Integrating Human and Machine Coding to Measure Political Issues in Ethnic Newspaper Articles

Jae Yeon Kim, *†

Forthcoming at the *Journal of Computational Social Science*

Abstract

The voices of racial minority groups have rarely been examined systematically with large-scale text analysis in political science. This study fills such a gap by applying an integrated classification framework to the analysis of the commonalities and differences in political issues that appeared in 78,305 articles from Asian American and African American newspapers from the 1960s to the 1980s. The automated text classification shows that Asian American newspapers focused on promoting collective gains more often than African American newspapers. Conversely, African American newspapers concentrated on preventing collective losses more than Asian American newspapers. The content analysis demonstrates that the issue priorities varied between the corpora, especially with respect to policy contexts. Gaining access to government resources was a more urgent issue for Asian Americans, while reducing or ending state violence, such as police brutality, was a more pressing matter for African Americans. It also helped avoid extreme interpretations of the machine coding, as the misalignment of political agendas between the two corpora widened up to 10 times when the training data were measured using the minimum, rather than the maximum, reliability threshold.

Quantitative scholarship on racial and ethnic politics in the U.S. has evolved based on the development of new surveys. The American National Election Studies (ANES) comprise high-quality panel data that go back to 1948 and have been central in investigating public opinion in U.S. politics (Campbell et al. 1980; Zaller et al. 1992; Bartels 1999). However, when it comes to studying racial minority groups' politics, the data have clear limitations because these groups take up only a very small portion of ANES data (Lien, Conway, and Wong 2004). Other prominent panel data on public opinion, such as the General Social Survey (GSS), are no exception. Racial and ethnic politics researchers have tried to overcome this data limitation by creating new surveys. To name a few, the National Black Election Study[1] was developed in 1984, the Latino National Political Survey[2] in 1989, the National Surveys of Latinos[3] in 2002, the Pilot National Asian American Political Survey[4] in 2000 and its official version[5] in 2008, and the Comparative Post-Election Survey[6] in 2008. These new data sets have enabled a large number of research to be conducted in African American, Latino, and Asian American politics (Gurin, Hatchett, and Jackson 1990; Tate 1993; Dawson 1994b; Fraga et al. 2011; Wong et al. 2011; McClain 2018). Nevertheless, most of these surveys are short lived and thus not comparable to the ANES or GSS in terms of longevity. More importantly, these surveys were developed mostly in the 1990s and early 2000s, so they are not useful to researchers interested in historical questions such as the origins and developments of minority political coalitions.

Automating text analysis provides a solution to this long-standing problem in the historical study of racial and ethnic politics in the U.S.. Text always has been the principal form of data in historical research. Archival materials and oral histories have provided insights into how minority groups involved in political activism and developed coalitions in the 1960s and 1970s (Muñoz 1989; Wei 1993; Joseph 2006; Maeda 2012; Ishizuka 2016; Linder

---

1. For more information, see https://www.icpsr.umich.edu/icpsrweb/ICPSR/series/163.
2. For more information, see https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/6841.
3. For more information, see https://www.pewresearch.org/topics/national-survey-of-latinos/.
4. For more information, see https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/3832.
5. For more information, see https://naasurvey.com/data/.
6. For more information, see https://cmpsurvey.org/.

et al. 2018). A particularly important source is ethnic newspapers because they have been an essential part of mobilization networks for racial minority groups. For instance, Marcus Garvey (1887-1940) founded and edited *The Negro World* for the Universal Negro Improvement Association. W. E. B. Du Bois (1868-1963) founded and edited *The Crisis* for the National Association for the Advancement of Colored People (NAACP) (Vincent 1973, 17). Similarly, Asian American activists founded more than 20 periodicals from the 1960s through the early 1970s (Espiritu 1992, 39). The National Council of La Raza, the nation's largest Latino organization, founded and circulated its official magazine, *Agenda*, among Latino activists nationwide (Mora 2014). Analyzing the commonalities and differences between ethnic newspapers' coverage of policies and politics can inform our understanding of the salient political issues for each racial group and whether they have common political agendas around which to mobilize. Nevertheless, ethnic newspapers have rarely been utilized for quantitative analysis because manually collecting, reading, and coding these documents is a labor-intensive process.

This is a missed opportunity because the cost of conducting a large-scale text analysis on ethnic newspapers' articles has decreased substantially. Recent advancements in computational text analysis tools and techniques have made analyzing large volumes of texts easier, faster, and more contextualized (Mikolov et al. 2013; Goth 2016; Joulin et al. 2016; Devlin et al. 2018). Political scientists have taken advantage of these tools and techniques to investigate the words of the powerful, such as U.S. congress members (Hopkins and King 2010; Grimmer, Messing, and Westwood 2012). In contrast, computational text analysis has rarely been applied for studying the voices of marginalized members, partly because different ethnic newspapers use different data structures to display and store articles. As data vendors (e.g., ProQuest) have digitized millions of ethnic newspapers' articles published in the U.S. and standardized their data structures[7], computationally analyzing ethnic newspapers at scale have become quite feasible.

---

7. For more information, see https://www.proquest.com/products-services/ethnic_newswatch.html.

However, opportunities come with challenges. To begin with, more is not always better. The advantage of large-scale computational text analysis over traditional content analysis is scale. A large sample makes hypothesis testing easier because it reduces the size of standard errors. Nevertheless, because ethnic newspaper articles are collected through non-probability sampling, the large size of the data also may increase the degree of bias in them (Meng 2018, 685-688). Some ethnic newspapers are more likely to be digitized and become available than others because of their financial status among other factors.

More fundamentally, measuring political issues in ethnic newspaper articles is difficult. Many ethnic newspapers were founded to offer alternative political voices. Because mainstream media did not cover minority issues, minority activists founded ethnic newspapers to discuss unique political issues (Espiritu 1992; Dawson 1994a; Rodriguez 1999; Dawson 2001; Kannegaard 2008; Harris-Lacewell 2010). These political issues could be conceptualized and measured at high (meta issues) or low (specific issues) levels. Meta issues, or broad issue frames, are useful in text classification, a supervised machine learning technique. Reducing the number of classes that need to be classified increases the number of observations for each class and simplifies the coding scheme for human coders, thereby enhancing training data's reliability (Mikhaylov, Laver, and Benoit 2012; Barberá et al. 2019). Nevertheless, these meta issues are valid conceptually only if they are associated closely with specific issues they are supposed to represent. In addition, different reliability thresholds could be applied to label training data. The minimum threshold indicates that at least one human coder must agree with the coding decision. The maximum threshold indicates that all human coders must agree with the coding decision.

I address these conceptual and methodological challenges by incorporating construct validity and reliability tests of the training data into the supervised machine learning pipeline. Technical guidelines exist for the automated part of the text analysis process, such as preprocessing documents, training algorithms, and evaluating their performance. However, computational social scientists have rarely engaged with how measurement decisions—humans in

3

the machine learning loop—affect machine learning outcomes (Mikhaylov, Laver, and Benoit 2012; Gitelman 2013; Geiger et al. 2020). I attempt to solve this problem by introducing an integrated text classification framework that examines how the quality of human-coded training data influences the performance and interpretation of machine-coded corpora.

As a demonstration of this framework, I analyzed 78,383 articles from Asian American and African American newspapers from the 1960s through the 1980s. I intentionally selected Asian American and African American newspapers based on the West Coast because Asian Americans and African Americans in this region shared strong social and political networks during the period under investigation. Such a case selection strategy also reduces alternative explanations. Meta issues among these groups are the broad frames that define issues relevant to deal with their unfair treatments in the U.S. racial hierarchy and can be divided into two categories: the promotion of collective gains (collective gain issue) and the prevention of collective losses (collective loss issue) The collective gain issue focuses on expanding opportunities and is related to beneficial government policies, such as minority business ownership. The collective loss issue focuses on reducing threats and is associated with oppressive government policies, such as police brutality. Content analysis assesses data quality by testing the training data's construct validity and reliability. Text classification demonstrates that Asian American newspapers published 110% more collective gain articles than African American newspapers. Conversely, African American newspapers published 133% more collective loss articles than Asian American newspapers. This gap between the two corpora widened up to 10 times when the training data were measured using the minimum, rather than the maximum, reliability threshold.

The article makes several contributions. Methodologically, the findings demonstrate why ensuring the quality of training data is vital in applying machine learning for social science. Data annotation's role in machine learning has been debated among scholars concerning bias and fairness in machine learning (Bender and Friedman 2018; Beretta et al. 2018; Gebru et al. 2018; Mitchell et al. 2019). Although computational text analysis has gained

popularity among social scientists (Grimmer and Stewart 2013; Wilkerson and Casas 2017; Brady 2019), empirical research that closely investigates human-coding decisions' impact on automated text classification remains lacking (Denny and Spirling 2017; Nelson et al. 2017; Nelson 2017; Elish and Boyd 2018; Nelson 2019). This article fills this gap in computational social science literature by showing how using less reliable data leads not only to weak predictions, but also to extreme interpretations. Substantially, the original large-scale evidence describes the underappreciated differences in the political issues between racial minority groups. Racial minorities in the U.S. are often coded altogether as people of color because they share racial marginalization. However, sharing the racial status does not automatically form an alignment on issue preferences among these groups (Omi and Winant 1986; Dawson 1994b; Kaufmann 2003). The missing element is policies. President Lyndon B. Johnson introduced civil rights legislation, expanded the welfare state (Orleck 2011; Bailey and Danziger 2013), and militarized policing (Hinton 2016). As his political legacy included the War on Poverty as well as the War on Crime, the ways in which minority groups responded to these contemporaneous policy challenges varied. Although many studies have shown strong ideological connections between the Asian American movement and the Black Power movement (Prashad 2002; Maeda 2005; Ho and Mullen 2008; Maeda 2012; Ishizuka 2016; Watkins 2012), I find that the two groups' priorities varied significantly. For Asian Americans, gaining access to government resources was a more urgent issue; however, for African Americans, reducing or ending state violence, such as police brutality, was a more important matter. Although these findings are limited in their temporal scope, they help scholars draw a linkage between racial marginalization and political preference formation within racial minority groups more carefully, in policy contexts.

1   Political Issues in Ethnic Newspapers

Political issues among racial minority groups in the U.S. focus on discrimination due to these groups' subordinate status in U.S. racial hierarchy (Omi and Winant 1986; Tate 1993; Dawson 1994b; Chae et al. 2008; Hwang and Goto 2008; Rothstein 2017; Trounstine 2018; Reardon, Kalogrides, and Shores 2019; Williams, Lawrence, and Davis 2019). Meta issues among these groups could be defined as broad frames that define issues that are relevant for dealing with unfair treatments. Dawson (1994b) tackled this conceptual problem in his analysis of the source of Black political solidarity. His Black utility heuristic theory argued that centuries of racial discrimination, including the continued legacy of slavery and Jim Crow, make it difficult for African Americans to separate their individual interests from group interests. Rich and poor African Americans vary little in their policy preferences and voting behavior because race ties their fate together and defines their meta issue: uniting African Americans against racism

I follow Dawson's footsteps but also go one step further by differentiating between the mechanisms that produce intraracial solidarity. Taking this step is useful in generalizing Dawson's framework and measuring how issue priorities between minority groups vary not only in degree, but also in kind. This extension is also important in examining the likelihood of interracial coalition formation. If issue priorities are misaligned between racial minority groups, it raises the coordination cost of forming a coalition among them.

I argue that although racial minority groups are often described collectively as people of color, their issues vary because they face different policy challenges. There are two ways in which a racial minority group can pursue its collective interests. When this article mentions "collective interests," it refers to common interests among individuals who share the same racial identification (e.g., Asian Americans, African Americans). One way is to emphasize the promotion of collective gain by expanding opportunities (collective gain issue), and the other way is to highlight the prevention of collective losses by reducing threats (collective loss issue).

Depending on the main policy challenge to which a group is exposed, these two meta issues' relevance and priority change. There are two types of government policies. Beneficial policies help a targeted group, while oppressive policies harm a targeted group. As summarized in Appendix A, what comprises beneficial and oppressive policies is not always clear. Policies such as police brutality and support for minority business ownership have clear directions as they either benefit or harm a targeted group. In contrast, policies like housing may benefit or harm a targeted group depending on the context. For instance, increasing affordable housing benefits a targeted group, while continuing housing discrimination harms a targeted group. When oppressive policies target a group, the collective loss frame becomes relevant and prioritized. Consider a situation in which a group is targeted disproportionately with police brutality. Protecting these human lives would become a theme through which to spur political mobilization. In contrast, when a group is being excluded from beneficial policies, the collective gain frame becomes relevant and prioritized. Consider a situation in which a minority group is not included as a beneficiary group for new social policies. Gaining access to state resources then would become the main focus of political advocacy.

Responding to policy changes, ethnic newspapers issue two different types of political articles to seek their group interests. Collective gain articles focus on collective gain issues, while collective loss articles focus on collective loss issues. If beneficial policies become salient, an ethnic newspaper would issue collective gain articles, but when oppressive policies become salient, an ethnic newspaper would issue collective loss articles. These theoretical expectations are summarized in Table 1.

	Beneficial policies	No beneficial policies
Oppressive policies	Mix of collective gain and loss articles	Collective loss articles
No oppressive policies	Collective gain articles	None

Table 1: How ethnic newspaper issue collective gain and loss articles in response to policy changes

2 Hypotheses

We can draw the following two hypotheses from the argument. If a racial minority group is exposed to beneficial policies, the newspaper aligned with the group will issue more collective gain articles. However, if a minority group is exposed to oppressive policies, the newspaper aligned with the group will issue more collective loss articles. Suppose that for class collective gain and class collective loss, "1" denotes membership in the class and "0" the opposite. $Pr(\text{Meta issue type} = 1|\text{Government policy type})$ indicates the probability of an ethnic newspaper issuing an article using a particular type of meta issue ($\in$ {collective gain, collective loss}) given that the group the newspaper is aligned with is exposed to a specific type of government policy ($\in$ {Beneficial, Oppressive}). We can define the hypotheses formally as follows:

- Hypothesis 1: $Pr(\text{Collective gain} = 1|\text{Beneficial policies}) >$
  $Pr(\text{Collective loss} = 1|\text{Beneficial policies})$

- Hypothesis 2: $Pr(\text{Collective loss} = 1|\text{Oppressive policies}) >$
  $Pr(\text{Collective gain} = 1|\text{Oppressive policies}).$

3 Case Selection

Domain knowledge helps identifying well-matched cases. For hypothesis testing, we ideally would like to compare a set of ethnic newspaper articles about two groups that are exactly alike, except that one group generally is exposed to beneficial policies and the other generally is exposed to oppressive policies. Unfortunately, no such identical cases exist in the real world (Holland 1986, 947). The next best option is to find cases in which the two minority groups' characteristics are as similar as possible, except for their differential policy treatments.

Asian Americans and African Americans on the West Coast are two of the most similar historical cases exposed to divergent policy treatments. Both groups were subject to

racial discrimination on the West Coast for decades. Racial discrimination against Asian Americans started in the late 19th century through a white supremacist alliance in Congress comprising Southern Democrats and Western Republicans (King and Smith 2005, 88). The Chinese Exclusion Act of 1882 singled out the Chinese as the first racially excluded group under U.S. immigration policy (Lee 2003; Ngai 2014). The California Alien Land Law of 1913 then deprived Asian Americans of property rights except those with birthright citizenship. Racial discrimination against African Americans soon appeared in the region as a large number of African Americans moved west during the Second Great Migration. Even though booming industrial cities on the West Coast provided unprecedented economic opportunities for African Americans, the postwar economy clearly had a racially charged hierarchical structure (Self 2005; Sides 2006).

More importantly, the physical and social boundaries between African American and Asian American neighborhoods on the West Coast were permeable. When African Americans moved to the West during World War II, they found a rare housing opportunity in vacant houses as a result of the internment of Japanese Americans. For this reason, from Los Angeles to Seattle, African American neighborhoods were located adjacent to Asian American ethnic enclaves. This proximity also may explain why and how the Black Power and Asian American movements co-evolved in inner cities on the West Coast (Prashad 2002; Maeda 2005; Ho and Mullen 2008; Maeda 2012; Ishizuka 2016; Watkins 2012). The Red Guard Party, a radical Asian American organization founded in San Francisco in 1969, was modeled on the Black Panther Party established in Oakland across the Bay in 1966. Like the Black Panther Party, the Red Guard Party had a 10-point program and ran a free breakfast program, and their members wore berets (Maeda 2005, 1079-1089).

Despite close social and political networks, along with their concurrent emergence and growth, these Asian American and African American political movements did not form an interracial coalition. Student activists of color focusing on racial justice certainly were present on college campuses (Umemoto 1989; Wei 1993; Maeda 2012), but outside university

9

environments, counterexamples abound (Brilliant 2010, 238-240). Interracial coalitions that emerged on college campuses were the exception rather than rule.

Recent historical studies and my own archival research show how the wide spectrum of policies introduced in the 1960s created the source of the discord. President Johnson expanded both oppressive and beneficial policies during the 1960s, and these policies raised different kinds of challenges for African Americans and Asian Americans.

The main policy challenge for African Americans was reducing the extent of state oppression. Although President Nixon was credited with spearheading the War on Crime in the U.S., a recent historical study revealed that Johnson already has sown the seeds of mass incarceration by expanding federal assistance to militarize policing through the establishment of the Office of Law Enforcement Assistance in 1965 to fight a war "within our boundaries" in inner cities across the U.S. (Hinton 2015, 102-105). For African American migrants from the Jim Crow South who migrated to the West, these punitive policies were all familiar. When African American activists protested the wrongful convictions of three African American inmates at California's Soledad Prison in 1970, they rallied around the following slogan: "Free the Soledad Brothers from Legal Lynching."[8]

Asian Americans wrestled with a different kind of policy challenge—expanding state support. The Immigration and Naturalization Act of 1965 resumed mass immigration from Asia. During the first wave of mass Asian immigration in the late 19th century, most Asian immigrants were cheap laborers, but the new wave of immigrants in the post-1960s included middle-class and college-educated people. Housing discrimination against Asian Americans weakened as the Cold War began, and some Asians, especially Chinese Americans, were viewed as allies (Brooks 2009). Taking advantage of this trend, these well-off Asian Americans bypassed old, run-down ethnic enclaves in inner cities and built new ethnic hubs in suburban areas (Kwong 1996; Lai 2003; Li 2006; Zhou 2010). Consequently, those left behind in old Chinatowns, Japantowns, and Manilatowns suffered from deteriorating

---

8. Greaves, Kay, "Davis Bail is Canceled; Poindexter Out on Bail," *Oakland Post*, October 22, 1970: 13

neighborhood conditions and desperately needed outside support. Johnson's other domestic political agenda, the War on Poverty programs, provided the kind of support that Asian American activists in these communities sought but with one serious limitation. From the perspective of government officials, Asian American communities were too small and fragmented (Kuramoto 1976). Michio Suzuki, deputy commissioner of the Administration for Public Services in Washington D.C., warned Asian American social service providers that they should "act together" because federal officials were unhappy when they had to fund various Asian groups separately as a result of a lack of coordination between them.[9] Asian Americans were a less-ideal subject for federal social policies, as the cost of delivering service to them was too high and the total number of people that the program served was too small. Thus, after decades-old exclusion from the state's social protection, Asian Americans at last received an opportunity to gain access to government resources, yet federal grant makers did not recognize them as a viable minority group.

The controlled case comparison (Skocpol and Theda 1979; Slater and Ziblatt 2013) helps reduce alternative explanations, such as neighborhood backgrounds or political ideologies. Nevertheless, some unobserved differences between the groups remain, and they might have influenced the relative production of collective gain and loss articles by Asian American and African American newspapers. For this reason, this approach is not suitable for causal inference. Rather, the aim of this case selection is mainly to strengthen descriptive inference from a large-scale collection of ethnic newspaper articles.

4 Data

Of the many newspapers and magazines written by and for Asian Americans and African Americans, I use the records of *Asian Week* (1983-1989), *International Examiner* (1976-1987), *The Sun Reporter* (1968-1979), *The Oakland Post* (1968-1981), and *The Sacramento*

---

9. Chin, Karen, "Pacific/Asian Elderly Conference: Social Service Providers Must Get Their 'Act Together'," *International Examiner*, April 30, 1979.

Observer (1968-1975). I selected these newspapers because they were all founded in inner cities on the West Coast and part of mobilization networks in their respective communities. They devoted substantial pages to describing community affairs and frequently shared news about community organizations. With the exception of the *International Examiner*, all of the other newspapers are based in Northern California, particularly the Bay Area, but the addition of a Seattle ethnic newspaper would make little difference as Asian American communities on the West Coast have shared close migration, settlement, and political development trajectories. For instance, most of the early Chinese settlers in San Francisco and Seattle came from a few counties in Guangdong province in China, and they considered the region between California and British Columbia the "Gold Mountain" (*Gam Sann* in Cantonese) (Chan 1983). These settlers founded the Chinese Benevolent Association, an umbrella organization made up of Chinese family and regional associations, in San Francisco in 1882 and in Seattle in 1892. Since the 1960s, these Chinese communities, both in San Francisco and Seattle, have been forming coalitions with other Asian ethnic groups to deal with their common political issues, such as gentrification and affordable housing, and the *International Examiner* and *Asian Week* emerged in these historical contexts (Kim 2020).

International Examiner, founded in Seattle in 1974, and *Asian Week*, based in San Francisco in 1979, were the two leading Asian American newspapers on the West Coast. Unlike other Asian ethnic newspapers that catered to national origin groups (e.g., Chinese, Japanese, Filipinos), these newspapers aimed to represent the united voice of Asian Americans. *Asian Week* was the first English-language Asian American publication and the largest until it closed in 2012 (Ling and Austin 2015, 29-31).*International Examiner* is the oldest and largest non-profit Asian American newspaper in the Northwest. The newspaper closely followed news related to local community-based organizations and paid special attention to the growing involvement of Asian Americans in electoral politics.

The use of English-langauge Asian newspapers carries immigration status, class, and political implications. Asian-language newspapers focused on home country politics. For

instance, during the first half of the 20th century, Chinese journalism in the U.S. mainly focused on the political situation back in China. The 1965 immigration legislation resumed a wave of mass immigration from Asia and increased the foreign-born Asian population. According to U.S. Census data, the percentage of foreign-born residents among the Chinese population was 52%, while the Japanese were 21%, and the Filipinos were 52% in the 1970s (Hirschman and Wong 1981, 499). The increase in the Asian population in the U.S. also led to the founding of new Asian-language newspapers. Asian-language newspapers, which were founded by major newsgroups in Asia, targeted these relatively affluent members. By contrast, English-language Asian newspapers such as *Asian Week* and the *International Examiner* focused on U.S. politics, with an emphasis on how the U.S. government treated its marginalized community members in traditional ethnic enclaves. Both newspapers paid close attention to policy support for the poor and elderly who remained in Chinatowns, Japantowns, and Manilatowns (D. Chin 2001; G. Chin 2015).

Similar to its Asian American counterpart, the African American corpus in the data was politically progressive. The *The Sun Reporter* was founded in San Francisco in 1944 as the city's first Black press. The founder, Dr. Carlton B. Goodlett, was an African American civil rights activist and former president of the local branch of the National Association for the Advancement of Colored People. Thomas Fleming, who worked as an editor, reporter, and columnist for the newspaper for 61 years from its founding, also held a strong political commitment to racial justice. He once argued that the mission of the Black press was to serve as the watchdogs in order to guarantee the constitutionally granted citizenship status of former chattel slaves [10]. *The Oakland Post* was founded in 1963 and was politically less vocal, as the newspaper intentionally focused on not making a social statement but instead just covering the news to increase its readership and thus its impact [11]. Nevertheless, as Oakland was a main center of African American political mobilization on the West Coast, it

---

10. Fleming, Thomas,"Thomas Fleming's Weekly Report," *Sun Reporter*, August 2, 1975:7
11. Berling, Lynn, "'Post' Tries to be Only Daily for Black Community'," *Oakland Post*, February 15, 1981: 6

was almost impossible for the newspaper not to address the topic of racial justice. Between 1968 and 1981, the newspaper covered The Black Panther Party 192 times. *The Sacramento Observer* was founded in Sacramento in 1962 and supported the Sacramento Urban League and the Sacramento Area Black Law Caucus. The newspaper also played a particularly important role in informing and alerting the African American community in Sacramento about the development of the Civil Rights and the Black Power Movements in other parts of the country (Covin 2009, 25). Other key African American newspapers on the West Coast, such as *Los Angeles Sentinel*, were not included, as the only articles published by that newspaper after 1991 were available in the database.

I downloaded data from these five newspapers from the Ethnic NewsWatch database, which has compiled more than 2.5 million articles published in U.S. ethnic newspapers and magazines.[12] ProQuest created this database and does not allow web scraping. However, one can still download articles and save them as hypertext markup language (HTML) files. I have turned these HTML files into a single dataset by using the Beautiful Soup library in Python. At first, there were more than 80,000 articles. In the data-cleaning process, I discovered that data from *The Sun Reporter* from early 1973 to late 1974 did not include the complete articles, only their abstracts. For this reason, I omitted such data. This reduced the total number of articles to 78,383. Table 2 summarizes the number of newspaper articles in the corpus.

Title	Asian Week	The International Examiner	The Sun Reporter	The Oakland Post	The Sacramento Observer
Period	1983-1989	1976-1987	1968-1979	1968-1981	1968-1975
N	11,883	2,719	24,793	29,560	9,428

Table 2: Number of each newspaper articles in the corpora

---

12. For more information, see https://www.proquest.com/products-services/ethnic_newswatch.html.

5 Empirical Strategies

The empirical strategy's goal was to find a systematic and cost-effective way to measure every article based on the conceptual framework proposed above. The integrated classification framework achieved this goal by examining the relationship between training data and automated text classification using content analysis (Berelson 1952). What classifying algorithms do is mimic human coder behaviors. Because the machine can follow only where humans lead, human-coding quality influences the quality of automated text classification. Therefore, it is crucial to assess the quality of human coding and its influence on machine learning performance and interpretations. Figure 1 describes the workflow from selecting cases to testing the hypotheses.
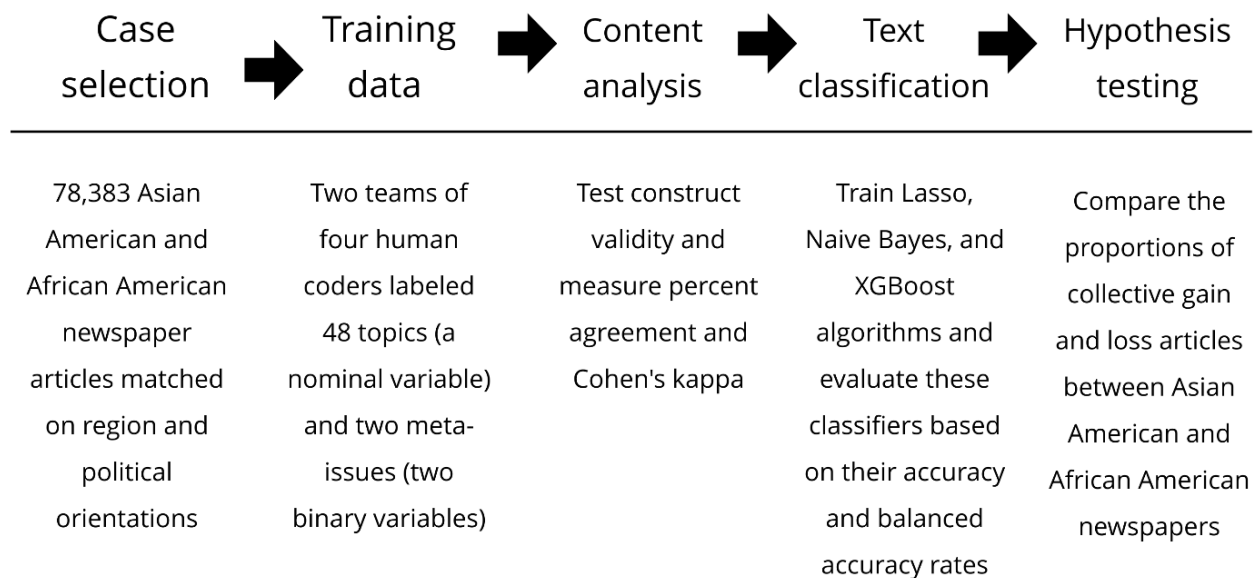
Case selection	Training data	Content analysis	Text classification	Hypothesis testing
78,383 Asian American and African American newspaper articles matched on region and political orientations	Two teams of four human coders labeled 48 topics (a nominal variable) and two meta-issues (two binary variables)	Test construct validity and measure percent agreement and Cohen's kappa	Train Lasso, Naive Bayes, and XGBoost algorithms and evaluate these classifiers based on their accuracy and balanced accuracy rates	Compare the proportions of collective gain and loss articles between Asian American and African American newspapers

Figure 1: Workflow

5.1 Training Data

I hired four undergraduate research assistants and coded the training data based on the following procedures. The human coders labeleded two meta issues and a list of topics because these topics are useful for testing construct validity. Throughout these procedures, none of the human coders were informed about the research hypotheses.

15

1. Detecting topics and distributing articles: I employed topic modeling using the stm package in R (Roberts, Stewart, and Tingley 2015) to inductively discover topics from each newspaper. I randomly divided these topics (n = 48) into two parts and assigned 100 articles from each topic in the first part to one team of two human coders and 100 articles from each topic in the second part to another team. I asked the human coders to label each topic based on these articles without consulting the other team member.

2. Topic coding: The two teams spent two weeks coding topics and then another week agreeing on the common labels through intergroup discussions. In the process, the human coders created a list of topics related to the articles.

3. Meta issue coding: After the topic coding, I randomly selected 1,008 articles from the Asian American corpus and 1,008 articles from the African American corpus stratifying on the year variable. Year was selected as a stratifying variable because key issues may change over time. I then paired the human coders into two groups again and assigned one group to the Asian American corpus sample and the other group to the African American corpus sample. To avoid demand effects, I did not inform human coders regarding the hypotheses I attempts to test. Each team coded whether the articles in the sample were about promoting collective gains (yes = 1, no = 0) or preventing collective losses (yes = 1, no = 0). The process created two binary variables—collective gain and collective loss. I do not create one binary variable and use collective gain and loss as its two exclusive classes because this ignores mixed or none classes identified in Table 3. Creating two separate binary variables is a more reasonable approach. In doing so, the human coders also coded each article with a relevant topic from the list they created in step 2.

	Collective gain	Non-collective gain
Collective loss	Mixed	Exclusive collective loss
Non-collective loss	Exclusive collective gain	None

Table 3: Conceptual Matrix

5.2   Content Analysis

I assessed the quality of the human-coded training data by measuring inter-coder reliability and construct validity.

Percentage agreement measures the percentage of the agreed coding decisions made by pairs of coders. The calculation of percentage agreement is fairly simple. Suppose two human coders measure a binary variable. Subtracting the values recorded by coder 1 from the values recorded by coder 2 returns some 0s. The number of 0s divided by the number of units provides the percentage agreement (McHugh 2012, 278). Therefore, percentage agreement is an intuitive method to assess the accuracy of human coding. Nevertheless, it is also a crude one, as it does not account for a certain degree of agreement that would simply arise by chance.[13]

Cohen's kappa coefficient provides measures for the degree of reliability in inter-coder agreement. Suppose p_a indicates the proportion of actual agreement and p_e the proportion of chance agreement. The difference between the two quantities, $p_a - p_e$, represents the proportion of units in which the agreement occurred beyond chance. $1 - p_e$ represents the remaining units when the chance agreement is excluded (Cohen 1960, 39-40). The coefficient (k) or kappa is defined as $k = \frac{p_a - p_e}{1 - p_e}$. The kappa can range from -1 and +1 and 0. A high kappa score is desirable because it represents a low degree of faulty evidence in the training data.[14] I also check whether the kappa score increases when I exclude the articles covering

---

13. Several studies have demonstrated how this measure tends to overestimate the true agreement among human coders (Birkimer and Brown 1979; Suen and Lee 1985; Lombard, Snyder-Duch, and Bracken 2002).

14. In practice, a kappa smaller than or equal to 0 indicates no agreement, a kappa in the 0.01-0.02 range indicates slight agreement, a kappa in the 0.21-0.40 range indicates fair agreement, a kappa in the 0.41-0.60 range indicates moderate agreement, a kappa in the 0.61-0.80 range indicates substantial agreement, and a kappa in the 0.81-1 range indicates an almost perfect agreement (McHugh 2012, 279).

non-political topics in the training data because these articles could confuse human coders and decrease inter-coder reliability.

Construct validity is about whether the measures measure what they are supposed to measure based on the underlying theory (Cronbach and Meehl 1955). At the conceptual level, collective gain and loss labels and topic labels are closely associated because the meta issues should represent these topics. Articles coded as collective gain are expected to cover topics on beneficial policies (convergent validation) more likely than oppressive government policies (discriminant validation). Articles coded as collective loss should behave in an opposite way (Campbell and Fiske 1959). I test these assumptions by calculating the difference between the number of collective gain articles and that of collective loss articles associated with particular topics.

Finally, I examine the relationship between reliability and construct validity by examining how these differences vary by measurement decisions. At the minimum, articles could be defined as collective gain or collective loss if one of the two human coders said so (minimum threshold). At the maximum, articles could be defined as such if all the human coders agreed (maximum threshold). The minimum threshold is a naive approach, and the maximum threshold provides more reliable data.

5.3 Text Classification

Using the coded articles, I trained and tested the least absolute shrinkage and selection operator (Lasso), naive Bayes, and extreme gradient boosting (XGBoost) algorithms. I put 70% of the sample articles into the training set and the rest into the test set by using the scikit-learn library in Python, again stratifying on the year variable (for more information, see Appendix C).

For the pre-processing, I tokenized the documents, removed special characters and white space, and turned these tokens into lower case. For the feature extraction, I used the bag-of-words model (Harris 1954). It is a simple representation of the text because it only

counts word frequencies in each article. These term frequencies are used as features to train algorithms and make predictions.[15] I also construct n-gram, a sequence of n items from a document, with a maximum length of two, because some key words, such as civil rights, are sensible in bigrams.

I evaluate machine learning performance by examining accuracy and balanced accuracy rates. The absolute difference between the two measures shows the effect of imbalanced training data on prediction accuracy. If the difference is substantially large, I address this problem by randomly oversampling the minority class in the training data with replacement (upsampling) and examining performance improvement using the accuracy and balanced accuracy rates. I also check the extent to which the machines performed better or worse than the human coders by comparing the percentage agreement and the two machine learning performance measures. I select a classifier that shows stable high performance (Yu 2013), I predict the unlabeled data, and I test the hypotheses by comparing the proportion of collective gain and loss articles in the Asian American and African American corpora.

6   Content Analysis

6.1   Percentage Agreement

To begin with, the percentage agreement displays the high reliability of the labels. Figure 2 shows that the inter-coder agreement reached 88% for the collective loss articles in both the African American and Asian American corpora. The metric is slightly lower for the collective gain articles: 7% down for the African American newspaper and 8% down for the Asian American one. However, this difference is marginal.
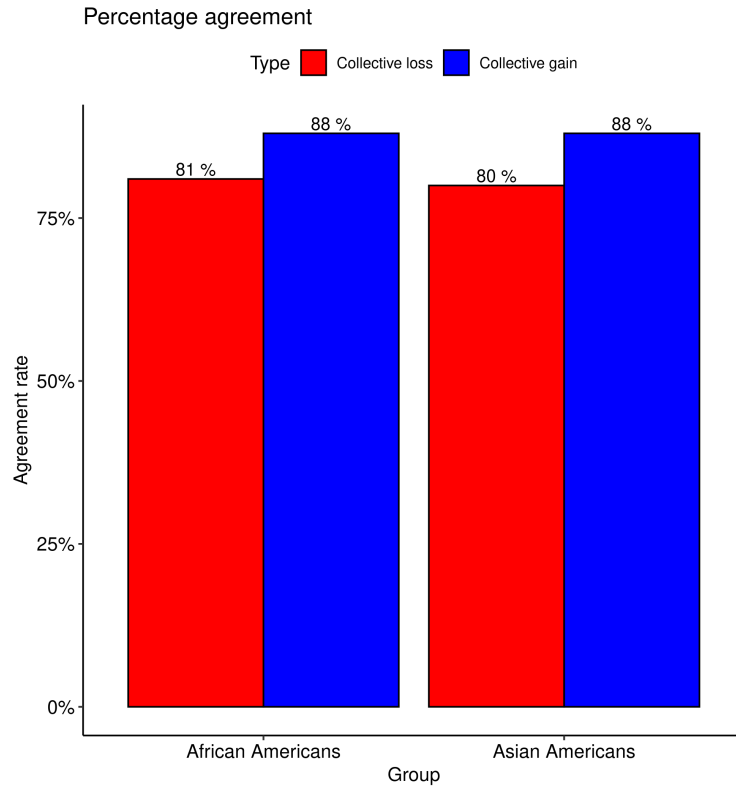
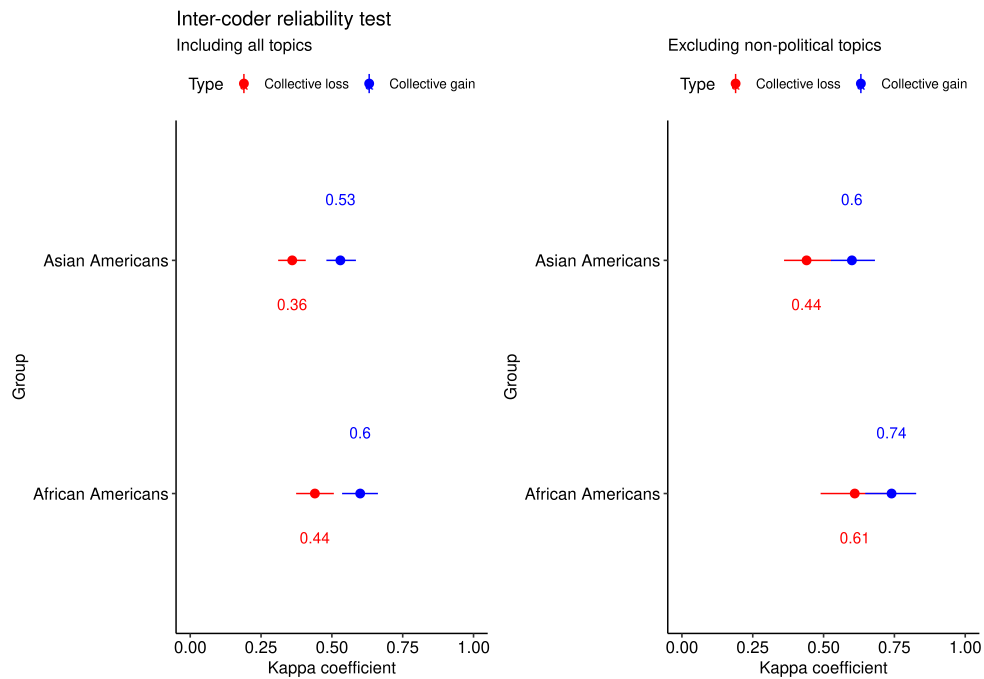Figure 2: Percentage agreement for collective gain and collective loss articles

Figure 3: Cohen's kappa with or without non-political topics

6.2 Cohen's Kappa

As the percentage agreement does not account for the stochastic element of the coding process, it is time to turn to Cohen's kappa. The kappa score is relatively higher for the collective loss articles in both newspapers (see the left panel in Figure 3). In addition, the labels in the African American newspapers are more reliable than those in their Asian American counterparts. Non-political issues, such as food and sports, could confuse human coders and increase disagreement between them. The right panel in the figure demonstrates how removing these articles increases inter-coder reliability. The results indicate that the training data are less reliable for non-political articles. Because unreliable data lead to weak prediction accuracy, I excluded articles related to sports, car sales, arts, recipes, and restaurants from the unlabeled data using a dictionary method. The specific terms used to create the dictionary are listed in Appendix B. This decision removed 46% (n = 6,831) of the Asian American and 44% (n = 32,820) of the African American corpus. I did not use this approach to exclude every non-political issue in the data because the dictionary method is insensitive to contexts. For instance, because churches have played an important role in African American politics (McDaniel 2009), some articles that are related to Christian advice and church events also might mention politics. When the *Sun Reporter* described the political struggles of Black people in Cairo, Illinois on March 27, 1981, it used deeply religious terms[16]:

> Millions of people across America are aware of the plight of Cairo's poor and
> Blacks because of an elaborate communication system developed by the front
> and because the Rev. Koen and other leaders of the front have traveled hundreds
> of thousands of miles to tell the Cairo story. The most important part of the
> struggle is found within its religious dimensions. The Black people of Cairo relate

---

15. I only used the top 5,000 most frequently appearing terms because Zipf's law expects frequently appearing features in documents to be a small fraction (Zipf 1936, 1949). The rest of the features will only increase sparsity in the training data and slow down the algorithmic process.

16. Anonymous, "Cairo, Illinois: From Exploitation To Freedom," *Sun Reporter*, March 27, 1971: 8

to a higher power.

The same argument can be made about Asian stereotypes, as overcoming the image of the model minority has been a key political issue in the Asian American community. For example, the *International Examiner* criticized the limited supply of social services for Asian Americans in Seattle's International District on December 31, 1979. In the article, the newspaper highlighted why being perceived as a model minority is a political disadvantage[17]:

> In 1969, there was no I.D. health clinic, no Asian Counseling and Referral Service, no Asian American Studies program, no Asian community newspapers, no Asian American Commission, no Asian elderly program, and no Asian American theatre. There was also no Kingdome. Low income housing for Asians was still a concept. Asians were the nice, quiet, studious, *model minority* [Italics added by the author] with no problems.

Similarly, no clear keywords exist that are related exclusively to health advice or lifestyle and not associated with social policies. To avoid these pitfalls, I did not remove articles related to Christian advice, Church events, Asian American stereotypes, health advice, and lifestyle topics using the dictionary method.

6.3   Construct Validity Test

The next step is to check construct validity. So far, we have tested how reliably the human coders labeled the training data. Now, we turn to whether the measures really measure what they are designed to measure. The theory assumes that collective gain articles are likely to be associated with beneficial policies. In contrast, collective loss articles are likely to be related to oppressive policies.

The analysis confirms that meta and specific issues hang together, as expected by the theory. This relationship holds regardless of whether one uses the maximum or minimum

---

17. Iwamoto, Gary, "A Picture of the 70's," *International Examiner*, December 31, 1979: 8

threshold to define the meta issues in the training data. In Figure 4, the X-axis is the difference between the number of collective gain articles and that of collective loss articles belonging to the identical topic. This calculation removes those articles labeled by a team of the two human coders as both collective gain and loss. I did not involve them in the analysis because their expected content is unclear. The Y-axis indicates these topics. The red bar plot indicates that the maximum threshold is used to define collective gain and loss labels. The blue bar plot indicates the use of the minimum threshold for the measurement. In the figure, the topics, a collection of documents, strongly associated with collective gain articles are at the top, and the topics highly related to collective loss articles are at the bottom.

In the Asian American case, healthcare and political representation are the top two issues related to collective gain, and hate crimes and Asian stereotypes are the top two issues associated with collective loss. After independent topic coding, the human coders also provided a brief summary of each topic. Based on these summaries, on the Asian American side, healthcare was about the physical and mental well-being of people of Asian descent, which consisted of access to health education, health services, and availability of translators and translations to access these services. Political representation was about the potential increase in political engagement of Asian Americans in politics at the local, state, and federal levels, especially in gaining recognition in U.S. Census data collection. Hate crimes were about instances of violence or discrimination against Asian Americans, and Asian American stereotypes were about generalizations of Asian men and women and cultural perceptions of outsiders as well as those within the Asian community. Issues about increasing government support were strong indicators of the collective gain articles, and issues about reducing outside threat, although not necessarily oppressive policies, were leading signs of collective loss articles.

The African American newspaper follows the same pattern. School programs were about public school programs and administration in the San Francisco Bay Area. This issue was counted as a top issue related to collective gain articles. Criminal justice was about dis-

crimination against African American defendants who were given unfair trials, legislators excluded from important committees, and citizens with higher imprisonment rates. Similarly, civil rights were about the overlapping topics of federal employees' failure (in terms of neglect of duty, abuse of power, or discrimination) and proposed laws, some of which would disproportionately affect African Americans. These were the top two issues related to collective loss articles.
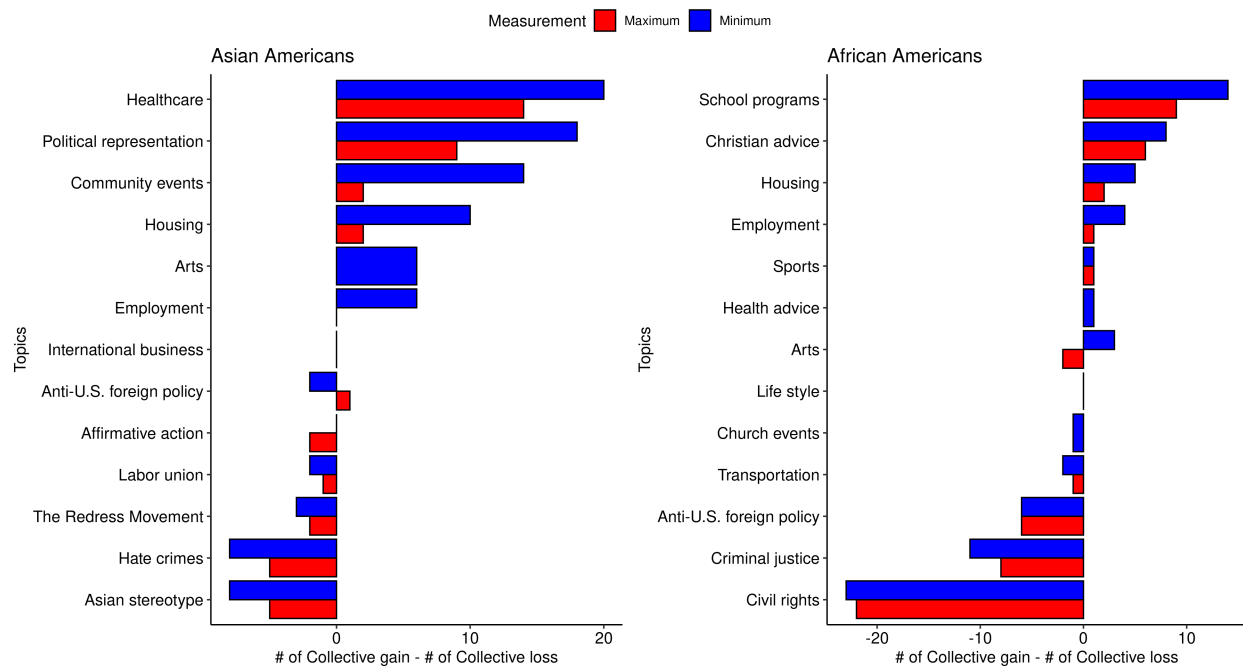

Figure 4: Construct validity test

One noticeable difference between the two corpora is the extent to which the threshold change affects the proportion of collective gain and loss articles. Using the maximum threshold increases the proportion of collective gain articles by 300% and that of collective loss articles by 140% in the Asian American corpus. The same change moves up the proportion of collective gain articles by 200% and that of collective loss articles by 46% in the African American corpus (see Figure 5). The Asian American corpus displays a greater degree of variability. This pattern is expected because the labels in the Asian American corpus are less reliable than their African American counterparts on the basis of their kappa scores. The content of the labels changes little regardless of whether the training data are measured
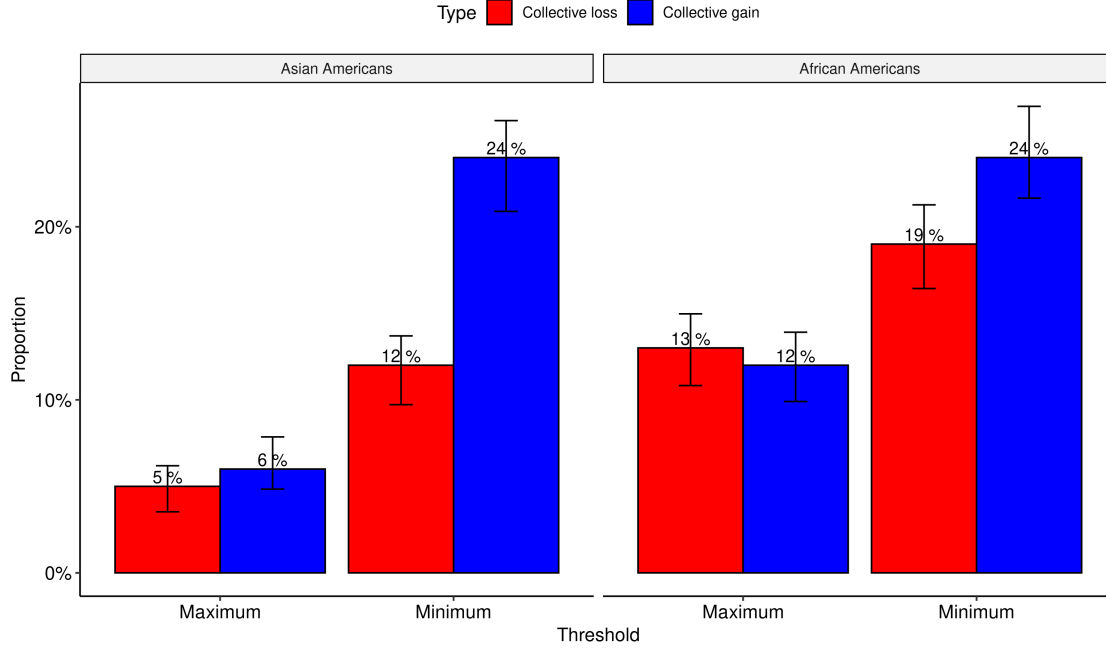
Figure 5: Threshold change and the proportion of collective gain and loss articles

using the minimum or the maximum threshold.

7   Automated Text Classification

Automated text classification scales up content analysis using machine learning. In Figure 6, the X-axis indicates the accuracy or the balanced accuracy rate. The Y-axis indicates different classifiers. The dotted lines represent the percentage agreement between the human coders.

The figure demonstrates that more reliable and balanced data produced better prediction outcomes. The worst performance is found in the bottom left panel, where the training data are measured using the minimum threshold and no resampling is used. The average accuracy rate is 77%, but the average balanced accuracy rate is 63%. The best performance is found in the top right panel, where the training data are measured using the maximum threshold and upsampling is used. The average accuracy rate goes up by 19%, and the average balanced accuracy rate moves up by 32%. Both Lasso and XGBoost algorithms

exhibit similar high performance. However, when these classifiers were used to predict the unlabeled data, XGBoost did not perform well. For example, when the training data were measured using the maximum threshold and upsampled, the algorithm was able to classify only four Asian American and 171 African American collective loss articles. This prediction is likely a strong underestimate of the true count of collective loss articles. For instance, hate crimes are a strong indicator of collective loss articles among Asian American newspapers (see Figure 4). Asian American newspapers all highlighted how an anti-Asian attitude led to the murder of Vincent Chin. *Asian Week* covered the Vincent Chin case 212 times and *International Examiner* 40 times between April 1983 and December 1989. These articles argued that Vincent Chin's murder was a hate crime. Even if we only counted these articles, the number of relevant cases for collective loss should be 63 times larger than the predicted estimate. I selected Lasso to classify the unlabeled data because it demonstrated the most stable high performance.
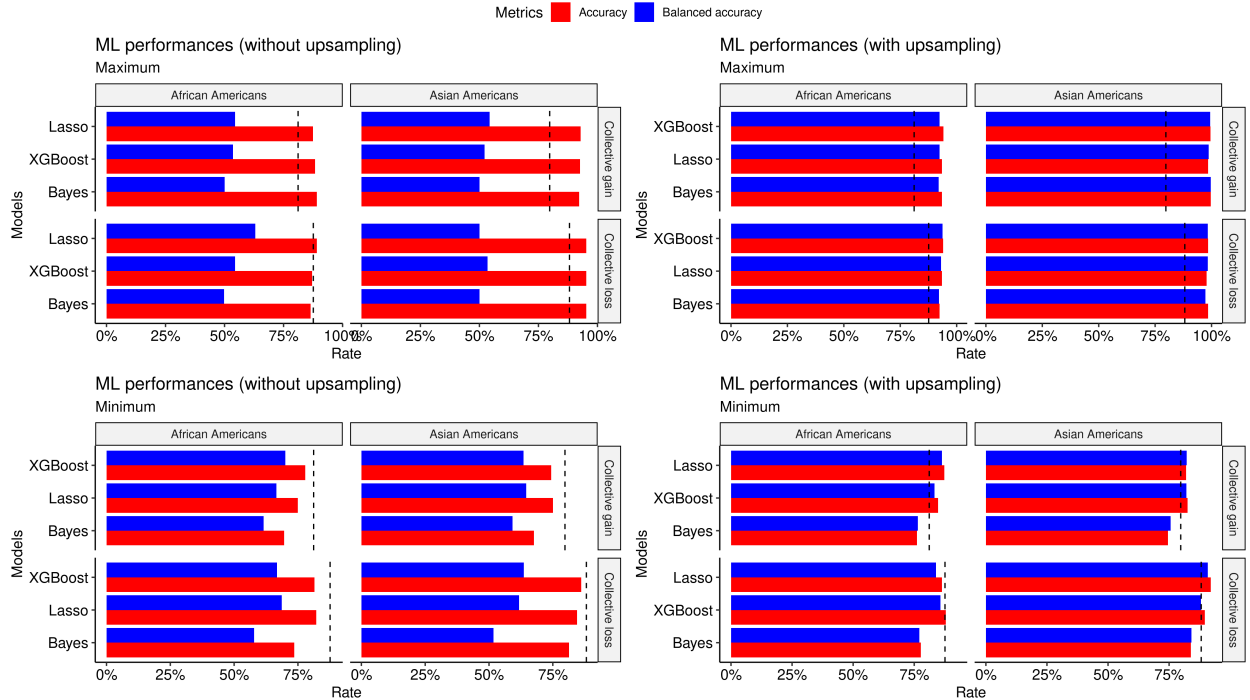

Figure 6: Classifier performance evaluations

Data quality affects not only prediction accuracy but also substantive results. Figure 7

displays how the proportion of collective gain and loss articles varies between the two corpora over time. In this case, the training data are measured using the maximum threshold. The Y-axis shows the percentage of articles in the corpus classified as the given meta issue type, and the X-axis indicates either publication years or months. The ribbons indicate 95% confidence intervals. The blue line indicates the proportion of exclusive collective gain articles, the red line indicates the proportion of exclusive collective loss articles, and the purple line indicates the proportion of articles classified as both collective gain and loss. Overall, the blue line is above the red line in the Asian American case, and the pattern is reversed in the African American case. Put differently, Asian American newspapers issued collective gain articles far more frequently than collective loss articles. This pattern is reversed in the African American case. In addition, the purple is below the blue line in the Asian American case and below the red line in the African American case. In either case, articles classified as both collective gain and loss articles did not take up the largest proportion.
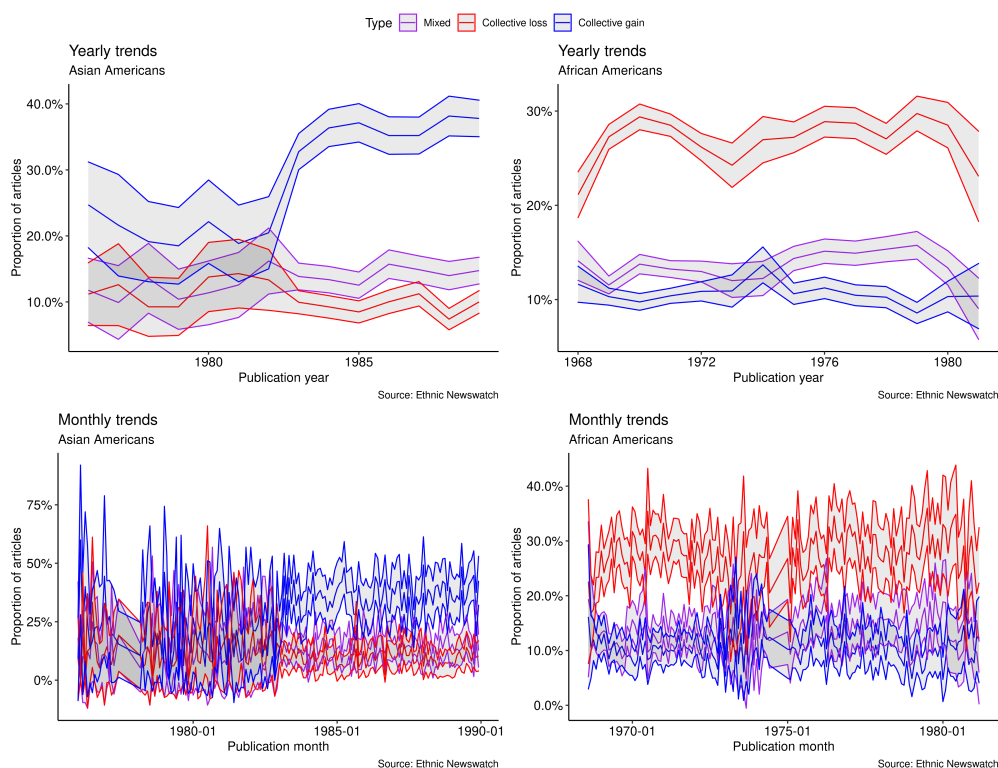

Figure 7: Time series trends (maximum threshold)

27

One limitation with the naive comparison between the two corpora is the fact that Asian American and African American newspapers cover different time periods. The Asian American corpus covers the 1976-1989 period, while the African American corpus covers the 1968-1981 period. This mismatch might cause issue priority differences in the two corpora. To address this concern, I matched the two data sets on their publication years (1976-1981) in the right panel and did not do the same in the left panel in Figure 8. This decision removes articles issued by *Asian Week* (1983-1989). The X-axis is the group, and the Y-axis is the proportion of exclusive collective gain and loss articles. I dropped the mixed category to simplify the comparison. The error bars indicate 95% confidence intervals. To begin with, matching their publication years does not change the general pattern that appeared in the two corpora. In both panels, Asian American newspapers clearly preferred collective gain articles, whereas African American newspapers preferred collective loss articles. As the differences are wider than the confidence intervals, the patterns are statistically significant. Nevertheless, the different time frames affect the magnitude of the differences between the two panels. Matching them on their publication years decreases the absolute difference in the average proportions of exclusive collective gain articles between these corpora from 30% to 18.5% and that of exclusive collective loss articles between them from 24% to 23%. In addition, measuring the training data using the minimum threshold widens the gap between the two corpora. When the threshold shifted to the minimum, Asian American newspapers exclusively reported on collective gain up to three times more than African American newspapers did. Conversely, African American newspapers exclusively covered collective loss up to 10 times more than their Asian American counterparts.

Finally, some unstable patterns deserve further explanations. In Figure 7, the proportion of exclusive collective gain articles in the Asian American corpus surged in the early 1980s. This increase is a likely result of combining two different data sources. Figure E.1 in Appendix E shows that the proportion of exclusive collective gain articles is higher in *Asian Week* than in the *International Examiner*. This difference does not simply derive from the fact that their
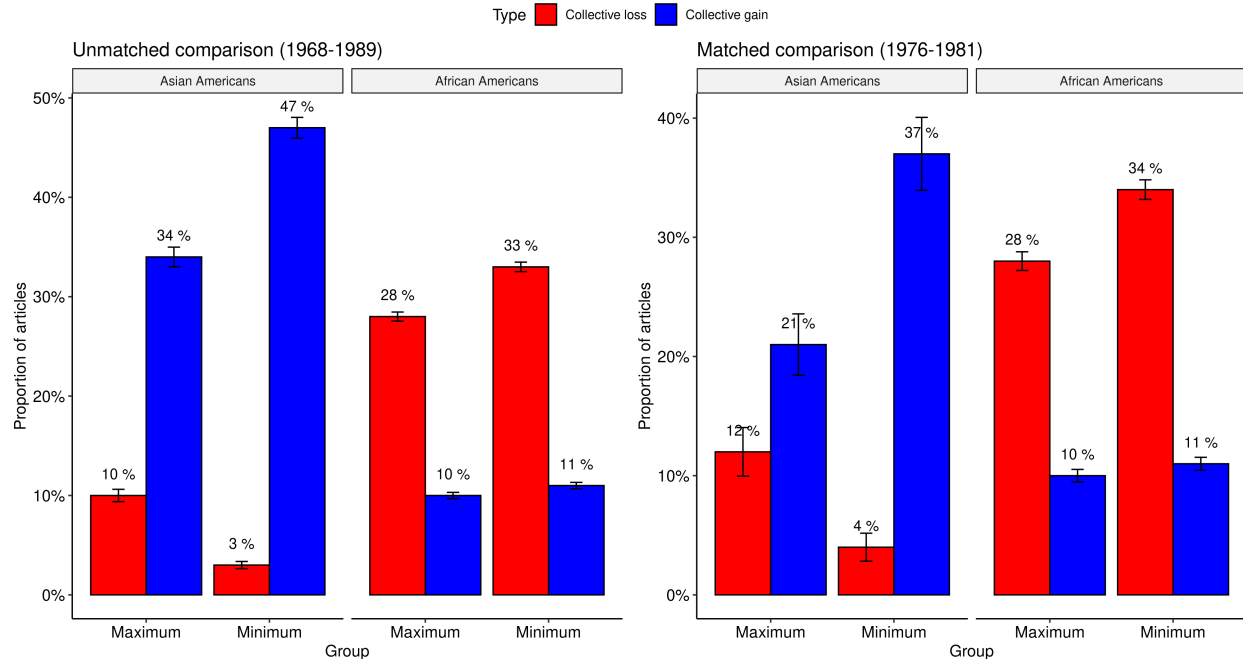
Figure 8: How text classification is sensitive to measurement decisions

publication years are different. Even if we matched the newspaper data on their publication years (1983-1987), the average proportion of exclusive collective gain articles in *Asian Week* is 40% greater than the same measure in the *International Examiner*. The two data sources were combined in the time series plot starting in 1983, which explains the abrupt increase in the proportion of exclusive collective gain articles appearing around that cut-off point. In addition, this surging trend seems conditional on the measurement decision, as it does not appear when the training data are measured using the minimum threshold (see Figure E.2 in Appendix E).

8 Discussion and Conclusion

Textual data are attractive because they allow political scientists to analyze politicians and citizen's words—the content of democracy (Brady 2019, 311). Nevertheless, the application of the method has been uneven, as most cases are limited to the study of the words of the powerful, such as U.S. congress members (Hopkins and King 2010; Grimmer, Messing, and

Westwood 2012). The voices of marginalized members have rarely been examined systematically with large-scale text analysis in political science. This study fills such a gap by setting an example of how we can apply machine learning to investigate how political issues vary among racial minority groups over time.

The findings demonstrate why careful and systematic content analysis is essential for automated text classification. Content analysis is labor intensive. More importantly, although content analysis is based on a probability sample, the descriptive statistics drawn from the analysis are still prone to error, as the sampling variance is unknown. Automated text classification is an efficient alternative to this traditional method, as machine learning algorithms can label a large collection of documents at an impressive speed. In addition, sampling variance is no longer a concern, as the analysis is based on the entire corpus. Nevertheless, a healthy dose of skepticism is still useful in embracing this new method. Uncertainty exists regarding how machine learning algorithms predict labels and how researchers can interpret the results. The findings clearly illustrate that unreliable training data lead not only to weak predictions, but also to extreme interpretations. When the maximum threshold was used to label the training data, the difference in issue priorities between Asian American and African American corpora were moderate. When the maximum threshold was used, the gap between the two corpora widened up to 10 times.

Although purely descriptive, the original large-scale evidence provides new insights into the differences in the political issues between racial minority groups. Whereas the African American corpus focused on opposing state oppression, the Asian American corpus concentrated on expanding state support in the 1970s—the period when their publication years were overlapped. The findings are limited in their temporal scope because U.S. policies have undergone important changes since the 1980s. Social policies have become subject to neoliberal reforms (Danziger and Haveman 1981; Wolman 1986; Soss, Hacker, and Mettler 2007; Pierson 2003), and criminal justice policies have further strengthened U.S. state's capacity to disproportionately criminalize and incarcerate African Americans (Alexander 2012;

30

Murakawa 2014; Gottschalk 2016; Grumbach 2018). Consequently, the exact differences in the political issues appearing in Asian American and African American newspapers might have changed, but the general theoretical insight is still valuable. Racial minorities are often coded collectively as people of color because they share racial marginalization. Race is not just all about the color of one's skin; it is a social construct often created by an oppressive historical process (Omi and Winant 1994; Dawson 1994b). Nevertheless, sharing the marginalization experience is often insufficient to form shared political agendas among racial minority groups (Kaufmann 2003, 201). From a distant perspective, the issues of non-White groups seemed similar, because they were rooted in such groupss disadvantaged statuses in U.S. society. However, in taking a closer look at their issues, we realized that these groups ranked their political issues in different ways because different racial groups had to deal with different policy challenges (Brilliant 2010).

Declarations

Funding

Not applicable

Availability of data and material

Not available due to copyright restrictions

Conflicts of Interest

The author states that there is no conflict of interest.

Code Availability

All replication files can be found at https://github.com/jaeyk/content-analysis-for-evaluating-ML-performances.

References

Alexander, Michelle. 2012. *The New Jim Crow: Mass Incarceration in the Age of Colorblindness.* The New Press.

Bailey, Martha J, and Sheldon Danziger. 2013. *Legacies of the War on Poverty.* Russell Sage Foundation.

Barberá, Pablo, Amber E Boydstun, Suzanna Linn, Ryan McMahon, and Jonathan Nagler. 2019. "Automated Text Classification of News Articles: A Practical Guide." *Political Analysis:* 1–24.

Bartels, Larry M. 1999. "Panel Effects in the American National Election Studies." *Political Analysis* 8 (1): 1–20.

Bender, Emily M, and Batya Friedman. 2018. "Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science." *Transactions of the Association for Computational Linguistics* 6:587–604.

Berelson, Bernard. 1952. *Content Analysis in Communication Research.* Free press.

Beretta, Elena, Antonio Vetrò, Bruno Lepri, and Juan Carlos De Martin. 2018. "Ethical and Socially-Aware Data Labels." In *Annual International Symposium on Information Management and Big Data,* 320–327. Springer.

Birkimer, John C, and Joseph H Brown. 1979. "Back to Basics: Percentage Agreement Measures Are Adequate, but There Are Easier W ays." *Journal of Applied Behavior Analysis* 12 (4): 535–543.

Brady, Henry E. 2019. "The Challenge of Big Data and Data Science." *Annual Review of Political Science* 22:297–323.

Breiman, Leo. 1997. *Arcing the Edge.* Technical report. Technical Report 486, Statistics Department, University of California, Berkeley.

Brilliant, Mark. 2010. *The Color of America Has Changed: How Racial Diversity Shaped Civil Rights Reform in California, 1941-1978.* Oxford University Press.

Brodersen, Kay Henning, Cheng Soon Ong, Klaas Enno Stephan, and Joachim M Buhmann. 2010. "The Balanced Accuracy and Its Posterior Distribution." In *2010 20th International Conference on Pattern Recognition,* 3121–3124. IEEE.

Brooks, Charlotte. 2009. *Alien Neighbors, Foreign Friends: Asian Americans, Housing, and the Transformation of Urban California.* University of Chicago Press.

Campbell, Angus, Philip E Converse, Warren E Miller, and Donald E Stokes. 1980. *The American Voter.* University of Chicago Press.

Campbell, Donald T, and Donald W Fiske. 1959. "Convergent and Discriminant Validation by the Multitrait-multimethod Matrix." *Psychological Bulletin* 56 (2): 81.

Chae, David H, David T Takeuchi, Elizabeth M Barbeau, Gary G Bennett, Jane Lindsey, and Nancy Krieger. 2008. "Unfair Treatment, Racial/ethnic Discrimination, Ethnic Identification, and Smoking among Asian Americans in the National Latino and Asian American Study." *American Journal of Public Health* 98 (3): 485–492.

Chan, Anthony B. 1983. *Gold Mountain: The Chinese in the New world.* New Star Books.

Chen, Tianqi, and Carlos Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." In *KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 785–794.

Chin, Doug. 2001. *Seattle's International District: The Making of a Pan-Asian American Community.* University of Washington Press.

Chin, Gordon. 2015. *Building Community, Chinatown Style: a Half Century of Leadership in San Francisco Chinatown.* San Francisco, California: Friends of Chinatown Community Development Center.

Cohen, Jacob. 1960. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement* 20 (1): 37–46.

Covin, David. 2009. *Black Politics After the Civil Rights Movement: Activity and Beliefs in Sacramento, 1970-2000.* McFarland.

Cronbach, Lee J, and Paul E Meehl. 1955. "Construct Validity in Psychological Tests." *Psychological Bulletin* 52 (4): 281.

Danziger, Sheldon, and Robert Haveman. 1981. "The Reagan Budget: A Sharp Break With the Past." *Challenge* 24 (2): 5–13.

Dawson, Michael. 1994a. "A Black Counterpublic?: Economic Earthquakes, Racial Agendas, and Black Politics." *Public Culture* 7 (1): 195–223.

———. 1994b. *Behind the Mule: Race and Class in African-American Politics.* Princeton University Press.

———. 2001. *Black Visions: The Roots of Contemporary African-American Political Ideologies.* University of Chicago Press.

Denny, Matthew, and Arthur Spirling. 2017. "Text Preprocessing for Unsupervised Learning: Why it Matters, When it Misleads, and What to Do About It." *Political Analysis.*

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. "Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding." *arXiv preprint arXiv:1810.04805.*

Elish, Madeleine Clare, and Danah Boyd. 2018. "Situating Methods in the Magic of Big Data and AI." *Communication monographs* 85 (1): 57–80.

Espiritu, Le Yen. 1992. *Asian American Panethnicity: Bridging Institutions and Identities.* Temple University Press.

Fraga, Luis R, John A Garcia, Rodney E Hero, Michael Jones-Correa, Valerie Martinez-Ebers, and Gary M Segura. 2011. *Latinos in the New Millennium: An Almanac of Opinion, Behavior, and Policy Rreferences.* Cambridge University Press.

Freund, Yoav, and Robert Schapire. 1999. "A Short Introduction to Boosting." *Journal of Japanese Society For Artificial Intelligence* 14 (771-780): 1612.

Friedman, Jerome H. 2001. "Greedy Function Approximation: a Gradient Boosting Machine." *Annals of Statistics:* 1189–1232.

Friedman, Jerome, Trevor Hastie, Robert Tibshirani, et al. 2000. "Additive Logistic Regression: a Statistical View of Boosting." *The Annals of Statistics* 28 (2): 337–407.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2018. "Datasheets for Datasets." *arXiv preprint arXiv:1803.09010.*

Geiger, R Stuart, Kevin Yu, Yanlai Yang, Mindy Dai, Jie Qiu, Rebekah Tang, and Jenny Huang. 2020. "Garbage In, Garbage Out? Do Machine Learning Application Papers in Social Computing Report Where Human-labeled Training Data Comes From?" In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency,* 325–336.

Gitelman, Lisa, ed. 2013. *"Raw data" Is an Oxymoron.* MIT press.

Goth, Gregory. 2016. "Deep or Shallow, NLP Is Breaking Out." *Communications of the ACM.*

Gottschalk, Marie. 2016. *Caught: The Prison State and the Lockdown of American Politics.* Princeton University Press.

Grimmer, Justin, Solomon Messing, and Sean J Westwood. 2012. "How Words and Money Cultivate a Personal Vote: The Effect of Legislator Credit Claiming on Constituent Credit Allocation." *American Political Science Review* 106 (4): 703–719.

Grimmer, Justin, and Brandon M Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267–297.

Grumbach, Jacob M. 2018. "From Backwaters to Major Policymakers: Policy Polarization in the States, 1970–2014." *Perspectives on Politics* 16 (2): 416–435.

Gurin, Patricia, Shirley Hatchett, and James S Jackson. 1990. *Hope and Independence: Blacks' Response to Electoral and Party Politics.* Russell Sage Foundation.

Harris-Lacewell, Melissa Victoria. 2010. *Barbershops, Bibles, and BET: Everyday Talk and Black Political Thought.* Princeton University Press.

Harris, Zellig S. 1954. "Distributional Structure." *Word* 10 (2-3): 146–162.

Hinton, Elizabeth. 2015. "A War Within Our Own Boundaries: Lyndon Johnson's Great Society and the Rise of the Carceral State." *The Journal of American History* 102 (1): 100–112.

———. 2016. *From the War on Poverty to the War on Crime.* Harvard University Press.

Hirschman, Charles, and Morrison G Wong. 1981. "Trends in Socioeconomic Achievement among Immigrant and Native-born Asian-Americans, 1960–1976." *The Sociological Quarterly* 22 (4): 495–514.

Ho, Fred, and Bill V Mullen. 2008. *Afro Asia: Revolutionary Political and Cultural Connections between African Americans and Asian Americans.* Duke University Press.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American statistical Association* 81 (396): 945–960.

Hopkins, Daniel J, and Gary King. 2010. "A Method of Automated Nonparametric Content Analysis for Social Science." *American Journal of Political Science* 54 (1): 229–247.

Hwang, Wei-Chin, and Sharon Goto. 2008. "The Impact of Perceived Racial Discrimination on the Mental Health of Asian American and Latino College Students." *Cultural Diversity and Ethnic Minority Psychology* 14 (4): 326.

Ishizuka, Karen. 2016. *Serve the People: Making Asian America in the Long Sixties.* Verso Books.

Joseph, Peniel E. 2006. *The Black Power Movement: Rethinking the Civil Rights-Black Power Era.* Taylor & Francis.

Joulin, Armand, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2016. "Bag of Tricks for Efficient Text Classification." *arXiv preprint arXiv:1607.01759.*

Kannegaard, Josef Sandoval. 2008. "The Press of a People: the Evolution of Spanish-language News and the Changing Political Community." PhD diss., Massachusetts Institute of Technology.

Kaufmann, Karen M. 2003. "Cracks in the Rainbow: Group Commonality as a Basis for Latino and African-American Political Coalitions." *Political Research Quarterly* 56 (2): 199–210.

Kim, Jae Yeon. 2020. "Racism Is Not Enough: Minority Coalition Building in San Francisco, Seattle, and Vancouver." *Studies in American Political Development:* 1–21.

King, Desmond S, and Rogers M Smith. 2005. "Racial Orders in American Political Development." *American Political Science Review* 99 (1): 75–92.

Kuramoto, Ford H. 1976. "Lessons Learned in the Federal Funding Game." *Social Casework* 57 (3): 208–218.

Kwong, Peter. 1996. *The New Chinatown.* Macmillan.

Lai, David Chuenyan. 2003. "From Downtown Slums to Suburban Malls: Chinese Migration and Settlement in Canada." In *The Chinese Diaspora: Space, Place, Mobility, and Identity,* edited by Laurence JC Ma and Carolyn L Cartier, 311–36. Rowman & Littlefield Publishers, Inc Lanham, Boulder, New York, Oxford.

Lee, Erika. 2003. *At America's Gates: Chinese Immigration During the Exclusion Era, 1882-1943.* University of North Carolina Press.

Li, Wei. 2006. *From Urban Enclave to Ethnic Suburb: New Asian Communities in Pacific Rim Countries.* University of Hawaii Press.

Lien, Pei-te, M. Margaret Conway, and Janelle Wong. 2004. *The Politics of Asian Americans: Diversity and Community.* Routledge.

Linder, Fridolin, Bruce Desmarais, Matthew Burgess, and Eugenia Giraudy. 2018. "Text as Policy: Measuring Policy Similarity through Bill Text Reuse." *Policy Studies Journal.*

Ling, Huping, and Allan W Austin. 2015. *Asian American History and Culture: An Encyclopedia: An Encyclopedia.* Routledge.

Lombard, Matthew, Jennifer Snyder-Duch, and Cheryl Campanella Bracken. 2002. "Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability." *Human Communication Research* 28 (4): 587–604.

Maeda, Daryl. 2005. "Black Panthers, Red Guards, and Chinamen: Constructing Asian American Identity Through Performing Blackness, 1969-1972." *American Quarterly* 57 (4): 1079–1103.

———. 2012. *Rethinking the Asian American Movement.* Routledge.

Maron, Melvin Earl. 1961. "Automatic Indexing: an Experimental Inquiry." *Journal of the ACM* 8 (3): 404–417.

Mason, Llew, Jonathan Baxter, Peter L Bartlett, and Marcus R Frean. 2000. "Boosting Algorithms as Gradient Descent." In *Advances in Neural Information Processing Systems,* 512–518.

McClain, Paula. 2018. *Can We All Get Along?: Racial and Ethnic Minorities in American Politics.* Routledge.

McDaniel, Eric L. 2009. *Politics in the Pews: The Political Mobilization of Black churches.* University of Michigan Press.

McHugh, Mary L. 2012. "Interrater Reliability: the Kappa Statistic." *Biochemia Medica* 22 (3): 276–282.

Meng, Xiao-Li. 2018. "Statistical Paradises and Paradoxes in Big Data (I): Law of Large Populations, Big Data Paradox, and the 2016 US Presidential Election." *The Annals of Applied Statistics* 12 (2): 685–726.

Mikhaylov, Slava, Michael Laver, and Kenneth R Benoit. 2012. "Coder Reliability and Misclassification in the Human Coding of Party Manifestos." *Political Analysis* 20 (1): 78–91.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." *arXiv preprint arXiv:1301.3781.*

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. "Model Cards for Model Reporting." In *Proceedings of the Conference on Fairness, Accountability, and Transparency,* 220–229.

Mora, G Cristina. 2014. *Making Hispanics: How Activists, Bureaucrats, and Media Constructed a New American.* University of Chicago Press.

Muñoz, Carlos. 1989. *Youth, Identity, Power: The Chicano Movement.* Verso.

Murakawa, Naomi. 2014. *The First Civil Right: How Liberals Built Prison America.* Oxford University Press.

Nelson, Laura K. 2017. "Computational Grounded Theory: A Methodological Framework." *Sociological Methods & Research:* 0049124117729703.

———. 2019. "To Measure Meaning in Big Data, Dont Give Me a Map, Give Me Transparency and Reproducibility." *Sociological Methodology* 49 (1): 139–143.

Nelson, Laura K, Derek Burk, Marcel Knudsen, and Leslie McCall. 2017. "The Future of Coding: A Comparison of Hand-coding and Three Types of Computer-assisted Text Analysis Methods." *Sociological Methods & Research:* 0049124118769114.

Ngai, Mae M. 2014. *Impossible Subjects: Illegal Aliens and the Making of Modern America.* Princeton University Press.

Omi, Michael, and Howard Winant. 1986. *Racial Formation in the United States: from the 1960s to the 1990s.* 2nd. New York: Routledge.

———. 1994. *Racial formation in the United States: from the 1960s to the 1990s.* 2nd. New York: Routledge.

Orleck, Annelise. 2011. "The War on Poverty from the Grass Roots Up." In *The War on Poverty: a New Grassroots History, 1964-1980,* edited by Annelise Orleck and Lisa Gayle Hazirjian. Athens: University of Georgia Press.

Pierson, Paul. 2003. "Big, Slow-Moving, and Invisible: Macrosocial Processes in the Study of Comparative Politics." Edited by James Mahoney and Dietrich Rueschemeyer: 177–207.

Prashad, Vijay. 2002. *Everybody Was Kung Fu Fighting: Afro-Asian Connections and the Myth of Cultural Purity.* Beacon Press.

Reardon, Sean F, Demetra Kalogrides, and Kenneth Shores. 2019. "The Geography of Racial/ethnic Test Score Gaps." *American Journal of Sociology* 124 (4): 1164–1221.

Roberts, Margaret E, Brandon M Stewart, and Dustin Tingley. 2015. *STM: R Package for Structural Topic Models. R Package Version 1.1. 0.*

Rodriguez, America. 1999. *Making Latino News: Race, Language, Class.* SAGE Publications.

Rothstein, Richard. 2017. *The Color of Law: A Forgotten History of How Our Government Segregated America.* Liveright Publishing.

Self, Robert O. 2005. *American Babylon: Race and the Struggle for Postwar Oakland.* Princeton University Press.

Sides, Josh. 2006. *LA City Limits: African American Los Angeles from the Great Depression to the Present.* University of California Press.

Skocpol, Theda, and Skocpol Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia and China.* Cambridge University Press.

Slater, Dan, and Daniel Ziblatt. 2013. "The Enduring Indispensability of the Controlled Comparison." *Comparative Political Studies* 46 (10): 1301–1327.

Soss, Joe, Jacob S Hacker, and Suzanne Mettler. 2007. *Remaking America: Democracy and Public Policy in an Age of Inequality.* Russell Sage Foundation.

Suen, Hoi K, and Patrick SC Lee. 1985. "Effects of the Use of Percentage Agreement on Behavioral Observation Reliabilities: A Reassessment." *Journal of Psychopathology and Behavioral Assessment* 7 (3): 221–234.

Tate, Katherine. 1993. *From Protest to Politics: The New Black Voters in American Elections.* Harvard University Press.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society* 58 (1): 267–288.

Trounstine, Jessica. 2018. *Segregation by Design: Local Politics and Inequality in American Cities.* Cambridge University Press.

Umemoto, Karen. 1989. ""On Strike!" San Francisco State College Strike, 1968–69: The Role of Asian American Students." *Amerasia Journal* 15 (1): 3–41.

Vincent, Theodore G. 1973. *Voices of a Black Nation: Political Journalism in the Harlem Renaissance.* Ramparts Press.

Watkins, Rychetta. 2012. *Black Power, Yellow Power, and the Making of Revolutionary Identities.* University Press of Mississippi.

Wei, William. 1993. *The Asian American Movement.* Temple University Press.

Wilkerson, John, and Andreu Casas. 2017. "Large-scale Computerized Text Analysis in Political Science: Opportunities and Challenges." *Annual Review of Political Science* 20:529–544.

Williams, David R, Jourdyn A Lawrence, and Brigette A Davis. 2019. "Racism and Health: Evidence and Needed Research." *Annual Review of Public Health* 40:105–125.

Wolman, Harold. 1986. "The Reagan Urban Policy and Its Impacts." *Urban Affairs Quarterly* 21 (3): 311–335.

Wong, Janelle S, S Karthick Ramakrishnan, Taeku Lee, Jane Junn, and Janelle Wong. 2011. *Asian American Political Participation: Emerging Constituents and Their Political Identities.* Russell Sage Foundation.

Yu, Bin. 2013. "Stability." *Bernoulli* 19 (4): 1484–1500.

Zaller, John R, et al. 1992. *The Nature and Origins of Mass Opinion.* Cambridge University Press.

Zhang, Harry. 2005. "Exploring Conditions for the Optimality of Naive Bayes." *International Journal of Pattern Recognition and Artificial Intelligence* 19 (02): 183–198.

Zhou, Min. 2010. *Chinatown: The Socioeconomic Potential of an Urban Enclave.* Temple University Press.

Zipf, George Kingsley. 1936. *The Psycho-Biology of Language: An Introduction to Dynamic Philology.* Routledge.

———. 1949. *Human Behavior and the Principle of Least Effort.* Addison-Wesley.

Online Appendix

Appendix A   Meta and Specific Issues

Specific issues can be categorized as one particular meta issue (clear boundary; single membership) or as one of the mixed meta issues (blurred boundary; mixed membership). Single membership is when a government policy clearly harms or benefits target groups. Mixed membership is when a government policy may harm or benefit target groups depending on contexts.

- Single membership

    - Collective gain: minority business ownership, minority political representation

    - Collective loss: police brutality, imperialist foreign policy, hate crimes, group stereotypes

- Mixed membership

    - housing, healthcare, education, employment, civil rights

Appendix B   Custom Dictionary for Non-political Issues

- Sports: "football", "basketball", "golf", "tennis", "swimming", "coach", "giants", "warriors", "raiders", "49ers", "track team", "track and field", "athletes", "soccer"

- Car sales: "engine", "powersteering", "windshields", "gasoline", "motors", "subcompact", "showrooms", "fuel prices"

- Arts: "art", "arts", "film", "films", "museum", "galleries", "painting", "paintings", "theater", "television", "circus", "opera", "orchestra", "symphony", "jazz", "disco",

"concert", "concerts", "festival", "festivals", "artists", "artist", "singer", "musician", "musicians", "pianist", "pianists", "guitarists", "guitarist", "ticket", "tickets", "violin", "lion dance"

- Recipes and restaurants: "recipe", "lunch", "lunch special", "dinner", "dinners", "entrees", "breakfast", "cooking", "teaspoon", "teaspoons", "quarts", "tablespoon", "tablespoons", "sugar", "fried"

Appendix C   Classifiers

Textual data are high-dimensional data because the number of features (e.g., words) is far greater than the number of documents. Lasso reduces this high number of features in an interpretable and stable way by shrinking some coefficients in absolute value and setting others to 0 (Tibshirani 1996, 267-273). Naive Bayes is naive because it assumes that the effect of a feature (X) on a class (C) or $P(C|X)$ is independent of the effects of other features (Maron 1961, 410). Despite this strong conditional independence assumption, the naive Bayes classifier has been proven successful in automated text classification (11-12) because when the dependencies among features are distributed evenly across classes, the naive Bayes classifier could still be optimal (Zhang 2005). XGBoost is a relatively recently developed algorithm that has demonstrated high performance in many machine learning competitions. Boosting improves machine learning performance by incrementally combining the performance of many weak classifiers (Freund and Schapire 1999; Friedman, Hastie, Tibshirani, et al. 2000). Gradient boosting is a flexible version of this approach (Breiman 1997; Mason et al. 2000; Friedman 2001), and XGBoost helps apply it to large-scale data efficiently (Chen and Guestrin 2016).

Appendix D   Performance Metrics

In a binary classification problem, machine learning performance is measured by a confusion matrix. A correctly predicted positive class is a true positive (TP), whereas an incorrectly predicted positive class is a false positive (FP). A correctly predicted negative class is a true negative (TN), whereas an incorrectly predicted negative class is a false negative (FN). The accuracy rate of a binary classifier is defined as $\frac{TP+TN}{TP+FN+TN+FP}$. This plain accuracy rate is not ideal when the training data suffer from a class imbalance problem. Consider the hypothetical case presented in Table D.1. The training data are imbalanced because the size of the positive class is 10 and the size of the negative class is 90. The algorithm failed to predict the positive class. Despite this, because the negative class was predicted perfectly, the accuracy rate was 90%. Nevertheless, if one looks at other scores, the performance is far less promising. Whereas the true negative rate (or specificity) is 100%, the true positive rate (or sensitivity) is 0. If the true positive rate (correctly coding collective gain and loss articles) is critical for hypothesis testing, then this situation should be avoided.

	Predicted positive	Predicted negative
Positive class	Number of TP results: 0	Number of FN results: 0
Negative class	Number of FP results: 10	Number of TN results: 90

Table D.1: Hypothetical imbalanced training data

The balanced accuracy rate, which averages specificity and sensitivity, is a more robust measure. The balanced accuracy rate of a binary classifier is defined as $\frac{1}{2}(\frac{TP}{P}(Sensitivity) + \frac{TN}{N}(Specificity))$ (Brodersen et al. 2010, 3122-3123).
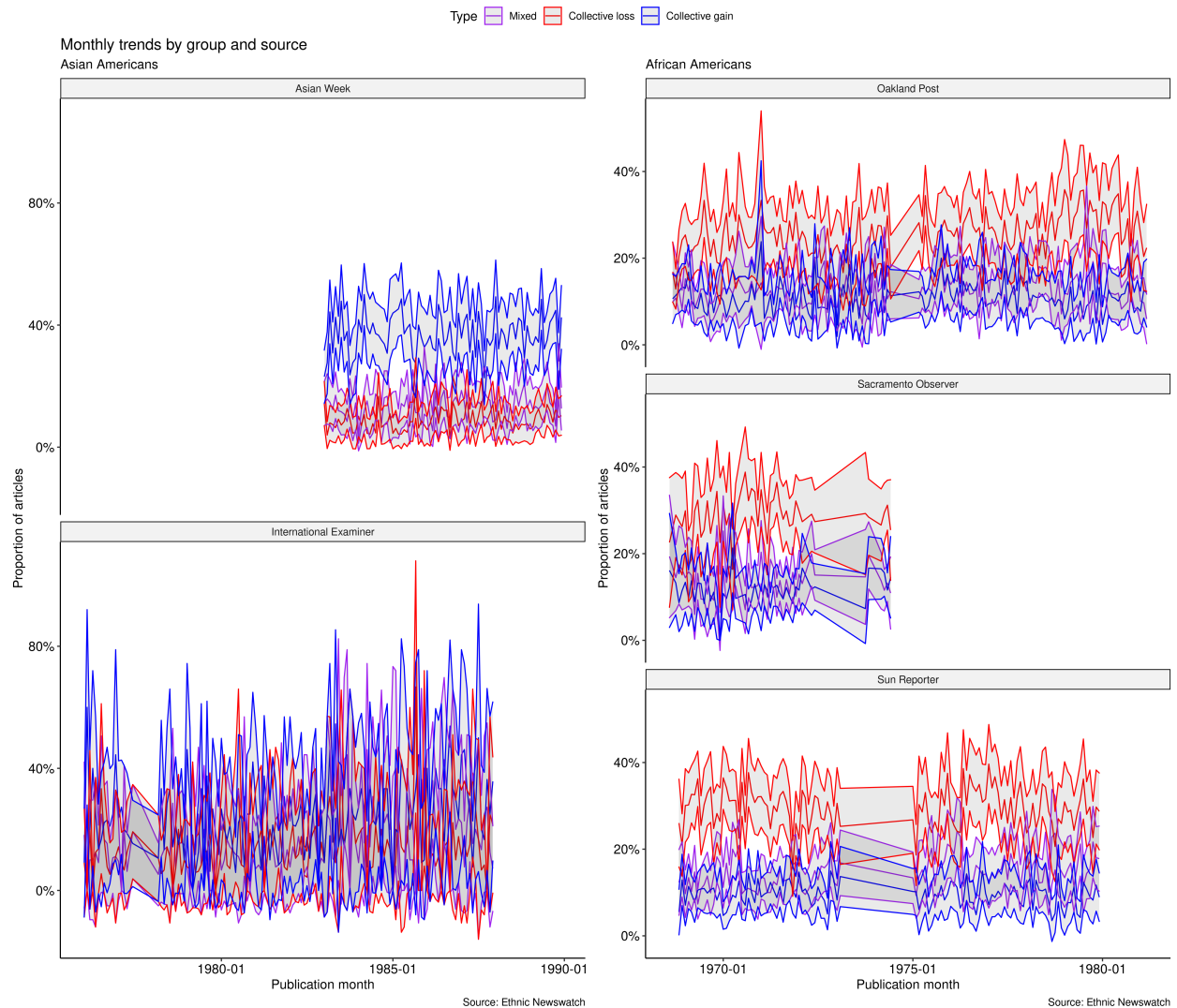
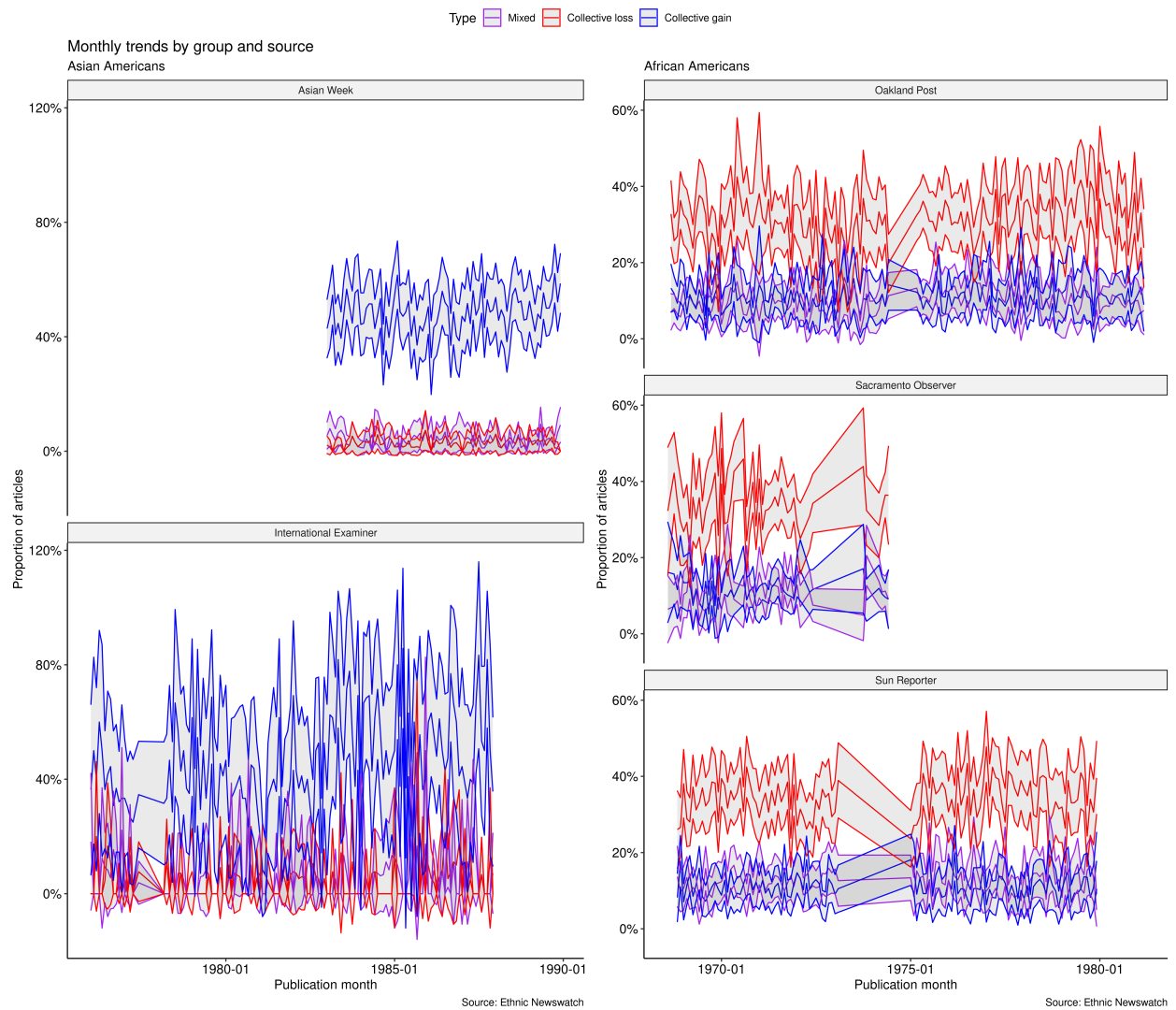# Appendix E   Time Series Plots

Figure E.1: Time series trends (maximum threshold)

Figure E.2: Time series trends (minimum threshold)