Do online newspapers promote or undermine nation# building in divided societies? Evidence from Africa

FINAL MANUSCRIPT TO *NATIONS AND NATIONALISM*

Do Online Newspapers Promote or Undermine Nation-Building in Divided Societies? Evidence from Africa

Evan Lieberman and Andrew Cesare Miller

ARTICLE HISTORY

Compiled August 13, 2020

ABSTRACT

Seminal contributions, including by Anderson (1983) and Deutsch (1953), emphasized the role of newspapers in the development of national consciousness. But do such theories apply in the modern context of online news sites, especially in diverse, post-colonial societies? Because *online* news sites contain forums for reader comments, this provides an opportunity to assess the relationship between exposure to media content and citizen sentiments. We investigate the extent to which a major online news site makes ethnic categories salient in sub-Saharan Africa's largest country, Nigeria. Analyzing more than 35,000 news articles and 300,000 comments, we find that commenters frequently broadcast strong expressions of sub-national ethnic animus in response to a wide variety of stories. In particular, the use of some ethnic categories in headlines is associated with a more than 40 percentage point increase in the probability of at least one reader making an ethnic-based comment. Extending the analysis to South Africa, we show that these patterns generalize beyond Nigeria. By amplifying ethnic animus, "national" newspapers may impede nation-building efforts.

KEYWORDS

ethnicity, race, online news, reader comments, Africa

Word Count: 9,452

Evan Lieberman is Total Professor of Political Science and Contemporary Africa and director of the Global Diversity Lab at the Massachusetts Institute of Technology, Cambridge, MA, USA.

Andrew Cesare Miller is Assistant Professor of Political Science at the United States Naval Academy, Annapolis, MD, USA.

Acknowledgments: We are grateful to Philip Abel, Reginald Best, Daniel Tobin, Meagan Jens, and Valerie Chen for research assistance; to participants in the comparative politics seminar at the University of Illinois; MIT's Political Experiments Research Lab (PERL); the CiFAR Boundaries, Membership and Belonging workshop, and to In Song Kim for helpful comments and feedback.

1. Introduction

Scholars have long pointed to the crucial role played by newspapers and the news media in the development of national identities. For example, in his seminal work on the origins of national consciousness, Anderson (1983)[1] theorized about the role of print newspapers, which afford members of nascent imagined communities the opportunity to learn about events in a common language, concurrently. In a similar manner, Deutsch (1953) focused on the role of newspapers, as well as other media, in the forging of national identities through shared and ongoing communication. Though less focused on newspapers *per se*, Gellner (1983) similarly highlighted the role of shared language and the transmission of ideas through print as a modernizing pre-requisite for nation-building. And more recently, Wimmer (2018) emphasized the role of communicative integration in the construction of nations across time and place.

Do such theories apply to online newspapers, which share similar types of content and frequently target a distinct national audience? Of course it is extremely difficult to parse out the effects of a broad media category in its entirety from all the other modernizing and technological influences that citizens have been exposed to over-time. But by looking closely at reader comments posted on such online news sites, and by analyzing how they are affected by specific stories contained on those sites each day, we can begin to assess the plausibility of the newspaper-as-nation-building theory.

In this article, we focus on the African context, at a time when there has been some evidence of growing national consciousness (Robinson 2014), but sub-national ethnic identification also remains strong in many countries. With respect to online news commentary, we find that readers frequently use disparaging and ethnically chauvinistic language—ethnic animus. And they do this to a significant degree in response to particular types of news stories, suggesting that it is the content of these online publications themselves that attracts such commentary, and not simply random placement of inflammatory statements ("trolling"). As such, we are less sanguine about the extent to which online newspapers—even with a national scope—can serve the nation-building purposes that earlier scholars theorized for the print media.

In addressing this larger theoretical question, we seek to build on what is still a

nascent scholarly literature on the influence of the media, and especially internet-based media, in the creation and reproduction of ethnic and national identity categories in the African context. At the extreme, broadcast and/or print media outlets have been implicated in dramatic events of ethnic violence, and the fracturing of nations, such as in the Rwandan genocide (Gourevitch, 2004; Straus, 2013; Yanagizawa-Drott, 2014) or violence following the 2007 national election in Kenya (Dercon and Gutiérrez-Romero, 2012; Ismail and Deane, 2008). Relatedly, scholars have studied, particularly with experimental approaches, the effects of broadcasting potentially unifying content to help overcome ethnic prejudice and related anti-social behaviors (Paluck, 2009). While a few studies have importantly identified possible connections between various forms of media and ethnic biases in several African countries (Mano 2015; Mpofu 2013; Nyamnjoh 2010) so far, these have been limited to fairly isolated case studies, based on relatively small samples of data. More recently, scholars have begun to develop approaches that had previously only been implemented in more mature media markets, including systematic analysis of the use of ethnic and racial appeals in news and social media (e.g.Mathe 2017). And Green (2018) provides a promising exploratory analysis of Google Trends data in the Nigerian context, demonstrating that for some ethnic groups, the number of searches of a group category is strongly correlated with state-level group size.[2] As more African citizens gain regular access to the internet, analyses of online media will become an important part of the social scientist's toolkit for measuring the prevalence of ethnic orientations within and across societies. Especially because these data are obtained unobtrusively and at high frequency, they hold great promise for improving our understanding of the role of the media in this dimension of political life.

In this article, we build on these contributions to study the role the online media plays in the everyday construction and reproduction of national and sub-national identities, including those organized in terms of race, ethnicity, and religion. As distinct from the approaches summarized by Brady and Kaplan (2009), in which news articles are treated as more "neutral" reflections of appeals and attitudes, we focus on the constitutive role of the media, in which journalists and editors affect the attitudes and

online behavior of their readers through the stories they write and how they choose to frame them.

While others have investigated the role of such online forums as deliberative civic spaces, that is not our concern here. As is the case in many countries around the world, such forums on the African news sites we study are similarly filled with a large share of quite uncivil comments.[3] But the extent and target of such comments is non-random, and such forums provide an otherwise "natural" setting to explore how citizens react to external prompts contained in the newspaper.

One of our central theoretical claims is that the choices made by media "gatekeepers" regarding whether to use ethnic categories in their coverage are highly consequential because they provide an authoritative reminder concerning relevant social categories. Lieberman and Singh (2017) find, for example, with respect to national *censuses*, typically conducted every few years, that the choice to enumerate ethnic categories was associated with substantial increases in the likelihood of ethnic conflict as compared with otherwise similar countries that did not enumerate along ethnic lines.[4] In an analogous manner, highly-regarded news outlets get to choose how they want to describe politics and society—for example, in terms of rich and poor, by gender, region, and/or in terms of particular ethnic groups. Their broadcasting of those categories and labels affects how readers come to think about their own identities and in relationship to other groups in society. For the high-traffic online newspapers we study in Nigeria and South Africa, we find that between one and three percent of all stories employ *explicit* ethnic frames—that is, through the use of a specific ethnic category (Brubaker 2002) or with reference to ethnicity or "tribalism" in the headline. In turn, such frames directly affect the likelihood that online readers will use ethnically-imbued language, overwhelmingly with animus, in their online comments, visible to other readers.

The consequences are substantively important. The quotidian challenge for coexisting in a shared society and polity already characterized by salient ethnic divides—what Horowitz (1985) called "divided societies—is the possibility that verbal attacks can spiral, hardening chauvinist sentiments, and increase the likelihood of destruc-

tive behaviors. Talk matters, because it is the basis for mobilization and recruitment, and at the extreme, can contribute to the dehumanization of group members, which can facilitate violence. Moreover, the framing of political and distributional issues in competitive ethnic terms can negatively affect the provision of public goods (Alesina, Baqir and Easterly 1999; Baldwin and Huber 2010; Miguel and Gugerty 2005), economic growth (Easterly and Levine 1997; Posner 2004), and contributes more generally to an uncivil environment.

On the other hand, when social and economic inequalities cleave along ethnic lines, particularly because of discrimination, the media may rightly put a spotlight on such problems, which in turn may facilitate the making of successful claims for better treatment.

Either way, we theorize that the news media consistently shapes how individuals think about who is "us," and who is "them," in its choices of stories to cover, and whether to include or to exclude particular labels when describing individuals, groups or broader patterns in society, representing potential cleavages or fragments of the nation.

We focus our investigation on the important case of Nigeria, the most populous African country, with high internet penetration, and a troubled history of failed national integration. Specifically, we analyze text from *The Vanguard*, Nigeria's most popular online news site with a purported national scope.[5] Using a simple "scraping" program created to obtain the site's comments and article content, we downloaded 306,203 comments from 35,937 articles posted between January 1 through November 2, 2016. In order to verify the robustness of our findings, we also investigate the use of ethnic categories in online news stories and comments in South Africa, a country in which both racial and ethnic categories have long been salient.

In both countries,[6] we find that the use of ethnic categories in headlines and in news stories is a very significant predictor of the use of ethnic language in online reader comments, the vast majority of which are expressed with a great deal of animus. As we detail, following other theories of the ethnic politics, several other types of news stories also increase the likelihood that readers will express ethnic animus.

2. Theory

Although we frame our study, as described at the outset, around a literature on the building of national identities, we also proceed from the premise that the most significant challenge to the development of such identities is the extent to which sub-national or ethnic identities are salient. Thus, the question of the nation-building potential of online news sites is directly related to whether it encourages or discourages the development of sub-national orientations to the news of the day.

Moreover, building on a large literature, we assume that the very use and embrace of such categories is not simply a longstanding or primordial "given," but the product of socially- and politically-constructed (Anderson, 1983; Chandra, 2012; Wimmer, 2008) ideas mobilized at particular times and places. Along these lines, prior research has demonstrated the role of state institutions (Laitin, 1986; Lieberman and Singh, 2012) and the dynamics of electoral competition (Chandra, 2004; Posner, 2005) in shaping the salience of sub-national (ethnic) categories. But national identities can strengthen relative to ethnic ones, and Robinson (2014) has found that in recent years, many "modernization" variables, such as urban location, education, and employment are associated with higher levels of national identification relative to ethnic identification. And in an analogous manner, Green (2020) finds that individuals from a dominant ethnic group are more likely to identify with the nation when a co-ethnic is president.

Within contexts where both national and sub-national identities are relevant, which includes most African countries, this still begs the question of what increases the salience of ethnicity relative to nationhood. Following the lead of those theorists of nationalism who took an interest in newspapers, we also consider how language and communication structure the degree of ethnic-orientation—the "mental models" (World Bank 2015)—that lead individuals to think and to act from the perspective of a particular group identity. We use such findings to re-visit the larger question of whether an ostensibly national news media leads people to discuss issues in terms of the nation, or in an internally competitive manner, with respect to sub-national groups.

Most prominently, we expect that explicit mentions of ethnic group categories will prime individuals to use in-group/out-group cognitive heuristics. We do not expect

that all individuals in even an ethnically-divided society will be inclined to speak with ethnic animus. But we expect that certain types of prompts will increase the likelihood of "activating" a subset of the population, to use these categories more frequently, and at times to do so with expressions of animus toward ethnic others. The news media outlets—as governed by the journalists, editors, and owners who control the content—provide one such context to periodically introduce the use of ethnic categories in the discussion of social, political and economic issues. Our point is not that they "invent" these categories, but that they do have a choice about whether to use them. As we depict in Figure 1, within the full universe of possible topics that a news media outlet might cover, there exists only a subset (a, excluding b) that could not plausibly be written with reference to domestic ethnic categories. That might include certain stories written entirely about foreign affairs, or perhaps particularly mundane special interest stories. At the other extreme, there is a subset of possible topics (c), that could not plausibly be written without reference to such categories—for example, a story about a hate crime that was committed explicitly in the name of one ethnic group against another. In between—what we believe are likely most possible news topics in an ethnically plural society—are those that *could* be described with reference to various ethnic groups, but could also be written without them.

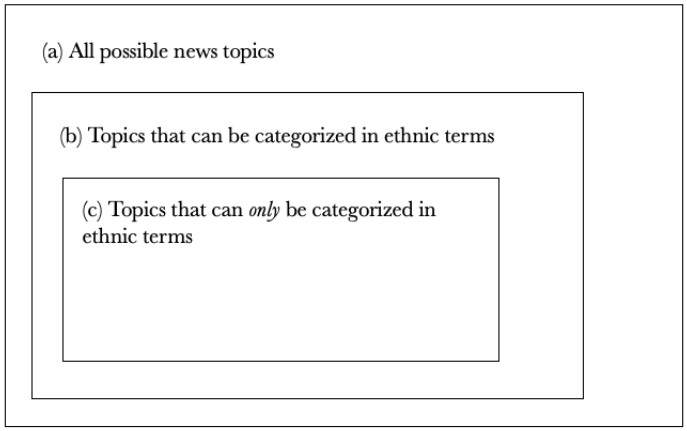


Figure 1. The Ethnic Orientation of the Topic Space for News

Because the question of which stories to cover and how they are covered is always discretionary, we view the introduction of ethnic labels in the news media text as a

choice, albeit one that is influenced by a range of other factors, including of course, the timing of particular events within society and how they present themselves in the real world.

From this discussion, we hypothesize that articles containing references to ethnicity or ethnic categories in the headline will increase the likelihood of ethnic-based commenting.

We also consider the possibility that other types of stories may prime an ethnic orientation among readers, and we draw on other theories of the drivers of ethnic competition to identify several other plausible triggers, which, in the context of an ethnically-divided society, might contribute to yet further expressions of animus.

For example, a significant literature on ethnic political competition has argued that when economic conditions deteriorate within an economically-diverse society, members of some groups are likely to blame other groups for their misfortune. For example, Olzak and Nagel (1986) argue that economic competition, particularly in labor markets, can lead to ethnic mobilization. Along these lines, seminal theories of the relationship between economic modernization and ethnic mobilization have argued that with modernization, groups may gain access to resources on unequal bases, increasing the likelihood of mobilization (Bates 1983;) 1995(1975), and while we investigate a narrower time frame than these scholars, if their theories are correct, we might observe some related patterns at the micro-level—specifically that references to economic development and downturn might put stress on ethnic tensions.

We consider awareness of political competition for national state power. Again, on the one hand, articles that focus on the national state could presumably prime national consciousness. On the other, Bates (1974), Chandra (2004), and Posner (2005) have all argued that ethnic identification may be shaped by elite efforts to form minimum winning coalitions in pursuit of political power and material benefits. In ethnically-divided societies, it is often assumed that ongoing competition and distributional battles are waged along ethnic lines (Cederman, Wimmer and Min, 2010; Horowitz, 1985). The mention of those who control state power may inspire frustration and resentment among some who perceive themselves to be excluded, in turn generating some defen-

sive reactions among those represented in power.

Relatedly, Cederman et al. (2010, p.88) conclude from a group-level study of ethnic rebellion that "ethnonationalist struggles over access to state power are an important part of the dynamics leading to the outbreak of civil wars." In order to address the concern that this variable could simultaneously affect both the content of news stories and of comments, we follow Eifert, Miguel and Posner (2010), who study ethnic salience (via surveys) as a function of proximity to an election, assuming that the closer to the election the more likely we will observe attention to ethnic competition. Given that our dataset in the South Africa case includes articles and comments in the lead up to the country's May 2019 general elections, we can incorporate electoral proximity into the South Africa analysis—but, not into the Nigeria analysis.

Moreover, we consider the role of scapegoating opportunities—the "cognitive and motivational processes by which people try to explain and to solve shared misfortunes" (Glick, 2008). In ethnically-divided societies, citizens and elites may engage in competition over moral virtue, and tend to blame the other for problems, mis-deeds, and non-normative behavior. Because such stories are, of course, routinely reported in the news media, we consider whether they lead to expressions of ethnic animus. The rationale here is that when confronted with such stories, this might increase the likelihood that a commenter would blame ethnic others for such behaviors. Specifically, we consider news stories that mention corruption or crime.

Finally, we consider the most plausible theory that might *mitigate* the use of ethnic animus in the text—and raise the use of national identification: the use of words that indicate national identity (in the case of Nigeria, "Nigeria" and "Naija," a popular slang term to reference Nigeria). We use the national keywords as comparison terms with which to put the effects of the ethnic keywords in perspective. As discussed above, we expect that these terms would be less potent triggers of ethnic discourse than is the case with ethnic terms themselves, and could even mitigate ethnic commenting.

3. Research Design and Context

We selected Nigeria as our primary case for analysis as a country with a long history of intense ethnic conflict, but one that has functioned in recent decades as a multi-party democracy. Nigeria is an important context to study in its own right. With an estimated population of more than 180 million in 2016, it is more than twice as large as sub-Saharan Africa's next most populous country, Ethiopia. Nigeria also enjoys a high level of internet penetration, and as of 2017, approximately 37.9% of Nigerians had internet access, a figure which grew to 47.1% in 2018.[7] As shown in Table 1, in 2015, roughly 17% of Nigerians consumed news online daily, 14% did so a few times a week, 7% did so a few times a month, 8% did so less than once a month, and 55% did not consume the news online.[8]

Table 1. Online News Consumption

Frequency	Percent
Daily	17
Weekly	14
Monthly	7
<Monthly	8
Never	55

Of course, one could study our hypotheses experimentally. This would involve randomly assigning subjects to be exposed to news stories with and without the use of ethnic categories, and through surveys or opportunities to write comments, investigate the resulting language from the treatment. But there are at least two problems with such an approach: First, given the nature of our proposition, that the use of ethnic language in this way would generate animus, we believe it would be unethical to do so. Second, there is too much temptation in a vignette experiment to introduce treatments that are sufficiently powerful to induce an effect, and the question of external validity would loom too large. Thus, we take up the directive of Kinder (2007), that researchers studying political communications pursue more observational research to understand the natural way in which issues are discussed and how citizens react. Instead, we focus our analysis on *The Vanguard* news site, from which we were able to download news articles for most of 2016.

While we believe that our approach is novel, we highlight some core similarities and

differences with recent research studying online communications to explore intergroup conflict and prejudice. Most of this research has been carried out in the United States, and we find that the patterns of behavior bear some similarities to what we find in the Nigerian context.[9]

Coe, Kenski and Rains (2014) undertake the analysis closest to the one we endeavor here: They consider a three-week census of articles and comments posted to the website of the *Arizona Daily Star*, analyzing the determinants of "incivility," which include name-calling, aspersion, lying, and vulgarity. Although they do not specifically identify racist or ethnically-prejudiced comments, we would classify the types of discourse of interest to us in terms of "name-calling incivility." They seek to describe patterns of incivility, looking for potential observable causes and consequences. They find that 22% of comments contained some form of incivility, and that of 1,073 commenters during the three week period, just over half (50.3%) posted at least one uncivil comment. They find that some topics and some authors are more likely to generate uncivil comments; they did not find any relationship between the frequency of incivility and the number of comments made; and they did not find that the most frequent commenters were the ones most likely to be uncivil.

Others have studied online reactions to both the texts and stories that readers encounter as well as to external events and circumstances. For example, Cheng, Bernstein, Danescu-Niculescu-Mizil and Leskovec (2017) explore the causes and consequences of online "trolling" behavior, seeking to provoke emotional reactions from others. They conduct an online experiment, and analyze over 16 million posts on cnn.com. They find effects from bad events as well as the particular order and duration of exposure to other negative comments. They find that time of day and day of week affect prevalence of antisocial comments. A pair of studies present effective interventions to reduce political polarization and racism on Twitter (Munger 2017*a*; Munger 2017*b*). To put into context our study, we focus on the extent to which seemingly neutral and "objective" journalistic use of ethnic categories leads to antisocial ethnic comments.

In Figure 2, we describe how data are generated and captured for our analyses.

As shown, we are not able to directly observe every step in the "real-world" process, but we believe we have created a credible strategy for carrying out our analyses with high-quality data.

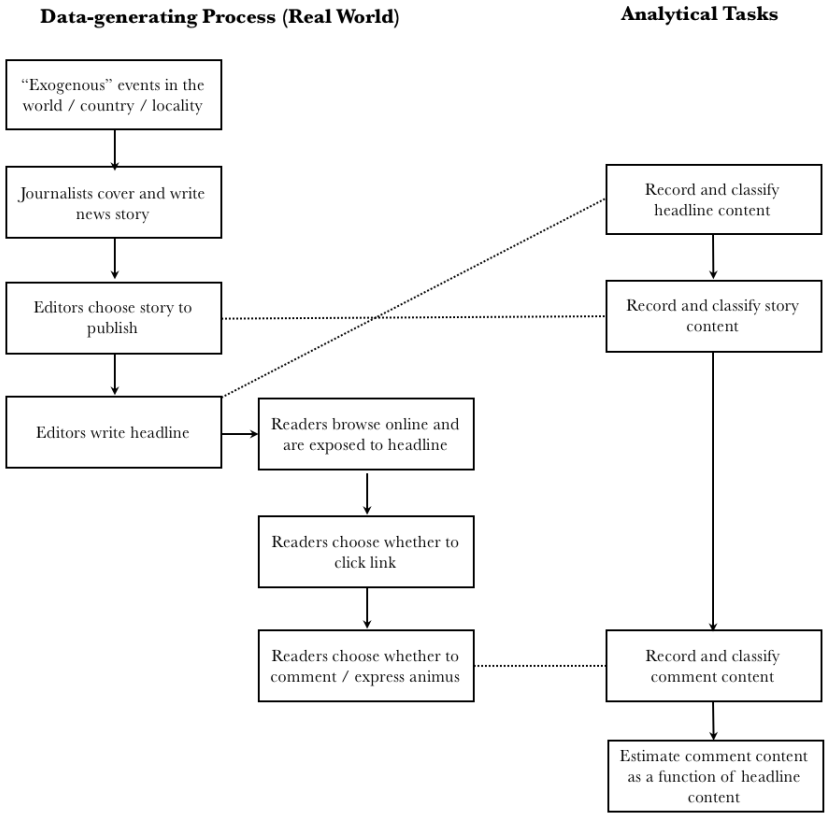


Figure 2. Capturing Data for Analysis

We conceive of each day's publication of news stories as being akin to the exogenous delivery of a set of "treatments" to online readers. First, events occur more-or-less randomly out in the world. In turn, journalists, under editorial guidance, cover some and not other such events and write stories about them. Prior to publication, editors select headlines that will succinctly capture the main point of the story, and when published online, this is what online readers see first. It is at that stage that individuals may decide that they are sufficiently interested in and/or have an emotional reaction to the headline that they will click on the link, providing them an opportunity to read the story and/or to make a comment. And it is at that stage, that the individual may

choose to contribute to the comments section and if so, whether to include mention of any ethnic groups.

We do not expect that all individuals in a society are equally likely to express ethnic animus or even to make reference to ethnic categories in a more even-handed manner—rather, within an ethnically-divided society, we assume that some sizeable minority, or perhaps even a majority—may make such expressions, while others opt against such expressions either because they do not harbor feelings of ethnic prejudice, because they have internalized social norms against such expressions, and/or they fear repercussions. Thus, our central concern is to understand, with respect to that subset of citizens who are inclined to express ethnic animus, what are the actual triggers of such behavior?

We are able to record the content of the headlines, the news stories, and the comments, and to link comments to individual commenters (whose identities are largely anonymous), because, as with many major media outlets, *The Vanguard* uses a Disqus commenting system, which requires individuals to log into their account before making comments. In turn, our analysis is premised on understanding the determinants of comments that contain ethnic animus. It is worth noting that, unlike most major American online news outlets, which are difficult-to-study with respect to racist speech because they tend to be moderated (Hughey and Daniels 2013), *The Vanguard* does not censor racist and ethnically-laden speech. The outlet's policy during the period we investigate was to censor only comments that attempt to sell a product or service.

In order to generate our dataset, we "scraped" articles and comments from just over ten months of online news content in 2016, downloading article content along with up to the first fifty comments made in a given article's comment section.[10]

3.1. Nigerian Ethnic Categories

As in these other studies, we recognize that by studying online commenters, we cannot claim that we are in any way describing a representative sample of Nigerians. Moreover, while high for Africa, internet penetration is still much lower in Nigeria than in the United States or other high-income countries. Among the factors that positively predict

who is likely consume news online are gender (males), university education, living in an urban environment, and reporting discussing politics occasionally. Interestingly, those who tend to consume news online are also more likely to be unemployed relative to those that report never consuming news online.[11]

Nigeria is comprised of a large number of ethnic groups, and in many cases, the contemporary social and political relevance of those groups can be be understood from the origins of the modern Nigerian state as a colonial construction (Ekeh 1975; Joseph 1987). We focus on the largest and most salient of these. The categories analyzed include: The Hausa and Fulani, who reside predominantly in the North, and are largely Muslim, as compared with Christians in the South. Moreover, in the central region, land conflict involving Fulani herdsman is regularly mentioned as a source of violence. Segments of the Igbo community, generally located in Southeastern Nigeria, historically have had aspirations for an independent state, which manifested itself in the bloody Biafran War (1967-1970). The Biafra secessionist movement gained renewed momentum under the Buhari administration from groups such as the Indigenous People of Biafra (IPOB). Further south, the Niger Delta with a large population of Ijaw, have often clashed with the Nigerian state around issues related to oil extraction in the region. The Yoruba, located mostly in the Southwest, are seen as a swing political group. The All People's Congress (APC), which has its stronghold in the North (along with the commercial capital Lagos) and the People's Democratic Party (PDP), strongest in the Southeast, vie for Yoruba support to win federal elections.

We provide estimates of online news consumption by ethnic group from the Afrobarometer data in Table 2. Daily news readers are somewhat disproportionately Yoruba and Igbo as those groups each comprise approximately 25% of online users, while Yoruba and Igbo represent approximately 21% and 18% of the Nigerian population respectively. Meanwhile Hausa-Fulani make up just under 13% of the daily internet consumers despite representing close to 30% of the population, and the Ijaw comprise about 2% of users while constituting about 10% of the population. We assume that such breakdowns might roughly characterize the online readership of a national "newspaper of record," such as *The Vanguard*.

Table 2. Online News Consumption by Ethnicity (Nigeria)

	$\mathbf{n}_{\text{Daily}}$	%	$\mathbf{n}_{\text{Weekly}}$	%	$\mathbf{n}_{\text{Monthly}}$	%	$\mathbf{n}_{<\text{Monthly}}$	%	$\mathbf{n}_{\text{Never}}$	%
Hausa-Fulani	50	12.3	55	17.1	30	19.4	44	23.5	424	32.9
Yoruba	110	27.1	67	20.8	32	20.6	30	16.0	299	23.2
Igbo	103	25.4	77	23.9	28	18.1	32	17.1	170	13.2
Ijaw	10	2.5	7	2.2	8	5.2	12	6.4	17	1.3
Other	133	32.8	116	36.0	57	36.8	69	36.9	380	29.5
All	406	100.0	322	100.0	155	100.0	187	100.0	1290	100.0

3.2. Measurement

We study the use of ethnic categories in news articles and reader comments through the automated detection of our relevant categories. In particular, we search for explicit references to ethnic categories, including "Hausa-Fulani," "Yoruba," "Ijaw," and "Igbo," in both headlines and comments. Table 3 shows a typical comment referencing each of the ethnic groups of interest. As demonstrated, these comments are often direct and hostile toward the given group and manual review of the comments makes clear that the vast majority are hostile in terms of tone and substance. We include a random selection of ten comments in the supplementary appendix to indicate the nature of animus typically associated with mentions of ethnic categories.

Table 3. Typical Ethnic Comments from Nigerian News Sites

Ethnicity	Comment
Hausa-Fulani	He goat, you didn't say anything ystday I thought it was you the hausa - fool-ani-mals burned to aches at the #orilegate lunatics everywhere.
Yoruba	yoruba the mother of all evil and the reason Nigeria is upside down.
Igbo	Go the easy way, top yourself. We'll have one less Igbo ape to worry about.
Ijaw	All kidnapping and bunkering takes its roots from the ijaws

For our analyses, we consider the content of headlines to be most important because it is essentially impossible to read an article in this online format without seeing the headline, and because the headline frames how readers will process the information contained in the story. And of course, by definition, the headline is intended to be a pithy summary of the article. We generate a series of variables that identify the presence or absence of theoretically-relevant keywords in each headline. Our key explanatory variables are created in terms of whether or not an ethnic term, disaggregated by ethnic group, is contained in the headline,[12] and whether other keywords

associated with other theories of ethnic politics are in the headline. The supplementary appendix includes a table with search strings used to generate the dummy variables for a given variable, along with its category mapping to theories of interest (Economic competition, Political competition, Scapegoating and Ethnic categorization).[13]

While we focus our analyses on headline over article content, we still consider the latter. However, measurement of article content in terms of theoretical categories is less straightforward: Because so many words appear in any story, including so many of our keywords simultaneously, a simple keyword search generates less valid results in terms of signaling the content of a story.

As such, we attempt to control for article content in order to focus our estimates on the role of headline keywords. We do this by estimating article content topics based on the commonly-used word-count generative model. The method allows us to inductively parse the corpus of articles into a relatively distinct set of topics via machine learning. In this framework, we set the number of topics at ten (though we attain similar results with various values for this parameter). Using the structural topic model (STM) algorithm, we identity the words that appear in each topic based on the frequency of their co-appearances within the documents (Roberts, Stewart and Airoldi 2015). Each article is assigned a proportion based on the extent to which the topic appears in the article. We include dummy variables for each topic with the article receiving a one if its contents contain 10% or more of a given topic. Although these topics do not line up with our set of theory-driven categories, they provide a useful contextualization for our analyses of headlines. By incorporating the resulting topics into our regression models, we can estimate the effects of different headline foci conditional on relatively similar article content. Thus, when conditioning on these variables, we can interpret the effects of headline keywords as framing devices, in which editors have the opportunity to put a focus on different aspects of otherwise similar types of news stories.

Particularly in ethnically-divided societies, many news stories may be so centrally concerned with ethnic leaders, ethnic associations, and/or ethnic conflict that they are much more likely than others to generate headlines with ethnic categories – to the

extent that it might even appear strange to readers to not have such categories in the headlines. Nonetheless, virtually any story could be described in sufficiently abstract terms so as to avoid explicit attention to ethnicity in the headline. For example, almost all headlines associated with 2020 *New York Times* coverage of the 'Black Lives Matter' movement included racial categories—but not all. For example, a July 9, 2020 article headline was, "A W.N.B.A. Owner Clashes With Players on Protests." To be clear our point is not that editors routinely engage in explicit deliberations about whether to include ethnic categories or not; but simply that they choose stories, and the headlines that go with them.

4. Findings

We begin by describing the data collected in our research. Next, we report on analyses of statistical models related to theories described above, employing a variety of approaches for estimation. Third, we provide evidence from a case study comparing the effect of headlines. Finally, we test if the findings generalize to South Africa.

4.1. Descriptive Statistics

Our corpus from *The Vanguard* contains a total of 35,937 articles.[14] A substantial 1.5% include at least one ethnic term in the headline, and of these, those mentioning "Igbo" account for the largest proportion (39.6%). Hausa-Fulani headlines also account for a large share (33.5%), followed by Ijaw (21.2%) and Yoruba (9.6%). Moreover, 9% of articles have at least one ethnic term in the article content itself (and of those, an ethnic term is mentioned seven times, on average). By comparison, 9% of headlines include a Nigerian national term.

Associated with this corpus of articles, we analyze 306,203 comments, made by 8,964 commenters. Of these, 8.4% contain at least one ethnic term—which we variously describe as an "ethnic comment," or, because a random sample of comments revealed themselves to be overwhelmingly negative and pejorative in tone, as expressions of

"ethnic animus." Moreover, we note that our reading of a random sample of 250 non-ethnic comments revealed a very high prevalence of negative or uncivil sentiment, suggesting that the primary distinction between ethnic and non-ethnic comments is the object of derision.[15] Of the commenters who made at least one comment of any kind in this corpus, 2,646 (25%) made at least one ethnic comment. About one-fifth (19.4%) of all articles are associated with at least one ethnic comment; and, when an article receives an ethnic comment, it begets more: an average of four ethnic comments are made in response to articles with at least one ethnic comment. Thirty-eight percent of articles have at least one comment mentioning a Nigerian national term.

Thus, descriptions of the news in terms of ethnic groups and expressions of ethnic animus by readers are extremely frequent occurrences in this context. These simple facts are at odds with the notion that this particular national news site is a strictly nation-building device. (And note the similarities in comment prevalence with respect to the study in the American context described earlier.)

4.2. Ethnic Headline Framing Triggers Ethnic Comments

In our analyses, we find that headlines with ethnic terms consistently increase the likelihood of commenters expressing ethnic animus. Moreover, we find that many other terms, including in some cases those that could plausibly increase national salience *also* frequently predict increased ethnic commenting, albeit to a lesser extent than in the presence of an ethnic keyword in the headline.

First, we test whether an ethnically-laden headline triggers at least one ethnic comment being made, and regress that outcome on the article headline keywords and content attributes. We control for explicit conflict-related keywords and other potential confounders. The Ordinary Least Squares (OLS) estimates in Table 4 should be interpreted as the percentage point (pp) increase in the likelihood of an article having at least one ethnic comment associated with it.[16] The models include the headline variables as well as the ten topic models derived from the articles' content. Again, we conceive of headlines as framing devices that give important clues about how to think about a particular story.[17]

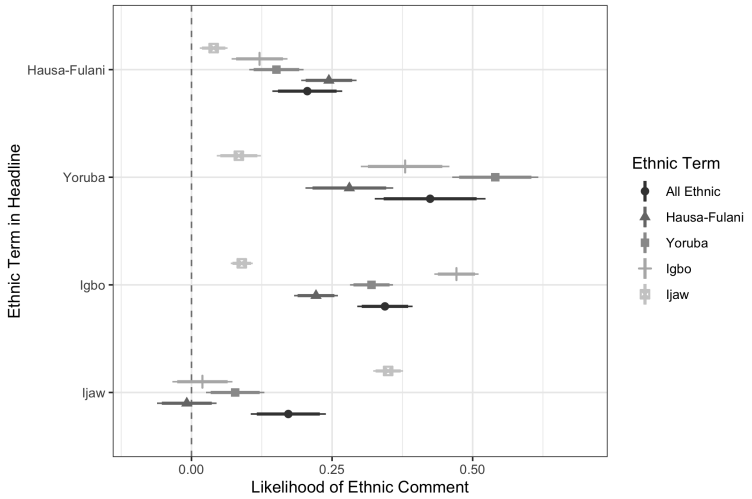
Indeed, we find strong patterns. For instance, as plotted in Figure 3, the mention of Yoruba in the headline increases the likelihood of a comment mentioning any ethnicity by 42.4pp and the likelihood of a comment mentioning Yorubas by 54pp. While this might appear to indicate a simple "parroting" effect of comments reflecting headline words, we also find that when ethnic terms for one group appear in a headline, this triggers commenters to make references to *other* groups. Notably, the Yoruba keyword in the headline significantly increases the likelihood of mentions of other groups, increasing the likelihood of Hausa-Fulani or Igbo groups being mentioned in the comments by 28.1pp and 38pp respectively.

Igbo and Hausa-Fulani similarly drive comments mentioning major ethnic groups. Igbo in the headline drives up the likelihood of an Igbo comment by 47.1pp and Hausa-Fulani boosts Hausa-Fulani comments by 22.1pp. We see less consistent results for the effect of Ijaw keywords, perhaps due to that group's relatively small population in Nigeria, but even still, an Ijaw headline word increases Ijaw comment mentions by 35pp.

Regression analysis at the comment-level, incorporating commenter fixed effects shows results consistent with the article-level analysis. (See Table 9 in the supplementary appendix.) This analysis is restricted to comments from commenters who made at least one ethnic and at least one non-ethnic comment. In turn, these analyses demonstrate that with respect to each commenter, when presented with an article with an ethnic headline, they are more likely to post a comment with an ethnic term. This outcome reinforces our central claim that it is the ethnic term in the headline that drives commenters' propensity to comment with an ethnic term. Indeed, we find in these fixed-effects analyses, that with respect to the other theories of ethnic politics, other article keywords are either unrelated or negatively associated with the likelihood of making ethnic comments, suggesting that the earlier patterns were driven by a self-sorting of commenters, rather than influencing what an individual commenter is likely to post.

Notably, we find religious terms in headlines having an analogous effect on comments as ethnic terms do. Mentions of Christian or Muslim terms in headlines increase com-

ments with religious terms for both the religion mentioned in the headline as well as the other religion not necessarily mentioned in the headline.[18]



95% (thick lines) and 90% (thin lines) confidence intervals shown.

Figure 3. Predicted Change in Likelihood of Ethnic Comment

We also analyze the effects of keywords on the *number* of ethnic comments. Table 5 shows the result of a poisson count model[19] where we test the effect of the headline keywords on the *number* of ethnic comments that are made. We find broadly consistent results with the OLS estimates, whereby the most number of ethnic comments come in response to ethnic headlines, in particular with respect to the Yoruba and Igbo headlines. A Yoruba headline, for instance, on average increases the number of Yoruba comments by on average 2.3 comments.

Some other theorized patterns are also evident in the data, but the associations are consistently weaker and more varied, relative to the effects of ethnic categories as headline keywords. Keywords associated with state power and political competition, the names of political leaders and parties ("Buhari," "Saraki," "PDP," and "APC") are strongly associated with ethnic comments. Economic variables have mixed effects; and scapegoating variables have a very weak association with the likelihood of ethnic comments (and the results are somewhat model dependent). It is worth noting that in all of these cases, the keywords increase the likelihood of commenters mentioning Nigeria, but this does not necessarily connote "nation-building," as such comments are also frequently expressed with great animus.

Moreover, we even find that mentions of Nigerian national terms increase the likelihood of any ethnic term by roughly 7pp, suggesting that it too boosts ethnic discourse, albeit not with the same potency as ethnic headlines. (And the likelihood of Nigeria appearing in the comments increases significantly more—to 16.1pp—when Nigeria is mentioned in the headline.)

We also consider the effects of the term "football" as a placebo test. We have no reason to believe that football is the source of ethnic conflict, but it is topic of great interest and online commentary. Thus, if "football" triggers a large amount of ethnic animus, it would suggest that the results are driven in part by random chance rather than the substance of the headline keywords. As shown in Table 4 and Table 5, we find that it does not significantly increase ethnic comments, albeit slightly decreasing the likelihood of (already infrequent) Igbo comments by 2.3pp. This lack of an effect thus increases confidence that it is the theoretically-relevant keywords that are more likely to trigger animus.

4.3. Case Study

Ideally, to demonstrate that the use of ethnic categories in headlines in turn triggers expressions of ethnic animus, we would assign otherwise identical headlines and article content to readers, but experimentally vary the inclusion of ethnic labels. For example, one could imagine a pair of headlines: "Economic downturn leads to job losses for urban traders," as compared with "Economic downturn leads to job losses for [ethnic group]." However, for ethical considerations (we would not want to deliberately promote ethnic animus), we do not engage in such experimentation, and we attempt to isolate such effects with analyses of observational data.

To further demonstrate the plausibility of our interpretation of the findings—the idea that *labelling* exerts a provocative effect on otherwise similar facts—we match headlines to create an approximation of this counter-factual experiment within the corpus of news articles.[20]

In the Nigerian context, the conflict between Fulani-identified cattle herders and farmers over grazing rights has been a source of persistent violence. In some headlines,

Table 4. Relationship between Headline Keywords and the Likelihood of At Least One Ethnic or National Comment in Response to Nigerian News Articles

	Dependent variable:					
	All Ethnic	Hausa-Fulani	Yoruba	Igbo	Ijaw	Nigeria
	(1)	(2)	(3)	(4)	(5)	(6)
Hausa-Fulani (Ethnic)	0.206***	0.244***	0.151***	0.121***	0.039	0.161***
	(0.038)	(0.040)	(0.043)	(0.043)	(0.026)	(0.039)
Yoruba (Ethnic)	0.424***	0.281***	0.540***	0.380***	0.084**	0.135**
	(0.060)	(0.065)	(0.061)	(0.064)	(0.042)	(0.068)
Igbo (Ethnic)	0.344***	0.221***	0.320***	0.471***	0.089***	0.150***
	(0.030)	(0.032)	(0.033)	(0.031)	(0.023)	(0.031)
Ijaw (Ethnic)	0.172***	−0.008	0.078**	0.019	0.350***	−0.001
	(0.046)	(0.038)	(0.039)	(0.037)	(0.045)	(0.047)
Nigeria (National)	0.069***	0.049***	0.038***	0.045***	0.002	0.181***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.003)	(0.009)
Unemployment (Economic)	0.017	0.009	0.013	−0.009	−0.001	0.093***
	(0.016)	(0.013)	(0.013)	(0.012)	(0.005)	(0.021)
Oil (Economic)	0.022	0.011	0.026**	0.001	0.009	0.088***
	(0.014)	(0.012)	(0.011)	(0.011)	(0.007)	(0.017)
Naira (Economic)	0.028	−0.00002	0.004	0.043***	−0.008*	0.140***
	(0.018)	(0.013)	(0.013)	(0.015)	(0.005)	(0.022)
Buhari (Political)	0.166***	0.128***	0.072***	0.086***	0.019***	0.231***
	(0.010)	(0.009)	(0.008)	(0.008)	(0.005)	(0.010)
Saraki (Political)	0.072***	0.036***	0.047***	0.045***	0.005	0.154***
	(0.017)	(0.013)	(0.014)	(0.014)	(0.006)	(0.019)
PDP (Political)	0.084***	0.054***	0.035***	0.024**	0.007	0.124***
	(0.012)	(0.010)	(0.009)	(0.009)	(0.005)	(0.013)
APC (Political)	0.074***	0.023**	0.063***	0.042***	−0.007	0.102***
	(0.015)	(0.011)	(0.012)	(0.012)	(0.005)	(0.016)
Federal (Political)	−0.011	−0.007	−0.005	−0.011	−0.003	0.048***
	(0.010)	(0.008)	(0.008)	(0.008)	(0.004)	(0.013)
Corrupt (Scapegoating)	0.027***	0.0003	0.008	0.020**	−0.008**	0.136***
	(0.010)	(0.008)	(0.008)	(0.008)	(0.003)	(0.012)
Crime (Scapegoating)	0.048***	0.042***	0.013*	0.018**	0.006	0.046***
	(0.010)	(0.008)	(0.008)	(0.008)	(0.004)	(0.011)
Football (Placebo)	−0.013	0.001	−0.014*	−0.023***	−0.0002	−0.007
	(0.011)	(0.009)	(0.008)	(0.008)	(0.004)	(0.016)
Observations	35,937	35,937	35,937	35,937	35,937	35,937
R^2	0.137	0.142	0.090	0.106	0.069	0.131
Adjusted R^2	0.137	0.141	0.089	0.105	0.068	0.130
Residual Std. Error (df = 35903)	0.368	0.291	0.286	0.294	0.146	0.453
F Statistic (df = 33; 35903)	173.211***	179.531***	107.954***	128.463***	81.049***	163.738***

Note: *p<0.1; **p<0.05; ***p<0.01

Models also include control variables for content topic developed with STM models.

Table 5. Relationship between Headline Keywords and the Average Change in the Number of Ethnic or National Comments in Response to Nigerian News Articles (Poisson)

	Dependent variable:					
	All Ethnic	Hausa-Fulani	Yoruba	Igbo	Ijaw	Nigeria
	(1)	(2)	(3)	(4)	(5)	(6)
Hausa-Fulani (Ethnic)	0.482***	0.685***	0.383*	0.333	0.027	0.329***
	(0.136)	(0.145)	(0.212)	(0.207)	(0.362)	(0.117)
Yoruba (Ethnic)	1.827***	1.386***	2.335***	1.523***	1.542***	0.478***
	(0.158)	(0.271)	(0.172)	(0.168)	(0.446)	(0.182)
Igbo (Ethnic)	1.387***	0.589***	1.410***	1.999***	1.122***	0.501***
	(0.100)	(0.134)	(0.133)	(0.109)	(0.291)	(0.089)
Ijaw (Ethnic)	0.398**	−0.308	0.381	−0.318	1.961***	−0.119
	(0.159)	(0.231)	(0.274)	(0.237)	(0.184)	(0.146)
Nigeria (National)	0.364***	0.379***	0.438***	0.341***	−0.071	0.641***
	(0.055)	(0.068)	(0.077)	(0.073)	(0.151)	(0.031)
Unemployment (Economic)	−0.168	−0.252*	−0.297*	−0.083	−0.050	0.232***
	(0.139)	(0.147)	(0.169)	(0.212)	(0.418)	(0.069)
Oil (Economic)	0.097	0.008	0.303**	0.049	0.180	0.274***
	(0.101)	(0.112)	(0.137)	(0.147)	(0.208)	(0.054)
Naira (Economic)	0.170	−0.037	−0.175	0.467***	−0.540	0.507***
	(0.133)	(0.163)	(0.170)	(0.168)	(0.499)	(0.069)
Buhari (Political)	0.479***	0.616***	0.369***	0.387***	0.344***	0.610***
	(0.048)	(0.056)	(0.071)	(0.068)	(0.122)	(0.028)
Saraki (Political)	0.436***	0.298**	0.534***	0.395***	0.210	0.546***
	(0.107)	(0.137)	(0.143)	(0.148)	(0.308)	(0.054)
PDP (Political)	0.070	0.195**	0.037	−0.075	0.268	0.320***
	(0.068)	(0.081)	(0.092)	(0.094)	(0.199)	(0.039)
APC (Political)	0.414***	0.161	0.651***	0.458***	−0.398	0.226***
	(0.085)	(0.107)	(0.115)	(0.110)	(0.247)	(0.050)
Federal (Political)	−0.066	0.058	0.026	−0.072	−0.168	0.244***
	(0.093)	(0.110)	(0.138)	(0.127)	(0.240)	(0.044)
Corrupt (Scapegoating)	−0.075	−0.180**	−0.252***	0.075	−0.287	0.316***
	(0.068)	(0.085)	(0.089)	(0.088)	(0.208)	(0.035)
Crime (Scapegoating)	0.077	0.275***	−0.036	−0.008	−0.001	0.150***
	(0.058)	(0.068)	(0.084)	(0.080)	(0.188)	(0.039)
Football (Placebo)	−0.115	0.144	−0.072	−0.398	0.198	−0.121
	(0.169)	(0.196)	(0.264)	(0.246)	(0.377)	(0.093)
Observations	35,937	35,937	35,937	35,937	35,937	35,937
Log Likelihood	−51,155.820	−21,258.970	−24,905.830	−30,189.300	−5,762.183	−74,739.870
Akaike Inf. Crit.	102,379.600	42,585.930	49,879.670	60,446.610	11,592.360	149,547.700

Note: *p<0.1; **p<0.05; ***p<0.01
Models also include control variables for content topic developed with STM models.

The Vanguard editors identify the attackers as "Fulani herdsmen" whereas in other headlines they identify the attackers as simply "herdsmen." The fact that the editors use similar headlines without the ethnic term suggests that, when they do use an ethnic term, they choose to use them as a way to frame the issue. And, take the case of two headlines about attacks in the northeastern state of Taraba to show how the use of the ethnic term triggers ethnic hostility in comments. One headline from February 2016 reads "Curfew in Taraba community as *Fulani herdsmen* kill 9, injure 10,"[21] whereas a headline from April 2016 reads "13 killed in Taraba *herdsmen* reprisal attack" (emphasis added).[22]

The articles, by the same journalist, discuss attacks in Taraba with similar language. For example, the ethnic-headline article states, "Saturday Vanguard reliably gathered that the attack on Ibi left five persons killed while five others were killed in Zando, Wusen and Gburucha villages." And, the non-ethnic-headline article states, "Sunday Vanguard gathered that ...some gunmen rushed to the area, attacked and chased out farmers at Kyaaior, Gindin Kerenya and Tse Gyo villages, leaving not less than 13 people killed."

The article without the ethnic headline, however, did not receive any comments whereas the ethnic headline article received twelve comments including ethnically-laden comments with harsh language. The first comment in response to the article tied the violence to Nigeria's president, who is of Fulani ethnicity: "BUHARI WILL NEVER OPENLY CONDEMN THE FULANI HERDSMAEN. HAHAHAHAHA". The comment thread proceeded with discussion of the Buhari's and the police's interest in tackling the conflict. In one case, a commenter projected the killings onto Fulanis as a whole, writing:

> These Fulani herdsmen seem to be having good run on senseless killing of humans, *apparently writing their name on the record of renowned murderous tribe on the face of the earth.* Apparently, security measures such as curfews and surveillances have tended not to contain or abate the incessant attacks, molestation and killing spree by the Fulani herdsmen. ... [Citizens] should take no chances upon sighting any Fulani herdsmen anywhere. It should be shoot-at-sight until they decide to put a stop their rubbish. Enough

of this nonsense! (emphasis added)

Comments such as these are clear expressions of ethnic animus, and perhaps most strikingly the commenter ties this animus to promote further violence on the ground. This is not to say that a comment such as this leads directly to violence but the aggregation of this sentiment likely creates tension that has the potential to boil over not just against herdsmen involved in attacks but against Fulanis more generally. The tense nature of the comment contextualizes the statistical findings, showing how ethnically-framed headlines can bring animus to the surface, and the potential for such expressions to have consequences that go beyond just words.

4.4. General Findings? The South African Case

In order to evaluate the generalizability of the findings from Nigeria to other African contexts, we replicate the core aspects of our analysis with news headlines and comments from the online version of the South African newspaper, *SowetanLIVE*.[23] South Africa presents a useful case for a replication exercise: As Africa's fifth-largest country, like Nigeria, it is populous as well as ethnically heterogeneous with an active online news readership. Moreover, with this case, we can extend our analysis, not only to ethnic cleavages largely based on language, region or religion, but also in terms of race—the most salient ethnic cleavage in South Africa.

We obtained 7,634 articles and 132,112 comments from the *SowetanLIVE* site using the webscraping tool Diffbot. All articles and comments were posted between October 3, 2017 and July 26, 2019. As with *The Vanguard*, editors sometimes include ethnic terms in the headlines. More specifically, 2.3% of headlines include an ethnic term (and we consider racial categories to be a subset of all ethnic categories). And, importantly, many do not require editors to use ethnic or racial terms to convey the article content, demonstrating that they have a choice to frame it in that way. Take, for instance, the March 20, 2019 headline: "Indian girl, 12, 'raped and killed' by brothers, uncle" The use of the descriptor term "Indian" is not an essential detail to convey the main content of the story. The headlines could have read simply "Girl, 12, 'raped and killed' by brothers, uncle" or perhaps specified the geographic location of the incident rather

than the ethnicity of the victim and alleged perpetrators.

As with Nigerian commenters, South African commenters are inclined to use ethnic terms: 23.6% of articles were associated with at least one ethnic comment. Some ethnic comments are short barbs at groups such as "typical uneducated Zulu" or "Zulu idiot." Other comments raise historical grievances such as "Many black people are poor because of 400 years of white looting, murder, rape etc." And others discuss culturally-grounded racial tensions: "Blacks and White weddings. We will never get the land back we dem idiots. Do white people do Zulu traditional weddings then have white weddings on same day like we do. jerrrrrr re dom mannnn. flippin makes me angry."

For the analysis, we disaggregate ethnic terms into the most common categories used in the headlines: white (including white ethnic), black, Indian, Zulu, and a fifth category capturing the remaining groups listed on the South Africa Afrobarometer survey. As in Nigeria, we control for economic, political, and scapegoating terms in headlines. We use the broader category control variables given that we are not analyzing country-specific political, economic, and scapegoating dynamics in South Africa. To control for the potential of elections driving ethnic/racial discourse, we also included a variable for electoral proximity surrounding the May 2019 South African general elections. The variable indicates the number of months before or after the election took place.

Table 6 indicates the results for a binary model in which the dependent variable is the article receiving at least one ethnic comment. Echoing the Nigeria findings, we see that headlines for each ethnic category increases the likelihood of an ethnic comment. Headlines that mention a white term, for instance, increase the likelihood that at least one comment will mention a white term by 43pp. Similarly, the use of the term "black" in the headline increases comments about "blacks" by 34.4pp. And headlines mentioning "Indian" or "Zulu" each boost the likelihood of an comment for their respective category by more than 40pp in both categories. As observed in Nigeria, these ethnic keywords in headlines also spark commenting toward groups not necessarily mentioned in the headline. For instance, an Indian term in an headline increases the likelihood of at least one comment having a white term by 26pp.[24]

Finally, Table 7 shows the results from a poisson model that demonstrates the association between the headline terms of interest and the change in the number of comments with the ethnic/racial terms. The magnitude of the results are largely consistent with the baseline analysis that uses at least one ethnic comment as the dependent variable. For example, the appearance of a Zulu term increases the number of comments with Zulu terms by more than three comments on average.

Other findings are similar to what we found in the Nigerian results: Political keywords are most likely to generate ethnic comments, but not to the same degree as ethnic categories. In line with Eifert et al. (2010), we find that the greater the proximity to the election, the more likely readers are to make ethnic comments. And similar to Nigeria, terms referencing the South African nation also positively predict that readers will make ethnic comments.

Table 6. Relationship between Headline Keywords and the Likelihood of At Least One Ethnic/Racial or National Comment in Response to South African News Articles

	At Least One Ethnic/Racial Comment					
	White	Black	Indian	Zulu	Other Ethnic	South Africa
	(1)	(2)	(3)	(4)	(5)	(6)
White (incl white ethnic)	0.430***	0.361***	0.176***	0.061	0.185**	0.198**
	(0.072)	(0.073)	(0.064)	(0.054)	(0.074)	(0.078)
Black	0.251***	0.344***	0.038	0.017	0.030	0.147***
	(0.051)	(0.050)	(0.028)	(0.024)	(0.037)	(0.051)
Indian	0.260**	0.194*	0.443***	0.078	0.235**	0.125
	(0.119)	(0.116)	(0.123)	(0.086)	(0.117)	(0.132)
Zulu	0.187	0.131	−0.046***	0.505***	0.149	0.216*
	(0.127)	(0.123)	(0.010)	(0.120)	(0.110)	(0.123)
Other Black Ethnic	0.130	0.124	0.111	0.079	0.474***	0.116
	(0.095)	(0.096)	(0.076)	(0.066)	(0.100)	(0.097)
South Africa	0.106*	0.056	0.065*	0.017	−0.002	0.195***
	(0.056)	(0.054)	(0.036)	(0.028)	(0.038)	(0.060)
Economic	0.042	0.060**	−0.005	−0.008	0.039*	0.093***
	(0.028)	(0.029)	(0.011)	(0.011)	(0.022)	(0.031)
Political	0.140***	0.115***	0.028***	0.009	0.057***	0.173***
	(0.022)	(0.022)	(0.011)	(0.010)	(0.017)	(0.023)
Scapegoating	−0.017	−0.024	−0.016***	0.009	−0.006	−0.002
	(0.016)	(0.016)	(0.006)	(0.008)	(0.012)	(0.017)
Election Proximity	−0.003***	−0.002***	−0.001***	−0.002***	−0.002***	−0.004***
	(0.001)	(0.001)	(0.0003)	(0.0003)	(0.0005)	(0.001)
Observations	7,634	7,634	7,634	7,634	7,634	7,634
R^2	0.023	0.021	0.024	0.022	0.015	0.018

Note: *p<0.1; **p<0.05; ***p<0.01

Table 7. Relationship between Headline Keywords and the Average Change in the Number of Ethnic/Racial or National Comments in Response to South African News Articles (Poisson)

	Number of Ethnic/Racial Comments					
	White	Black	Indian	Zulu	Other Ethnic	South Africa
	(1)	(2)	(3)	(4)	(5)	(6)
White	1.964***	1.645***	1.817***	0.922	1.149***	0.866***
	(0.374)	(0.353)	(0.425)	(0.579)	(0.361)	(0.321)
Black	1.159***	1.528***	0.673*	0.361	0.261	0.665***
	(0.213)	(0.214)	(0.392)	(0.420)	(0.285)	(0.212)
Indian	1.213**	0.897*	3.226***	1.070	1.371***	0.561
	(0.534)	(0.511)	(0.509)	(0.869)	(0.516)	(0.567)
Zulu	0.871*	0.622	−12.453***	3.208***	0.984*	0.936*
	(0.529)	(0.532)	(0.321)	(0.487)	(0.555)	(0.493)
Other Ethnic	0.635	0.594	1.456**	1.210*	2.424***	0.536
	(0.425)	(0.429)	(0.614)	(0.642)	(0.421)	(0.415)
South Africa	0.531**	0.287	1.116***	0.359	−0.014	0.862***
	(0.254)	(0.262)	(0.401)	(0.524)	(0.378)	(0.244)
Economic	0.227	0.315**	−0.146	−0.245	0.350**	0.444***
	(0.143)	(0.139)	(0.379)	(0.366)	(0.178)	(0.135)
Political	0.696***	0.574***	0.653***	0.229	0.490***	0.789***
	(0.099)	(0.100)	(0.199)	(0.229)	(0.129)	(0.096)
Scapegoating	−0.101	−0.139	−0.626**	0.231	−0.056	−0.011
	(0.095)	(0.094)	(0.276)	(0.183)	(0.124)	(0.088)
Election Proximity	−0.017***	−0.012***	−0.023***	−0.051***	−0.021***	−0.019***
	(0.004)	(0.004)	(0.008)	(0.008)	(0.005)	(0.004)
Observations	7,634	7,634	7,634	7,634	7,634	7,634

Note: *p<0.1; **p<0.05; ***p<0.01

5. Conclusions

The role of national news media in the creation of national consciousness in Africa's diverse societies appears to be double-edged. News stories, written by journalists and overseen by editorial staff at professional news organizations, provide an important foundation for the spread of facts and informed opinions. These media outlets certainly remind readers of national scope and the relevance of events and personalities that are shared by members of those nations; and such ideas are propagated through various social media channels.

But the content within these outlets does not always promote nation-building. What those media organizations decide to post online affects how their readers will come to view their own societies. A central finding is that when news stories are reported with explicit reference to sub-national ethnic categories, not only does this cause more readers to make comments about ethnic relations, this tends to be done so in virulent tones.

But there is also the reality that even many news stories which are explicitly

national in scope—and do increase the national orientation of reader comments—simultaneously also increase the likelihood of expressions of ethnic animus. It is not contradictory that both patterns are at work: this simply reminds us that the process of nation-building in ethnically-divided societies is frequently fraught with conflict, and theories of the role of the national news media previously under-emphasized this important dynamic.

It is true that uncivil discourse has been problematic for online news sites practically since their origins. Nonetheless, violent histories of ethnic conflict and violence make the occurrence of ethnically-oriented, uncivil commenting particularly problematic in many African countries. In turn, our finding that the use of ethnic categories in news headlines substantially contributes to this outcome in two key African countries ought to weigh heavily on editorial decisions. Of course, putting a spotlight on instances of ethnic discrimination, ethnically-oriented hate speech or crimes, are important subjects of inquiry for ethical news organizations, and it may be unavoidable that the publication of such news will attract hateful instances of ethnic commenting.

Ethnic categories are not simply social "facts" that must be used to describe important phenomena of interest.[25] In an environment such as Nigeria's, in which ethnic cleavages are socially and politically salient, we recognize that in the current context it can be difficult to imagine *not* framing certain issues in ethnic terms, but given the fluidity of ethnic identification—even in this context—as well as the ambiguous labeling of actors as being members of particular ethnic groups, we conclude that ethnic framing in headlines exerts an autonomous impact on how readers come to understand information. And such framing clearly increases the likelihood of intergroup competition, often expressed in quite nasty and uncivil language. Thus, so long as the news media broadcasts information in such ways, their impact on nation-building may be more limited than prominent scholars initially imagined.

Notes

[1]See also, discussion in Breuilly (2016).

[2]This work references a related study by Stephens-Davidowitz (2014), which demonstrates that google

searches of racially-charged search terms were good predictors of state-level vote share for Barack Obama in 2008 and 2012.

[3]See, for example, Coe et al. (2014) in the American context, and Da Silva (2013) for the Brazilian context.

[4]See also, Nobles (2000).

[5]As of March 2018, *The Vanguard* was Nigeria's most trafficked news website and the country's twelfth most trafficked website overall (Amazon.com 2018).

[6]For further discussion of the challenges of nation-building in these two diverse societies, see Osaghae (1999).

[7]https://www.statista.com/statistics/484918/internet-user-reach-nigeria/, accessed January 28, 2020.

[8]Consumption figures from Afrobarometer data collected in 2015, the year prior to the articles posted for our study (Afrobarometer 2015).

[9]Recently, Mueller and Rauh (2017) use newspaper texts to predict conflict around the world, which, while an important contribution, continues to be at a macro-level of analysis, distinct from the approach we take here.

[10]We faced technical constraints in downloading more than fifty comments for any single article.

[11]See Table 4 in the supplementary appendix for a demographic breakdown by online news consumption.

[12]Because the histories of the Hausa and Fulani are largely intertwined and these groups are often referred to collectively as "Hausa-Fulani," we code mentions of either group as an indication of the collective group in hyphenated form.

[13]See Table 1 in the supplementary appendix.

[14]Articles are considered unique if they have different URLs. In a small number of cases, articles were re-posted with different URLs.

[15]We tried several automated sentiment analyses, but none provided remotely valid results, so we relied only on our own reading of comments to establish sentiment values.

[16]We use robust standard errors to protect against heteroskedasticity and provide a more conservative estimate.

[17]The supplementary appendix presents the same regressions without the topic model controls for the content, showing the results hold. Table 7 of the supplementary appendix also shows consistent results employing a logit model.

[18]See Table 8 in the supplementary appendix for the model results.

[19]Results are substantively unchanged when estimated with a Negative binomial regression that accounts for overdispersion of the data.

[20]Although promising techniques exist for using topics to match complex texts (Roberts, Stewart and Nielsen, 2018), we conduct matching manually, drawing on our understanding of the case and context to identify a relevant case study.

[21]Access the article here: https://www.vanguardngr.com/2016/02/curfew-in-taraba-community-as-fulani-herdsmen-kill-9-injure-10. Note that the comments have been removed since they were originally scraped.

[22]Access the article here: https://www.vanguardngr.com/2016/04/13-killed-taraba-herdsmen-reprisal-attack.

[23]Although comprehensive readership data were not available, *SowetanLIVE* is likely the largest black read-

ership online news site in the country.

[24]Table 12 in the supplementary appendix shows the results without control variables. Table 13 in the supplementary appendix shows the results using a logit model.

[25]Perhaps the most important evidence of this in the Nigerian case comes from Laitin's seminal work in Yorubaland (Laitin 1986).

References

Afrobarometer (2015), 'Afrobarometer data, nigeria round 6 data (2015)'.
URL: *http://www.afrobarometer.org*

Alesina, A., Baqir, R. and Easterly, W. R. (1999), 'Public goods and ethnic divisions', *Quarterly Journal of Economics* **114**(4), 1243–1284.

Amazon.com (2018), 'Alexa - top sites in nigeria'.
URL: *https://www.alexa.com/topsites/countries/NG*

Anderson, B. (1983), *Imagined Communities: Reflections on the Origin and Spread of Nationalism*, Verso, London.

Baldwin, K. and Huber, J. D. (2010), 'Economic versus cultural differences: Forms of ethnic diversity and public goods provision', *The American Political Science Review* **104**(4), 644–662.

Bates, R. (1974), 'Ethnic competition and modernization in contemporary africa', *Comparative Political Studies* (Jan. 1974), 457–484.

Bates, R. H. (1983), 'Modernization, ethnic competition, and the rationality of politics in contemporary africa', *State versus ethnic claims: African policy dilemmas* **152**, 171.

Brady, H. E. and Kaplan, C. S. (2009), Conceptualizing and measuring ethnic identity, *in* 'Measuring identity: A guide for social scientists', pp. 33–71.

Breuilly, J. (2016), 'Benedict anderson's imagined communities: A symposium', *Nations and Nationalism* **22**(4), 625–659.

Brubaker, R. (2002), 'Ethnicity without groups', *European Journal of Sociology* **43**(2), 163–189.

Cederman, L.-E., Wimmer, A. and Min, B. (2010), 'Why do ethnic groups rebel? new data and analysis', *World Politics* **62**(01), 87–119.

Chandra, K. (2004), *Why Ethnic Parties Succeed: Patronage and Ethnic Headcounts in India*, Cambridge University Press, Cambridge, UK ; New York.

Chandra, K. (2012), *Constructivist theories of ethnic politics*, Oxford University Press.

Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C. and Leskovec, J. (2017), 'Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions', *Proceedings of CSCW* .

Coe, K., Kenski, K. and Rains, S. A. (2014), 'Online and uncivil? Patterns and determinants of incivility in newspaper website comments', *Journal of Communication* **64**(4), 658–679.

Da Silva, M. T. (2013), 'Online forums, audience participation and modes of political discussion: readers' comments on the brazilian presidential election as a case study.', *Comunicación*

y Sociedad **26**(4), 175 – 193.

Dercon, S. and Gutiérrez-Romero, R. (2012), 'Triggers and characteristics of the 2007 kenyan electoral violence', *World Development* **40**(4), 731 – 744.

Deutsch, K. (1953), *Nationalism and Social Communication*, Wiley, New York.

Easterly, W. and Levine, R. (1997), 'Africa's growth tragedy: Policies and ethnic divisions', *The Quarterly Journal of Economics* **112**(4), 1203–1250.

Eifert, B., Miguel, E. and Posner, D. N. (2010), 'Political Competition and Ethnic Identification in Africa', *American Journal of Political Science* **54**(2), 494–510.

Ekeh, P. P. (1975), 'Colonialism and the two publics in Africa: a theoretical statement', *Comparative Studies in Society and History* **17**(01), 91–112.

Gellner, E. (1983), *Nations and Nationalism*, Cornell University Press, Ithaca.

Glick, P. (2008), *When Neighbors Blame Neighbors: Scapegoating and the Breakdown of Ethnic Relations*, Blackwell Publishing Ltd., pp. 121–146.

Gourevitch, P. (2004), *We wish to inform you that tomorrow we will be killed with our families : stories from Rwanda*, 1st edn, Picador/Farrar, Straus Distributed by Holtzbrinck Publishers, New York.

Green, E. (2018), 'Using google trends to measure ethnic and religious identity in sub-saharan africa: Potentials and limitations'.

URL: *https://blogs.lse.ac.uk/africaatlse/2018/ 06/04/using-google-trends-to-measure-ethnic-and- religious-identity-in-sub- saharan-africa-potentials-and-limitations/*

Green, E. (2020), 'Ethnicity, national identity and the state: evidence from sub-saharan africa', *British Journal of Political Science* **50**(2), 757–779.

Hechter, M. (1995(1975)), *Internal Colonialism: The Celtic Fringe in Bristish National Development*, Transaction Publishers.

Horowitz, D. L. (1985), *Ethnic Groups in Conflict*, University of California Press, Berkeley.

Hughey, M. W. and Daniels, J. (2013), 'Racist comments at online news sites: a methodological dilemma for discourse analysis', *Media, Culture & Society* **35**(3), 332–347.

Ismail, J. A. and Deane, J. (2008), 'The 2007 general election in kenya and its aftermath: The role of local language media', *The International Journal of Press/Politics* **13**(3), 319–327.

Joseph, R. A. (1987), *Democracy and prebendal politics in Nigeria : the rise and fall of the Second Republic*, Cambridge University Press, Cambridge Cambridgeshire ; New York.

Kinder, D. R. (2007), 'Curmudgeonly advice', *Journal of communication* **57**(1), 155–162.

Laitin, D. D. (1986), *Hegemony and Culture: Politics and Change among the Yoruba*, University

of Chicago Press.

Lieberman, E. S. and Singh, P. (2012), 'The institutional origins of ethnic violence', *Comparative Politics* **45**(1), 1–24.

Lieberman, E. S. and Singh, P. (2017), 'Census enumeration and group conflict: A global analysis of the consequences of counting', *World Politics* **69**(1), 1–53.

Mano, W. (2015), 'Racism, ethnicity and the media in africa', *Racism, ethnicity and the media in Africa: Mediating conflict in the twenty-first century* **47**, 1.

Mathe, L. (2017), 'A network analysis of conversational news readers on political and economic zimbabwe online media', *Journal of Communication* **8**(1), 44–55.

Miguel, E. and Gugerty, M. K. (2005), 'Ethnic diversity, social sanctions, and public goods in kenya', *Journal of Public Economics* **89**(11-12), 2325–2368.

Mpofu, S. (2013), 'Social media and the politics of ethnicity in zimbabwe', *Ecquid Novi: African Journalism Studies* **34**(1), 115–122.

Mueller, H. and Rauh, C. (2017), 'Reading between the lines: Prediction of political violence using newspaper text', *American Political Science Review* pp. 1–18.

Munger, K. (2017*a*), 'Experimentally reducing partisan incivility on twitter', p. 37.

Munger, K. (2017*b*), 'Tweetment effects on the tweeted: Experimentally reducing racist harassment', **39**(3), 629–649.

Nobles, M. (2000), *Shades of citizenship: Race and the census in modern politics*, Stanford University Press.

Nyamnjoh, F. B. (2010), 'Racism, ethnicity and the media in africa: Reflections inspired by studies of xenophobia in cameroon and south africa', *Africa Spectrum* **45**(1), 57–93.

Olzak, S. and Nagel, J. (1986), *Competitive ethnic relations*, Academic Press, Orlando.

Osaghae, E. E. (1999), 'Democracy and national cohesion in multiethnic african states: South africa and nigeria compared', *Nations and Nationalism* **5**(2), 259–280.

Paluck, E. L. (2009), 'Reducing intergroup prejudice and conflict using the media: a field experiment in rwanda.', *Journal of personality and social psychology* **96**(3), 574.

Posner, D. (2004), 'Measuring ethnic fractionalization in africa', *American Journal of Political Science* **48**(4), 849–863.

Posner, D. N. (2005), *Institutions and ethnic politics in Africa*, Cambridge University Press, New York.

Roberts, M. E., Stewart, B. M. and Airoldi, E. M. (2015), 'A model of text for experimentation in the social sciences'.

URL: *http://scholar.harvard.edu/files/bstewart/files/stm.pdf*

Roberts, M. E., Stewart, B. M. and Nielsen, R. A. (2018), 'Adjusting for Confounding with Text Matching', p. 50.

Robinson, A. L. (2014), 'National versus ethnic identification in africa: Modernization, colonial legacy, and the origins of territorial nationalism', *World Pol.* **66**, 709.

Stephens-Davidowitz, S. (2014), 'The cost of racial animus on a black candidate: Evidence using Google search data', *Journal of Public Economics* **118**, 26–40.

Straus, S. (2013), *The order of genocide: Race, power, and war in Rwanda*, Cornell University Press.

Wimmer, A. (2008), 'The making and unmaking of ethnic boundaries: A multilevel process theory', *American Journal of Sociology* **113**(4), 970–1022.

Wimmer, A. (2018), *Nation Building: Why Some Countries Come Together and Others Fall Apart*, Princeton University Press.

World Bank (2015), *World Development Report: Mind, Society, and Behavior*, World Bank, Washington D.C.

Yanagizawa-Drott, D. (2014), 'Propaganda and conflict: Evidence from the rwandan genocide', *The Quarterly Journal of Economics* **129**(4), 1947–1994.