objects (door and clothes), occupations (teachers and professors), and places (Africa and Ethiopia).

3.3 | Aim 3: Agency and communion-related topics

We examined the relationship between the STM and human-coded themes to determine the agency and

TABLE 2 The uniquely associated words with each topic across the seven ethnic groups.

Topics	Black/African American	Asian	Latinx/Hispanic	Native American/ Indigenous	Multiracial	White	Noncategorizable
Cultural difference	Scared, vanished, cheek	Store, soup, aunt	Venezuelan, old, send	Phrase, chicks, exchange	Frat, jokes, stereotyped	Bus, minorities, scholarship	Sikh, bothers, interests
Peer dynamics	Music, puzzled, sleeping	Disrespectful, listing, meat	English, napa, crushed	Lions, suburban, white	Nights, arms, assimilate	Jokes, Persian, hurtful	Eastern, Phillipino, extreme
Othering	Spirit, bus, secluded	Game, teams, song	Displayed, desk, Chinese	Patrol, plate, unit	Melody, Dallas, rural	Teenage, awestruck, names	Old, bright, schoolmates
Prejudice	Proof, paper, crushed	Vietnamese, mother, lady	Bright, sad, illegally	Hawaii, paralyzed, unwelcome	Self-esteem, send, disbelief	Assignment, Russian, kiss	Confusing, shoved, because
Underrepresentation	Shoved, scores, environment	Esl, five, tan	Letters, side, signed	King, longest, crush	Transferred, worded, pursue	Read, dominated, songs	Proof, disheartened, rural
Ethnic celebration	Rally, invite, protested	Ceremony, keynote, skits	Principle, listening, afterschool	–	Middles, awestruck, hostile	Irish, joy, boards	Save, signed, surprise
Cultural shock	Schedule, option, interacting	Seat, institute, row	Mexican-American, Arabs, independence	Reported, Caucasian, switch	Christmas, rituals, celebrated	Birth, slang, supportive	Spicy, Egyptian, Arabic
Violence	Guard, matched, drug	Ran, crowded, crazy	Dared, ring, assist	Bothering, happen, shoving	Jews, holocaust, please	Car, Latino, attire	Hopes, chuckled, ruin
Racial discr	Stress, crisis, schoolmates	Slurs, woman, hurtful	Flight, plate, grandpa	Guess, camp, atmosphere	Quit, willingness, defense	Team, game, gangster	Nigeria, secluded, room
Negative affect	Paused, matters, report	Hmong, expensive, folks	Friends, apologized, represented	Military, removed, nights	Disheartened, involvement, hateful	Thrown, apartments, types	Pledge, challenge, secretary
Cultural space	–	Pork, dogs, Taiwanese	Nostalgic, resolution, conflict	Saint, pleased, holiday	Insight, left, restaurant	Speak, claiming, confusion	Blushed, intentionally, funny
School	Middles, unwanted, color	China, social, letter, shocking	Spanish, teaching, fluent	Innercity, more, fights	–	Jewish, observing, social	Overheard, assist, scores
Structural Ineq	Blind, buddy, dared	India, roommate, Pakistan	Immigration, benefits, illegally	–	Shit, Phillipino, intentionally	Relieved, faith, questioning	Shyed, ocean, serve
Whiteness	Class	Japanese, dads, ten	Challenge, exams, Latina	Proposed, effecting, disagree	–	Norwegian, Scandinavian, session	–

Note: The blank cells indicate that there are not sufficient narratives by ethnicity for a meaningful association between the words and that topic.

FIGURE 3 High probability words with awareness of underrepresentation between Black and Latinx individuals. The words larger in size indicate a higher probability to occur.

communion-related topics. These relationships were estimated within a regression framework; we report here the slope coefficients[9] indexing the relationship between topic prevalence and theme (see Figure 4). High levels of agency were positively associated with topics *Awareness of underrepresentation* (unstandardized $b = 0.02$, $p = 0.020$), *Ethnic/cultural celebration* ($b = 0.05$, $p < 0.001$), *Gained perspective of structural inequality* ($b = 0.01$, $p = 0.030$), and *Lack of belonging in cultural space* ($b = 0.01$, $p = 0.034$). Agency was negatively associated with *Othering* ($b = -0.02$, $p = 0.004$), *Reactance to prejudice* ($b = -0.02$, $p = 0.011$), *Violence* ($b = -0.04$, $p < 0.001$), and *Awareness of cultural difference* ($b = -0.02$, $p = 0.046$).

Communion was positively associated with *Awareness of underrepresentation* ($b = 0.02$, $p = 0.015$), *Ethnic/cultural celebration* ($b = 0.06$, $p < 0.001$), and *Peer dynamics* ($b = 0.02$, $p = 0.026$). Communion was negatively associated with *Othering* ($b = -0.02$, $p = 0.001$), *Reactance to prejudice* ($b = -0.03$, $p < 0.001$), *Violence* ($b = -0.03$, $p < 0.001$), and *Awareness of cultural difference* ($b = -0.02$, $p = 0.004$). Notably, both agency and communion were negatively associated with *Othering*, *Reactance to prejudice*, *Awareness of cultural differences*, and *Violence*.

In examining the relationship between agency, communion, and the topics, we included the interaction between agency and communion. This interaction was significant in predicting the prevalence of *Ethnic/cultural celebration* (unstandardized, $b = 0.05$, $p < 0.001$). Figure 5 shows the effect of agency on topic prevalence at different levels of communion. The prevalence of *Ethnic/cultural celebration* was positively associated with agency among high communion narratives. When communion is low, there was no significant relationship between *Ethnic/*

cultural celebration prevalence and agency. There was a significant interaction effect for *Peer dynamics* ($b = -0.03$, $p = 0.001$). In Figure 5b, agency was negatively associated with the topic *Peer dynamics* when communion was high, but unassociated with this topic otherwise. *Experience of cultural shock* ($b = 0.02$, $p = 0.040$) and *Lack of belonging in cultural space* ($b = -0.01$, $p = 0.010$) also had significant interaction effects; these effects are reported in the supplemental material (see Figure S3). Overall, the presence of communion modified the relationship between agency and topic prevalence across these four topics.

As a descriptive analysis, we tested whether the motivational themes differed across ethnicity, and we found that there was no significant difference across groups for communion ($F(5) = 1.40$, $p = 0.220$, $\eta^2 = 0.01$). We did find that agency significantly differed based on ethnicity ($F(5) = 3.07$, $p = 0.009$, $\eta^2 = 0.01$). However, after using a Bonferroni correction, none of the pairwise comparisons were significant. In Table 3, we ran ANOVAs to compare topic prevalence across ethnicity. We found that across the STM topics, topic prevalence varied across ethnicity only for *Cultural shock* ($F(5) = 3.30$, $p = 0.006$, $\eta^2 = 0.01$). Following the Bonferroni correction, Asian and White individuals significantly differed in the prevalence of describing experiences of cultural shock ($t(1125) = 4.00$, $p = 0.001$, $d = 0.28$). Native American/Indigenous participants were excluded from the analyses examining group differences, as <50 participants identified with this group.

4 | GENERAL DISCUSSION

The current study applied STM to narratives of ethnicity-related experiences to better understand ethnic identity
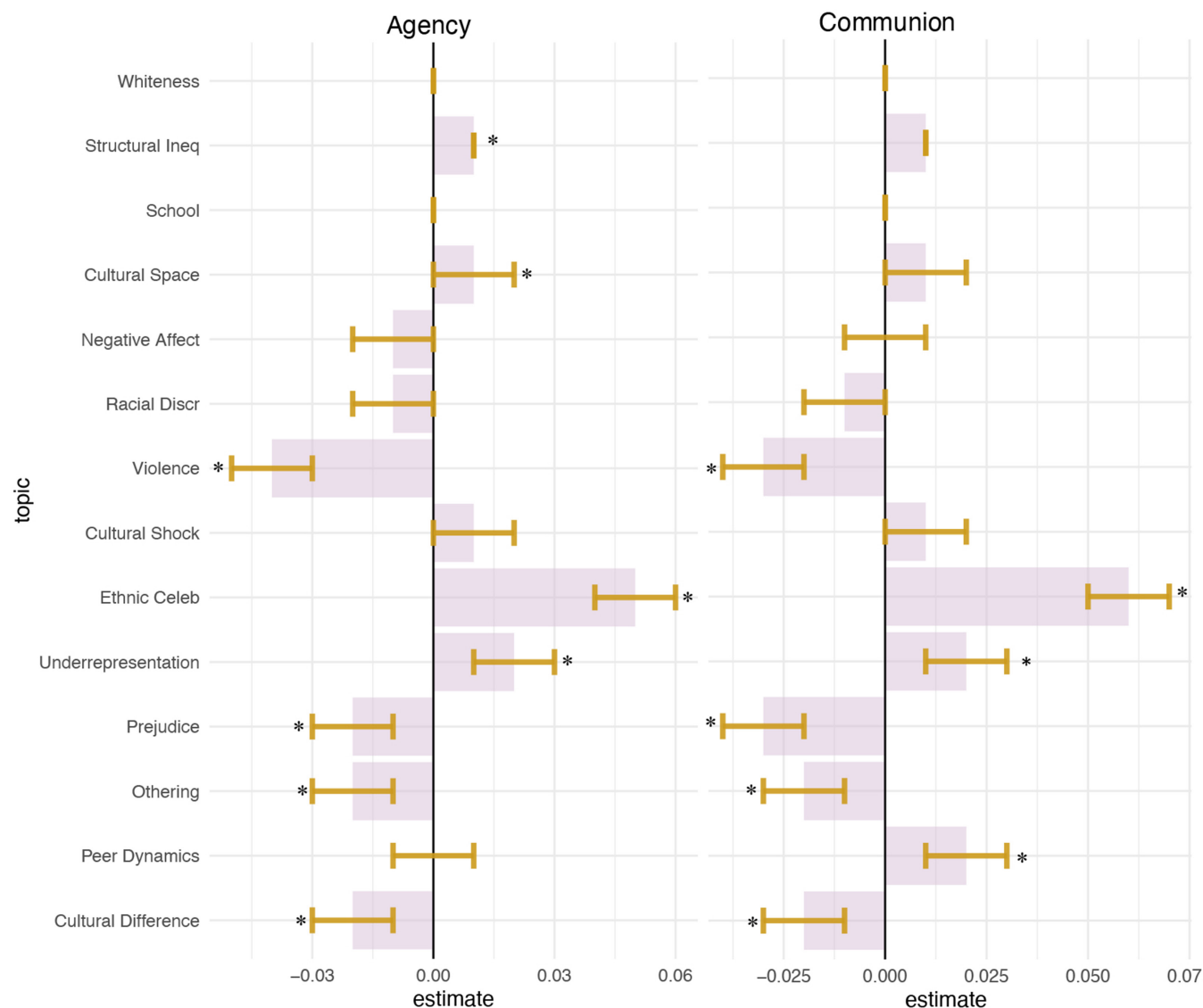
FIGURE 4 The estimates between the structural topic modeling topics with agency and communion.

content. We found 14 major topics of discussion that provided insights into lived experiences of ethnicity. The most prevalent topics were *Awareness of cultural difference*, *Peer dynamics*, *Othering*, and *Reactance to prejudice*. One topic, *Experience of culture shock*, was more prevalent in narratives written by Asian American participants compared to White participants; however, there were no other differences in topic prevalence across ethnic groups. In addition, topics varied in the degree to which they were agentic and communal, with *Ethnic/cultural celebration* being the most prevalent among narratives rated high in agency and communion. In what follows, we discuss (1) the value of incorporating STM into the study of ethnic identity and narratives, (2) the unique topics uncovered and their relationship to ethnicity, and (3) the relationship of STM-derived topics to human-coded themes of agency and communion.

4.1 | Using STM to understand ethnic narratives

Prior work examining narratives about ethnicity has identified broad content themes: (1) awareness of difference, (2) awareness of underrepresentation, (3) feelings of prejudice, and (4) feelings of connection (Syed & Azmitia, 2008, 2010). Interestingly, many of the STM topics identified in the current study largely map onto these themes, although in some cases, the topics we identified were narrower. The first two themes—awareness of difference and awareness of underrepresentation—appear to be replicated in the present study. Meanwhile, STM topics *Reactance to prejudice* and *Negative affect* were more specific examples of feelings of prejudice. Similarly, the theme of feelings of connection is represented by the STM topics of *Ethnic/cultural celebration* and *Peer dynamics*.
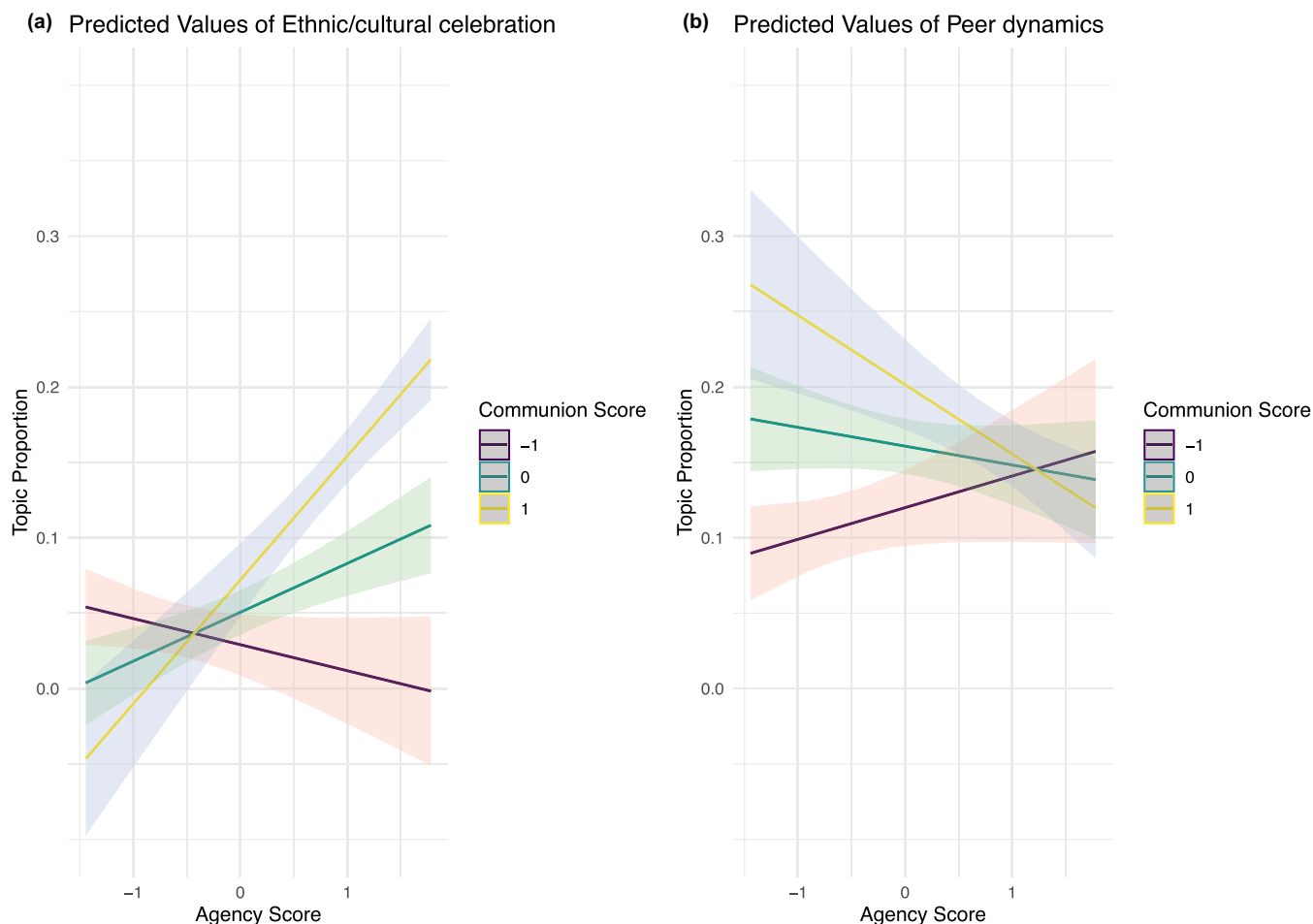
(a) Predicted Values of Ethnic/cultural celebration   **(b)** Predicted Values of Peer dynamics

FIGURE 5 (a, b) Agency and communion scores predicting topic proportions for Ethnic/cultural celebration and peer dynamics.

Crucially, the distinction between topics allowed us to identify subtle differences between participants. For example, among the topics related to feelings of connection, *Ethnic/cultural celebration* was most prevalent in narratives from Latinx participants, while *Peer Dynamics* was most prevalent in narratives from Multiracial participants. Had we coded the narratives using only the four broad content themes identified in past research, we may have missed this nuance because both Latinx and Multiracial participants wrote narratives about feeling connected. Although these differences were not statistically significant, likely due to small sample sizes of individual ethnic groups, we believe this observation illustrates the potential for STM to identify more nuanced and specific themes that better distinguish groups. Although we did not code narratives for broader content themes identified by past researchers, the strong overlap between previously identified themes and our STM-derived topics suggests topic modeling serves as an accurate and efficient method for quantifying narratives. Future work may estimate the degree to which human coding of these themes is captured by STM topics.

Our STM analysis also was able to capture how individuals navigated structures related to ethnicity, including

racism and white supremacy. Participants incorporated ecological topics, such as *Gained perspective of structural inequality* and *Whiteness*, into their narratives of ethnicity. Past research indicates awareness of structural factors is a touchstone for identity development (McLean et al., 2018). Similarly, recent narrative research has found structural factors, such as white supremacy and racial justice, appear in personal experiences described by emerging adults (Moffitt et al., 2022; Rogers et al., 2021). Our findings are consistent with the past research demonstrating the importance of assessing structural factors. Subsequent work may evaluate their long-term psychological impacts, including how participants process experiences involving white privilege and how their perspective on such events unfolds over time.

4.2 | Prevalence in STM topics based on ethnic group

Topic prevalence did not substantially differ across ethnic groups; in other words, participants' backgrounds were not associated with the likelihood of including specific topics in their narratives. This was surprising given that prior

TABLE 3 The mean agency and communion score and topic proportions across ethnicity.

	Group means						η^2	*p*-value
	African American	Asian	Latinx/ Hispanic	Multiracial	White	Noncategor.		
Themes								
Agency	1.10	1.30	1.20	1.30	1.50	1.30	0.01	**0.009***
Communion	0.80	1.20	1.10	1.20	1.20	1.20	0.01	0.220
Topics								
Cultural difference	0.10	0.14	0.15	0.16	0.15	0.15	0.00	0.865
Peer dynamics	0.15	0.13	0.16	0.13	0.13	0.18	0.00	0.552
Othering	0.12	0.09	0.12	0.13	0.07	0.10	0.01	0.137
Prejudice	0.07	0.07	0.10	0.08	0.10	0.07	0.00	0.548
Underrepresentation	0.09	0.07	0.08	0.05	0.08	0.10	0.00	0.812
Ethnic celebration	0.04	0.09	0.06	0.10	0.09	0.05	0.01	0.301
Cultural shock	0.08	0.12	0.08	0.09	0.06	0.07	0.01	**0.006***
Violence	0.10	0.08	0.05	0.12	0.09	0.03	0.01	0.091
Racial discr.	0.11	0.04	0.05	0.04	0.06	0.04	0.01	0.147
Negative affect	0.07	0.06	0.04	0.04	0.05	0.06	0.00	0.859
Cultural space	0.03	0.05	0.02	0.04	0.05	0.06	0.00	0.502
School	0.02	0.02	0.04	0.00	0.03	0.05	0.01	0.228
Structural ineq.	0.00	0.02	0.03	0.02	0.02	0.03	0.00	0.652
Whiteness	0.01	0.02	0.02	0.01	0.02	0.01	0.01	0.273

Note: The *p*-values and effect sizes (η^2) for the ANOVAs are presented. The *p*-values where the themes and topics significantly differ across ethnicity are asterisked and in bold (*p*-value <0.05).

work has found such differences across ethnic groups. For example, compared to participants of color, White participants engaged in less meaning-making of ethnic identity when reflecting on societal inequity (Moffitt et al., 2022). Hence, we expected differences between participants of color and White participants for individual-level STM topics, such as *Negative Affect* and *Violence*. This departure in findings may be because Moffitt et al. (2022) prompted participants to reflect on a recent conversation about ethnicity, rather than an experience in which their ethnicity was made salient. Our broader prompt and finer partitioning of ethnic groups (rather than grouping participants of color into a single category) may also contribute to the different outcomes.

Still, we found one significant difference in topic prevalence across ethnic groups: Asian American participants described *Experience of culture shock* significantly more compared to White participants. Among the relevant narratives from Asian participants, we found bicultural participants reporting visiting their heritage country either for the first time or after a long period away. Thus, it may be that the "Asian American" label encompasses both Asian Americans and Asian immigrants. We would expect the latter group to potentially report experiences of initial exposure to host cultural norms (Benet-Martínez &

Haritatos, 2005), while the former group may report exposure to heritage cultural norms (Altweck & Marshall, 2015). Given this, we suggest that future research focus on immigrant status as a source of topic prevalence.

4.3 | Topics vocabulary by ethnic group

We identified vocabulary differences across ethnic groups to better understand variations in ethnic experiences. Importantly, these results are descriptive in nature only, but they can help motivate future research questions. We found that Black/African American participants used more political language ("rally", "protest") while Asian American participants discussed conferences and celebrations for *Ethnic/cultural celebration*; future research might investigate how organized cultural events influence ethnic identity. People, objects, and places surfaced while reflecting on ethnic group identity. For example, among Asian American participants, references to people, specifically, family terms—"mother," "aunt," "dads," "folks"— and food—as "soup," "meat," and "pork"—differentiated topics. It may not be unusual to see such terms, given the importance of family (Umaña-Taylor et al., 2013) and food (Reddy & Van Dam, 2020) to ethnic socialization, but the

separation of topics by the presence of specific people or cultural elements is notable. More broadly, these findings may inform future hypothesis-driven analyses exploring language differences using a word count approach (e.g., LIWC, Pennebaker et al., 2003). For example, researchers might test whether Asian American participants are more likely to discuss family or food compared to other participants. The current study's exploration of ethnic-specific language elucidates the diverse cultural landscapes that shape ethnic identity, suggesting avenues for future research to delve deeper into how cultural events, family, food, and settings contribute to group identity formation.

4.4 | Agentic and communal STM topics

Ethnicity-related experiences were sometimes agentic, sometimes communal, and occasionally both. Narratives high in agency often described differentiation from others, such as *Awareness of underrepresentation* and *Lack of belonging in cultural space*. In the face of isolating experiences, the individual may be positioned toward greater autonomy and independence rather than feelings of loneliness (Henning et al., 2021). The framing of agentic narratives often demonstrated acceptance and self-insight that worked to empower the individual (McAdams et al., 1996). Independent thinking aligns with another highly agentic topic: *Gained perspective of structural inequality*. Overall, these narratives suggest that agentic ethnicity-related experiences motivate the individual to exert control over their thoughts, behaviors, and decision-making, potentially buffering against what could be a negative ethnic experience.

In tandem, highly communal topics were *Peer dynamics* and *Awareness of underrepresentation*. Although past research found that instances of underrepresentation may contribute to feeling isolated and excluded (Ackerman-Barger et al., 2020; Campbell, 2023), the narratives in this study suggest that underrepresentation may also make salient one's connection with their own ethnic group (Mehra, 2023). For emerging adults, it may be that underrepresentation narratives are highly communal because they emphasize belonging to one's ethnic group rather than disconnecting from an out-group.

Some topics were more prevalent in narratives with high levels of both agency and communion. For example, the topic *Ethnic/cultural celebration* had a stronger positive relation with the agency when communion was moderate and high. In line with prior work, individuals described a strong sense of belonging during ethnic celebrations (Abbasian, 2022; Rokach, 2020). Thus, ethnic celebration may be discussed more in agentic and communal narratives because individuals derive empowerment from belonging. Discerning instances where agency and

communion significantly interact with one another allows us to better understand the role of the individual and the community across these experiences.

In another instance, communion moderated the relationship between agency and *Peer dynamics*, such that topic proportion for peer dynamics was only negatively related to agency when narratives were rated as high in communion. Experiences involving peers often involve a source of connection (Camarena et al., 1990; McLean et al., 2018). When communal identity is high in the presence of others, this may sometimes be at the expense of autonomy (Clasen & Brown, 1985; Connolly & McIsaac, 2009). Agentic and communal experiences solidify the individual's self-understanding as part of a larger group identity, demonstrating resilience and adaptability in navigating cultural landscapes.

4.5 | Limitations and future directions

The current study aimed to improve our understanding of ethnic experiences related to social connectedness as well as autonomy. Future ethnicity-related work should expand this analysis to include other narrative elements, such as emotional valence or redemption (Breen & McLean, 2017; Haimi, 2023). This would increase specificity in our understanding of ethnic experiences. For example, STM analysis may provide insight into how individuals process adversity and in what cases they become empowered as a result. Future studies can use STM to investigate sociocultural factors and identity processes (e.g., bicultural identity integration). Leveraging STM in narrative research can greatly enhance our comprehension of how different groups articulate and derive meaning from their experiences.

The study had several strengths: STM results in empirically derived topics, which is efficient—saving time, money, and energy while coding hundreds of narratives in a matter of minutes—and reduces subjectivities that may be introduced with human coding. However, STM is still subjective in topic interpretation (Finch et al., 2018) and limits researchers' opportunities to conduct confirmatory analyses. Those wishing to examine the generalizability of these findings may wish to use a bag-of-words approach, like LIWC (Ashokkumar & Pennebaker, 2022), or human coding to identify pre-determined topics and themes.

4.6 | Constraints on generality

Our sample contained college students in the U.S., limiting generalizability to older or younger participants or those who do not enter college. We encourage future researchers to collect a larger sample of ethnic minority

participants to foster an understanding of intra- and inter-group differences in ethnic experiences.

5 | CONCLUSION

The contribution of this study is methodological and theoretical. At the methodological level, STM is a useful tool to derive content topics within the narrative context, like content domains and keywords, and estimates relationships of these topics to other variables, including narrative themes or demographic variables. At the theoretical level, we identified key topics that arise in ethnic-identity narratives, including cultural celebrations and awareness of underrepresentation.

We found that these topics varied in the extent that they were present in agentic and communal narratives, demonstrating the importance of these motivational coordinates for the understanding of social group experiences, more broadly. These findings allow for a greater understanding of when individuals feel connected to a larger community, when they experience independence, and when both are relevant.

AUTHOR CONTRIBUTIONS
Rachel Jacobson: Conceptualization, computational and narrative analysis, methodology, writing—original draft. Dulce Wilkinson Westberg: Conceptualization, narrative analysis, methodology, writing—review and editing. Edward Chou: Conceptualization, methodology, narrative analysis, writing—review and editing. Moin Syed: Conceptualization, data curation, supervision, writing—review and editing. Sara J. Weston: Conceptualization, methodology, supervision, writing—review and editing.

CONFLICT OF INTEREST STATEMENT
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS STATEMENT
This research was conducted following the Ethical Standards of the American Psychological Association, with approval from the Institutional Review Board at each data collection institution.

ORCID
Rachel Jacobson 🔘 https://orcid.org/0000-0002-8667-5804
Edward Chou 🔘 https://orcid.org/0000-0003-2045-8388

ENDNOTES
[1] We use 'ethnicity,' to refer to both ethnicity and race. According to the Ethnic-Racial Identity Framework (Umaña-Taylor et al., 2014), ethnicity and race are overlapping and often inseparable social constructs. Thus, our use of the term 'ethnicity' can be interpreted as 'ethnic/racial.'

[2] For those interested in a comprehensive explanation of STM, we recommend the tutorials by Roberts et al. (2019) and Weston et al. (2023).

[3] Portions of these data have been published previously to examine a subset of the narrative data (Syed & Azmitia, 2008), as well as ethnic group typicality and theoretical models of ethnic identity related to identity coherence and psychological functioning (Mitchell et al., 2018; Syed & Juang, 2014).

[4] Participants who selected more than one category or self-identified as Multiracial were labeled as 'Multiracial.'

[5] Within this category, we manually coded an open-ended item asking these participants to describe their ethnicity/race into one of the six aforementioned ethnic groups. For example, participants in this subset who identified as "Irish" and nothing else were categorized as "White." We manually coded 27 out of the 1149 total participants. The breakdown of these participants is the following: 8 White, 11 Asian/Asian American, 6 Latinx, 1 Black/African American, and 1 Multiracial. Still, 6.5% were not able to be categorized based on their responses and made up the Other or 'non-categorizable' category.

[6] This code is available on the OSF Repository: https://osf.io/7krxn/.

[7] We initially preregistered to use held-out likelihood as a metric and instead used exclusivity based on recommendations from Weston et al. (2023).

[8] This is not dissimilar to exploratory factor analysis, in which multiple numbers of factors are supported by the fit statistics. In both cases (STM and EFA), researchers subjectively evaluate the solutions to determine whether one or several provide theoretical insight and potential application.

[9] By using unstandardized coefficients, the interpretation remains in the original scale of the covariates and topics, making it easier to understand how actual changes in covariates relate to changes in topic proportions. Standardizing coefficients can obscure the relationship between the covariates and the raw data, making it less intuitive to relate the model results back to the original data.

REFERENCES
Abbasian, S. (2022). Disparate emotions as expressions of well-being: Impact of festival participation from the participants' subjective view. *International Journal of Environmental Research and Public Health*, *20*(1), 329. https://doi.org/10.3390/ijerph20010329

Ackerman-Barger, K., Boatright, D., Gonzalez-Colaso, R., Orozco, R., & Latimore, D. (2020). Seeking inclusion excellence:

Understanding racial microaggressions as experienced by underrepresented medical and nursing students. *Academic Medicine*, *95*(5), 758–763. https://doi.org/10.1097/ACM.0000000000003077

Adler, J. M. (2018). Bringing the (disabled) body to personality psychology: A case study of Samantha. *Journal of Personality*, *86*(5), 803–824. https://doi.org/10.1111/jopy.12364

Adler, J. M., Dunlop, W., Fivush, R., Lilgendahl, J. P., Lodi-Smith, J., McAdams, D. P., McLean, K. C., Pasupathi, M., & Syed, M. (2017). *Research methods for studying narrative identity: A primer* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/cz3q3

Adler, J. M., & Poulin, M. J. (2009). The political is personal: Narrating 9/11 and psychological well-being. *Journal of Personality*, *77*(4), 903–932. https://doi.org/10.1111/j.1467-6494.2009.00569.x

Altweck, L., & Marshall, T. C. (2015). When you have lived in a different culture, does returning 'home' not feel like home? Predictors of psychological readjustment to the heritage culture. *PLoS ONE*, *10*(5), e0124393. https://doi.org/10.1371/journal.pone.0124393

Ashokkumar, A., & Pennebaker, J. W. (2022). Tracking group identity through natural language within groups. *PNAS Nexus*, *1*(2), 1–9. https://doi.org/10.1093/pnasnexus/pgac022

Bakan, D. (1966). *The duality of human existence: Isolation and communion in Western man*. Beacon Press.

Benet-Martínez, V., & Haritatos, J. (2005). Bicultural identity integration (BII): Components and psychosocial antecedents. *Journal of Personality*, *73*(4), 1015–1050. https://doi.org/10.1111/j.1467-6494.2005.00337.x

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*, 77–84.

Breen, A. V., & McLean, K. C. (2017). The intersection of personal and master narratives. In B. Schiff, A. E. McKim, & S. Patron (Eds.), *Life and narrative: The risks and responsibilities of storying experience*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190256654.003.0011

Brondolo, E., Gallo, L. C., & Myers, H. F. (2009). Race, racism and health: Disparities, mechanisms, and interventions. *Journal of Behavioral Medicine*, *32*(1), 1–8. https://doi.org/10.1007/s10865-008-9190-3

Camarena, P. M., Sarigiani, P. A., & Petersen, A. C. (1990). Gender-specific pathways to intimacy in early adolescence. *Journal of Youth and Adolescence*, *19*(1), 19–32. https://doi.org/10.1007/BF01539442

Campbell, K. M. (2023). Mitigating the isolation of minoritized faculty in academic medicine. *Journal of General Internal Medicine*, *38*(7), 1751–1755. https://doi.org/10.1007/s11606-022-07982-8

Chae, D. H., Yip, T., Martz, C. D., Chung, K., Richeson, J. A., Hajat, A., Curtis, D. S., Rogers, L. O., & LaVeist, T. A. (2021). Vicarious racism and vigilance during the COVID-19 pandemic: Mental health implications among Asian and black Americans. *Public Health Reports*, *136*(4), 508–517. https://doi.org/10.1177/00333549211018675

Clasen, D. R., & Brown, B. B. (1985). The multidimensionality of peer pressure in adolescence. *Journal of Youth and Adolescence*, *14*(6), 451–468. https://doi.org/10.1007/BF02139520

Connolly, J. A., & McIsaac, C. (2009). Romantic relationships in adolescence. In *Handbook of adolescent psychology: Contextual influences on adolescent development* (Vol. *2*, 3rd ed., pp. 104–151). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470479193.adlpsy002005

Denny, M., & Spirling, A. (2017). *Replication data for: Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it* [dataset]. Harvard Dataverse. https://doi.org/10.7910/DVN/XRR0HM

Dunlop, W. L., Wilkinson, D., Harake, N., Graham, L. E., & Lee, D. (2020). The redemption and contamination research form: Exploring relations with narrative identity, personality traits, response styles, and life satisfaction. *Memory*, *28*(10), 1219–1230. https://doi.org/10.1080/09658211.2020.1828926

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in R. *Journal of Statistical Software*, *25*(5), 1–54. https://doi.org/10.18637/jss.v025.i05

Finch, W. H., Hernández Finch, M. E., McIntosh, C. E., & Braun, C. (2018). The use of topic modeling with latent Dirichlet analysis with open-ended survey items. *Translational Issues in Psychological Science*, *4*(4), 403–424. https://doi.org/10.1037/tps0000173

Fish, J., & Syed, M. (2021). Digital storytelling methodologies: Recommendations for a participatory approach to engaging underrepresented communities in counseling psychology research. *Journal of Counseling Psychology*, *68*(3), 271–285. https://doi.org/10.1037/cou0000532

Fu, Q., Zhuang, Y., Gu, J., Zhu, Y., & Guo, X. (2021). Agreeing to disagree: Choosing among eight topic-modeling methods. *Big Data Research*, *23*, 100173. https://doi.org/10.1016/j.bdr.2020.100173

Galliher, R. V., McLean, K. C., & Syed, M. (2017). An integrated developmental model for studying identity content in context. *Developmental Psychology*, *53*(11), 2011–2022. https://doi.org/10.1037/dev0000299

Haimi, M. (2023). The tragic paradoxical effect of telemedicine on healthcare disparities- a time for redemption: A narrative review. *BMC Medical Informatics and Decision Making*, *23*(1), 95. https://doi.org/10.1186/s12911-023-02194-4

Henning, G., Segel-Karpas, D., Bjälkebring, P., & Berg, A. I. (2021). Autonomy and loneliness—Longitudinal within- and between-person associations among Swedish older adults. *Aging & Mental Health*, *26*(12), 2416–2423. https://doi.org/10.1080/13607863.2021.2000937

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, *33*(2), 125–143.

Kalsner, L., & Pistole, M. C. (2003). College adjustment in a multiethnic sample: Attachment, separation-individuation, and ethnic identity. *Journal of College Student Development*, *44*(1), 92–109. https://doi.org/10.1353/csd.2003.0006

Koltcov, S., Ignatenko, V., & Koltsova, O. (2019). Estimating topic modeling performance with sharma–mittal entropy. *Entropy*, *21*(7), 660. https://doi.org/10.3390/e21070660

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kuhn, K. D. (2018). Using structural topic modeling to identify latent topics and trends in aviation incident reports. *Transportation Research Part C: Emerging Technologies*, *87*, 105–122.

Light, R., & Odden, C. (2017). Managing the boundaries of taste: Culture, valuation, and computational social science. *Social Forces*, *96*(2), 877–908. https://doi.org/10.1093/sf/sox055

Lilgendahl, J. P., Benet-Martinez, V., Bishop, M., Gilson, K., Festa, L., Levenson, C., & Rosenblum, R. (2018). "So now, I wonder, what Am I?": A narrative approach to bicultural identity integration. *Journal of Cross-Cultural Psychology*, *49*(10), 1596–1624. https://doi.org/10.1177/0022022118801555

Loveys, K., Torrez, J., Fine, A., Moriarty, G., & Coppersmith, G. (2018). Cross-cultural differences in language markers of depression online. *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, 78–87. https://doi.org/10.18653/v1/W18-0608

Loyd, A. B., Westberg, D. W., Williams, L., Humphries, M., Meca, A., & Rodil, J. C. (2023). "I just want to be me, authentically": Identity shifting among racially and ethnically diverse young adults. *Journal of Youth and Adolescence*, *52*(4), 701–718. https://doi.org/10.1007/s10964-023-01744-3

McAdams, D. P. (1995). What do we know when we know a person? *Journal of Personality*, *63*(3), 365–396. https://doi.org/10.1111/j.1467-6494.1995.tb00500.x

McAdams, D. P. (2001). The psychology of life stories. *Review of General Psychology*, *5*(2), 100–122. https://doi.org/10.1037/1089-2680.5.2.100

McAdams, D. P., Anyidoho, N. A., Brown, C., Huang, Y. T., Kaplan, B., & Machado, M. A. (2004). Traits and stories: Links between dispositional and narrative features of personality. *Journal of Personality*, *4*(72), 761–784. https://doi.org/10.1111/j.0022-3506.2004.00279.x

McAdams, D. P., Hoffman, B. J., Day, R., & Mansfield, E. D. (1996). Themes of agency and communion in significant autobiographical scenes. *Journal of Personality*, *64*(2), 339–377. https://doi.org/10.1111/j.1467-6494.1996.tb00514.x

McCabe, A., & Dinh, K. T. (2016). Agency and communion, ineffectiveness and alienation: Themes in the life stories of Latino and southeast Asian adolescents. *Imagination, Cognition and Personality*, *36*(2), 150–171. https://doi.org/10.1177/0276236616648648

McLean, K. C., Lilgendahl, J. P., Fordham, C., Alpert, E., Marsden, E., Szymanowski, K., & McAdams, D. P. (2018). Identity development in cultural context: The role of deviating from master narratives. *Journal of Personality*, *86*(4), 631–651. https://doi.org/10.1111/jopy.12341

McLean, K. C., Shucard, H., & Syed, M. (2017). Applying the master narrative framework to gender identity development in emerging adulthood. *Emerging Adulthood*, *5*(2), 93–105. https://doi.org/10.1177/2167696816656254

McLean, K. C., & Syed, M. (2015). Personal, master, and alternative narratives: An integrative framework for understanding identity development in context. *Human Development*, *58*(6), 318–349. https://doi.org/10.1159/000445817

McLean, K. C., Syed, M., Pasupathi, M., Adler, J. M., Dunlop, W. L., Fivush, R., Graci, M., Lilgendahl, J., Lodi-Smith, J., McAdams, D. P., Drustrup, D., & McCoy, T. (2019, March 8). Coding manuals. Retrieved from osf.io/s2ewr/

McLean, K. C., Syed, M., Pasupathi, M., Adler, J. M., Dunlop, W. L., Drustrup, D., Fivush, R., Graci, M. E., Lilgendahl, J. P., Lodi-Smith, J., McAdams, D. P., & McCoy, T. P. (2020). The empirical structure of narrative identity: The initial big three. *Journal of Personality and Social Psychology*, *119*(4), 920–944. https://doi.org/10.1037/pspp0000247

Mehra, A. (2023). At the margins: A distinctiveness approach to the social identity and social networks of underrepresented groups. *Academy of Management Journal*, *41*, 441–452. https://doi.org/10.2307/257083

Mitchell, L. L., Kathawalla, U.-K., Ajayi, A. A., Fish, J., Nelson, S. C., Peissig, L. H. M., & Syed, M. (2018). Ethnic-racial typicality and its relation to ethnic identity and psychological functioning. *Cultural Diversity and Ethnic Minority Psychology*, *24*(3), 400–413. https://doi.org/10.1037/cdp0000193

Moffitt, U., Rogers, L. O., Mzizi, Y., & Charlson, E. (2022). Race talk during the 2020 U.S. presidential election: Emerging adults' critical consciousness and racial identity in context. *Journal of Adolescent Research*, *39*, 1048–1085. https://doi.org/10.1177/07435584221145009

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, *54*(1), 547–577. https://doi.org/10.1146/annurev.psych.54.101601.145041

R Core Team. (2021). *R: A language and environment for statistical computing* [manual]. https://www.R-project.org/

Reddy, G., & Van Dam, R. M. (2020). Food, culture, and identity in multicultural societies: Insights from Singapore. *Appetite*, *149*, 104633. https://doi.org/10.1016/j.appet.2020.104633

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm: R package for structural topic models. *Journal of Statistical Software*, *91*, 1–40.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, *58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Rogers, L. O., Niwa, E. Y., Chung, K., Yip, T., & Chae, D. (2021). M(ai)cro: Centering the macrosystem in human development. *Human Development*, *65*(5–6), 270–292. https://doi.org/10.1159/000519630

Rokach, A. (2020). Belonging, togetherness and food rituals. *Open Journal of Depression*, *9*(4), 77–85. https://doi.org/10.4236/ojd.2020.94007

Singer, J. A. (2004). Narrative identity and meaning making across the adult lifespan: An introduction. *Journal of Personality*, *72*(3), 437–459. https://doi.org/10.1111/j.0022-3506.2004.00268.x

Sterling, J., Jost, J. T., & Hardin, C. D. (2019). Liberal and conservative representations of the good society: A (social) structural topic modeling approach. *SAGE Open*, *9*(2), 1–13. https://doi.org/10.1177/2158244019846211

Syed, M. (2015). Theoretical and methodological contributions of narrative psychology to ethnic identity research. In *Studying ethnic identity: Methodological and conceptual approaches across disciplines* (pp. 27–54). American Psychological Association. https://doi.org/10.1037/14618-003

Syed, M., & Azmitia, M. (2008). A narrative approach to ethnic identity in emerging adulthood: Bringing life to the identity status model. *Developmental Psychology*, *44*(4), 1012–1027. https://doi.org/10.1037/0012-1649.44.4.1012

Syed, M., & Azmitia, M. (2010). Narrative and ethnic identity exploration: A longitudinal account of emerging adults'

ethnicity-related experiences. *Developmental Psychology*, *46*(1), 208–219. https://doi.org/10.1037/a0017825

Syed, M., & Juang, L. P. (2014). Ethnic identity, identity coherence, and psychological functioning: Testing basic assumptions of the developmental model. *Cultural Diversity and Ethnic Minority Psychology*, *20*(2), 176–190. https://doi.org/10.1037/a0035330

Syed, M., & McLean, K. C. (2022). Who gets to live the good life? Master narratives, identity, and well-being within a marginalizing society. *Journal of Research in Personality*, *100*, 1–7. https://doi.org/10.1016/j.jrp.2022.104285

Syed, M., & Nelson, S. C. (2015). Guidelines for establishing reliability when coding narrative data. *Emerging Adulthood*, *3*(6), 375–387. https://doi.org/10.1177/2167696815587648

Tajfel, H., & Turner, J. C. (1979). An integrative theory of inter-group conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of inter-group relations*. Brooks/Cole.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Thompson, B., Roberts, S. G., & Lupyan, G. (2020). Cultural influences on word meanings revealed through large-scale semantic alignment. *Nature Human Behaviour*, *4*(10), 1029–1038. https://doi.org/10.1038/s41562-020-0924-8

Umaña-Taylor, A. J., Quintana, S. M., Lee, R. M., Cross, W. E., Rivas-Drake, D., Schwartz, S. J., Syed, M., Yip, T., Seaton, E., & Ethnic and Racial Identity in the 21st Century Study Group. (2014). Ethnic and racial identity during adolescence and into young adulthood: An integrated conceptualization. *Child Development*, *85*(1), 21–39. https://doi.org/10.1111/cdev.12196

Umaña-Taylor, A. J., Zeiders, K. H., & Updegraff, K. A. (2013). Family ethnic socialization and ethnic identity: A family-driven, youth-driven, or reciprocal process? *Journal of Family Psychology*, *27*(1), 137–146. https://doi.org/10.1037/a0031105

Wallach, H. M. (2006). Topic modeling: Beyond bag-of-words. In *Proceedings of the 23rd International Conference on Machine Learning—ICML'06* (pp. 977–984). https://doi.org/10.1145/1143844.1143967

Watts, R. J., Diemer, M. A., & Voight, A. M. (2011). Critical consciousness: Current status and future directions. *New Directions for Child and Adolescent Development*, *134*, 43–57. https://doi.org/10.1002/cd.310

Webber, M., & Waru-Benson, S. (2022). The role of cultural connectedness and ethnic group belonging to the social-emotional wellbeing of diverse students. In T. A. Olsen & H. Sollid (Eds.), *Indigenising education and citizenship* (pp. 294–312). Scandinavian University Press. https://doi.org/10.18261/9788215053417-2022-16

Westberg, D. W. (2022). Stories off the beaten path: Exploring the content and correlates of cultural deviation narratives. *Journal of Research in Personality*, *100*, 104266. https://doi.org/10.1016/j.jrp.2022.104266

Weston, S. J., Shryock, I., Light, R., & Fisher, P. A. (2023). Selecting the number and labels of topics in topic modeling: A tutorial. *Advances in Methods and Practices in Psychological Science*, *6*(2), 251524592311601. https://doi.org/10.1177/25152459231160105

Weststrate, N. M., & McLean, K. C. (2023). "You gotta give them hope": A structural psychobiography of Harvey Milk (1930–1978). *Journal of Personality*, *91*(1), 105–119. https://doi.org/10.1111/jopy.12744

Wiberg, M., Molenaar, D., González, J., Böckenholt, U., & Kim, J.-S. (Eds.). (2021). *Quantitative psychology: The 85th annual meeting of the psychometric society, virtual* (Vol. *353*). Springer International Publishing. https://doi.org/10.1007/978-3-030-74772-5

Williams, C. D., Byrd, C. M., Quintana, S. M., Anicama, C., Kiang, L., Umaña-Taylor, A. J., Calzada, E. J., Pabón Gautier, M., Ejesi, K., Tuitt, N. R., Martinez-Fuentes, S., White, L., Marks, A., Rogers, L. O., & Whitesell, N. (2020). A lifespan model of ethnic-racial identity. *Research in Human Development*, *17*(2–3), 99–129. https://doi.org/10.1080/15427609.2020.1831882

Yip, T., Cheah, C. S. L., Kiang, L., & Hall, G. C. N. (2021). Rendered invisible: Are Asian Americans a model or a marginalized minority? *American Psychologist*, *76*(4), 575–581. https://doi.org/10.1037/amp0000857

Yip, T., Seaton, E. K., & Sellers, R. M. (2006). African American racial identity across the lifespan: Identity status, identity content, and depressive symptoms. *Child Development*, *77*(5), 1504–1517. https://doi.org/10.1111/j.1467-8624.2006.00950.x

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.