Article

Social Science Computer Review
1-24

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/08944393211031452
journals.sagepub.com/home/ssc
SAGE

Prevalence of Prejudice-Denoting Words in News Media Discourse: A Chronological Analysis

David Rozado[1], Musa Al-Gharbi[2], and Jamin Halberstadt[3]

Abstract

This work analyzes the prevalence of words denoting prejudice in 27 million news and opinion articles written between 1970 and 2019 and published in 47 of the most popular news media outlets in the United States. Our results show that the frequency of words that denote specific prejudice types related to ethnicity, gender, sexual, and religious orientation has markedly increased within the 2010–2019 decade across most news media outlets. This phenomenon starts prior to, but appears to accelerate after, 2015. The frequency of prejudice-denoting words in news articles is not synchronous across all outlets, with the yearly prevalence of such words in some influential news media outlets being predictive of those words' usage frequency in other outlets the following year. Increasing prevalence of prejudice-denoting words in news media discourse is often substantially correlated with U.S. public opinion survey data on growing perceptions of minorities' mistreatment. Granger tests suggest that the prevalence of prejudice-denoting terms in news outlets might be predictive of shifts in public perceptions of prejudice severity in society for some, but not all, types of prejudice.

Keywords

computational social science, word frequency counts, communications, sociology, the great awokening, prejudice, news media

Computational content analysis of large bodies of text can be illuminating to elucidate trends embedded in such corpora (Caliskan et al., 2017; Kozlowski et al., 2019; Rozado, 2019, 2020b). Simply charting word frequencies in a diachronic corpus of written news articles accurately tracks the time course of historical events and can highlight the dynamics of social trends within the cultural context where the texts were produced (Rozado, 2020a). This work examines word usage frequencies in 27 million news and opinion articles published between January 1, 1970, and December 31, 2019, in 47 of the most popular news media outlets in the United States such as

[1] ECL—Otago Polytechnic, Dunedin, Otago, New Zealand
[2] Columbia University, New York, NY, USA
[3] University of Otago, Dunedin, Otago, New Zealand

Corresponding Author:
David Rozado, ECL—Otago Polytechnic, Forth Street, Dunedin, Otago 9054, New Zealand.
Email: david.rozado@op.ac.nz

The New York Times, *The Washington Post*, *The Wall Street Journal*, and *Fox News*. Our analysis focuses primarily on tracking the prevalence of words that describe prejudice such as *racism*, *sexism*, *islamophobia*, *anti-Semitism*, and *homophobia*.

The usage frequency of prejudice-denoting words in news media discourse is a question of significant sociological interest. Previous research by the lead author of this work identified a sharp increase in the prevalence of prejudice-denoting terms in *New York Times* articles within the 2010–2019 time frame (Rozado, 2020a). Public perceptions about the severity and ubiquity of prejudice seem to have risen dramatically over the same period, particularly among Democrats (Pew Research Center, 2017, 2018). Yet, U.S. social surveys appear to show a marked reduction in expressed prejudicial attitudes with regard to demographic identities such as gender, sexual orientation, and ethnicity since the 1960s (Gao, 2015; Krysan & Moberg, 2016; Marsden et al., 2020; Meagher & Shu, 2019). Despite these trends in surveys and polling, systematic disparities in social outcomes continue to persist—particularly along ethnic lines (al-Gharbi, 2019). Some scholars interpret the persistent disparities as evidence that American society has not grown less prejudiced; but that instead, discrimination and antipathy against protected minorities have merely become less overt (Bonilla-Silva, 2017). Other scholars contradict these accounts and report that prejudice has at least partially decreased (Charlesworth & Banaji, 2019; Hopkins & Washington, 2020; Krysan & Moberg, 2016). Consequently, the nature of the relationships between news media coverage of prejudice, public perceptions of prejudice, and the actual levels of prejudice within society at a given time remains unclear.

The purpose of this work is to characterize the prevalence of prejudice-denoting words in a comprehensive and representative sample of news media outlets popular in the United States and to test whether the increasing prevalence of such words previously observed for *The New York Times* (Rozado, 2020a) generalizes across the media landscape. We then examine the relationship between the usage of prejudice-denoting terms in written news media and other factors, such as news outlets' ideological leanings or the prevalence of prejudice-signifying words in cable news. We also attempt to elucidate whether the prevalence of prejudice-denoting words in news media discourse changed before or after 2015, a significant year that marked the beginning of the 2016 U.S. presidential election campaign. Our analysis continues by investigating whether some outlets preceded others on the usage dynamics of prejudice-denoting words. Finally, we examine the relationship between the prevalence of prejudice-denoting words in news media discourse and public opinion perceptions about prejudice severity in the wider society.

To our knowledge, this work is the first comprehensive scholarly attempt at describing the prevalence of prejudice-denoting words across a large and representative set of written articles from news media outlets. Although it is beyond the scope of this study to make strong causal claims, our descriptive analyses provide detailed insights into the prevalence of prejudice-denoting words in news media discourse and its relationship with other potentially relevant factors, providing a strong methodological and empirical foundation for subsequent research.

Methods

List of News Media Outlets Analyzed

The list of 47 news media outlets used in this work was taken from the AllSides organization (AllSides Media Bias Ratings, 2019). The external human ratings of outlets ideological leanings were also taken from the AllSides organization . The original 2019 Media Bias Chart Version 1.1 from AllSides is provided as Supplementary Material, see Figure S1.

Raw Data and Temporal Coverage

The textual content of news and opinion articles from the outlets analyzed is available in the outlet's online domains and/or public cache repositories such as *Google cache*, *The Internet Wayback* machine (Notess, 2002), and *Common Crawl* (Mehmood et al., 2017). Textual content included in our analysis is circumscribed to the articles' headlines and main text and does not include other article elements such as figure captions. Targeted textual content was located in HTML raw data using outlet specific XPath expressions. Tokens were lowercased prior to estimating frequency counts. To prevent outlets with sparse text content for a year from distorting aggregate yearly frequency counts across outlets, we only include outlet frequency counts from years for which the outlet has at least 1.25 million words of articles' textual content available. This threshold was chosen to maximize inclusion in our analysis of outlets with low article volume per year such as *Reason*, *AlterNet* or *The American Spectator* while simultaneously maintaining a large enough sample size of words to obtain accurate frequency counts per outlet/year.

The temporal coverage of articles availability in different online news outlets is not uniform. For most media organizations, substantial news articles availability in online domains or Internet cache repositories becomes sparse for earlier years. This is not the case for a few news outlets, where online availability of news articles goes back as far as the 1970s. Still, frequency data of news media word usage is constrained in its representativeness since most news outlets do not have online availability of news article content in their online domains prior to the year 2000. Figures S2 and S3 in the Supplementary Material illustrate the time ranges of article data analyzed based on news outlets articles online availability and the amount of words and articles per outlet/year.

Frequency of Word Usage

Yearly frequency usage of a target word in an outlet in any given year was estimated by dividing the total number of occurrences of the target word in all articles of a given year by the number of all words in all articles of that year. This method of estimating frequency accounts for variable volume of total article output over time.

Figure 1 shows the min–max scaled yearly frequencies of several sample words in *New York Times* (NYT) content during the past 50 years. Min–max scaling is a common way to normalize time series. The method rescales the range of the data to a scale between 0 and 1 using the formula shown in Equation 1 where y is the original frequency count for a given word and y' is its normalized/scaled value. Min–max scaling of frequency counts allows comparison of minimum and maximum temporal prevalence across terms in the corpus irrespective of their absolute frequencies.

$$y' = \frac{y - \min(y)}{\max(y) - \min(y)}. \tag{1}$$

In Figure 1, the interested reader can notice how the presidencies of Bush Senior and son are apparent in the corresponding word usage frequency plot on the upper left of the figure. Frequency counts of target words also capture the increasing prevalence of China in journalistic discourse, the disappearance of the Soviet Union or the transient nature of some technological artifacts such as cassettes and DVDs. Due to missing data for the year 1980 in the NYT domain and Internet cache repositories, year 1981 data is used to interpolate approximate frequency counts for 1980.

In a small percentage of articles, outlet specific XPath expressions failed to properly capture the content of the article due to the heterogeneity of HTML elements and Cascading Style Sheets styling

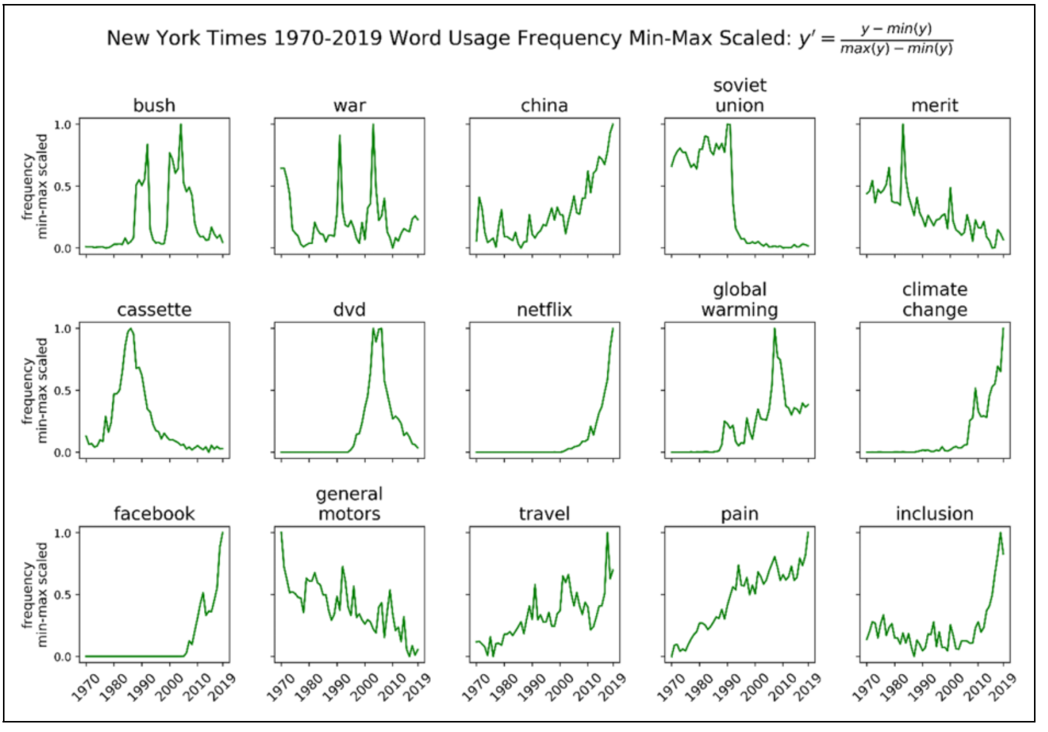

Figure 1. Min–max scaled yearly frequency of word usage in *New York Times* articles.

combinations with which articles text content is arranged in outlets online domains. As a result, the total and target word counts metrics for a small subset of articles are not precise. In a random sample of articles and outlets, manual estimation of target words counts overlapped with the automatically derived counts for over 90% of the articles.

Most of the incorrect frequency counts were only minor deviations from the actual counts such as, for instance, counting the word *Facebook* in an article footnote encouraging article readers to follow the journalist's Facebook profile and that the XPath expression mistakenly included as the content of the article main text. Some additional outlet-specific inaccuracies that we could identify occurred in *The Hill* and *Newsmax* news outlets where XPath expressions had some shortfalls at precisely capturing articles' content. For *The Hill*, in years 2007–2009, XPath expressions failed to capture the complete text of the article in about 40% of the articles. This does not necessarily result in incorrect frequency counts for that outlet but in a sample of articles' words that is about 40% smaller than the total population of articles words for those 3 years. In the case of *NewsMax*, the issue was that for some articles, XPath expressions captured the entire text of the article twice. Notice that this does not result in incorrect frequency counts. If a word appears x times in an article with a total of y words, the same frequency count will still be derived when our scripts count the word $2x$ times in the version of the article with a total of $2y$ words. To conclude, in a data analysis of 27 million articles, we cannot manually check the correctness of frequency counts for every single article and 100% accuracy at capturing articles' content is elusive due to the small number of difficult to detect boundary cases such as incorrect HTML markup syntax in online domains. Overall, however, we are confident that our frequency metrics are representative of word prevalence in print news media content (see Figures 1 and 2 for supporting evidence).

Comparing the frequency of words across different outlets can be informative to illustrate the different saliency of themes across outlets. Figure 2 shows similar and distinct patterns of word occurrence across four different news outlets that target different news consumption market segments. The prevalence of terms such as *opioids* or *fitness* is similar across outlets. In contrast, words such as *guys* or *baseball* are particularly prominent in the *New York Post* while terms such as *art* or *cuisine* are more prevalent in *The New York Times*. The distinction could be due to differences in socioeconomic status among outlets' readerships. Similarly, mentions of Blacks/African Americans tends to be consistently higher for *The Washington Post* than the Manhattan-based outlets (*The New York Times*, *New York Post*, and *The Wall Street Journal*)—likely a product of the demographic differences between these cities. Roughly 46% of DC residents are African American, as compared to about 18% of Manhattan residents (*U.S. Census Bureau QuickFacts*, 2020a; *U.S. Census Bureau QuickFacts: New York County (Manhattan Borough), New York*, 2020b).

Public Opinion Surveys and the Dyad Ratios Algorithm

We have gathered publicly available survey data about Americans' perceptions on the prevalence of different types of prejudice in the country. This is not to be confused with survey data attempting to assess the existence of prejudicial attitudes among survey participants. Rather, we collected surveys assessing participants' subjective perceptions about the severity of prejudice itself in the wider society.

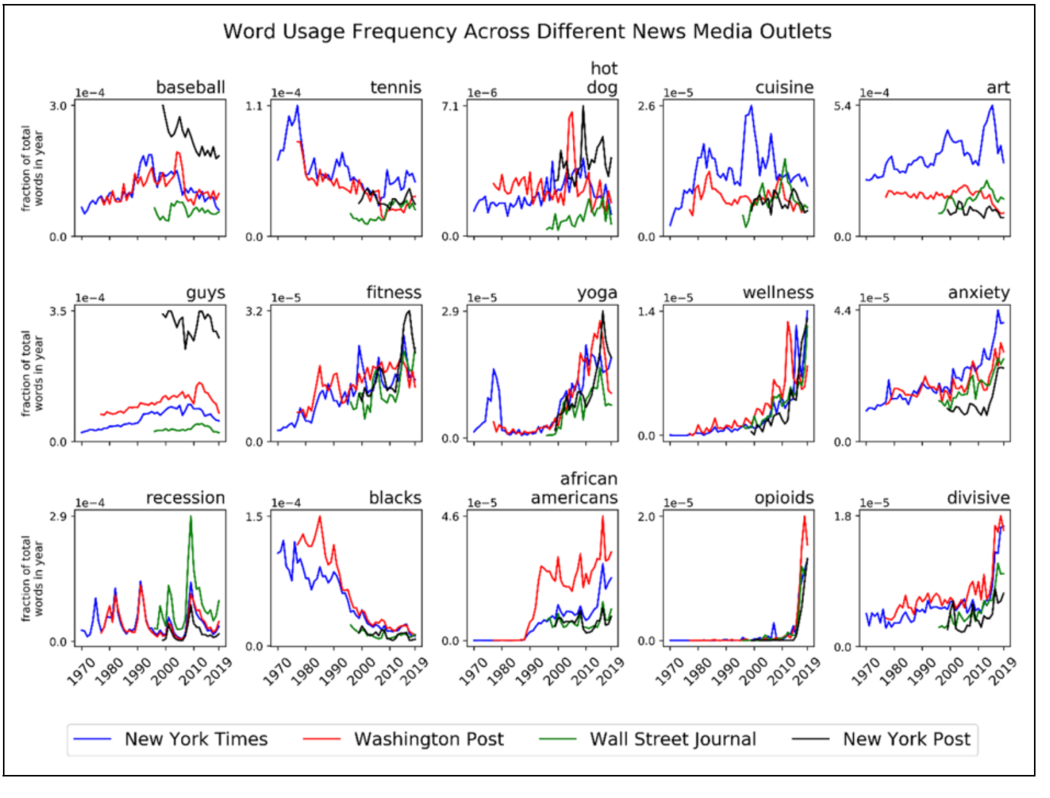

Figure 2. Yearly frequency of word usage across four different news media outlets.

To create robust indices on perceptions of prejudice severity that encompass the entire time range analyzed (2000–2019), we have "stitched" together different longitudinal surveys, each measuring a common latent variable of specific prejudice severity perceptions, using the *Dyad Ratios* algorithm (Stimson, 2018). The *Dyad Ratios* algorithm is a method for the extraction of a common dimension in longitudinal data such as survey marginal responses over time that are massively incomplete. That is, most variables (survey questions) do not exist for most time samples. The starting assumption of the *Dyad Ratios* algorithm is that the ratio of responses to the same survey question is empirical evidence of change in time for the underlying latent variable. If there are several distinct item questions measuring the same latent factor, the degree to which ratios at particular times covary across items is evidence of shared variance. Thus, the algorithm tries to determine how much of the questions' variance is shared with the latent concept.

For instance, if we want to estimate the temporal dynamics of a latent factor such as public perceptions of sexual orientation prejudice severity in society, we can collect several longitudinal surveys from different organizations such as *Gallup* or *Pew Research*, all attempting to measure with different questions and scale of responses the same latent factor: public perceptions on sexual orientation prejudice severity. For a longitudinal survey question to qualify as an input variable i for the *Dyad Ratios* algorithm, survey data for the same question using the same scale of responses needs to exist for at least two different points in time, p and q. The *Dyad Ratios* algorithm then uses the ratio r_{ipq}(see Equation 2) of survey responses for survey question i at two distinct time points, p and q, as an estimate of change in the underlying latent factor. Since ratios lack intrinsic units, different questions $(i, j, k, \ldots)$ corresponding to different longitudinal surveys attempting to measure the same latent factor can then be aggregated and interpolated. This accomplishes the double purpose of reducing noise and bridging the substantial temporal gaps between each independent longitudinal survey question i, where data for most years are often missing. The output of the *Dyad Ratios* algorithm is a single aggregate time series estimate of the latent factor dynamics covering continuously the entire studied time range. A more detailed description of the algorithm is provided in the Supplementary Material or in the original source (Stimson, 2018).

$$r_{ipq} = \frac{x_{ip}}{x_{iq}}. \quad (2)$$

Aggregating several survey questions i from different organizations about the severity of each prejudice type studied in this work minimizes single longitudinal survey bias and helps create robust indexes on public perceptions of prejudice severity across time. For example, for tracking perceptions on severity of ethnic prejudice in U.S. society, The General Social Survey (GSS) *racdif1* variable has systematically asked with yearly or biannual frequency the question: "On the average (Blacks/African Americans) have worse jobs, income, and housing than White people. Do you think these differences are mainly due to discrimination?" Similarly, Gallup has asked "For each of the following groups, please say whether you are very satisfied, somewhat satisfied, somewhat dissatisfied, or very dissatisfied with the way they are treated. How about Blacks." The *Dyad Ratios* algorithm has the advantage of combining both of these longitudinal surveys and similar others into a single index that provides a more robust overall measurement of sentiment across the population than either longitudinal survey in isolation. We use mainly the *Roper iPoll* database to find survey data but we also utilize other surveys search engines such as those from *Gallup*, *GSS*, or *YouGov* databases to search for relevant surveys.

The main criteria we use for the inclusion of a survey question as an input variable to the *Dyad Ratios* algorithm is that the same question (or very similarly worded question) and same or very similarly worded scale of answers was asked at least at two different years, since the *Dyad Ratios* algorithm needs at least two sample points to estimate a ratio. Another requirement for the inclusion of surveys in our sample is the usage of a national representative sample with no oversampling or exclusive sampling of a particular demographic group. Additionally, we excluded surveys asking about personal and subjective lived experiences of discrimination such as “Have you ever been discriminated in your job because of your gender?” instead focusing on more general society–wide perceptions on the severity of specific prejudice types.

Survey data on public perceptions of ethnic prejudice severity are relatively abundant. We have located 108 survey sample points from 19 different question types (i.e., input variables) spanning the studied time range and that we aggregate into a common index using the *Dyad Ratios* algorithm to create a single time series index of the latent variable: public perceptions on severity of racism in society. The same procedure is applied for the other prejudice types. For perceptions on gender prejudice, we aggregate 33 sample points from six different question types. For perceptions on sexual orientation prejudice, we aggregate 44 sample points from four different question types.

The biggest methodological limitation of our approach has been the sparsity of longitudinal survey data about people’s perception on the severity of anti-Semitism or Islamophobia. We have not been able to find consistent survey questions that have regularly asked a nationally representative sample on a yearly or biannual basis questions regarding perceptions of anti-Semitism or islamophobia, using the same scale of responses for the time range 2000–2019. Since only survey questions that have been asked at least at 2 different years can be fed into the *Dyad Ratios* algorithm, this excluded surveys with only one sample point. The limiting inclusion criteria of only using surveys with a nationally representative sample—excluding those gathering data exclusively from a specific demographic subpopulation (Jews, Muslims, etc.), or those with oversampling of a particular demographic group—further limited the availability of valid survey data.

To get around the data sparsity challenges with respect to perceptions of prejudice against Muslims, we relaxed the inclusion criteria in order to have better coverage of the 2006–2013 time interval that was otherwise sparsely populated. Thus, we included a longitudinal survey question from Gallup about Americans’ perceptions of how Arabs are treated. This question deviates from precisely asking survey participants about the severity of discrimination against Muslims, but we still use it as a proxy for latent perceptions of Islamophobia due to the strong associations among Americans between Arab ethnicity and Muslim religion (“Racially Profiling ‘Jihadists’ Sounds Like Common Sense. Here’s Why It Doesn’t Work,” 2016; Zogby, 2018), despite the fact that a majority of U.S. Arabs are actually Christian (Telhami, 2002), and most American Muslims are non-Arab (Johnson, 2011). We acknowledge that this relaxation of the inclusion criteria is suboptimal but it was also necessary to properly track the sparsely populated 2000–2013 time interval for public perceptions on Islamophobia. The specific question reads:

> Next, we’d like to know how you feel about the way various groups in society are treated. For each of the following groups, please say whether you are very satisfied, somewhat satisfied, somewhat dissatisfied, or very dissatisfied with the way they are treated. How about . . . Arabs?

Despite the inclusion of this question, the time interval 2000–2005 is still very poorly covered with survey data only available for the years 2000 and 2005. In total, for the Islamophobia index, we aggregate 22 sample points from three different question types.

Finally, longitudinal survey data on perceptions of anti-Semitism that fulfilled our inclusion criteria were also extremely sparse. We have only been able to locate 15 sample points from three different question types. The time interval 2000–2012 is particularly poorly covered with sample data only available for 2003, 2005, and 2009. We could not find survey data on public perceptions of transgender prejudice prior to 2011, so we dropped this category from the analysis to maintain the consistent time frame 2000–2019 across all the other prejudice types. Survey data prior to the year 2000 becomes even more scant. The suboptimal availability of survey data for perceptions of some prejudice types, particularly Islamophobia and anti-Semitism, warrants caution when interpreting results derived from those surveys. The full set of survey questions, dates, reference responses, survey IDs, and URL locations are provided as Supplementary Material in electronic form.

Analysis Scripts and Data Availability

The analysis scripts, surveys data, cable news prevalence metrics, human ratings of outlet ideological leanings, list of written articles URLs analyzed, and the counts of target words and total words per article are available in the following repository: https://doi.org/10.5281/zenodo.5073079

Results

Prevalence of Prejudice-Denoting Words in Written News Media

Figure 3 illustrates the increasing types of prejudice in two influential newspapers in the United States: *The New York Times* (in blue) and *The Washington Post* (in red). A clear trend of increasing prevalence of prejudice-related terms is apparent with words such as *racist* or *sexist* increasing in usage between

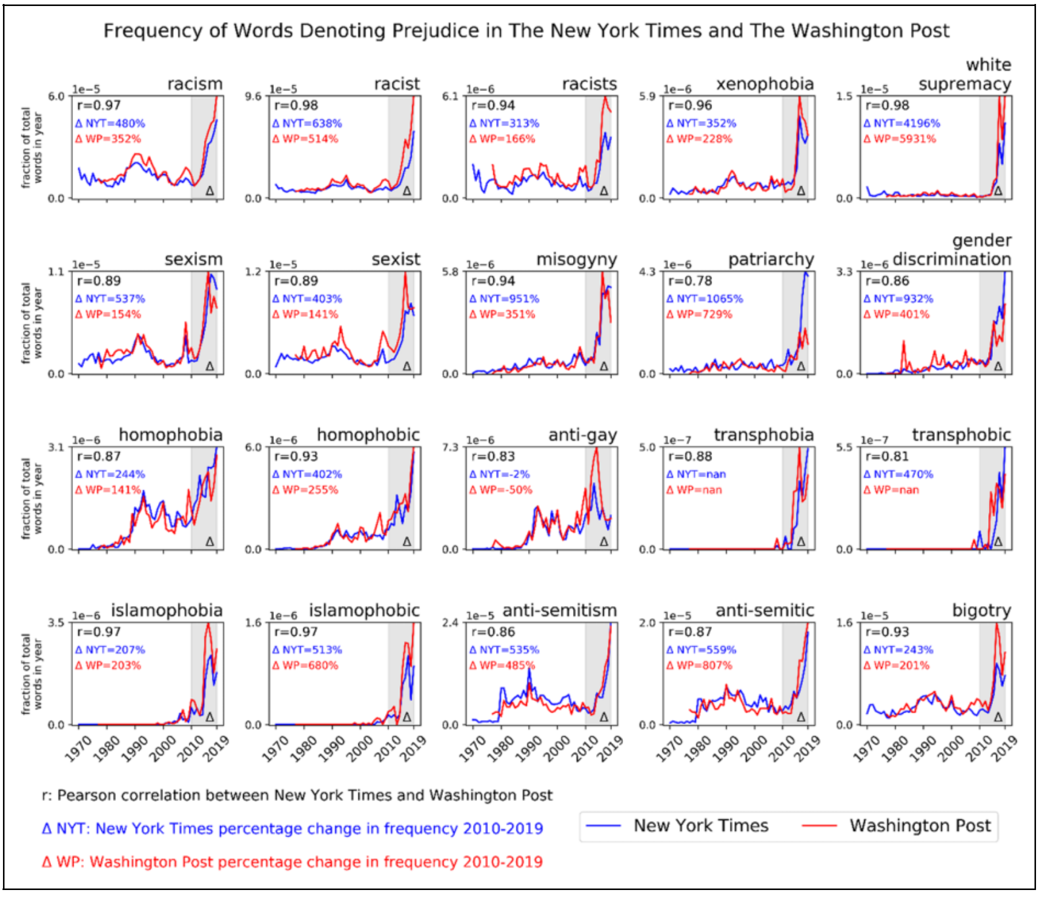

Figure 3. Yearly usage frequency of prejudice-denoting terms in *The New York Times* and *The Washington Post* news and opinion articles. *Note*. The Pearson correlation coefficient, *r*, between both time series and the percentage change in frequency usage, Δ, between 2010 and 2019 are shown in the upper-left corner of each plot.

2010 and 2019 by 638% and 403% in *The New York Times* or 514% and 141%, respectively, in *The Washington Post*. The yearly usage of prejudice-related words is highly correlated between both outlets as shown by the Pearson correlation coefficient, *r*, in the upper-left corner of each plot.

Aggregating yearly frequency counts across the 47 news media outlets analyzed shows that the trend from Figure 3 is not circumscribed to *The New York Times* and *The Washington Post* but it is the general tendency across most news media outlets (see Figure 4). The pattern highlighted in Figure 4 is not exclusive to the specific set of words shown in the figure. A different set of prejudice-related terms also shows a similar trend (see Figure S4 in Supplementary Material). Additionally, the pattern in Figure 4 is not due to larger availability of news outlets content in recent years. A similar pattern is apparent when replicating this experiment using 12 outlets with news and opinion articles availability exceeding our inclusion threshold of 1.25 million words for all years since 2000 (see Figure S5 in Supplementary Material). A very subtle trend in Figure 4 (highlighted by the gray dashed vertical bar) is that the prevalence of a reduced set of prejudice-denoting words such as *racism*, *sexism*, or *bigotry* also experienced a milder usage peak in the 1990s.

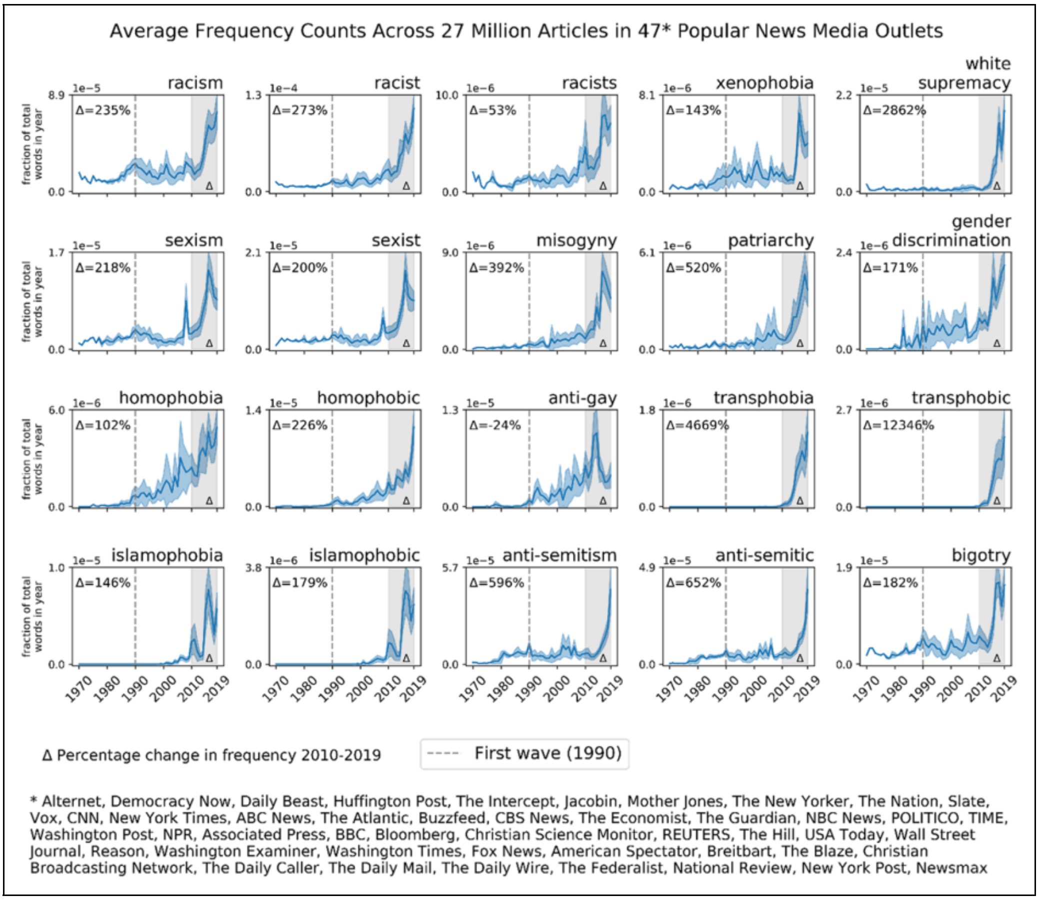

Figure 4. Average frequencies of prejudice-denoting terms across 47 popular news media outlets. *Note*. The percentage change in frequency usage, Δ, between 2010 and 2019 is shown in the upper-left corner of each plot. The shaded area around the trend line indicates the 95% confidence interval.

We next use factor analysis to quantify shared variability among the studied prejudice-denoting terms shown in Figure 4. Factor analysis allows elucidation of whether an underlying latent factor captures most of the variance observable in the individual dynamics of each prejudice-denoting term time series. Systematic interdependence among a set of variables is relevant for characterizing the latent factors that gives rise to joint variation. Factor analysis of all the prejudice-signifying words time series in Figure 4 does indeed show that a single factor accounts for over 78% of all the variance (Cronbach's $\alpha = 0.99$; see Figure S6 in Supplementary Material). Kaiser–Meyer–Olkin test to measure the suitability of data for factor analysis is 0.76. Redoing the analysis by combining the 20 prejudice terms in Figure 4 with the set of 20 additional prejudice-denoting terms in Figure S4 of the Supplementary Material generated similar results with a single factor accounting for over 75% of all the variance in the set of 40 terms (Cronbach's $\alpha = 0.99$; see Figure S7 in Supplementary Material. Repeating the analysis of Figure S6 using just the 12 outlets with volume of article content fulfilling our inclusion criteria since the year 2000 generates similar results (see Figure S8 in Supplementary Material).

Figure 5 plots the prevalence of prejudice-related words in news outlets aggregated by ideological leanings using human annotations of media political bias from the 2019 AllSides Media Bias Ratings Version 1.1 (AllSides Media Bias Ratings, 2019). The figure shows that the growing usage of prejudice-related words in news articles has been consistent across news outlets regardless of their ideological leanings, but overall, prejudice-denoting words appear to be less prevalent in centrist outlets as shown by the green trend line (representing centrist outlets) being consistently below the blue (left-leaning outlets) and red (right-leaning outlets) trend lines.

We next compare overall news media prevalence across prejudice types. Figure 6A shows the average prevalence of related word pairs denoting six distinct types of prejudice. Both historically and in recent years, the *racism* theme displays the highest absolute prevalence in news written articles followed by *anti-Semitism*, *sexism* and *homophobia*. *Islamophobia* and *transphobia* appear to be the least prevalent prejudice-related themes in recent journalistic discourse.

Figure 6B shows the min–max scaled average frequencies of prejudice-specific word pairs to highlight times of maximum relative usage irrespective of overall prevalence. Notice that for the year 2019, four different types of prejudice cluster in the top right of the plot, denoting maximum usage over the entire time range. A dashed gray vertical bar indicates the year 2015, when Donald Trump entered the contest for the nomination of the Republican Party to the presidency of the United States. The figure shows that in the previous year, 2014, the usage of words denoting racism, homophobia, transphobia, or sexism were at or near, up to that year, all-time highs. These results suggest that the trend of increasing prevalence of prejudice-related words in media discourse precedes the political emergence of Donald Trump—although Trump's presidency and subsequent reactions to it may have exacerbated these trends. Repeating these analyses using just the 12 outlets with volume of article content fulfilling our inclusion criteria since the year 2000 generates similar results (see Figure S9 in Supplementary Material).

To discern whether the increasing usage trend of prejudice-denoting words accelerates after 2015, we carried out a paired t test of the word pairs frequencies slopes between 2010–2014 and 2015–2019 for the target words in Figure 4 only for the 30 outlets with article text volume exceeding our inclusion criteria since the year 2010. Results appear to indicate an acceleration of the trend after 2015, though statistical significance was marginal, $t(19) = -2.59$, p value $\approx .02$ (see Table S1 in Supplementary Material for details).

Correlations of Prejudice-Denoting Word Usage Across Written News Media Outlets

Correlations of yearly frequency counts for specific prejudice themes across the 47 news media outlets in the 2000–2019 time range suggest that left-leaning and centrist news media outlets tend to

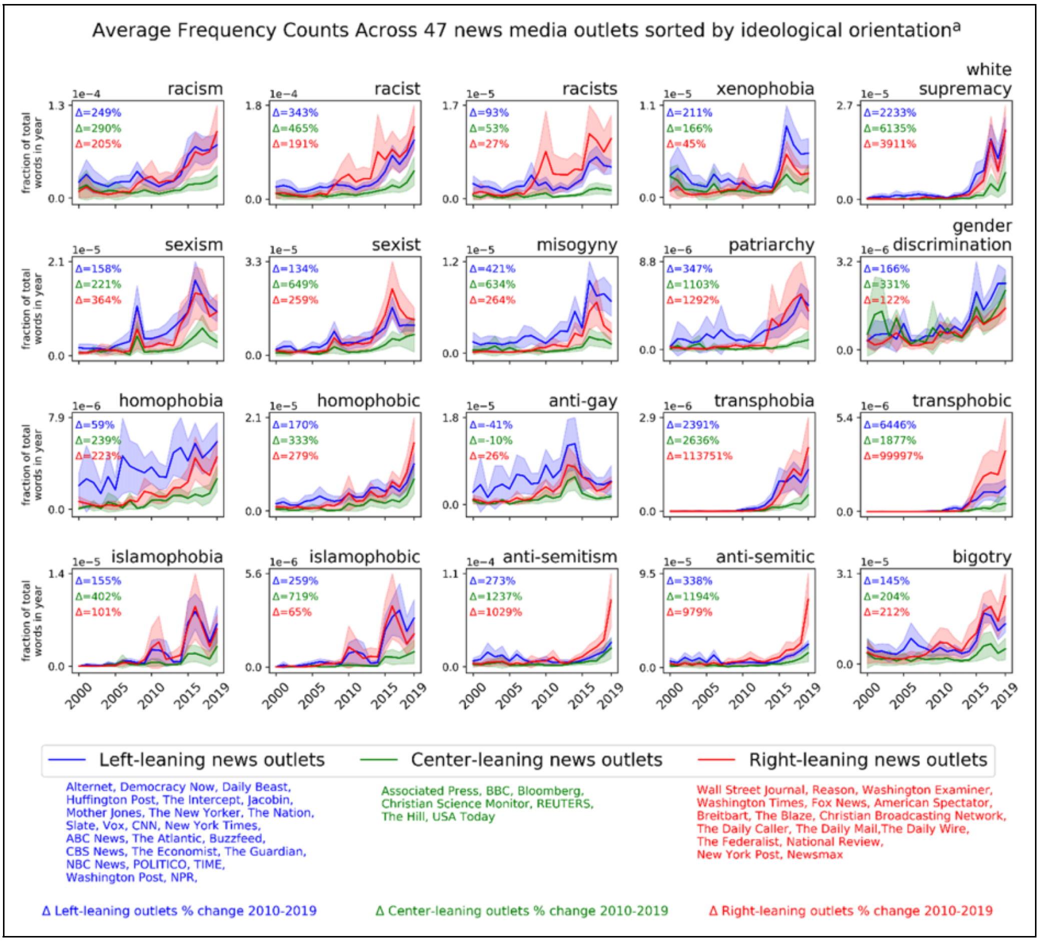

Figure 5. Average frequencies of prejudice-denoting terms across media organizations sorted by the ideological leanings of news outlets. *Note*. The percentage change in frequency usage, Δ, between 2010 and 2019 is shown in the upper-left corner of each plot. The shaded areas around trend lines indicate the 95% confidence intervals. [a] Political leanings labels from AllSides Media Bias Chart Version 1.1 (https://www.allsides.com/blog/updated-allsides-media-bias-chart-version-11).

be highly synchronized in the yearly usage of words that denote racism. In contrast, moderate left-leaning, centrist, and right-leaning outlets tend to be highly correlated in the usage of terms that denote anti-Semitism. Centrist and right-leaning outlets also seem to be correlated in their usage of terms that denote homophobia (see Figure 7).

An important limitation of Figure 7 is that several of the outlets used in the analysis do not contain article data all the way back to the year 2000. This can be due to some outlets being created after the year 2000 like, for instance, *Breitbart* or the *Huffington Post*. Alternatively, it can also be due to lack of online availability of news articles content above our minimum tokens per outlet/year threshold for inclusion in our analysis. Thus, word usage frequency time series for some outlets are very short and their correlation metrics are less certain. Figure S10 in the Supplementary Material displays correlation matrices for the 12 outlets in our data set fulfilling our minimum volume of article words

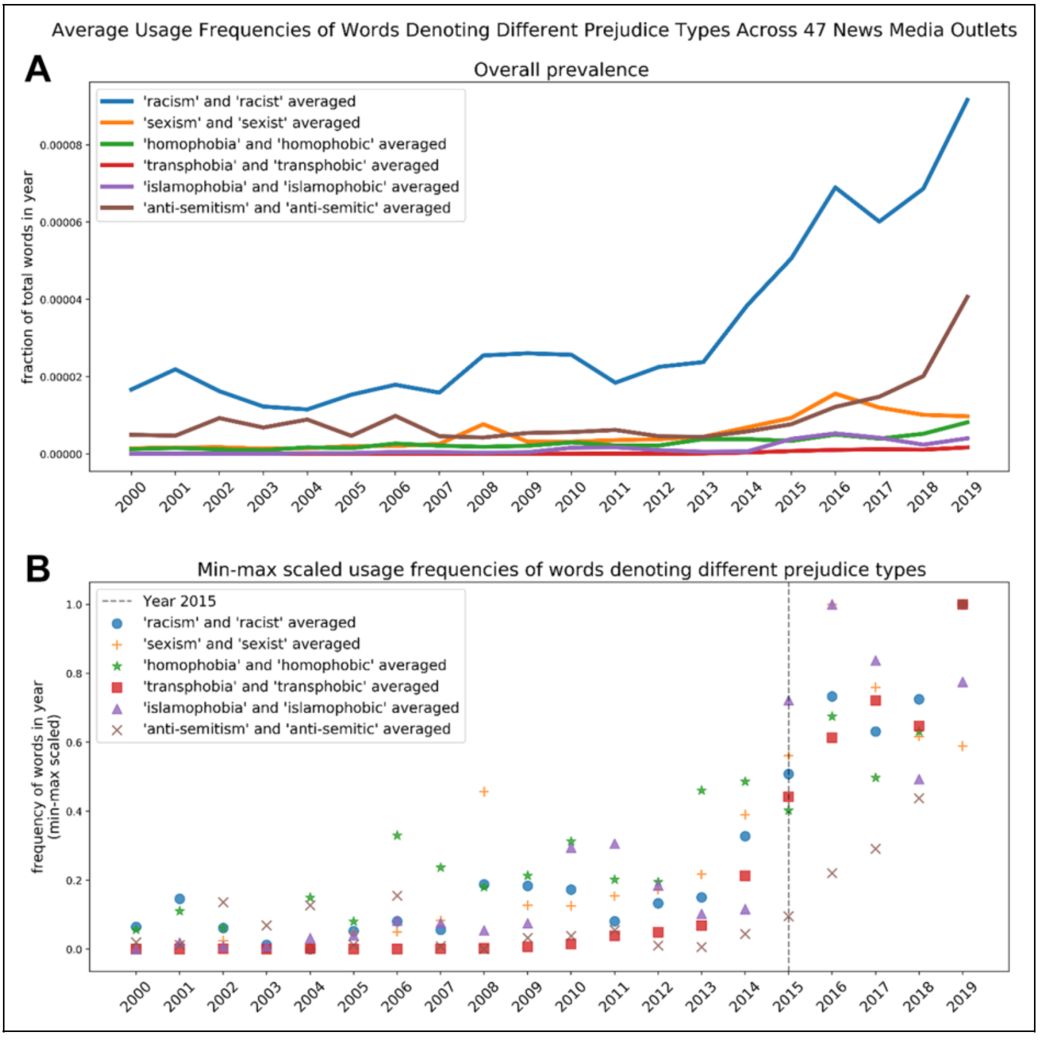

Figure 6. Subfigure A shows the overall prevalence of different types of prejudice themes in media discourse. Subfigure B shows min–max scaled yearly frequencies to visualize years of maximum and minimum word usage irrespective of overall prevalence. *Note*. The year 2015, marking the entrance of Donald Trump in the U.S. presidential contest, is denoted with a vertical dashed gray line.

per outlet/year since at least the year 2000. The figure also displays patterns of high and low correlation between outlets with respect to their usage of prejudice-denoting words.

Correlation Between Written News Media and TV Cable News

Using word prevalence data from Stanford Cable TV News Analyzer (Computer Graphics Lab at Stanford University, n.d.; Hong et al., 2020; containing data all the way back to 2010), we compare the prevalence of prejudice-related words between written news media and TV cable news (CNN, Fox, and MSNBC) for the 2010–2019 time frame (see Figure 8). The degree of correlation between prejudice-denoting words in written news media and TV cable news is very high as indicated by the large r Pearson correlation coefficients in the upper-left corner of each plot.

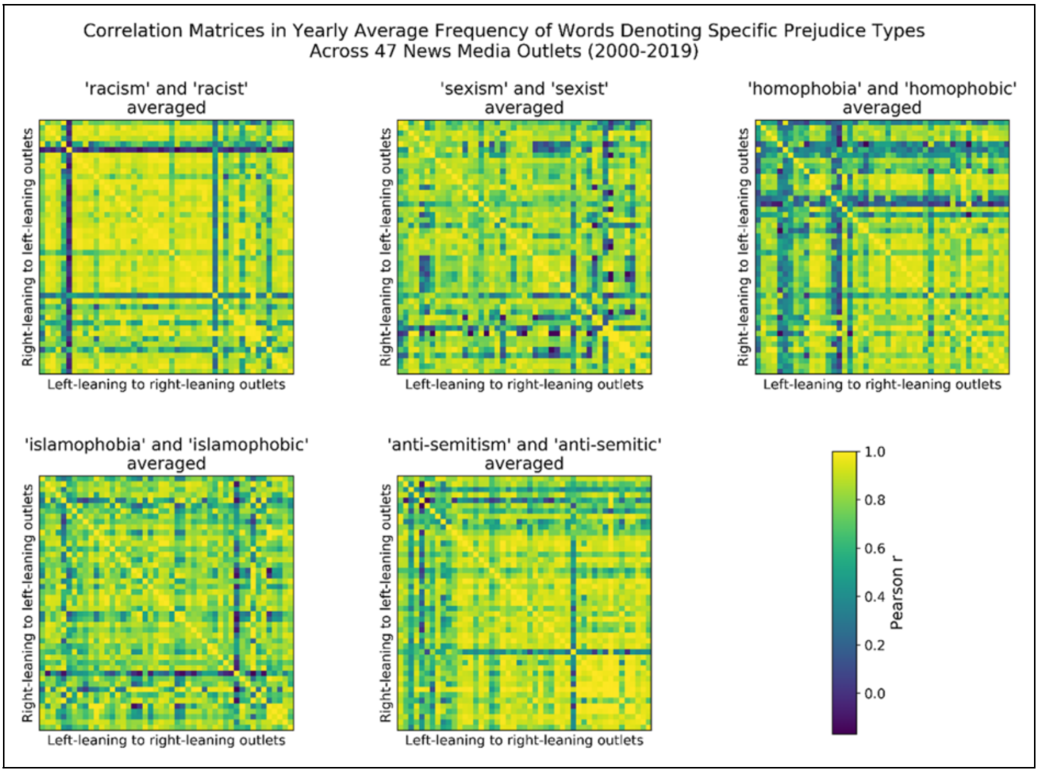

Figure 7. Correlation across outlets in yearly usage frequency of different prejudice themes. *Note*. Outlets are arranged in the axis according to ideological leaning ratings from the AllSides organization (AllSides Media Bias Ratings, 2019).

News Outlets Pioneers and Followers in the Usage of Prejudice-Denoting Words

Granger-causality tests can be used to determine whether a time series of prejudice-denoting words yearly frequency in one outlet is predictive of future frequency counts of those words in another outlet (Granger, 1969). The term *causality* in Granger *causality* is misleading because one time series preceding another is a necessary but not sufficient condition for establishing causation. Thus, we use the Granger-causality test to simply describe statistically significant precedence in time of an independent time series that is predictive of a dependent time series using lagged values of the independent and dependent time series. We use the sum-of-squared-residuals based χ^2 test for determining effect size. We adjust p values to control for multiple comparisons using the Bonferroni adjustment.

Prior to testing Granger causality, time series of word frequencies and survey data (used in the next section) need to be assessed for stationarity. Commonly used methods to assess stationarity are the Kwiatkowski–Phillips–Schmidt–Shin test (KPSS; Kwiatkowski et al., 1992) and the Augmented Dickey–Fuller (ADF; Dickey & Fuller, 1979) unit root test. Unfortunately, the tests often generated contradictory results in our time series data due to each test using opposite null hypotheses. ADF null hypothesis is that the time series has a unit root. In contrast, KPSS null hypotheses is that the time series is stationary. The short nature of the time series analyzed ($N = 20$ years) renders estimations of stationarity underpowered for either test. This impasse is resolved by applying two distinct approaches to assess and enforce stationarity and we report the results of both. One method uses

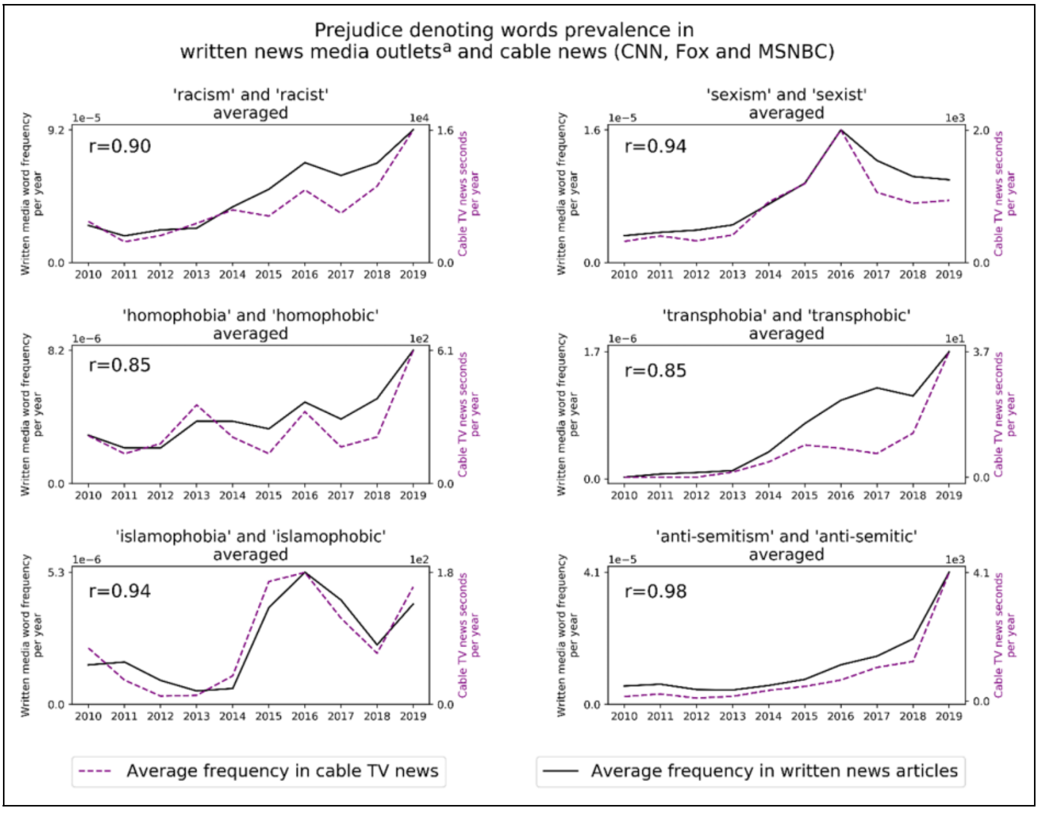

Figure 8. Word prevalence in written news media articles and cable news. *Note*. Pearson correlation coefficient between both time series is shown in the upper-left corner of each plot. The source for cable news data is Stanford Cable TV News Analyzer (Computer Graphics Lab at Stanford University, n.d.; Hong et al., 2020). [a] *Alternet*, *Democracy Now*, *Daily Beast*, *Huffington Post*, *The Intercept*, *Jacobin*, *Mother Jones*, *The New Yorker*, *The Nation*, *Slate*, *Vox*, *CNN*, *New York Times*, *ABC News*, *The Atlantic*, *Buzzfeed*, *CBS News*, *The Economist*, *The Guardian*, *NBC News*, *POLITICO*, *TIME*, *Washington Post*, *NPR*, *Associated Press*, *BBC*, *Bloomberg*, *Christin Science Monitor*, *REUTERS*, *The Hill*, *USA Today*, *Wall Street Journal, Reason*, *Washington Examiner*, *Washington Times*, *Fox News*, *American Spectator*, *Breitbart*, *The Blaze*, *Christian Broadcasting Network*, *The Daily Caller*, *The Daily Mail*, *The Daily Wire*, *The Federalist*, *National Review*, *New York Post*, and *Newsmax*.

the KPSS test to assess stationarity using the null hypothesis that the data are stationary around a constant. If the time series are deemed nonstationary, differencing is applied until KPSS suggests stationarity. In the other approach, both the ADF and the KPSS stationarity tests are interpreted simultaneously. For time series for which the presence of a unit root cannot be rejected, the series is differentiated. For time series for which the presence of a trend cannot be rejected, the series is de-trended. This process continues until the time series is made stationary.

Figure 9 shows a matrix of color-coded p values (Bonferroni corrected for multiple comparisons) representing Granger-causality tests for time series of average frequency of prejudice-denoting words in an outlet (columns) being predictive of future time series (lag $=$ 1 year) of the same set of words in another outlet (rows). The KPSS test was used to assess the stationarity of the word frequencies time series. The columns with the largest amount of p values below the .01 significant threshold are color-coded in red to highlight the outlets that have led the way in the usage of prejudice-denoting words. The matrix shows that some influential outlets such as *The New York*

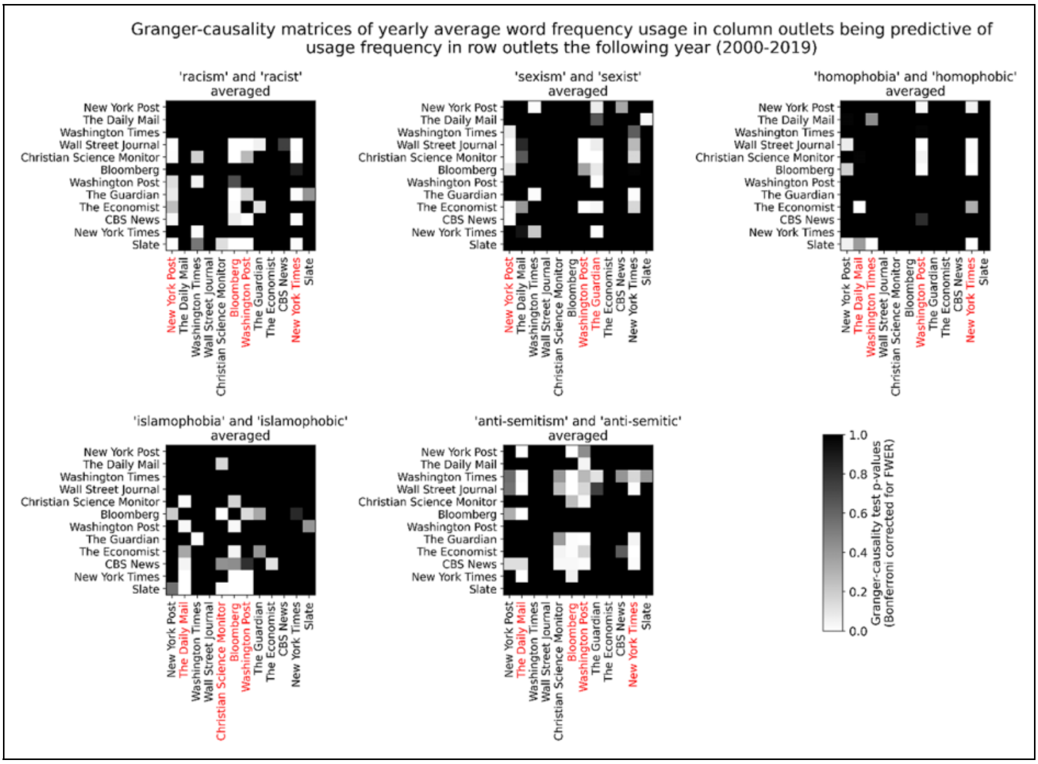

Figure 9. Granger-causality tests for whether prejudice-denoting terms yearly usage in column outlets are predictive (low p value) of same prejudice-denoting terms usage in row outlets the following year. *Note.* Each matrix shows color-coded Granger causality p values, Bonferroni adjusted for multiple comparisons. Column outlets highlighted in red are the column with the largest amount of p values below the .01 threshold indicating that the usage of prejudice signifying words by those outlets are most predictive of usage frequencies in other news outlets in the subsequent year.

Times, *Bloomberg*, or *The Washington Post* have been trendsetters in the usage of prejudice-related terms while other outlets have followed the trend in the subsequent year. The figure only displays data for the 12 outlets for which online availability of news articles content since the year 2000 is above the minimum words per year inclusion threshold listed in the Method section. Redoing this analysis using the ADF and KPSS methods to assess stationarity of time series generated resembling results, with outlets such as *The New York Times*, *Bloomberg*, or *The Washington Post* also appearing to sometimes predict usage of prejudice-denoting words in other outlets in the subsequent year. However, statistical significance was substantially more tenuous (see Figure S11 in Supplementary Material).

Prevalence of Prejudice-Denoting Words in Written News Media and Public Perceptions of Prejudice Severity

We next compare indexes derived from survey data on public opinion perceptions about the severity of different types of prejudice in U.S. society and news media frequency usage of prejudice-denoting words. The correlation between both time series for most types of prejudice is very high (see Figure 10). A notable exception is the relationship between homophobia denoting words prevalence in news media and public perceptions on severity of homophobia, where the correlation is negative.

We next test whether prejudice words frequency usage in news media predicts shifts in public opinion or the other way around. The same limitations described in the previous section regarding the tests used to determine stationary of time series being underpowered due to the short nature of the time series analyzed apply to this analysis. We again report results for the two different approaches used to test stationarity.

We first use the KPSS test to assess stationarity and apply differencing if needed to stationarize the time series. Granger-causality tests, Bonferroni adjusted for multiple comparisons, for test lags of 1, 2, and 3 years, show that word usage of ethnic and gender prejudice-denoting terms in news media is predictive of shifts in public opinion about the severity of ethnic and gender prejudice (see Figure 10). All Granger causality tests in the reverse direction (public opinion → word frequency) for all types of prejudice are not significant. Repeating this analysis using just outlets with data availability above our inclusion threshold since the year 2000 reproduced these results (see Figure S12 in Supplementary Material). All Granger-causality tests in the reverse direction (public opinion

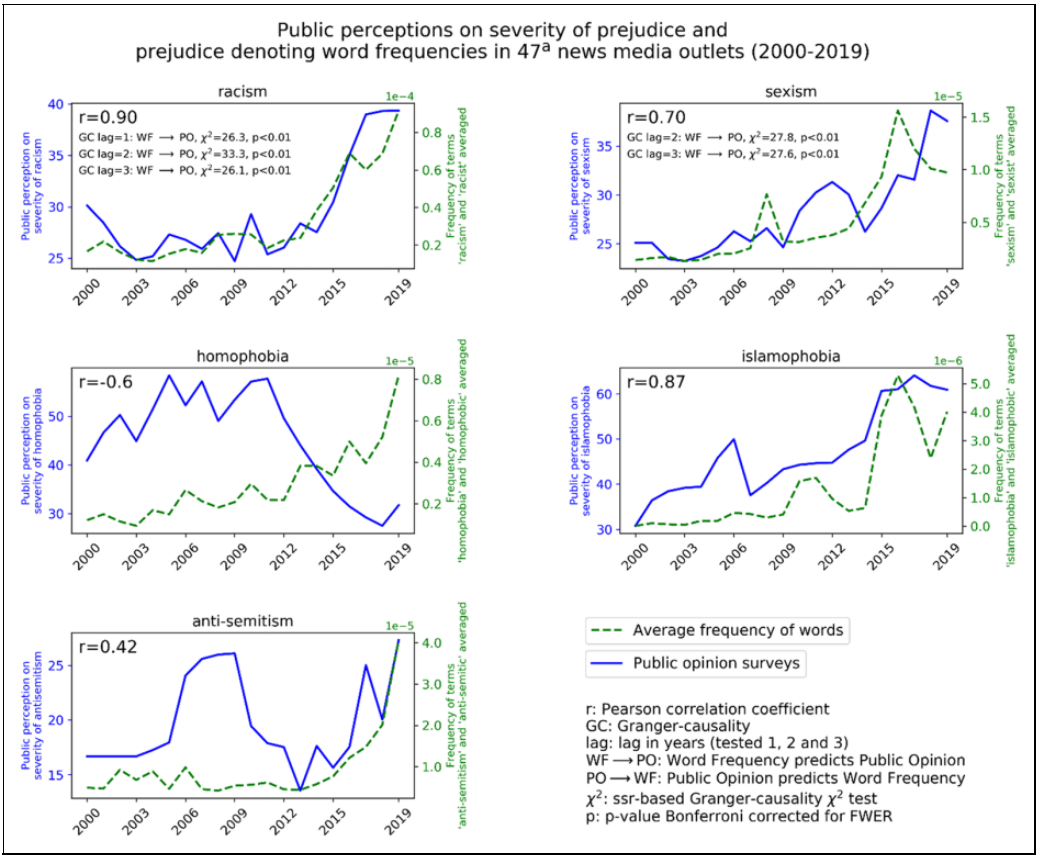

Figure 10. Relationship between frequency of specific prejudice terms in 47 popular news outlets and public opinion survey data about people's perceptions on severity of prejudice in society. [a] *Alternet*, *Democracy Now*, *Daily Beast*, *Huffington Post*, *The Intercept*, *Jacobin*, *Mother Jones*, *The New Yorker*, *The Nation*, *Slate*, *Vox*, *CNN*, *New York Times*, *ABC News*, *The Atlantic*, *Buzzfeed*, *CBS News*, *The Economist*, *The Guardian*, *NBC News*, *POLITICO*, *TIME*, *Washington Post*, *NPR*, *Associated Press*, *BBC*, *Bloomberg*, *Christin Science Monitor*, *REUTERS*, *The Hill*, *USA Today*, *Wall Street Journal, Reason*, *Washington Examiner*, *Washington Times*, *Fox News*, *American Spectator*, *Breitbart*, *The Blaze*, *Christian Broadcasting Network*, *The Daily Caller*, *The Daily Mail*, *The Daily Wire*, *The Federalist*, *National Review*, *New York Post*, and *Newsmax*.

→ word frequency) for all types of prejudice are again not significant. When using both KPSS and ADF to test for stationarity, Granger causality tests suggest again ethnic and gender prejudice word usage in news media being predictive of shifts in public opinion about perceptions of prejudice severity in society. However, the effect sizes are markedly smaller than with the previous analysis and the *p* values are marginal (between 0.01 and 0.05; see Figure S13 in Supplementary Material). All Granger-causality tests in the reverse direction (public opinion → word frequency) for all types of prejudice are again not significant. Results of repeating this analysis using just outlets with data availability above our inclusion threshold since the year 2000 fail to reach statistical significance for Granger-causality tests in both directions: word frequency → public opinion and public opinion → word frequency (see Figure S14 in Supplementary Material).

Discussion

This work has characterized a substantial increase on the prevalence of prejudice-denoting words in news media within the 2010–2019 period. It is natural to ponder about the causes underlying this trend and its relationship with concomitant rising concern about prejudice severity in public opinion. Previous theoretical frameworks and empirical results offer potential candidate explanations for the rising thematic prevalence of prejudice in news media content and its link with public opinion. We attempt to enumerate the most salient ones below in light of our findings.

The "Agenda Setting" Hypothesis

Previous research has shown that news media can play an important "agenda setting" role with respect to public opinion (McCombs & Valenzuela, 2021). For instance, increased intensity of media coverage about terrorism or crime has been shown to precede increased public concern about terrorism or crime, irrespective of the actual prevalence of terror incidents or empirical trends in crime rates (Callanan, 2012; Smith et al., 2019). Indeed, media coverage regularly outstrips or outright defies empirical trends with respect to incidents of terrorism and crime and trends in public perception often follow suit (Lowry et al., 2003). The "agenda-setting" literature also finds that certain core outlets seem to drive the conversation for most other media, as writers across the political spectrum react to or strive to emulate coverage in prestige media outlets. Coverage trends in prestige outlets therefore tend to echo throughout the media landscape, irrespective of other outlets' ideological lean—and coverage trends in print, online, and television media tend to overlap considerably (McCombs, 2005). The political and financial incentives of media organizations and journalists seem to play an important role in driving the "agenda" set by the media (al-Gharbi, 2020).

At first blush, the findings of this study seem highly compatible with other work on news media "agenda setting" (al-Gharbi, 2020; Lowry et al., 2003; McCombs, 2005; Smith et al., 2019). We find significant overlap between the prevalence of prejudice-denoting words in written news media and trends in public perception of prejudice severity in the United States. Furthermore, the prevalence of prejudice-denoting words in certain influential outlets such as *The Washington Post*, *Bloomberg*, or *The New York Times* can be predictive of usage frequency of such words in other outlets.

Our results also reveal that TV cable news usage of prejudice signifying words follows a similar trend to that of print media, revealing significant synchronicity between both media types regarding the prevalence of prejudice-denoting themes. This tight coupling between cable and print news is in line with a substantial body of previous literature noticing or predicting convergence of print, television, and online news media (Cooke, 2005; Huang & Heider, 2007; Kawamoto, 2003; Klinenberg, 2005; Singer, 2004, 2009).

Despite the methodological limitations of an observational study to determine causal links, our results are consistent with the central arguments in the agenda-setting literature. Prejudice could function similarly to terrorism or crime themes within the media ecosystem, suggesting that media coverage on these topics may be informed by a similar set of financial and political considerations. Exploring this hypothesis could be a fruitful area for further research.

The Role of the 2016 U.S. Presidential Election

It may be the case that instead of media coverage driving public perceptions of prejudice, both media discourse *and* public opinion are responding to some other factor—such as an increase in prejudicial language or behaviors within the United States. Some authors have argued that the 2016 U.S. presidential election and what they called the "Trump effect" have had a substantial impact on growing overt expression of prejudice toward specific protected groups (Crandall et al., 2018; Quinton, 2019). If true, this could partly predict the recent upward tick in the usage of prejudice-denoting words in news media, if we assume that news outlets have quickly noticed growing prejudicial attitudes or behaviors and are often informing the public about them. However, the evidence for the "Trump effect" hypothesis is contradictory, with other work reporting decreases in prejudicial attitudes until at least 2016 (Charlesworth & Banaji, 2019; Moberg et al., 2019) and during the first 2 years of the Trump presidency (Hopkins & Washington, 2020).

Our results show that the upward trends in the prevalence of prejudice-denoting words in news media began *prior to* Donald Trump running for, and subsequently serving as, President of the United States. Thus, Trump seems to have emerged on the political scene in a context of already-heightened news media discussion about prejudice and discrimination. That said, our findings do appear to indicate that the trend accelerated after 2015.

Other Potential Explanatory Factors Underlying the Described Results

It is also plausible that modern Western societies have simply grown more intolerant of discrimination and more fine-tuned to oppose prejudice against protected groups in recent years (Mallett & Monteith, 2019). Whereas in previous decades, overtly prejudicial societies would avoid denouncing prejudice or be constitutionally incapable of recognizing prejudice as such, contemporary news media may be more adept at identifying and denouncing prejudice against protected groups. Growing sensitivity to mistreatment of protected groups and assertiveness of egalitarian attitudes could predict increased prevalence of prejudice-denoting words in news media. A relaxation of the criteria used to define prejudice itself (Haslam, 2016; Levari et al., 2018) may have exacerbated this trend.

Our factor analysis results showed that over 78% of the variance in the prevalence of the 20 prejudice-denoting words analyzed is accounted for by a single factor. This is suggestive of a latent thread that permeates journalistic discursive patterns across the topics of ethnicity, gender, and sexual or religious orientation and from which the increased prevalence of prejudice-denoting terms could have flown. This is consistent with a story of growing sensitivity within the media and among the public with respect to prejudice and inequality irrespective of reference group—or perhaps it may reflect increased perceptions that different forms of prejudice across reference groups are interrelated.

Alternatively, the observed patterns in media discourse and public opinion could also be partially explained in terms of cultural shifts which have purportedly increased the incentives to appeal to group or victimhood identity in situations of social conflict (Campbell & Manning, 2014; Fassin & Rechtman, 2009; Leong, 2021; Lukianoff & Haidt, 2018) and to offer public expressions of moral judgment as a means of gaining moral prestige (Tosi & Warmke, 2016).

The Outlier Trend on Perceptions of Sexual Orientation Prejudice

It must be emphasized that the link between media prevalence of prejudice-denoting words and public opinion about prejudice severity defies a simplistic overarching interpretation. For instance, despite consistent increases in media discussion of sexual orientation prejudice over the last 20 years, public perceptions about the prevalence or severity of homophobia have decreased sharply over the last decade as documented here and previously elsewhere (Charlesworth & Banaji, 2019). This is the only observed prejudice type with a negative correlation between decreasing perceptions of prejudice severity in society and rising prevalence of words denoting this form of prejudice in written news media.

There are many plausible explanations for this contrarian trend. Over the last 2 decades, there have been dramatic changes for gay rights in America such as the Obama Administration repealing "Don't Ask, Don't Tell", the U.S. Attorney General dropping the Defense of Marriage Act in court, many states legislatures recognizing civil unions and the U.S. Supreme Court ruling legalizing gay marriage nationwide (Gallup, 2020; MAP, 2020; Savage et al., 2011; U.S. Department of Justice, 2011). In popular culture, LGBTQ characters have gone from underrepresented to statistically overrepresented in U.S. movies and television shows (Associated Press, 2020; Bahr, 2021; Dawson, 2020) and a growing number of Americans support gay marriage (Gallup, 2020) and are identifying as LGBTQ (Gallup, 2021). These rapid and dramatic changes for the LGBTQ community over the last decade may have led many Americans to perceive homophobia as less of a problem, despite increased media discussion on sexual orientation prejudice.

Alternatively, news media discourse on sexual orientation prejudice could have actually focused on *declines* in anti-LGBTQ sentiment in the United States or prejudice against LGBTQ citizens in *other* countries instead of the United States. Either possibility could potentially explain the decoupling of trends in levels of media coverage and public perceptions of prejudice severity.

Limitations of This Work

The variety of potential explanatory factors through which our results can be interpreted points to a fundamental limitation of this work, namely, frequency counts of prejudice-denoting terms in news media lack critical information about the context in which the terms are being used. Thus, our results cannot elucidate whether prejudice terms are being used in the context of describing increasing or decreasing prejudice in the United States or elsewhere. It could be the case that different news media outlets, perhaps according to political leanings, use prejudice signifying words in different contexts to imply increasing or decreasing prejudice. It is also conceivable that descriptions of changing prejudice severity in society are not uniform across prejudice types. That is, perhaps some prejudice types are described in the media in terms of increasing severity/prevalence in society while others are described in terms of decreasing severity/prevalence.

Another limitation of our analysis is the sparsity and heterogeneity of public opinion time series data regarding perceptions on severity of some types of prejudice. This motivated the usage of the *Dyad Ratios* algorithm to aggregate and interpolate massively incomplete surveys measuring the same latent construct on perceptions of prejudice severity. While the algorithm has been widely used in the literature (Stimson, 2018), it cannot provide high-resolution temporal dynamics for sparsely populated temporal ranges. Thus, our results based on public opinion surveys for those sparsely tracked public opinion perceptions should be interpreted with caution.

A further limitation of this work is that the time series survey data used is statistically underpowered (i.e., it is very short) due to the decaying availability of survey data for earlier years. The short nature of the time series analyzed creates substantial ambiguity about how to test and remove nonstationary prior to applying Granger causality. The potential confound of lesser news outlets article data availability for earlier years also creates uncertainty about how to model word prevalence dynamics and their relationships with other factors. That is, it is uncertain whether it is better to use a varying set of outlets over time in order to have a more representative sample of news media in recent years or whether it is better to sacrifice recent years representativeness to have a reduced but fixed set of news outlets with consistent data availability since the year 2000. In effect, different assumptions did occasionally result in different experimental effect sizes. We have chosen to thoroughly report experimental results under varying assumptions to provide a comprehensive overview of the occasional sensitivity of our experiments to different assumptions.

Our analysis of prejudice words usage by ideological leanings of news outlets has one important validity limitation due to the small sample size ($N = 7$) of the news outlets categorized as *centrist* by the AllSides 2019 media bias chart Version 1.1 (AllSides Media Bias Ratings, 2019). Furthermore, several of the outlets listed as centrists might not be representative of stereotypical news media organizations, as two of them are exclusively news agencies (*Associated Press* and *Reuters*) and have a significantly smaller ratio of editorial content relative to straightforward reporting. Another outlet classified as *centrist* by AllSides, Bloomberg, focuses primarily on economic and financial news.. Interestingly, despite the prevalence of prejudice-denoting words in centrist outlets being lower than in their more partisan counterparts, their rate of growth in the usage of prejudice-denoting terms is similar to all other outlets.

Conclusion

To our knowledge, this work is the first comprehensive scholarly attempt at describing the prevalence of prejudice-denoting words across a large and representative set of written articles from news media outlets. Our analyses reveal a sharp, substantial, and ubiquitous rise in the usage of words that denote prejudice against protected groups across a diverse set of news media outlets with many prejudice-denoting words in prominent outlets displaying increases in frequency of more than 500% within the 2010–2019 time frame. The trend precedes the emergence of Donald Trump in the political landscape for most of the terms analyzed but appears to accelerate after 2015.

We have also detected a strong correlation between rising prejudice words' prevalence in news media discourse and growing public opinion concern regarding the severity of prejudice in society. Granger-causality statistical tests offer limited evidence that the prevalence of prejudice-related words in media discourse might be predictive of shifts in U.S. public opinion regarding the perceived severity of some, but not all, types of prejudice. Despite these findings being highly resonant with antecedent research on news media shaping public opinion, our Granger-causality tests cannot conclusively established causality but mere predictive precedence in time. That is, we cannot rule out the effect of potential confounding factors influencing both media discourse and public opinion with different latency. Furthermore, the short nature of the time series analyzed makes statistical estimations underpowered.

To conclude, this work has documented a marked increase in the prevalence of prejudice-denoting words in news media discourse within the 2010–2019 time frame. The abrupt and dramatic changes in word frequencies suggest the existence of powerful underlying social dynamics at play. Whatever the ultimate cause for the rising prevalence of prejudice-signifying words in news media, it is noteworthy that such words are markedly increasing in prevalence, but also that such increases appear to occur alongside long-term *decreases* in overt expression of prejudice yet recent *increases* in the perceived prevalence of such prejudice among the general public. It is our hope that the detailed characterization of the phenomena presented here can pave the way for future studies looking in-depth at potential causal factors for the trends described herein as well as the impact of news media rhetoric on public consciousness and the social implications of growing perceptions of prejudice severity among the general population.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Supplemental Material

The supplemental material is available in the online version of the article.

References

al-Gharbi, M. (2019). Resistance as sacrifice: Toward an ascetic antiracism. *Sociological Forum*, *34*(S1), 1197–1216. https://doi.org/10.1111/socf.12544

al-Gharbi, M. (2020, September 8). *Cable news profits from its obsession with Trump. Viewers are the only victims*. Columbia Journalism Review. https://www.cjr.org/politics/cable-news-trump-obsession.php

AllSides Media Bias Ratings. (2019). AllSides. https://www.allsides.com/blog/updated-allsides-media-bias-chart-version-11

Associated Press. (2020, July 16). *Diversity of LGBTQ characters in film declines, study finds*. https://www.voanews.com/arts-culture/diversity-lgbtq-characters-film-declines-study-finds

Bahr, S. (2021, January 14). L.G.B.T.Q. TV representation falls for first time in 5 years, Glaad finds. *The New York Times*. https://www.nytimes.com/2021/01/14/arts/television/tv-lgbtq-glaad.html

Bonilla-Silva, E. (2017). *Racism without racists: Color-blind racism and the persistence of racial inequality in America* (5th ed.). Rowman & Littlefield Publishers.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186. https://doi.org/10.1126/science.aal4230

Callanan, V. J. (2012). Media consumption, perceptions of crime risk and fear of crime: Examining race/ethnic differences. *Sociological Perspectives*, *55*(1), 93–115. https://doi.org/10.1525/sop.2012.55.1.93

Campbell, B., & Manning, J. (2014). Microaggression and moral cultures. *Comparative Sociology*, *13*(6), 692–726. https://doi.org/10.1163/15691330-12341332

Charlesworth, T. E. S., & Banaji, M. R. (2019). Patterns of implicit and explicit attitudes: I. Long-term change and stability from 2007 to 2016. *Psychological Science*, *30*(2), 174–192. https://doi.org/10.1177/0956797618813087

Computer Graphics Lab at Stanford University. (n.d.). *Stanford cable TV news analyzer*. Stanford Cable TV News Analyzer. Retrieved September 2, 2020, from https://tvnews.stanford.edu/

Cooke, L. (2005). A visual convergence of print, television, and the Internet: Charting 40 years of design change in news presentation. *New Media & Society*, *7*(1), 22–46. https://doi.org/10.1177/1461444805049141

Crandall, C. S., Miller, J. M., & White, M. H. (2018). Changing norms following the 2016 U.S. presidential election: The trump effect on prejudice. *Social Psychological and Personality Science*, *9*(2), 186–192. https://doi.org/10.1177/1948550617750735

Dawson, L. (2020, December 16). Presence vs. representation: Report breaks down LGBTQ visibility on TV. *NBC News*. https://www.nbcnews.com/feature/nbc-out/presence-vs-representation-report-breaks-down-lgbtq-visibility-tv-n1251153

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(366), 427–431. https://doi.org/10.2307/2286348

Fassin, D., & Rechtman, R. (2009). *The empire of trauma: An inquiry into the condition of victimhood* (R. Gomme, Trans.; 1st ed.). Princeton University Press.

Gallup. (2020, June 1). *U.S. support for same-sex marriage matches record high. Gallup.Com*. https://news.gallup.com/poll/311672/support-sex-marriage-matches-record-high.aspx

Gallup. (2021, February 24). *LGBT identification rises to 5.6% in latest U.S. estimate*. Gallup.Com. https://news.gallup.com/poll/329708/lgbt-identification-rises-latest-estimate.aspx

Gao, G. (2015, June 29). Most Americans now say learning their child is gay wouldn't upset them. *Pew Research Center*. https://www.pewresearch.org/fact-tank/2015/06/29/most-americans-now-say-learning-their-child-is-gay-wouldnt-upset-them/

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*(3), 424–438. https://doi.org/10.2307/1912791

Haslam, N. (2016). Concept creep: Psychology's expanding concepts of harm and pathology. *Psychological Inquiry*, *27*(1), 1–17. https://doi.org/10.1080/1047840X.2016.1082418

Hong, J., Crichton, W., Zhang, H., Fu, D. Y., Ritchie, J., Barenholtz, J., Hannel, B., Yao, X., Murray, M., Moriba, G., Agrawala, M., & Fatahalian, K. (2020). Analyzing who and what appears in a decade of US cable TV news. *ArXiv:2008.06007 [Cs]*. http://arxiv.org/abs/2008.06007

Hopkins, D. J., & Washington, S. (2020). The rise of trump, the fall of prejudice? Tracking white Americans' racial attitudes via a panel survey, 2008–2018. *Public Opinion Quarterly*, *84*(1), 119–140. https://doi.org/10.1093/poq/nfaa004

Huang, J. S., & Heider, D. (2007). Media convergence: A case study of a cable news station. *International Journal on Media Management*, *9*(3), 105–115. https://doi.org/10.1080/14241270701521717

Johnson, T. (2011). *Muslims in the United States*. Council on Foreign Relations. https://www.cfr.org/backgrounder/muslims-united-states

Kawamoto, K. (2003). *Digital journalism: Emerging media and the changing horizons of journalism*. Rowman & Littlefield Publishers.

Klinenberg, E. (2005). Convergence: News production in a digital age. *The ANNALS of the American Academy of Political and Social Science*, *597*(1), 48–64. https://doi.org/10.1177/0002716204270346

Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review*, *84*(5), 905–949. https://doi.org/10.1177/0003122419877135

Krysan, M., & Moberg, S. (2016, August 25). *Trends in racial attitudes*. University of Illinois Institute of Government and Public Affairs. https://igpa.uillinois.edu/programs/racial-attitudes#section-5

Kwiatkowski, D., Phillips, P., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*(1–3), 159–178.

Leong, N. (2021). *Identity capitalists: The powerful insiders who exploit diversity to maintain inequality* (1st ed.). Stanford University Press.

Levari, D. E., Gilbert, D. T., Wilson, T. D., Sievers, B., Amodio, D. M., & Wheatley, T. (2018). Prevalence-induced concept change in human judgment. *Science*, *360*(6396), 1465–1467. https://doi.org/10.1126/science.aap8731

Lowry, D. T., Nio, T. C. J., & Leitner, D. W. (2003). Setting the public fear agenda: A longitudinal analysis of network TV crime reporting, public perceptions of crime, and FBI crime statistics. *Journal of Communication*, *53*(1), 61–73. https://doi.org/10.1111/j.1460-2466.2003.tb03005.x

Lukianoff, G., & Haidt, J. (2018). *The coddling of the American mind: How good intentions and bad ideas are setting up a generation for failure* (Illustrated ed.). Penguin Press.

Mallett, R. K., & Monteith, M. J. (2019). *Confronting prejudice and discrimination: The science of changing minds and behaviors*. Academic Press.

MAP. (2020). *Movement advancement project | 2020 Tally report*. https://www.lgbtmap.org/2020-tally-report

Marsden, P. V., Smith, T. W., & Hout, M. (2020). Tracking US social change over a half-century: The general social survey at fifty. *Annual Review of Sociology*, *46*(1), 109–134. https://doi.org/10.1146/annurev-soc-121919-054838

McCombs, M. (2005). A look at agenda-setting: Past, present and future. *Journalism Studies*, *6*(4), 543–557. https://doi.org/10.1080/14616700500250438

McCombs, M., & Valenzuela, S. (2021). *Setting the agenda: Mass media and public opinion* (3rd ed.). Polity.

Meagher, K. D., & Shu, X. (2019). Trends in U.S. gender attitudes, 1977 to 2018: Gender and educational disparities. *Socius*, *5*. https://doi.org/10.1177/2378023119851692

Mehmood, M. A., Shafiq, H. M., & Waheed, A. (2017). Understanding regional context of World Wide Web using common crawl corpus. In *2017 IEEE 13th Malaysia international conference on communications (MICC)* (pp. 164–169). https://doi.org/10.1109/MICC.2017.8311752

Moberg, S. P., Krysan, M., & Christianson, D. (2019). Racial attitudes in America. *Public Opinion Quarterly*, *83*(2), 450–471. https://doi.org/10.1093/poq/nfz014

Notess, G. R. (2002). The wayback machine: The web's archive. *Online*, *26*(2). https://www.elibrary.ru/item.asp?id=4370136

Pew Research Center. (2018, October 22). Democrats increasingly view racism and sexism as very big national problems; larger shares in both parties say drug addiction is a major problem. *Pew Research Center*. https://www.pewresearch.org/fact-tank/2018/10/22/more-in-u-s-see-drug-addiction-college-affordability-and-sexism-as-very-big-national-problems/ft_18-10-22_nationalproblems_democrats-views-racism-sexism/

Pew Research Center, 1615 L. St. (2017, August 29). More now see racism as major problem, especially democrats. *Pew Research Center*. https://www.pewresearch.org/fact-tank/2017/08/29/views-of-racism-as-a-major-problem-increase-sharply-especially-among-democrats/

Quinton, W. J. (2019). Unwelcome on campus? Predictors of prejudice against international students. *Journal of Diversity in Higher Education*, *12*(2), 156–169. https://doi.org/10.1037/dhe0000091

Racially profiling "jihadists" sounds like common sense: Here's why it doesn't work. (2016, September 21). *Musa Al-Gharbi*. https://musaalgharbi.com/2016/09/21/profiling-commonsense-badidea/

Rozado, D. (2019). Using word embeddings to analyze how universities conceptualize "diversity" in their online institutional presence. *Society*, *56*(3), 256–266. https://doi.org/10.1007/s12115-019-00362-9

Rozado, D. (2020a). Prejudice and victimization themes in *New York Times* discourse: A chronological analysis. *Academic Questions*, *33*(1), 89–100. https://doi.org/10.1007/s12129-019-09857-7

Rozado, D. (2020b). Wide range screening of algorithmic bias in word embedding models using large sentiment lexicons reveals underreported bias types. *PLoS One*, *15*(4), e0231189. https://doi.org/10.1371/journal.pone.0231189

Savage, D. (writer and activist). (2011). *2011: An extraordinary year for gay rights*. NPR.Org. https://www.npr.org/2011/12/31/144520538/2011-an-extraordinary-year-for-gay-rights

Singer, J. B. (2004). Strange bedfellows? The diffusion of convergence in four news organizations. *Journalism Studies*, *5*(1), 3–18. https://doi.org/10.1080/1461670032000174701

Singer, J. B. (2009). Convergence and divergence. *Journalism*, *10*(3), 375–377. https://doi.org/10.1177/1464884909102579

Smith, B. K., Stohl, M., & al-Gharbi, M. (2019). Discourses on countering violent extremism: The strategic interplay between fear and security after 9/11. *Critical Studies on Terrorism*, *12*(1), 151–168. https://doi.org/10.1080/17539153.2018.1494793

Stimson, J. A. (2018). The dyad ratios algorithm for estimating latent public opinion: Estimation, testing, and comparison to other approaches. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, *137–138*(1), 201–218. https://doi.org/10.1177/0759106318761614

Telhami, S. (2002, December 1). Arab and Muslim America: A snapshot. *Brookings*. https://www.brookings.edu/articles/arab-and-muslim-america-a-snapshot/

Tosi, J., & Warmke, B. (2016). Moral grandstanding. *Philosophy & Public Affairs*, *44*(3), 197–217. https://doi.org/10.1111/papa.12075

U.S. Census Bureau QuickFacts: District of Columbia. (2020a, April 1). https://www.census.gov/quickfacts/DC

U.S. Census Bureau QuickFacts: New York County (Manhattan Borough), New York. (2020b, April 1). https://www.census.gov/quickfacts/fact/table/newyorkcountymanhattanboroughnewyork/EDU635219

U.S. Department of Justice. (2011, February 23). *Statement of the attorney general on litigation involving the Defense of Marriage Act*. https://www.justice.gov/opa/pr/statement-attorney-general-litigation-involving-defense-marriage-act

Zogby, J. (2018). *Underreported, under threat: Hate crime in the United States and the targeting of Arab Americans*. Decode Hate—Arab American Institute. https://www.decodehate.com/reporthate

Author Biographies

David Rozado holds a Ph.D. in Computer Science from the Autonomous University of Madrid. He has been teaching at New Zealand's Otago Polytechnic since 2015. His research interests are computational content analysis and accessibility software for computer users with severe motor impairment.

Musa al-Gharbi is a Paul F. Lazarsfeld Fellow in Sociology at Columbia University. His research interests include sociology of knowledge, social psychlogy, cognitive sociology, and social epistemology. His book, *We Have Never Been Woke: Social Justice Discourse, Inequality and the Rise of a New Elite,* is forthcoming with Princeton University Press. Readers can learn more about this project and Musa's other research at musaalgharbi.com

Jamin Halberstadt received his B.A. in philosophy from Swarthmore College and his Ph.D. in social cognition at Indiana University, Bloomington. He is currently a full professor and Head of Department at the University of Otago's Department of Psychology. His eclectic research interests include emotion and decision making, social categorization and aesthetics, and religious cognition. He has authored and co-authored over 100 articles and book chapters and is currently a Senior Editor at Psychological Science, the field's premier outlet for empirical research.