Article

The good, the bad, and the ugly: A corpus linguistics analysis of US newspaper coverage of Latinx, 1996–2016

Erik Bleich iD
Middlebury College, USA

James P Callison
Middlebury College, USA

Georgia Grace Edwards
Middlebury College, USA

Mia Fichman
Middlebury College, USA

Erin Hoynes
Middlebury College, USA

Razan Jabari
Middlebury College, USA

A Maurits van der Veen
William & Mary, USA

Abstract

We examine how the US print media portray Latinos and Hispanics, drawing on computer-assisted coding of 185,244 articles in 17 American newspapers between

Corresponding author:
Erik Bleich, Department of Political Science, Middlebury College, Middlebury, VT 05753, USA.
Email: ebleich@middlebury.edu

1996 and 2016. We explore the prevalence of themes of criminality, immigration, illegal immigration, and economic threats. Among these themes, we find that criminality and illegal immigration are associated with the greatest degree of negativity. Yet, the overall tone of articles is neutral rather than negative. Using collocation analysis, we examine the topics associated with positivity within Latinx articles. Stories relating to Latinx achievement and culture have the strongest positive associations with the tone of newspaper coverage. Our research thus identifies the themes associated with both negativity and positivity, and shows that coverage of Latinx has been relatively neutral rather than predominantly negative.

A large body of research demonstrates that media representations have consequential effects for identity groups. Media coverage can influence group members' conceptions of their own status and place in society (Tukachinsky, 2015). It can also affect specific public attitudes (Boomgaarden and Vliegenthart, 2009), broad societal constructions (Schneider and Ingram, 1993), and preferences over public policies toward different groups (Kellstedt, 2000). Given the demographic and social importance of Latinos/Hispanics in the United States, scholars have devoted substantial attention to media coverage of these groups (Barreto et al., 2012; Branton and Dunaway, 2009; Chuang and Chin Roemer, 2015; Dixon and Linz, 2000; Dunaway et al., 2010; Kim et al., 2011; McConnell, 2011; Padín, 2005; Santa Ana, 2002).

Much research focuses on the negative coverage of Latinos/Hispanics (henceforth Latinx) based on associations with criminality, with immigration and illegal immigration, and with economic threats to non-Latinx. To demonstrate these connections, scholars analyzing the print media examine dozens (McConnell, 2011; Santa Ana, 1999), hundreds (Chuang and Chin Roemer, 2015; Dunaway et al., 2011; Padín, 2005), or even over a thousand articles (Branton and Dunaway, 2009; Dunaway et al., 2010). While this research illustrates negative media reporting about Latinx, it does not reveal whether negativity dominates the media as a whole. Moreover, existing scholarship does not provide a consistent picture of which topics are associated with the greatest degree of negativity. Furthermore, while some scholars touch on positive images of Latinx (Barreto et al., 2012; Padín, 2005), positive coverage in the media is not studied as frequently as negative coverage. We know less about which positive characteristics are most commonly associated with Latinx, or about the degree of positivity associated with Latinx as a whole.

We address these issues through computer-assisted coding of a representative swath of the US print media. We examine 185,244 articles mentioning Latinx in 17 national and regional American newspapers between 1 January 1996 and 31 December 2016. We use lexical sentiment analysis and regression analysis to extend scholarly understandings of

how the print media portray Latinx. We explore the prevalence of themes of criminality, immigration, illegal immigration, and threats to economic well-being. In addition, we use collocation analysis to identify positive themes associated with Latinx.

Our research reinforces findings that associations between Latinx and two identifiable themes – criminality and illegal immigration – make the tone of newspaper articles more negative. Yet, coverage of economic threats and coverage of immigration with no references to illegality are not strongly associated with negativity. Most surprisingly, our analysis shows that the overall tone of Latinx articles is not negative. In fact, the tone does not differ substantially from that of a randomly selected cross-section of American newspaper articles. The positivity in our Latinx articles thus balances the effect of negative articles. Themes of *achievement* and *culture* are most strongly associated with positivity in newspaper coverage of Latinx.

Our research makes several key contributions to scholarship on media coverage of Latinx. We confirm but nuance findings about negativity in newspaper articles, while also presenting new information about positive coverage of Latinos and Hispanics. Given the significant impact that the media can have for identity groups, developing this deeper understanding of articles about Latinx in American newspapers is vital to identifying how the media influence self-perceptions as well as attitudes and policies that directly affect Latinx in the United States.

Theoretical building blocks

Communications theorists argue that whether and how the media choose to cover issues influences the way the public considers them. As Cohen notes, the press 'may not be successful much of the time in telling people what to think, but it is stunningly successful in telling its readers what to think about' (cited in Gamson and Modigliani, 1989: 2). For instance, when asked to name the leading causes of death in the United States, most people guessed natural disasters and shootings instead of more common causes such as heart attacks and strokes (Fiske and Taylor, 1984: 271). The more *accessible* knowledge about a topic is, the more pressing that topic seems (McCombs, 2005; Scheufele and Tewksbury, 2007).

When the media devote attention to an issue, they also offer ways to think about the topic. How issues are framed and associated with overarching themes or symbols can drive public perception about those topics (Entman, 1993; McCombs, 2005; Scheufele and Tewksbury, 2007). News media can formulate 'problem definition, causal interpretation, moral evaluation and/or treatment recommendation for the item described' (Entman, 1993: 52), rooted in knowledge and positive or negative associations the audience already has (Scheufele and Tewksbury, 2007). Issue framing can help construct a cognitive map for the audience of how to interpret the key elements of the topic (Entman, 1993; Gamson et al., 1992; Scheufele and Tewksbury, 2007).

Researchers also explore the related concept of *mental models*, defined as malleable and personalized mental representations of knowledge, shaped by subjective and objective elements, and formed, at least in part, subconsciously (Mastro et al., 2007; North and Denzau, 1994). By offering simplified versions of reality, the media contribute to the construction of mental models that help shape social relations. In particular, exposure to

stereotypes can influence judgments about a group, even if the associations are implicit rather than explicit (Banaji et al., 1993). Studies show that both positive and negative exemplars of individuals from an outgroup can affect perceptions toward that entire group (Bodenhausen et al., 1995; Mastro and Tukachinsky, 2011; Ramasubramanian, 2011).

These dynamics have prompted scholars to focus on portrayals of immigrants and minorities in the media (Bleich et al., 2015; Entman and Rojecki, 2000; Van Dijk, 1991). Such depictions can affect intergroup relations and can inform self-perceptions among members of the marginalized group (Tukachinsky, 2015). In general, research shows a connection between how the news media discuss minority groups and how the general public perceives these groups (Dunaway et al., 2011; McLaren et al., 2017). When the media socially construct particular groups (Schneider and Ingram, 1993), the public responds with politically directed solutions, particularly if the messaging emphasizes the dangers of minority communities (Boomgaarden and Vliegenthart, 2007, 2009). Media coverage can thus shape public attitudes and actions as well as policy preferences toward that group.

Media coverage of Latinx

Researchers have conducted numerous content analyses of the coverage of Latinos and Hispanics in the print media (Branton and Dunaway, 2009; Chuang and Chin Roemer, 2015; Dunaway et al., 2010; Kim et al., 2011; McConnell, 2011; Padín, 2005; Santa Ana, 1999, 2002). According to Chuang and Chin Roemer (2015), news media representations of Latinx are infused with 'distortions and negative stereotyping' (p. 1046). Santa Ana (1999, 2002) illustrates the prevalence of negative metaphors in media coverage and public discourse about Latinx. For Dunaway et al. (2011), the tone of coverage of Latino immigrants is 'known to be generally negative, sensational, and stereotypical' (p. 918). In short, the bulk of the existing literature suggests that articles mentioning Latinx are on balance quite negative.

To understand the construction of these negative associations, it is important to pinpoint the factors that account for this negativity. The most prominent themes scholars associate with negative coverage of Latinx are (1) criminality (Barreto et al., 2012; Dixon and Linz, 2000; Fishman and Casiano, 1969; Rivadeneyra, 2006), (2) immigration and illegal immigration (Barreto et al., 2012; Branton and Dunaway, 2009; Chuang and Chin Roemer, 2015; Kim et al., 2011), and (3) economic threats related to employment, welfare, and taxes (Barreto et al., 2012; Menjívar, 2016: 604; Padín, 2005).

Criminality is one of the most commonly cited themes in print content analysis studies. As early as the 1960s, Fishman and Casiano (1969) found that when newspapers reported on Puerto Ricans, it was usually regarding crime, looting, or rioting. In more recent years, Dunaway et al. (2010: 362) note that news outlets systematically cover violent crime, and also disproportionately feature stories with non-White perpetrators; Menjívar (2016: 604) observes that Latino immigrants are often equated with criminals and potential terrorists; and McConnell (2011: 189–190) highlights a newspaper's discussion of gang activity in a multiethnic county outside of Atlanta, while quoting only

Latino gang members. In short, scholars using widely divergent types of evidence demonstrate that Latinx are frequently associated with violent crimes and illegal activities.

Immigration is another theme commonly linked to Latinx coverage. For Chuang and Chin Roemer (2015: 1046), media representations of Latinx have virtually 'rendered members of this group synonymous with "immigrant" and even "illegal immigrant"', a finding that receives additional support from surveys undertaken by Barreto et al. (2012: 22). Moreover, Dunaway et al. (2011) argue that articles about immigration serve 'as important precursors in the development of racial tension and resentment toward newly arrived Latino immigrant populations' (p. 919). Historical moments of prominent immigration legislation not only cemented the link between Latinos and immigration (Santa Ana, 2002; Valentino et al., 2013: 160), but more importantly, between Latinx and *illegal* immigration (Kim et al., 2011: 293; see also Abrajano and Singh, 2009: 5). It is therefore crucial to analyze not only how immigration coverage influences discussions about Latinx in the news, but also how illegal immigration does.

Finally, some scholars discuss media portrayals of Latino immigrants as an economic challenge (often framed as a burden or a threat) related to use of the welfare system, job security, or taxes. Barreto et al.'s (2012) study finds that just over half of survey respondents believe that the term 'welfare recipient' describes Latinx somewhat or very well, and that over a third of respondents believe that Latinos are 'taking jobs from Americans'(p. 3–4). Menjívar (2016) notes that Latinx immigrants are frequently portrayed as 'seeking access to social benefits in the United States that are paid for by US taxpayers', and that this concern manifests itself in discussions of 'anchor babies' (p. 604). We therefore examine whether articles referencing economic threats associated with immigrants and Latinx have a more negative tone than articles without such references.

While scholars repeatedly focus on negative themes associated with Latinx, there has been far less attention to positive portrayals. The few studies that identify positive associations with Latinx in the print media are limited either by a narrow range of sources or by a restricted topic and timeline. For example, Padín's (2005) study examines 14 years of coverage in just one newspaper (the *Oregonian*) and finds that more than half of all collected stories (51.8%) portrayed Latinx as assets to society. These sympathetic depictions include stories about community and civic engagement, vignettes of success, cultural contributions, and articles about entertainment (Padín, 2005). Even in this study, however, Padín argues that negative news portrayals still carry greater weight than positive ones. Chuang and Chin Roemer (2015) analyze coverage of the DREAM act from 2010–2012 in newspapers like the *Los Angeles Times*, the *San Jose Mercury News*, and the *Contra Costa*, concluding from 232 articles that 'hard work, academic achievement, self-determination, and other traditionally "American" cultural codes, juxtaposed with signifiers of poverty and financial need, constitutes a stereotypically selective "success story"' (p. 1045). Given the existing fragmentary attention to positive themes associated with Latinx, we pursue a strategy of inductively identifying potentially positive themes associated with Latinx in our dataset using collocation analysis (described below). We then estimate the associations between these themes and the tone of coverage of Latinx in US newspapers.

Data and methods

To explore the factors associated with the valence of coverage of Latinx, we assemble a corpus of articles from US newspapers published during the 21-year period between 1 January 1996 and 31 December 2016. We examine 185,244 articles drawn from 17 newspapers to obtain a representative sample of US print media coverage of this group. The newspapers are diverse in their ideological orientation, geographic representation, and format, and include high-circulation national newspapers that influence elite opinion and other media outlets. We selected every article containing the words Latino, Latina, Latinx, and Hispanic (and plurals) in the title or in the article text across all sections of each newspaper.[1] We recognize that these terms have multiple meanings and various connotations. We believe that the mainstream media are unlikely to take into account these nuances when representing these groups, and because earlier research has argued that Latinos/Hispanics are a 'pan-ethnic construct' in newspaper reporting, and indeed that 'imputations are made about Latinos/Hispanics as a collectivity' across a wide range of institutional settings (Padín, 2005: 51–52). Without privileging any single term, we use 'Latinx' to stand for these groups as portrayed by US newspapers.

Table 1 displays the distribution of articles across newspapers containing at least one of our four terms. As expected, some newspapers have devoted more coverage to Latinx than others. The *New York Times* and the *Washington Post* published 14.9 percent and 13.8 percent of the articles in our dataset, respectively. The fewest articles are found in the *Philadelphia Daily News* (1.7%) and the *Minneapolis Star-Tribune* and *Richmond Times Dispatch* (2.4% each). We also examined the rate of coverage over time. As a percentage of the total articles, the individual years range between 3.4 percent in 1996 per-cent to 6.4 percent in 2006. While our data are not perfectly evenly distributed across newspaper or time, no single newspaper or time period has an overriding effect on our results.[2]

To analyze the tone of coverage, we use automated sentiment analysis techniques similar to those applied by social scientists to study a variety of topics (Grimmer and Stewart, 2013; Hopkins and King, 2010; Young and Soroka, 2012). Sentiment analysis aims to identify whether the valence of a text is positive or negative. In this article, we develop a lexicon-based approach that uses externally validated dictionaries of positive and negative words to assess the tone of each article. Our approach is easily reproducible and generalizable across topics. It has the advantage of providing information not just on whether a text is positive or negative, but also about *how* positive or negative it is. Automated sentiment analysis cannot replace more fine-grained interpretive coding by trained researchers. However, automated sentiment analysis may better capture subtle underlying positivity and negativity systematically associated with groups. Research on implicit cognition demonstrates the effect of such unconscious elements on behavior and attitudes toward groups of people, which further suggests the value of automated analy-sis as a complement to human-based coding (Freeman and Ambady, 2011; Greenwald and Banaji, 1995).

We base our analysis on eight widely used general-purpose lexica, constructed by a range of different methods for a variety of tasks, each of which lists positive and negative words.[3] To assess the valence of an article in our Latinx corpus, each word in the article

Table 1. Articles included in our corpus, by newspaper.

	Description	Political leaning	Latinx articles	Percentage of total
New York Times	National	Left	27,511	14.9
Washington Post	National	Left	25,464	13.8
USA Today	National	Centrist	5966	3.2
Wall Street Journal	National	Centrist	7450	4.0
Arizona Republic	Regional (west)	Right	15,341	8.3
Atlanta Journal-Constitution	Regional (south)	Centrist	11,622	6.3
Boston Globe	Regional (northeast)	Left	12,536	6.8
Denver Post	Regional (mountain)	Left	10,205	5.5
Las Vegas Review-Journal	Regional (west)	Right	4848	2.6
Minneapolis Star-Tribune	Regional (Midwest)	Left	4495	2.4
Philadelphia Inquirer	Regional (mid-Atlantic)	Left	7845	4.2
Richmond Times Dispatch	Regional (south)	Right	4500	2.4
San Jose Mercury News	Regional (Pacific)	Left	17,832	9.6
Tampa Bay Times	Regional (south)	Left	8981	4.9
New York Daily News	Tabloid (NYC-based)	Centrist	11,413	6.2
New York Post	Tabloid (NYC-based)	Right	6035	3.3
Philadelphia Daily News	Tabloid (mid-Atlantic)	Left	3200	1.7

is compared to the words in each individual lexicon. If the word in the article exists in the lexicon, the (positive or negative) valence score assigned by the lexicon is added to a running count. At the end of each article, we divide the sum of the valence score by the total number of words in the article and average the scores for each article across all eight lexica, producing an initial valence score for each article.

Yet, even if a Latinx article has a negative valence, it may be no more negative than the average newspaper article; research suggests that the media devote more coverage to negative stories than to positive ones, and that coverage of negative stories is more intense than coverage of positive ones (Garz, 2014; Lengauer et al., 2011; Soroka, 2006). Therefore, we calibrate the valence measures for our Latinx corpus against a body of texts that is (on average) representative of US newspaper coverage. We select a random set of articles from our 17 newspapers.[4] This produces a 'representative corpus' of 48,135 articles.[5] After conducting automated sentiment analysis as described above, we assign the average valence of this corpus a value of 0 and standardize the valence measures so that the standard deviation of the representative corpus is set to 1. These calibration parameters are then applied to each article in our Latinx corpus. This allows us to say precisely how positive or negative each Latinx article is relative to our representative corpus of newspaper articles.

In addition, our computer-assisted method allows us to identify articles that contain root words associated with particular themes. To test the association between negative themes identified in existing scholarship and the valence of Latinx corpus articles, we tag each article that contains a root word related to criminality, immigration, illegal immigration, and economic threat. For example, articles containing the root words 'migration'

or 'immigration' were coded for the theme *immigration*. Any article containing the root words 'unemployment', 'welfare', 'poverty', 'poor', or 'homeless' was coded for *economic threat*. Root word lists for each theme were identified through a combination of author team discussion and collocation analysis.[6]

Because existing scholarship has not identified a consistent set of positive themes associated with Latinx, we explore positivity within our corpus using collocation analysis. Collocates are words that are more commonly located in proximity to one another than would occur by chance across the entire corpus (Baker et al., 2013; Blinder and Allen, 2016). Collocation analysis thus allows us to identify words frequently associated with Latinx words within our corpus. Because our aim is to identify positive themes, however, we applied collocation analysis to the subset of our corpus containing only the 93,923 articles with a positive valence. In addition, we identified words in our corpus that were more commonly proximate not only to Latinx words, but also to a set of highly positive adjectives.[7] These steps enable us to identify words associated both with positivity and with Latinx.

Collocation analysis requires researchers to define a window of words around terms of interest within which to search for collocates. For our analysis of the 93,923 articles with a positive valence, we follow standard practices using five words each to the left and right of our Latinx target words (the 'L5R5 collocates') and by examining what is called the 'L1 collocate', which captures just the single word to the left of the target word (Blinder and Allen, 2016: 12), often an adjective that modifies the term of interest. Because Latinx words can be both nouns modified by an adjective (e.g. 'talented Latinos'), but also adjectives modifying a noun (e.g. 'Hispanic cuisine'), we also examine a two-word window encompassing the one word to the left and the one word to the right (the 'L1R1 collocates') of the target words in question. For our analysis of the collocates of positive adjectives combined with Latinx words, we follow the same process, although we omit the L1 collocate analysis, given that keywords preceding positive adjectives are unlikely to yield relevant results. To identify words associated with Latinx and positivity, we thus undertake multiple collocation analyses.[8]

We examined sets of words from these analyses to inductively identify seven potentially positive themes: achievement, culture, groupness, leaders, growth, origin, and size. For example, the *achievement* theme captures words that likely indicate excellence and contributions to a broader group; it includes the root words achievement, contribution, award, excellence, qualified, nominated, scored, talent, wise, and stride. The *culture* theme indicates elements that are commonly associated with group culture or Latinx communities; it includes the root words flair, pride, quinceañera, flavor, fest, tradition, culture, heritage, and authentically. *Groupness* captures words such as community or population; *leaders* includes words such as scholar, athlete, or entrepreneur; *growth* focuses on words such as increasing or growing. The theme of *origin* includes words such as ethnic, descent, or ancestry; and *size* is represented by words such as predominantly, large, or mainly.[9]

Our collocation analyses reveal that these words are somehow associated with positivity and Latinx, but that does not guarantee that articles mentioning these thematic words are positive within our Latinx corpus. It may be the case, for example, that words like 'increasing' or 'predominantly' are even more highly concentrated in Latinx articles

with a negative tone; themes like *growth* or *size* may therefore be associated with neutrality or even negativity rather than with positivity. We use regression analysis to investigate whether the identified themes are in fact associated with positivity in the Latinx article corpus. In addition, to check whether the root words associated with the two most positive themes identified through regression analysis are indeed used in a positive way according to normal language use, we randomize all sentences containing at least one thematic root word and at least one Latinx word. The authors collectively identified and independently assessed 50 *achievement* sentences and 50 *culture* sentences to comprehend the ways in which the presumptively positive words were used in the context of actual articles.

Because there is no uniformly accepted method for identifying words associated with themes of interest, there is admittedly a degree of subjectivity in generating lists of root words and in identifying positive themes. Our lists do not exhaust the words that may be logically associated with each theme, nor are themes mutually exclusive categories. We provide all results in appendices so that future research may test different combinations of root words and themes. Our goal in this article is simply to identify positive themes that are logically consistent, in order to explore their potential relationship to the tone of Latinx articles in our corpus.

Analysis and discussion

We begin by assessing the average tone of Latinx articles. Our method allows us to estimate not only whether Latinx articles are broadly positive or negative on the whole, but also to identify any divergence in tone between our Latinx corpus and our representative corpus of articles. A negative average valence score within our Latinx corpus would support the proposition that US newspaper coverage of Latinx is on balance negative.

However, the mean valence score for Latinx articles is not negative. In fact, at 0.02, it is marginally positive, although only slightly more positive than the mean of the representative corpus itself. In effect, the average Latinx article has a tone essentially equivalent to that of the average article drawn from a random sample of the 17 national and regional newspapers in our dataset. Moreover, the vast majority of Latinx articles have a tone that ranges between −3 and 3, as is true for the articles in our representative corpus. Figure 1 provides a visualization of the distribution of valences in our Latinx corpus compared to the representative corpus based on kernel density estimates. The *y*-axis of this smoothed histogram represents probability densities, with the area under the curve equal to 1. Figure 1 confirms that articles about Latinx are not substantially more negative than the articles in our representative corpus. This is a noteworthy finding, given that existing scholarship focuses largely on negative themes associated with coverage of Latinx in the US media.

This result raises two questions. First, do the negative themes highlighted by existing scholarship strongly affect the valence of newspaper articles? The overall neutrality of our data raises the possibility that the negative themes identified by the existing literature are not present in large numbers of articles, or that they have a substantially smaller effect on article tone than is widely believed. Second, what themes best account for the positivity in our Latinx corpus? If negative themes are meaningful but the average
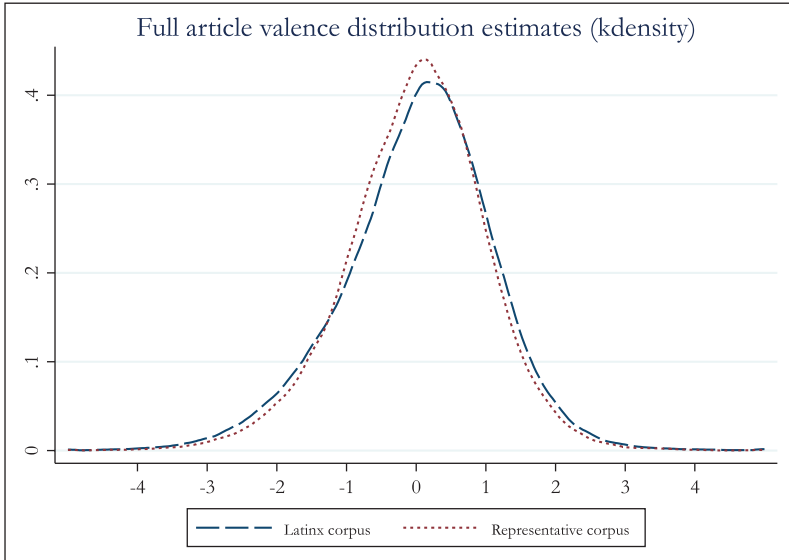
Figure 1. Distribution of articles in Latinx and representative corpora.

valence of the Latinx corpus is neutral, what positive themes counterbalance this negativity?

To address these questions, we first assess each article for the presence of our four presumptively negative and seven potentially positive themes. We find that words associated with each negative theme are present in a substantial number of articles in our dataset. Root words associated with criminality are present in 37 percent of all articles in the Latinx corpus. That adds support to research suggesting that criminality is a widespread theme associated with Latinx in the media. By comparison, immigration root words are present in 25 percent of articles, economic threat words in 21 percent of articles, and illegal immigration, a subset of immigration, in 12 percent of all articles. Because these categories are not mutually exclusive, 60 percent of our articles contain at least one root word associated with a negative theme, and 40 percent contain no negative-themed root words. Among the seven potentially positive themes identified through our collocation analysis, groupness is the largest category (53%), followed in descending order by growth (33%), culture (32%), leaders (27%), achievement (23%), origin (14%), and size (9%). Overall, 83 percent of our articles contain one or more root word associated with a potentially positive theme, and 17 percent contain none of these root words.

Knowing that words associated with each theme are relatively common within our Latinx corpus, however, does not indicate the relationship between specific themes and a positive or negative tone of Latinx newspaper articles. To gauge the extent to which each theme is associated with the tone of Latinx articles, we estimate linear ordinary least squares (OLS) regressions with the valence of Latinx articles as a continuous dependent variable. Table 2 provides the results of three regressions. Model 1 isolates the negative themes, Model 2, the positive themes, and Model 3 combines both sets of

Table 2. Effects of negative and positive themes on Latinx article valence.

	(1)	(2)	(3)
Criminality	−0.804		−0.788
Immigration	0.028		−0.063
Illegal immigration	−0.343		−0.256
Economic threat	−0.026		−0.124
Achievement		0.526	0.488
Culture		0.462	0.434
Groupness		0.185	0.228
Leaders		0.179	0.207
Growth		0.047	0.077
Origin		−.007*	0.031
Size		−0.152	−0.123
Constant	0.365	−0.392	−0.057
Adj-R^2	0.1447	0.1240	0.2658
Observations	185,244	185,244	185,244

Each cell gives the estimated coefficients from OLS regressions. All variables are significant at the $p < 0.01$ level based upon two-tailed hypothesis tests, unless otherwise indicated.
*Not significant at the $p < 0.10$ level.

themes. All coefficients are statistically significant at the $p < 0.01$ level unless otherwise indicated.[10]

Model 1 indicates that keywords associated with *criminality* have by far the strongest association with the tone of articles in our corpus. The coefficient of −0.804 is larger in absolute terms than the coefficient for any other variable. The predicted value of articles mentioning criminality is thus eight-tenths of a standard deviation more negative, on average, than that of articles without criminality root words. Once again, this adds evidence that research on media associations between Latinx and criminality has identified a powerful negative factor in coverage of this group. Articles containing references to *illegal immigration* also have, all else being equal, a more negative tone than those without such references, although the coefficient is less than half that of criminality. By contrast, the analysis shows that *immigration* is not strongly associated with a negative tone in Latinx articles. This offers an initial indication that negative immigration coverage about Latinx groups is linked to themes of illegal immigration rather than to immigration in general. While the coefficient for *economic threat* is negative and statistically significant, its size is quite small, suggesting a limited association with the tone of articles about Latinx.

Among the potentially positive themes, Model 2 demonstrates that *achievement* and *culture* have the strongest associations with article valence in our Latinx corpus. The predicted value of articles containing root words from one of these themes is approximately half a standard deviation more positive than the predicted value of those without them, all else being equal. While this is smaller than the magnitude of the criminality coefficient, it is greater than that of illegal immigration. *Groupness* and *leaders* have a more modest positive association with article valence, while *growth* has a substantively

Table 3. Predicted valences associated with selected themes.

	Predicted valence	Relative to all representative corpus articles
Criminality	−0.845	<82%
Criminality & Illegal Immigration	−1.101	<87%
Achievement	0.431	>67%
Achievement & Culture	0.865	>82%

smaller correlation. The coefficient for each of these positive themes is statistically significant at the $p < 0.01$ level. The coefficient for *origin* is negative, but not statistically significant, whereas the coefficient for *size* is statistically significant and negative, indicating that *size* root words are also present in substantial numbers of articles with a negative tone. In short, the results suggest that articles containing references to Latinx *achievement* or *culture* are likely to account for a substantial portion of the positivity embedded in our Latinx corpus.

The results from Model 3 are substantively similar to those from Models 1 and 2 in almost every respect, confirming the direction and magnitude of the most important variables. Criminality remains the principal theme associated with a negative tone in our Latinx corpus. Illegal immigration also remains associated with negativity, though to a much lower degree than criminality, and to a somewhat lower degree than in Model 1. The estimated coefficients of the positive themes are similar to those in Model 2. The largest shift in the magnitude of the coefficient for any positive theme is 0.04. Achievement and culture remain the most substantively important positive themes, with groupness and leaders also associated with positive coverage, though at a lower level. In addition, there are a few noteworthy changes: immigration has a negative and statistically significant coefficient in the full model as does economic threat, although the magnitude of both remains relatively small; among positive themes, the origin variable is statistically significant and positive, though its size is minimal. Overall, therefore, criminality – and, to a lesser degree, illegal immigration – are the principal factors associated with the negative tone of Latinx articles. Achievement and culture – and, to a lesser degree, groupness and leaders – are the principal themes associated with a positive tone of Latinx articles.

Beyond identifying the themes that have the strongest association with the valence of our articles, we also want to provide a more intuitive sense of the magnitude of these effects than is conveyed simply by reporting coefficients in standard deviation units. To do so, we estimate the predicted valence of articles from a series of Model 3 categories and compare that value with the actual distribution of articles our representative corpus, as illustrated in Table 3. For example, the predicted valence of a Latinx article with a criminality root word and no other thematic words is −0.845 (the sum of the constant and the coefficient for criminality, or −0.057 + −0.788). An article with that valence is more negative than 82 percent of all articles in our representative corpus. If an article in our Latinx corpus has root words for both criminality and illegal immigration (which is true for 11,421 of the articles in our corpus, or 6% of all articles), its predicted valence is −1.101. An article with that valence would be more negative than 87 percent of all

articles in our representative corpus. An average Latinx article containing references to illegal immigration and criminality would thus likely have a more negative tone than approximately six of every seven randomly selected articles in US newspapers.

At the other end of the spectrum, articles containing only an achievement root word have a predicted valence of 0.431, a level that is more positive than 67 percent of representative corpus articles. If an article has both an achievement and a culture keyword (which is true for 10 percent or 18,717 of the articles in the Latinx corpus), its predicted valence is 0.865, which would make it more positive than 82 percent of articles in the representative corpus. These examples thus illustrate the significance of negative and positive themes – particularly in combination – on the tone of Latinx articles.

Because our exploration of potentially positive themes relies on the inductive identification of themes from collocation analysis, we seek to verify that root words associated with these themes are indeed used in ways that are positive. To do this, we isolate sentences containing *achievement* and *culture* words alongside Latinx words. We then randomly select and review 50 sentences from each thematic pool to introduce human coding that complements our computer-assisted coding.

Sentences with achievement words include straightforward examples such as 'In 1985 he received the White House Hispanic-heritage award—it is the highest award given by the Mexican government to a Mexican-American for work done in both countries', and 'Party lines: the Latino Health Institute celebrated its 5th annual health awards night Friday at the World Trade Center with Latin food and live music'. While these types of distinctively positive sentences were among the most common, two additional observations emerged from reviewing sentences with achievement words. First, this set also contained some sentences that conveyed a mixed or even potentially negative tone, such as 'The lack of Latino talent is very unfortunate said Felix Sanchez executive director of the National Hispanic Foundation for the Arts here'. These types of sentences serve as a reminder that there is a distribution of texts associated with any particular theme, rather than a uniformly positive or negative tone associated with any given set of root words. Second, readers may interpret reporting of Latinx achievements in multifaceted ways. Several achievement sentences revolved around awards specifically designated for Latinx, rather than society-wide awards that happen to have been won by Latinos, Latinas, or Hispanics. While such awards are clearly positive and signal achievement by Latinx, their community-based nature may lead some readers to interpret them as limited in scope, rather than as unambiguously positive.

If sentences related to the achievement theme are broadly positive with some complexities, culture theme texts are more straightforwardly positive. Typical sentences include 'Hispanic heroes hailed: Latino stars shine on bomber roster, stoking ethnic pride', and 'Cornucopia of art at Latino fest: the AT&T Latino arts festival is in full swing with art exhibitions films and performances of music dance and theater through August at the Latin American Cultural Center of Queens theatre'. Once again, as expected, not all sentences are unambiguously positive. For example, the following sentence may imply that culture is acting as a brake on success: 'And the lower rate among Hispanics was attributed to an unfamiliarity with available services as well as the lack of access or a cultural focus on the needs of the family rather than the individual'. Nonetheless, this type of sentence is an exception rather than the rule. In short, close

readings of sentences containing achievement and culture root words support the finding that these themes are associated with positivity in US newspaper articles.

Conclusion

Research on media representations of Latinx has largely focused on negative associations. Using a method that allows us to survey over 185,000 articles drawn from 17 US newspapers over a 21-year period, we assess the extent of this negativity and the prevalence and significance of different negative themes identified by existing scholarship. Our results indicate that criminality is the most powerful theme within our corpus of newspapers, and that illegal immigration is another potent source of negativity. However, our analysis also suggests that immigration on its own as well as economic factors such as poverty, unemployment, and welfare use are not associated with as much negativity as is often assumed.

While negativity is common, US newspaper coverage of Latinx is not overwhelmingly negative on the whole. In fact, articles about Latinx are no more negative than a representative corpus of randomly selected articles. This finding led us to investigate the positive themes that are most prevalent within American newspaper articles. Using collocation analysis coupled with human coding as a validation strategy, we identify achievement and culture as the two most influential positive themes in US newspaper coverage of Latinx. Not every article with a root word associated with achievement or culture is unambiguously positive, and nuances and complexities associated with individual sentences or articles require closer and sustained analysis. Nonetheless, articles containing words associated with both themes constitute over 10 percent of all Latinx articles, and are, on average, more positive than over 80 percent of articles in our representative corpus.

Using computer-assisted tools such as lexical sentiment analysis and collocation analysis allows us to test prevailing scholarly propositions in a way that extends earlier studies. Yet, the limitations of our methods also suggest avenues for future research. For instance, our approach does not distinguish between the tone of headlines compared with that of the main body of text, nor is it able to discern the location of individual stories within the newspaper, nor whether the article is accompanied by an image, and whether that image is anodyne or inflammatory. If articles with a particular valence are concentrated in high-visibility and high-impact locations and are accompanied by strong images, they may have a greater influence on media consumers' views of any given group.

This is especially likely with regard to negative messages. Research has demonstrated that negativity has a greater impact on memory and on public opinion than positivity (Baumeister et al., 2001; De Vreese et al., 2011; Soroka, 2006). Moreover, given social psychological findings about the 'ultimate attribution error' with respect to views about minorities (Pettigrew, 1979), the overall neutrality and presence of positive themes in Latinx articles may not be as meaningful as the negativity embedded in articles about criminality and illegal immigration that can trigger processes of racial priming (Valentino et al., 2013). These observations may themselves account for why so much of the existing literature on Latinx emphasizes media negativity. Nonetheless, research on attitudes toward Latinx demonstrates that exposure to positive media coverage is correlated with

a higher opinion of Latinx across a wide series of attributes (Barreto et al., 2012: 12–22), just as social psychological research has shown that frequent positive contact with an outgroup can counterbalance the effects of negative interactions (Graf et al., 2014). Positive coverage matters, and it is important to know that there has been a significant amount of positivity in coverage of Latinx in a broad cross-section of US newspapers over the past two decades.

Finally, to understand whether our results about Latinx coverage in newspapers are indicative of a common pattern requires the application of computer-assisted methods to the study of additional media types, such as television coverage, online news sources, and social media. Since computer-assisted coding has limitations, human coding helps provide a deeper comprehension of how both positive and negative themes are conveyed through different media. This combined approach has the potential to reveal much more about media coverage of Latinx, and it can also contribute to knowledge about media coverage of minorities more generally. As observations from large-scale analyses are refined and validated, it will be increasingly possible to examine the relationship between the tone of media coverage and group self-perceptions, broader public attitudes, societal constructions, or policy preferences toward Latinx, as well as toward a wider array of groups that constitute important parts of our societies.

Acknowledgements

Funding

Supplemental material

Supplemental material for this article is available online.

Notes

1. We used the following search terms: 'Latino! OR Latina! OR Latinx! OR Hispanic!'. The search terms thus capture variations such as Hispanicization, but exclude words such as Latin, and therefore articles on Latin America. Our sample does not include words such as surnames or images that some readers may associate with Latinx, as we sought a sample of articles that all readers unambiguously associate with Latinx.
2. Additional tests show that differences in mean valence scores across newspapers with diverse political leanings are minimal (right papers mean = 0.05, SD = 1.18; left papers mean = 0.03, SD = 1.04; centrist papers mean = −0.03, SD = 1.08).
3. See Appendix B.
4. Because our newspaper databases do not permit random sampling of articles, we conduct searches based on a list of 17 words designated as neutral by our lexica (such as 'boot', 'throat', or 'fourth') to generate articles on 'random' topics that contain one or more of these words. It is a representative corpus not because the words in the articles are neutral, but

because the articles deal with a wide variety of topics and range from highly negative to highly positive.

5. Corpus construction details are available in Appendix A.
6. See Appendix D for root words associated with each theme.
7. See Appendix E.
8. See Appendix F.
9. Appendix G lists root words associated with each theme.
10. We also estimated each of the models including controls for paper ideology; as this resulted in minimal changes to the coefficients of interest, we omit these variables from our table for ease of interpretation.

ORCID iD

Erik Bleich (iD) https://orcid.org/0000-0002-0887-7013

References

Abrajano M and Singh S (2009) Examining the link between issue attitudes news source: The case of Latinos and immigration reform. *Political Behavior* 31: 1–30.

Baker P, Gabrielatos C and McEnery T (2013) *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press*. Cambridge: Cambridge University Press.

Banaji MR, Hardin C and Rothman AJ (1993) Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology* 65: 272–281.

Barreto MA, Manzano S and Segura G (2012) *The Impact of Media Stereotypes on Opinions and Attitudes Towards Latinos*. Pasadena, CA: National Hispanic Media Coalition and Latino Decisions.

Baumeister RF, Bratslavsky E, Finkenauer C, et al. (2001) Bad is stronger than good. *Review of General Psychology* 5: 323–370.

Bleich E, Bloemraad I and de Graauw E (2015) Introduction: Migrants, minorities, and the media: Information, representations, and participation in the public sphere. *Journal of Ethnic and Migration Studies* 41: 857–873.

Blinder S and Allen WL (2016) Constructing immigrants: Portrayals of migrant groups in British national newspapers, 2010–2012. *International Migration Review* 50: 3–40.

Bodenhausen GV, Schwarz N, Bless H, et al. (1995) Effects of atypical exemplars on racial beliefs: Enlightened racism or generalized appraisals? *Journal of Experimental Social Psychology* 31: 48–63.

Boomgaarden HG and Vliegenthart R (2007) Explaining the rise of anti-immigrant parties: The role of news media content. *Electoral Studies* 26: 404–417.

Boomgaarden HG and Vliegenthart R (2009) How news content influences anti-immigration attitudes: Germany, 1993–2005. *European Journal of Political Research* 48: 516–542.

Branton RP and Dunaway J (2009) Spatial proximity to the U.S.-Mexico border and newspaper coverage of immigration issues. *Political Research Quarterly* 62: 289–302.

Chuang A and Chin Roemer R (2015) Beyond the positive–negative paradigm of Latino/Latina news-media representations: DREAM Act exemplars, stereotypical selection, and American otherness. *Journalism* 16: 1045–1061.

De Vreese CH, Boomgaarden HG and Semetko HA (2011) (In)direct framing effects: The effects of news media framing on public support for Turkish membership in the European Union. *Communication Research* 38: 179–205.

Dixon TL and Linz D (2000) Overrepresentation and underrepresentation of African Americans and Latinos as lawbreakers on television News. *Journal of Communication* 50: 131–154.

Dunaway J, Branton RP and Abrajano MA (2010) Agenda setting, public opinion, and the issue of immigration reform. *Social Science Quarterly* 91: 359–378.

Dunaway J, Goidel RK, Kirzinger A, et al. (2011) Rebuilding or intruding? Media coverage and public opinion on Latino immigration in post-Katrina Louisiana. *Social Science Quarterly* 92: 917–937.

Entman RM (1993) Framing: Toward clarification of a fractured paradigm. *Journal of Communication* 43: 51–58.

Entman RM and Rojecki A (2000) *The Black Image in the White Mind: Media and Race in America*. Chicago, IL: The University Of Chicago Press.

Fishman JA and Casiano H (1969) Puerto Ricans in our press. *The Modern Language Journal* 53: 157–162.

Fiske ST and Taylor SE (1984) *Social Cognition*. Boston, MA: Addison-Wesley Pub. Co.

Freeman JB and Ambady N (2011) A dynamic interactive theory of person construal. *Psychological Review* 118: 247–279.

Gamson WA and Modigliani A (1989) Media discourse and public opinion on nuclear power: A constructionist approach. *American Journal of Sociology* 95: 1–37.

Gamson WA, Croteau D, Hoynes W, et al. (1992) Media Images and the social construction of reality. *Annual Review of Sociology* 18: 373–393.

Garz M (2014) Good news and bad news: Evidence of media bias in unemployment reports. *Public Choice* 161: 499–515.

Graf S, Paolini S and Rubin M (2014) Negative intergroup contact is more influential, but positive intergroup contact is more common: Assessing contact prominence and contact prevalence in five central European countries. *European Journal of Social Psychology* 44: 536–547.

Greenwald AG and Banaji MR (1995) Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review* 102: 4–27.

Grimmer J and Stewart BM (2013) Text as data: The promise and pitfalls of automated content analysis methods for political texts. *Political Analysis* 21: 267–297.

Hopkins D and King G (2010) A method of automated nonparametric content analysis for social science. *American Journal of Political Science* 54: 229–247.

Kellstedt PM (2000) Media framing and the dynamics of racial policy preferences. *American Journal of Political Science* 44: 245–260.

Kim S-h, Carvalho JP, Davis AG, et al (2011) The view of the border: News framing of the definition, causes, and solutions to illegal immigration. *Mass Communication and Society* 14: 292–314.

Lengauer G, Esser F and Berganza R (2011) Negativity in political news: A review of concepts, operationalizations and key findings. *Journalism* 13: 179–202.

McCombs M (2005) A look at agenda-setting: Past, present and future. *Journalism Studies* 6: 543–557.

McConnell ED (2011) An 'incredible number of Latinos and Asians': Media representations of racial and ethnic population change in Atlanta, Georgia. *Latino Studies* 9: 177–197.

McLaren L, Boomgaarden H and Vliegenthart R (2017) News coverage and public concern about immigration in Britain. *International Journal of Public Opinion Research* 30: 173–193.

Mastro D and Tukachinsky R (2011) The influence of exemplar versus prototype-based media primes on racial/ethnic evaluations. *Journal of Communication* 61: 916–937.

Mastro D, Behm-Morawitz E and Ortiz M (2007) The cultivation of social perceptions of Latinos: A mental models approach. *Media Psychology* 9: 347–365.

Menjívar C (2016) Immigrant criminalization in law and the media: Effects on Latino immigrant workers' identities in Arizona. *American Behavioral Scientist* 60: 597–616.

North DC and Denzau AT (1994) Shared mental models: Ideologies and institutions. *Kyklos* 47: 3–31.

Padín JA (2005) The normative mulattoes: The press, Latinos, and the racial climate on the moving immigration frontier. *Sociological Perspectives* 48: 49–75.

Pettigrew TF (1979) The ultimate attribution error: Extending Allport's cognitive analysis of prejudice. *Personality and Social Psychology Bulletin* 5: 461–476.

Ramasubramanian S (2011) The impact of stereotypical versus counterstereotypical media exemplars on racial attitudes, causal attributions, and support for affirmative action. *Communication Research* 84: 497–516.

Rivadeneyra R (2006) Do you see what I see? Latino adolescents' perceptions of the images on television. *Journal of Adolescent Research* 21: 393–414.

Santa Ana O (1999) 'Like an animal I was treated': Anti-immigrant metaphor in US public discourse. *Discourse & Society* 10: 191–224.

Santa Ana O (2002) *Brown Tide Rising: Metaphors of Latinos in Contemporary American Public Discourse*. Austin, TX: University of Texas Press.

Scheufele DA and Tewksbury D (2007) Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of Communication* 57: 9–20.

Schneider A and Ingram H (1993) Social construction of target populations: Implications for politics and policy. *The American Political Science Review* 87: 334–347.

Soroka SN (2006) Good news and bad news: Asymmetric responses to economic information. *Journal of Politics* 68: 372–385.

Tukachinsky R (2015) Where we have been and where we can go from here: Looking to the future in research on media, race, and ethnicity. *Journal of Social Issues* 71: 186–199.

Valentino NA, Brader T and Jardina AE (2013) Immigration opposition among U.S. whites: General ethnocentrism or media priming of attitudes about Latinos? *Political Psychology* 34: 149–166.

Van Dijk TA (1991) *Racism and the Press*. London: Routledge.

Young L and Soroka S (2012) Affective news: The automated coding of sentiment in political texts. *Political Communication* 29: 205–231.

Author biographies

Erik Bleich is a Charles A. Dana Professor of Political Science, Middlebury College, where he directs the Media Portrayals of Minorities Project. He co-edited *Migrants, Minorities, and the Media: Information, Representations, and Participation in the Public Sphere* (Routledge, 2017), and has published articles on media portrayals of minorities in journals such as *Ethnic and Racial Studies*, the *International Journal of Comparative Sociology*, and *Politics, Groups, and Identities*.

James P Callison is a Member of Middlebury College Class of 2017.

Georgia Grace Edwards is a Member of Middlebury College Class of 2018.

Mia Fichman is a Member of Middlebury College Class of 2019.

Erin Hoynes is a Member of Middlebury College Class of 2019.

Razan Jabari is a Member of Middlebury College Class of 2018.

A Maurits van der Veen is an Associate Professor of Government, William & Mary, where he directs the STAIR lab, which uses text mining and machine learning tools to analyze and address political issues, with a particular focus on international relations. He is the author of *Ideas, Interests and Foreign Aid* (Cambridge University Press, 2011).