Why Is All COVID-19 News Bad News?

Bruce Sacerdote, Ranjan Sehgal and Molly Cook*

Abstract

We analyze the tone of COVID-19 related English-language news articles written since January 1, 2020. Eighty seven percent of stories by U.S. major media outlets are negative in tone versus fifty percent for non-U.S. major sources and sixty four percent for scientific journals. The negativity of the U.S. major media is notable even in areas with positive developments including school re-openings and vaccine trials. Media negativity is unresponsive to changing trends in new COVID-19 cases or the political leanings of the audience. As evidenced by most viewed and most shared major media readers in the U.S. and U.K. strongly prefer negative stories about COVID-19, and negative stories in general. But the U.S. major media is more willing to satisfy this demand for negativity in both COVID and pre-COVID years. We suggest that this American exceptionalism stems from the lack of fair and balanced media laws and a lack of a large public option in the U.S. media. The causal impacts of this negative COVID coverage are less obvious; counties in the U.S. that rely more heavily on the major media are as likely to re-open schools as other similar counties.

* Sacerdote is at Dartmouth College, National Bureau of Economic Research, and Jameel Poverty Action Lab., Sehgal is at Dartmouth College and Cook is at Brown University. Corresponding author: bruce.sacerdote@dartmouth.edu, 6106 Rockefeller, Hanover NH 03755. We thank seminar participants at MIT for excellent suggestions which improved the paper. We thank Maxwell Grozovsky, Lauren Paer and Nashe Mutende for superb research assistance.

Introduction

On February 18th, the *Oxford Daily Mail* published a story that Professor Sarah Gilbert and her colleagues at Oxford's Jenner Institute were working on a vaccine for the novel coronavirus *and* that rapid vaccine development could be possible given the scientists' existing work and experience with a possible MERS vaccine. In contrast to *Oxford Mail*'s reporting, the U.S. major media outlets of Fox News, CNN, *The New York Times*, and *The Washington Post* did not begin coverage of Professor Gilbert's COVID-19 related work until late April.[1] The U.S. based stories emphasized caveats from health officials and experts downplaying the optimistic timeline and past success of the Oxford researchers. The earliest available (major outlet) U.S. story is from CNN on April 23rd and begins with a quote from England's Chief Medical Officer Chris Whitty saying that the probability of having a vaccine or treatment "anytime in the next calendar year" is "incredibly small."

There is a similar disconnect between positive research findings on school re-openings and U.S. major media reporting on the same topic; the reporting is overwhelmingly negative. Oster (2020) collects data on school reopenings and COVID-19 infections within schools and districts.[2] She finds that infection rates among students remain low (at 0.14 percent) and schools have not become the super-spreaders many feared.[3] Goldhaber et al (2020) and Harris et al (2020) both find no association between school reopenings and COVID cases when community levels of infections

---

[1] We base this statement on a LexisNexis search for the terms "Sarah Gilbert" or "Sarah Gilbert and vaccine" since January 1, 2020.

[2] https://statsiq.co1.qualtrics.com/public-dashboard/v0/dashboard/5f62eaee4451ae001535c839#/dashboard/5f62eaee4451ae001535c839?pageId=Page_1ac6a6bc-92b6-423e-9f7a-259a18648318.

[3] https://www.theatlantic.com/ideas/archive/2020/10/schools-arent-superspreaders/616669/.

per capita are below the 75th percentile.[4] Guthrie et al (2020) and Viner et al (2020) review the available evidence and reach similar conclusions. However, eighty six percent of school reopening articles from U.S. mainstream media are negative and only fifty four percent for the English-language major media in other countries.

The tone of media coverage impacts both human health and attitudes towards preventative measures including vaccination, mask wearing, and social distancing (Bursztyn et al 2020, Van Bavel and Baicker et al 2020, Simonov et al 2020, Kearney and Levine 2015, Ash et al 2020)[5]. The proportion of U.S. adults who exhibit depression symptoms has risen threefold since the start of the novel coronavirus pandemic (Etman et al 2020, Fetzer et al 2020). In discussing this increase in mental health problems, U.S. Centers for Disease Control and Prevention recommend against heavy consumption of news stories about the pandemic[6].

Our results suggest the CDC's warning is prescient. We categorize by topic over 9.4 million published news stories on COVID-19 since January 1, 2020. We then conduct several forms of textual analysis on roughly 20,000 COVID-19 news stories to examine levels of negativity by subtopic, source of the news, and time period. We have five major findings. First, COVID-19 stories published by media outlets in the U.S. top 15 (by readership/viewership) are 25 percentage points more likely to be negative in content than more general U.S. sources or major media outlets outside the U.S[7]. Second, the time pattern in observed negativity is at most weakly related to the

[4] Harris et al (2020) uses hospitalizations for COVID as the outcome measure.
[5] Bannerjee et al (2020) find that text messaging can significantly increase reporting of COVID symptoms and use of social distancing and other health promoting measures. Nyhan et al (2014) find that it's difficult to correct misperceptions around vaccine safety.
[6] https://www.cdc.gov/coronavirus/2019-ncov/daily-life-coping/managing-stress-anxiety.html

[7] This regression-based estimate controls flexibly for article length and week of publication. The unadjusted probability of an article being negative is 87 percent for US major media versus 50 percent for English-language non-US major media.

actual time trend in new weekly cases of COVID-19 in the U.S. Third, the most popular stories in *The New York Times*, CNN, and the BBC have high levels of negativity for all types of articles but particularly for COVID-19-related articles.[8] Fourth, negativity appears to be unrelated to the political leanings of the newspaper's or network's audience (Niven 2001). Finally, the strong negative correlation (across counties) between school reopenings and consumption of U.S. major media appears to be driven by selection rather than causality.

Overall, we find that relative to other media sources, the most influential U.S. news sources are outliers in terms of the negative tone of their coronavirus stories and their choices of stories covered.[9] We are unable to explain these patterns using differential political views of their audiences or time patterns in infection rates. This is analogous to Niven (2001) which finds a strong negative bias in the U.S. media when covering unemployment and limited evidence of partisanship. U.S. major outlets do demonstrate an above- average interest in promoting prosocial behavior like mask wearing and social distancing. Consistent with the existing literature (Gentzkow and Shapiro 2010 and Gentzkow, Glaeser and Goldin 2006), our results suggest that U.S. major outlets publish unusually negative COVID-19 stories in response to reader demand and interest. The U.S. versus non-U.S. difference in negativity among major media outlets may stem from the lack of a major publicly owned player in the U.S. media or the absence of fair and balanced reporting regulations.

Data Description

[8] This is consistent with the findings of Gentzkow and Shapiro (2010) who find that media respond strongly to consumer preferences. Eshbaugh-Soha (2010) finds that negativity media coverage of the President responds to local support for the President.

[9] Puglisi and Snyder (2016) have labelled the choice of what to cover as "agenda setting" and this effect is distinct from tone. The "filtering" of news modeled by Gentzkow, Shapiro and Stone (2015) could appear in our analysis as either a difference in tone (by filtering what facts are included) or as agenda setting by filtering which stories to cover.

We obtain counts of COVID-19 articles and separately the text of COVID-19 articles using the LexisNexis database. We use all English news sources and a date range of January 1, 2020 to December 31, 2020. We divide our universe of sources into the top (most widely read or watched) sources and all other sources. We further stratify by U.S. versus non-U.S. sources. The top non-TV sources for the U.S. that are also included in LexisNexis are *Newsweek*, the *New York Post*, *Los Angeles Times*, *USA Today*, *Politico*, *The Hill*, and the *New York Times*. For the top television sources we include both written articles and television transcripts from ABC, CBS, CNN, Fox News, MSNBC and NBC. We include the text of articles discussing COVID-19 vaccines from five widely read scientific and medical journals namely *Science*, *JAMA*, *The New England Journal of Medicine, The Lancet*, and *Nature*. We gather the *New York Times* most popular articles from their website from September 4-November 1 2020. This is a sample of 416 articles. We rely on the NYT most read articles. We supplement this with a separate sample of the most shared articles from CNN and the BBC.

We analyze the text of 43,000 articles that fall within three subtopics regarding COVID-19: vaccines, increases and decreases in case counts, and reopenings (of businesses, schools, parks, restaurants, government facilities, etc). We limit ourselves to roughly 43,000 articles given the legal requirement to "manually" download the articles from LexisNexis 100 articles at a time[10]. We classify all articles using two different but related methods. First, we measure the fraction of words that are negative according to established dictionaries of negative words. See Liu 2012, Tetlock 2007, Loughran and McDonald 2011 for canonical examples of this approach.[11] The

[10] LexisNexis does not permit automated downloading of the text of stories. We manually downloaded articles in batches of 100 articles.
[11] Riffe Lacy Fico and Watson (2019) is an in depth presentation of these methods. Grimmer and Stewart (2013) review the value of text analysis for summarizing political documents and transcripts.

results reported here use the Hu-Liu (2004) dictionary of positive and negative words.[12] We compute the fraction of total words that are negative according to the dictionary and, for ease of interpretation, standardize this variable to be mean 0 variance 1.

Second, we create a predicted probability that an article is negative in tone. We identify characteristics of negative and positive media reports in a set of 200 articles classified as strongly positive or negative by human readers. We use the most common phrases (typically 2-5 words in length) appearing in the training articles combined with machine leaning techniques to find the phrases that best predict whether the human reader will classify an article as strongly negative. We implement a Naïve Bayes classification scheme (Zhang 2004 , Pazzani 1996, Antweiler and Frank 2004)[13]. Naïve Bayes assumes that each phrase in the article contributes independently to the probability that the article is negative and maximizes the number of correct predictions given the phrases.

We use the resulting model to predict whether each of the articles in our sample are negative. For example, the inclusion of the phrases "clinical trial" and "Jenner Institute" are strong predictors of an article being positive while "death toll" and "stay at home" are strong predictors of a negative article.

Appendix Table 1 reports summary statistics at the article level for our main sample of articles that cover COVID during January 1, 2020-December 31st 2020. On average the articles have 1512 words. This count and our subsequent statistics are measured after we apply a truncation procedure

[12] We have conducted the same analysis using the Harvard General Inquirer dictionary of positive and negative words and obtain qualitatively similar results. http://www.wjh.harvard.edu/~inquirer/

[13] To extract phrases and implement the Naïve Bayes classification scheme we use WordStat software created by Provalis research.

to limit the text to be within 10 lines of the words “COVID” or “coronavirus”. We applied this truncation to deal with very long television transcripts that switched to non-COVID topics in the middle of the transcript. However, results are quite similar with or without truncation.

The share of negative words (using the Hui-Liu dictionary) is 4 percent. As mentioned above we standardize this variable to aid in interpreting the coefficients. By construction, our articles are divided roughly equally between articles on increases/decreases in cases, reopenings, and vaccines. The division among US major media, US General media, International Major Media, and International General media is also roughly equal.[14]

Results

Figure 1 plots the time trend in media negativity for major media outlets in the U.S. (green line) and outside the U.S. (blue line) using the scale on the left. The most striking fact is that 87 percent of the U.S. stories are classified as negative whereas 51 percent of the non-U.S. stories are classified as negative. Figure 1 uses our estimated probability that an article is negative. Similar results obtain using the Hu-Liu dictionary and the fraction of words in the article that are negative.

The red line plots the weekly average of daily new cases of COVID-19 in the U.S. using the scale on the right. The x-axis is the week of the year within 2020. New cases per day rise sharply from March through mid-April. Cases decline until about June 15th, then rise rapidly until late July, when cases begin to decline again. Average media negativity over time is not correlated with new case counts, as our regression results confirm.[15]

[14] We don’t have exactly 25% of articles in each major category because our initially drawn sample included many articles that were repeats which we then eliminated to arrive at this final sample.

[15] Not reported here in the interest of space.

In Table 1 we regress our estimated probability that an article is negative on indicator variables for whether the source is from the U.S. major media, U.S. general sources, or international general sources. The omitted category is other non-U.S. major media sources. In the regressions we control flexibly for the length of the article and the week the article was published. We run a linear probability model, though results from probit and logit models are similar to those reported here. The non-U.S. major media sources have a baseline rate of negativity of 51 percent and the coefficients are relative to this baseline. In column (1) we show that relative to the omitted category, articles in the U.S. major media are 25 percentage points more likely to be negative. In contrast, U.S. general and non-U.S. general sources have about the same level of negativity as non-U.S. major media.

In column (2) we switch the dependent variable to the share of negative words in the article. We standardize the outcome to be mean 0 standard deviation 1.[16] The U.S. major media publish stories that are .22 standard deviations more negative relative to non-U.S. major media. U.S. general media are significantly less negative than all other categories. In columns (3)-(5) we examine media negativity by subtopic within COVID-19. Relative to both types of international media, U.S. major media are particularly negative in their vaccine articles. Vaccine stories in the U.S. major media are 38 percentage points more likely to be negative relative to stories in the non-U.S. general media.

In Figure 2 we present the mean share of negative words (standardized) by source and topic (COVID-19 versus not). Starting with the bars at the bottom of the chart, we see that in a sample of non-COVID-19 stories (pre-January 2020), the U.S. major media are noticeable more negative

[16] We standardize within our broad sample that includes a pre-COVID sample of articles from 2019.

than the rest of the sample at .34 standard deviations. In covering COVID-19 (the second bar from the bottom), U.S. major media negativity is .48 standard deviation above the average while the non-U.S. major media are .22 standard deviation below average. Notably, scientific media articles on COVID-19 vaccines are .60 standard deviations below average in negativity. In contrast, the *New York Times'* most popular articles are .83 standard deviations above the sample mean in negativity for non-COVID-19 stories and 1.66 standard deviations above the mean when covering COVID-19 topics. Readers of the U.S. major media (as represented by the *New York Times*) are attracted to negative stories in general and negative stories about COVID-19 in particular.

In Appendix Figures 1 and 2 we show the share of words that are negative *within* vaccine articles and within school reopening articles. We standardize across the entire sample (all topics) and hence are comparing the negativity in the vaccine articles to the overall sample mean. For vaccine articles, all media categories are meaningfully below the overall sample mean for negativity, except for the U.S. major media which produces articles on COVID-19 vaccines that are .37 standard deviations higher on negativity. In results not reported, we find that the gap in vaccine article negativity between U.S. major media and all other sources remained even after vaccines are approved for use starting in November 2020.

For school re-opening articles, the U.S. major media is .30 standard deviations more negative than the overall sample mean. All other media categories are less negative than the sample mean. The U.S. general media produces school reopening articles that are .23 standard deviations less negative.

A natural question is whether media negativity varies greatly by the specific news source and whether that variation is related to the political beliefs of the readership. Our results are perhaps surprising. COVID-19 stories from all the major U.S. outlets have high levels of negativity and the variation that does exist is not correlated with readers' political leanings. See Figure 3. We plot the share of negative words (standardized) by U.S. media source versus the probability that conservative-leaning people say that this is a "trusted media source." The latter comes from a 2019 Pew survey of 12,000 people about their consumption of election news[17].

The estimated probability that a COVID-19 article is negative varies from 60 percent to 100 percent among major U.S. outlets. These probabilities do not align with the likelihood that conservative consumers of news trust the source. COVID-19 stories from Fox News are more negative than those from CNN. We obtain similar results using the share of negative words in the article.

Demand for negative news appears to be strong in other countries too. We used the WebHoseIO service to extract a sample of all English language news articles with more than 5000 Facebook shares during 2019 and 2020. The heavily shared CNN, Yahoo!, MSN and BBC articles are all very negative in tone with an average of 5 percent of the words being negative in the Hu-Liu dictionary for all these sources. For the three U.S. sources we are able to create a time series of the negativity of most shared articles. These articles are just as negative in 2019 (pre-COVID) as in 2020.

We now take a broader look at which COVID-19 topics the media choose to emphasize. Appendix Table 2 provides an overview of the number of COVID-19-related articles during the first half our sample period (January-July 2020) and counts of articles by topic, where one article can cover multiple topics. Overall, we found 2.6 million articles from U.S.-based sources and 6.4 million from non-U.S. sources. The rows represent different search terms we included while the columns

---

[17] https://www.pewresearch.org/fact-tank/2020/01/24/qa-how-pew-research-center-evaluated-americans-trust-in-30-news-sources/

represent four broad categories of sources, namely U.S. versus non-U.S. interacted with major media outlet versus general media. We are most interested in the relative coverage of different topics. For example, among the U.S. major media, 15,000 stories mention increases in caseloads while only 2,500 mention decreases, or a 6 to 1 ratio. During the period when caseloads were falling nationally (April 24th to June 27th), this ratio remains a relatively high 5.3 to 1.

We also count mentions of COVID-19 vaccines and any names of the top ten institutions or companies working on a COVID-19 vaccine. The U.S. major outlets ran 1,371 such stories. During the same period they ran 8,756 stories involving Trump and mask wearing and 1,636 stories about Trump and hydroxychloroquine.

A natural question is whether the media is promoting prosocial behaviors (Simonov et al 2020 and Burstyn et al 2020). While we cannot answer whether the U.S. media are "doing enough" to promote helpful (transmission-reducing) behavior in absolute terms, we can compare how emphasis of the benefits of mask wearing or social distancing varies across media categories. Five percent of COVID-19 articles in U.S. major outlets mention the benefits of mask wearing compared to .6 percent for non-U.S. outlets and 2 percent for general U.S. sources. U.S. major media outlets are also much more likely to discuss the benefits of social distancing (4 percent of stories) than their non-U.S. counterparts (1 percent of stories). This suggests the U.S. media are outperforming the non-U.S. media in promoting prosocial behavior, though perhaps because such messages are more needed in the U.S.[18]

A deeper question is whether U.S. media negativity has causal impacts on American's COVID behavior such as vaccine hesitancy or reluctance to reopen schools. We address impacts on school

[18] See Della Vigna and La Ferrara (2015) for a summary which discusses more generally the impact of media consumption on human behavior.

reopening by examining the relationship between school reopening patterns and consumption of U.S. major media. We obtain from Burbio weekly county level data on the percent of public schools in the county that offering traditional in-person schooling, virtual schooling, or a hybrid model. Burbio collects these data at the county level by examining tens of thousands of public school and government websites each day.

We measure reliance on the U.S. major media by using Google trends data on searches for all the major TV news channels including CNN, Fox, MSNBC, ABC News, CBS News, and NBC News. We use Google trends scores over five years available at the designated market area (DMA) level.[19]

In Figure 4 we show the percent of schools that are in-person by week starting in the second week of August 2020. We stratify the data by the quartile of the county's reliance on the major news media.[20] Two facts are immediately apparent. First, counties that rely less on major news media are much more likely to have their schools open. However, all of the quartiles of news reliance exhibit the same parallel trends. This suggests that counties relying more heavily on the major U.S. news have level differences in school openings but their changes in openings are no better or worse than other counties.

We also test these points formally in Appendix Table 4. In the county-week panel data, the trend coefficient on week*national news google score is -.0064. A one standard deviation move in the google trend score is .14. Hence 10 weeks of one standard deviation higher additional exposure

---

[19] Google trends data correlate strongly with newspaper circulation data and TV ratings data. For example we found that the correlation between Google Trends search data on Wall Street Journal correlates .88 with circulation. We use the following crosswalk https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/IVXEHT to assign counties to each of the 210 DMAs. Our regression estimates discussed here cluster at the DMA level.

[20] For this graph we used county level data to create quartiles but we obtain a similar picture if we collapse the Burbio data to the DMA level and create DMA level quartiles.

to national news media lowers the fraction of schools that are open by 10*.14*(.006)=.009 which is a small impact. We can rule out negative impacts on school opening (from 10 weeks of a 1 std higher exposure to national news media) of greater than -.03. We also consider the county cross sectional differences in school reopening behavior. The bivariate regression (with no controls) shows that a one standard increase in national media exposure is associated with a 14 percent decrease in the percent of schools that are open. However, adding simple demographic controls for the county and Democratic vote share in 2016 reduces this effect by almost two thirds. We conclude that there is little evidence that the negativity of the national news media causes a reduction in school reopenings.

Discussion and Conclusion

Overall, we find that COVID-19 stories from U.S. major media outlets are much more negative than similar stories from other U.S. outlets and from non-U.S. sources. The negativity does not respond to changes in new cases. Potentially positive developments such as vaccine stories receive less attention from U.S. outlets than do negative stories about Trump and hydroxychloroquine. Overall, we are unable to explain the variation in negativity by appealing to political affiliation or case count changes, but we do find that U.S. readers demand negative stories (as evidenced by article popularity).

An obvious question is, why are the U.S. major media so much more negative than international media and other outlets? We show that demand for negative stories (as proxied by Most Read and Most Facebook shared stories) is quite strong in the U.S. and the U.K. among readers of the New York Times, CNN, and BBC. Yet U.S. outlets are more likely to cater to the demand for negativity than are international outlets. We suggest three possible explanations which deserve further exploration. First, most of the non-U.S. markets in our sample include a dominant publicly owned news source. The U.K. has the BBC, while Canada has CBC and Australia has the ABC. Each of these news outlets is the number one news source in its respective country and may be following a different objective function than private news providers. This could potentially alter the behavior of all news providers.

Second, U.S. media markets are notably less concentrated that media markets in other OECD countries (Noam 2016). This higher level of competition may cause U.S. major media companies to use negativity as a tool to attract viewers.

Finally, the U.S. Federal Communication Commission eliminated its fairness doctrine regulation in 1987. This regulation required broadcasters to provide adequate coverage of public issues and to fairly represent opposing views. In contrast the U.K. and Canada still maintain such regulations. On the surface, the fairness doctrine would appear most relevant to partisan bias as opposed to negativity. It may be that profit maximizing U.S. news providers realized that they should provide not only partisan news to serve their consumers tastes but also negative news which is in high demand. We hope that our results spur additional investigation of U.S. media negativity and its causes and consequences.

References and Notes

Antweiler, Werner, and Murray Z. Frank. "Is all that talk just noise? The information content of internet stock message boards." *The Journal of finance* 59, no. 3 (2004): 1259-1294.

Ash, Elliott, Sergio Galletta, Dominik Hangartner, Yotam Margalit, and Matteo Pinna. "The Effect of Fox News on Health Behavior During COVID-19." *Available at SSRN 3636762* (2020).

Banerjee, Abhijit, Marcella Alsan, Emily Breza, Arun G. Chandrasekhar, Abhijit Chowdhury, Esther Duflo, Paul Goldsmith-Pinkham, and Benjamin A. Olken. *Messages on covid-19 prevention in india increased symptoms reporting and adherence to preventive behaviors among 25 million recipients with similar effects on non-recipient members of their communities*. No. w27496. National Bureau of Economic Research, 2020.

Bursztyn, Leonardo, Aakaash Rao, Christopher Roth, and David Yanagizawa-Drott. "Misinformation during a pandemic." *University of Chicago, Becker Friedman Institute for Economics Working Paper* 2020-44 (2020).

Chadha, Kalyani. "The Indian news media industry: structural trends and journalistic implications." *Global Media and Communication* 13.2 (2017): 139-156.

Djankov, Simeon, et al. "Who owns the media?." *The Journal of Law and Economics* 46.2 (2003): 341-382.

Eshbaugh-Soha, Matthew. "The tone of local presidential news coverage." *Political Communication* 27, no. 2 (2010): 121-140.

Ettman, Catherine K., Salma M. Abdalla, Gregory H. Cohen, Laura Sampson, Patrick M. Vivier, and Sandro Galea. "Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic." *JAMA network open* 3, no. 9 (2020): e2019686-e2019686.

Fetzer, Thiemo, Lukas Hensel, Johannes Hermle, and Christopher Roth. "Coronavirus perceptions and economic anxiety." *Review of Economics and Statistics* (2020): 1-36.

DellaVigna, Stefano, and Eliana La Ferrara. "Economic and social impacts of the media." In *Handbook of media economics*, vol. 1, pp. 723-768. North-Holland, 2015.

Gentzkow, Matthew, and Jesse M. Shapiro. "Competition and Truth in the Market for News." *Journal of Economic perspectives* 22.2 (2008): 133-154.

Gentzkow, Matthew, Edward L. Glaeser, and Claudia Goldin. "The rise of the fourth estate. How newspapers became informative and why it mattered." In *Corruption and reform: Lessons from America's economic history*, pp. 187-230. University of Chicago Press, 2006.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. "Text as data." *Journal of Economic Literature* 57, no. 3 (2019): 535-74.

Gentzkow, Matthew, and Jesse M. Shapiro. "What drives media slant? Evidence from US daily newspapers." *Econometrica* 78, no. 1 (2010): 35-71.

Gentzkow, Matthew, Jesse M. Shapiro, and Daniel F. Stone. "Media bias in the marketplace: Theory." In *Handbook of media economics*, vol. 1, pp. 623-645. North-Holland, 2015.

Groseclose, Tim, and Jeffrey Milyo. "A measure of media bias." *The Quarterly Journal of Economics* 120, no. 4 (2005): 1191-1237

Gurun, Umit G., and Alexander W. Butler. "Don't believe the hype: Local media slant, local advertising, and firm value." *The Journal of Finance* 67, no. 2 (2012): 561-598.

Guthrie, Brandon L., Jessie Seiler, Lorenzo Tolentino, Wenwen Jiang, Molly Fischer, Rodal Issema, Sherrilynne Fuller, Dylan Green, Diana M. Tordoff, Julianne Meisner, Ashley Tseng, Diana Louden, Jennifer M. Ross, Alison L. Drake "Summary of Evidence Related to Schools During the COVID-19 Pandemic" Report from COVID-19 Literature Report Team, Washington State Department of Public Health. October 2020.

Grimmer, Justin, and Brandon M. Stewart. "Text as data: The promise and pitfalls of automatic content analysis methods for political texts." *Political analysis* 21, no. 3 (2013): 267-297.

Kearney, Melissa S., and Phillip B. Levine. "Media influences on social outcomes: The impact of MTV's 16 and pregnant on teen childbearing." *American Economic Review* 105, no. 12 (2015): 3597-3632.

Kennedy, Patrick J., and Andrea Prat. "Where do people get their news?." *Economic Policy* 34.97 (2019): 5-47.

Larcinese, Valentino, Riccardo Puglisi, and James M. Snyder Jr. "Partisan bias in economic news: Evidence on the agenda-setting behavior of US newspapers." *Journal of public Economics* 95, no. 9-10 (2011): 1178-1189.

Liu, Bing. "Sentiment analysis and opinion mining." *Synthesis lectures on human language technologies* 5.1 (2012): 1-167.

Loughran, Tim, and Bill McDonald. "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks." *The Journal of Finance* 66, no. 1 (2011): 35-65.

Minqing Hu and Bing Liu. "Mining and Summarizing Customer Reviews." Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004), Aug 22-25, 2004, Seattle, Washington, USA

Martin, Gregory J., and Ali Yurukoglu. "Bias in cable news: Persuasion and polarization." *American Economic Review* 107, no. 9 (2017): 2565-99.

Niven, David. "Bias in the news: Partisanship and negativity in media coverage of presidents George Bush and Bill Clinton." *Harvard International Journal of Press/Politics* 6.3 (2001): 31-46.

Noam, Eli M., ed. *Who owns the world's media?: Media concentration and ownership around the world*. Oxford University Press, 2016.

Nyhan, Brendan, Jason Reifler, Sean Richey, and Gary L. Freed. "Effective messages in vaccine promotion: a randomized trial." *Pediatrics* 133, no. 4 (2014): e835-e842.

Oster, Emily, "Schools Aren't Super-Spreaders", *The Atlantic*, October 9, 2020.

Pazzani, M. J. 1996. Search for dependencies in Bayesian classifiers. In Fisher, D., and Lenz, H. J., eds., Learning from Data: Artificial Intelligence and Statistics V. Springer Verlag.

Puglisi, Riccardo. "Being the New York Times: the political behaviour of a newspaper." *The BE journal of economic analysis & policy* 11, no. 1 (2011).

Puglisi, Riccardo, and James M. Snyder Jr. "Empirical studies of media bias." In *Handbook of media economics*, vol. 1, pp. 647-667. North-Holland, 2015.

Riffe, Daniel, Stephen Lacy, Frederick Fico, and Brendan Watson. *Analyzing media messages: Using quantitative content analysis in research*. Routledge, 2019.

Simonov, Andrey, Szymon K. Sacher, Jean-Pierre H. Dubé, and Shirsho Biswas. *The persuasive effect of fox news: non-compliance with social distancing during the covid-19 pandemic*. No. w27237. National Bureau of Economic Research, 2020.

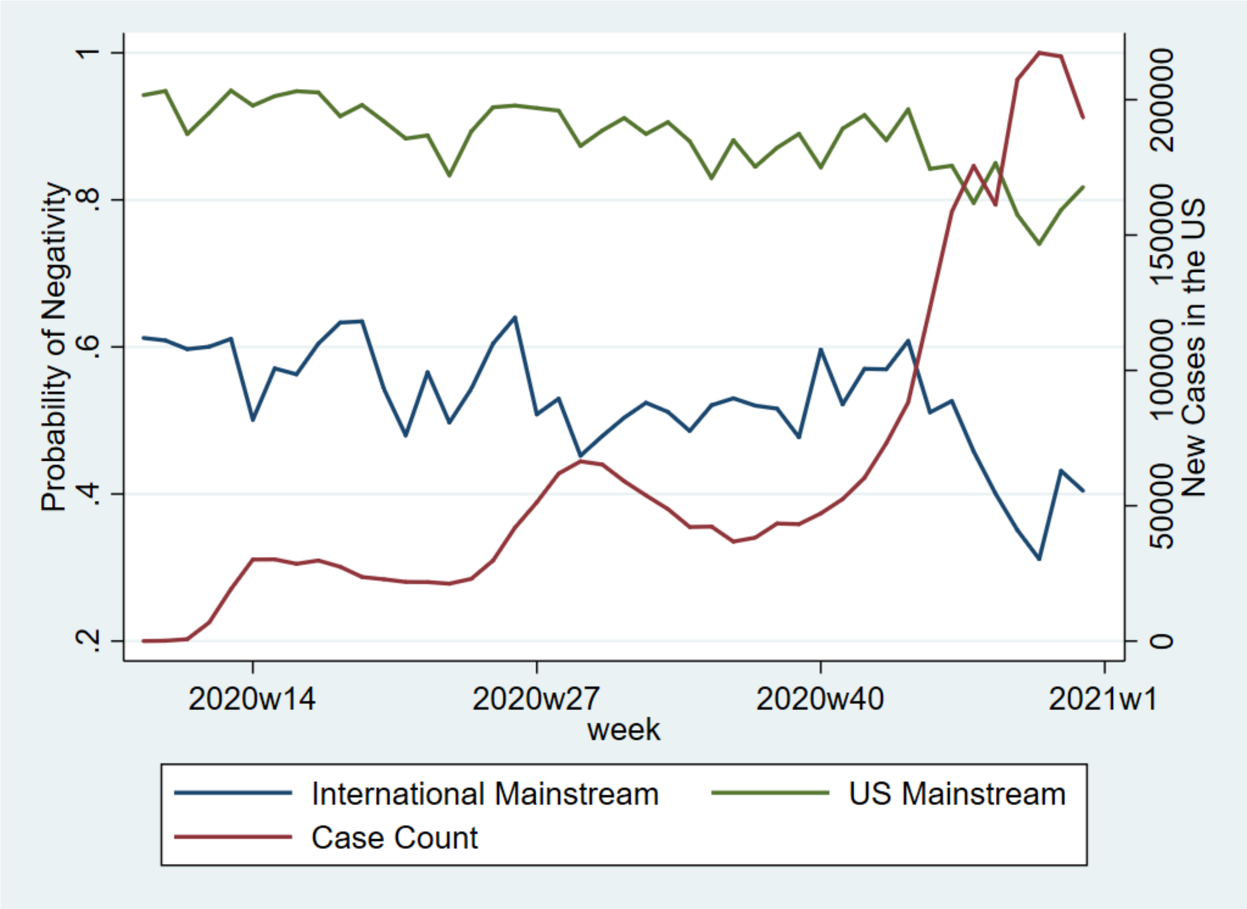
Tetlock, Paul C. "Giving content to investor sentiment: The role of media in the stock market." *The Journal of finance* 62, no. 3 (2007): 1139-1168.

Van Bavel, Jay J., Katherine Baicker, Paulo S. Boggio, Valerio Capraro, Aleksandra Cichocka, Mina Cikara, Molly J. Crockett et al. "Using social and behavioural science to support COVID-19 pandemic response." *Nature Human Behaviour* (2020): 1-12.

Viner, Russell M., et al. "School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review." *The Lancet Child & Adolescent Health* (2020).

Zhang H. (2004). The optimality of Naive Bayes. Proc. FLAIR.

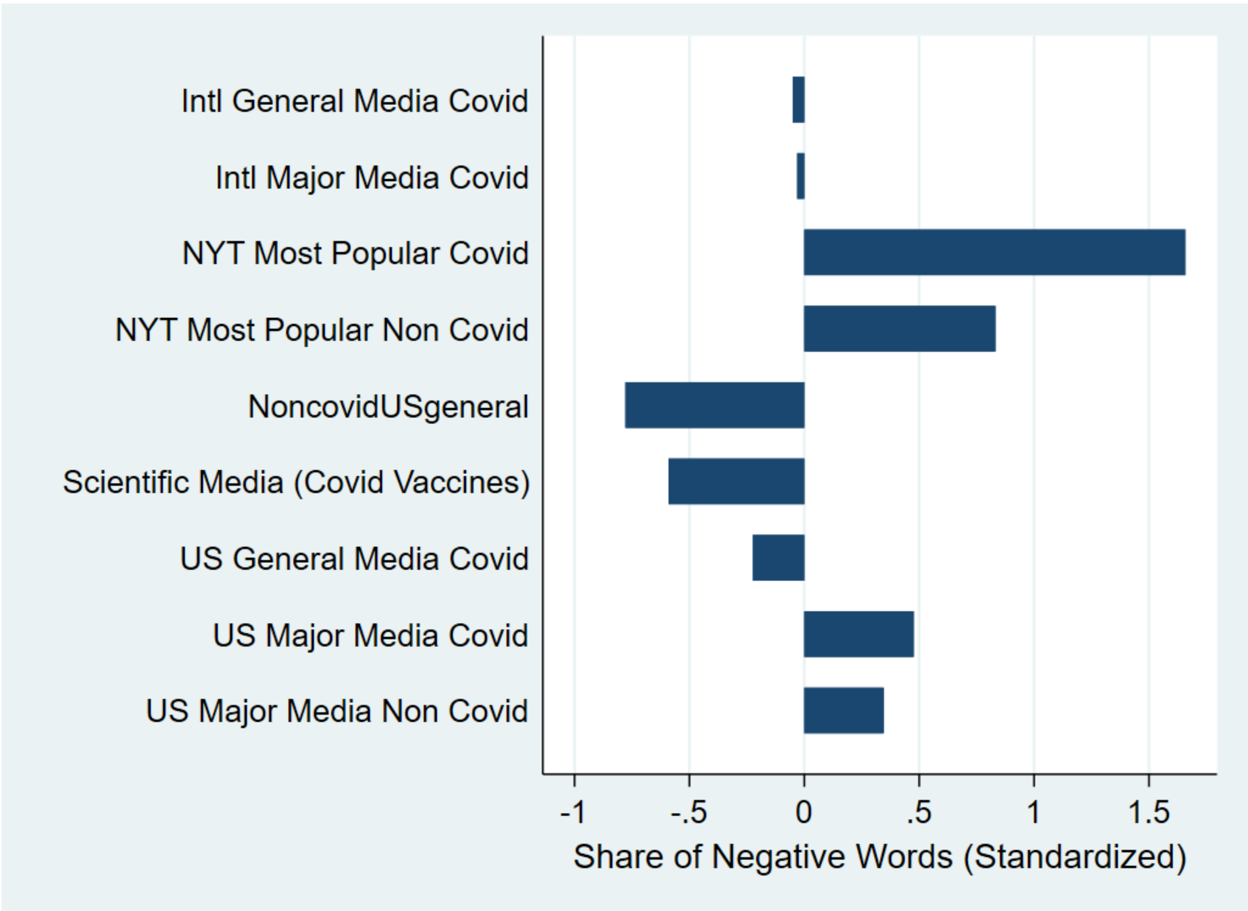
Figure 1: Media Negativity and New COVID-19 Cases Over Time



Notes: Negativity is estimated using supervised machine learning on article phrases coupled with a training data set. Articles are manually downloaded from LexisNexis for the period January 1st, 2020 to July 31st, 2020. The red line shows the weekly average of daily confirmed new COVID-19 cases and is accessed from the *New York Times* website.

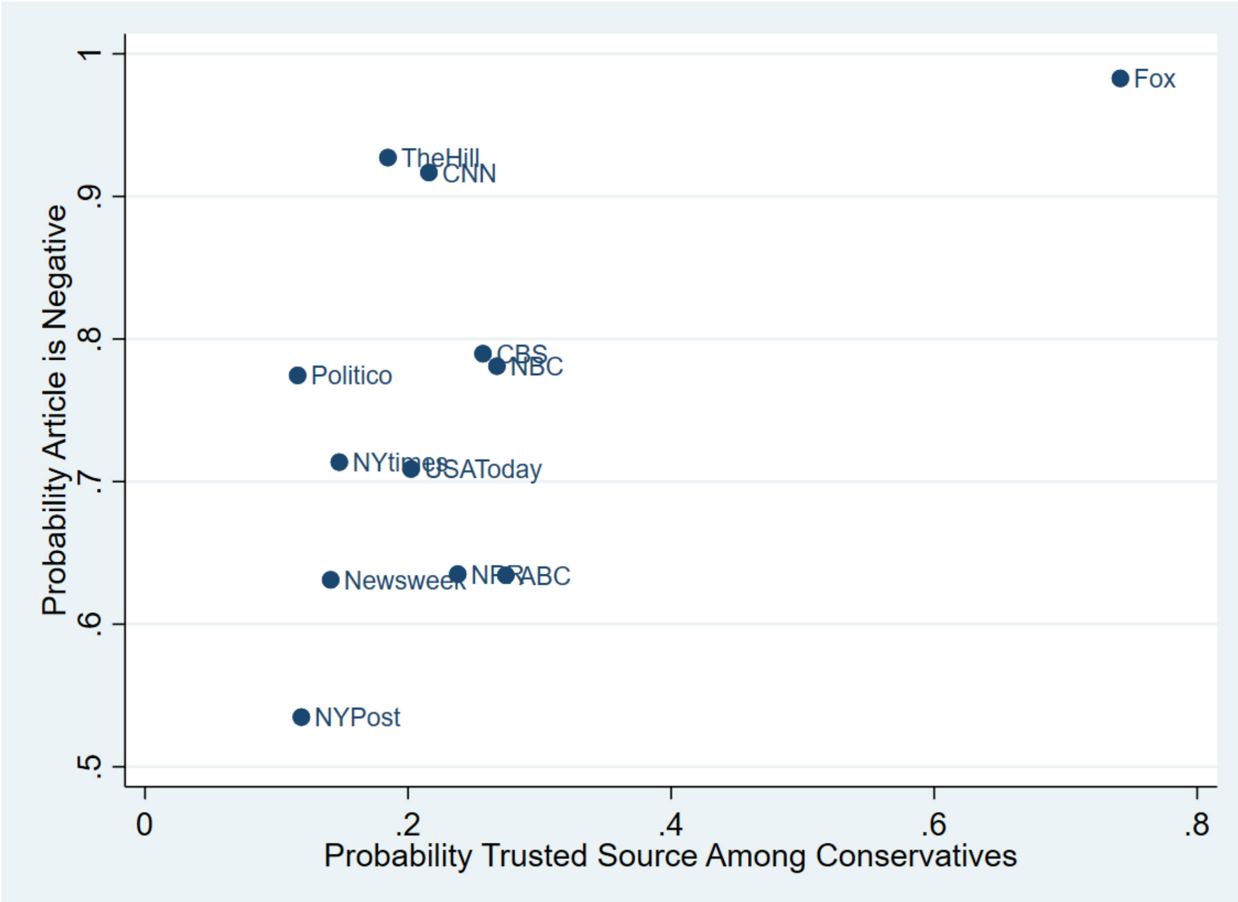
Figure 2:

Media Negativity by Source for COVID-19 and Non-COVID-19 Articles



Notes: Negativity is estimated as the fraction of negative words in the article and is standardized. Dark blue bars are for COVID related articles and light blue bars are for non-COVID related articles. The raw share of negative words is .043 with a standard deviation of .021. Negative words are defined by the Hu-Liu (1997) dictionary. Articles and transcripts are manually downloaded from LexisNexis for the period January 1st, 2020 to July 31st, 2020 and websites for *Science*, *JAMA*, *The New England Journal of Medicine*, *The Lancet*, and *Nature*. The *New York Times* website is used for the list and text of the most popular articles.

Figure 3: Media Negativity and Audience Political Leanings



Notes: Probability of an article being negative is estimated using Naïve Bayes and a training sample of articles hand classified by humans as negative or positive. Articles and transcripts are manually downloaded from LexisNexis for the period January 1st, 2020 to December 31st, 2020. "Trusted source" is measured in a 2019 Pew Survey of U.S. adults.

Figure 4: County Level School Reopenings By Quartile of National Media Consumption

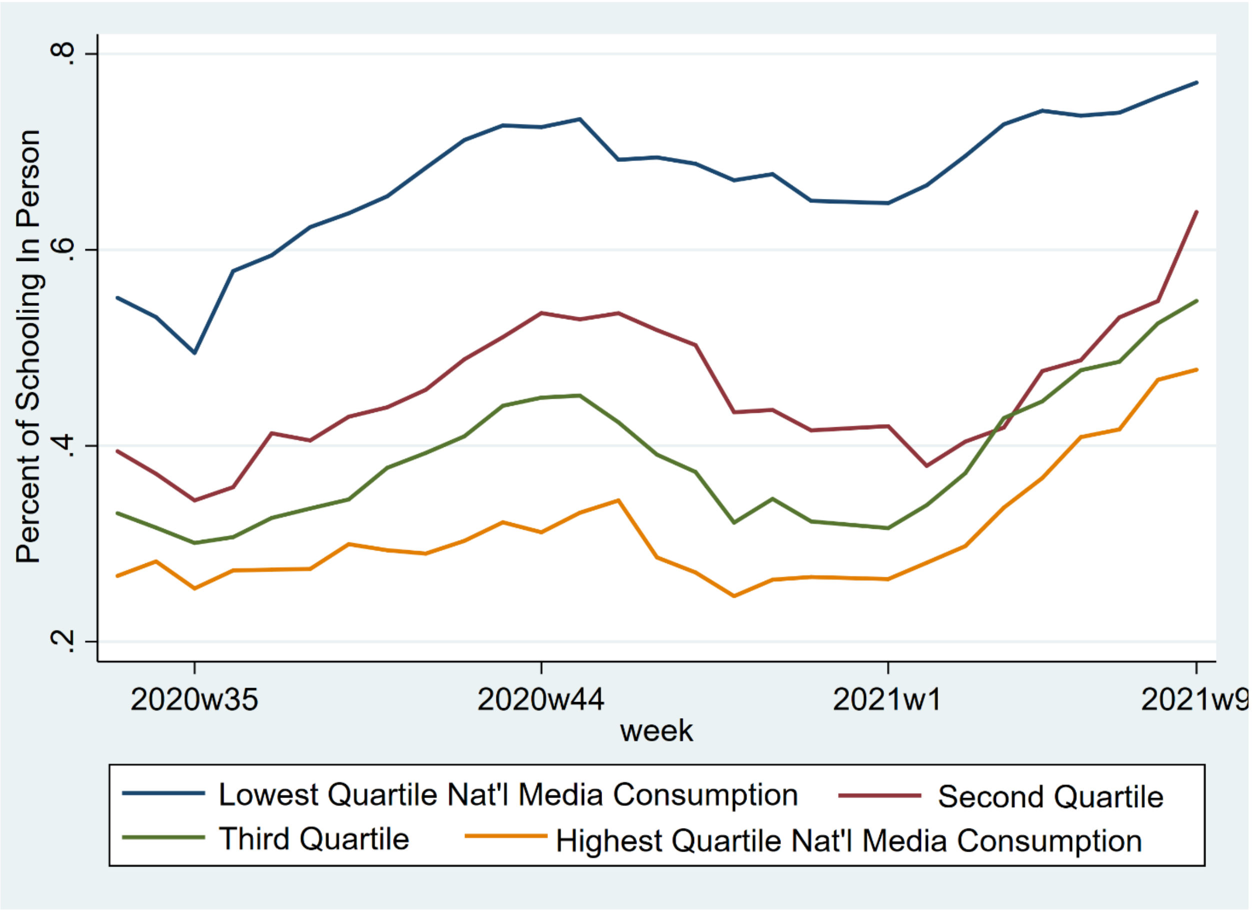


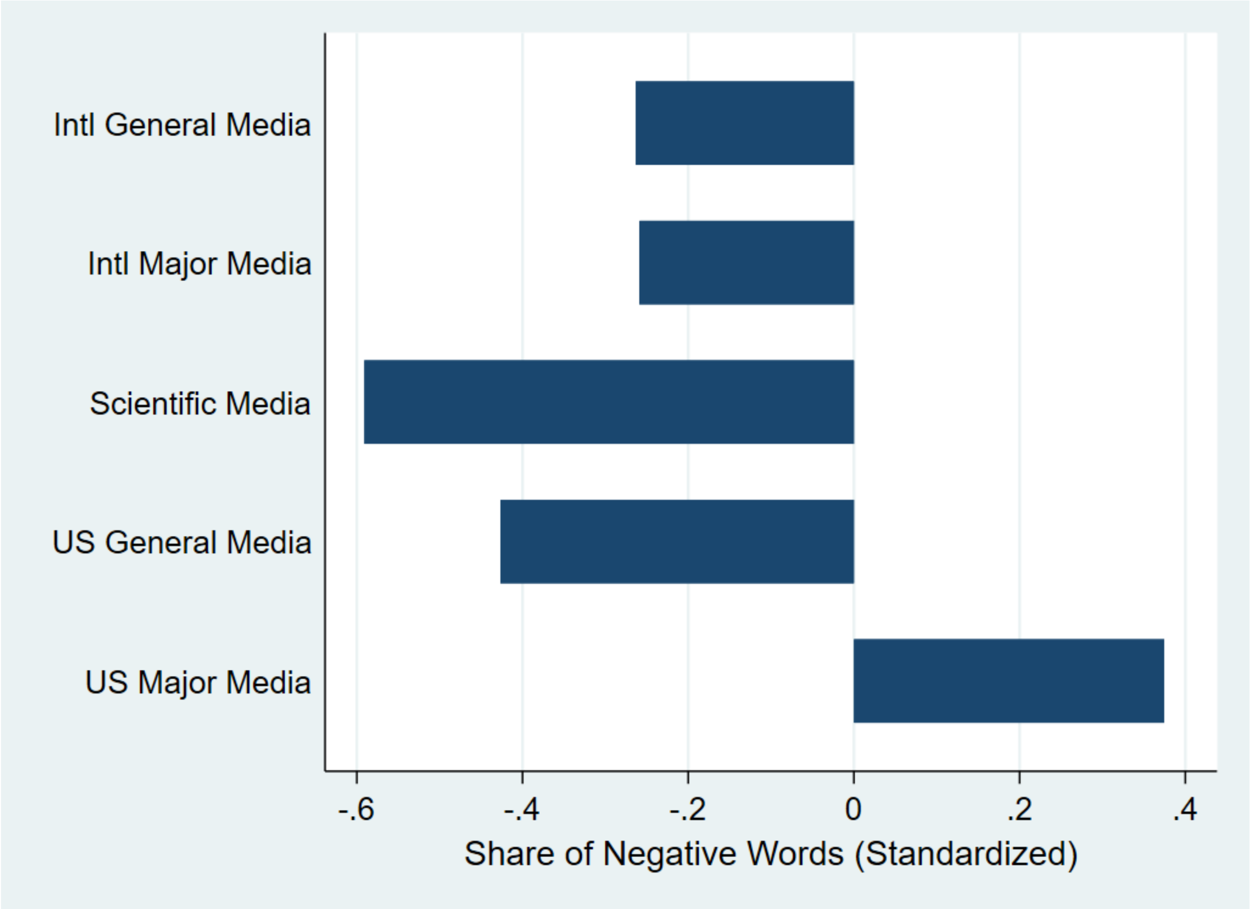
Table 1: Negativity by Media Category and Topic

	(1) Probability Article is Negative (All Articles)	(2) Share of Negative Words (Standardized)	(3) Prob Article is Negative (Vaccine Articles)	(4) Prob Article is Negative (Case Count Articles)	(5) Prob Article is Negative (Reopening Articles)
US Major Media	0.250***	0.218***	0.379***	0.195***	0.171***
	(0.00496)	(0.0125)	(0.00857)	(0.00628)	(0.00595)
US General Media	0.0241***	-0.315***	-0.00588	0.0118	0.104***
	(0.00564)	(0.0133)	(0.00882)	(0.00769)	(0.00637)
International General Media	-0.0222***	-0.0975***	-0.0157**	0.0319***	0.0215***
	(0.00455)	(0.0119)	(0.00659)	(0.00668)	(0.00551)
Mean for Omitted Category (International Major Media)	.506	-.024	.236	.645	.618
Observations	42,753	42,753	14,540	13,367	14,691
R-squared	0.364	0.232	0.503	0.402	0.393

Notes: Probability of the article being negative is estimated using supervised machine learning on article phrases coupled with a training data set. Share of negative words is estimated as the fraction of negative words in the article and is standardized. The raw share of negative words is .043 with a standard deviation of .021. Negative words are defined by the Hui-Lu (1997) dictionary. Articles are manually downloaded from LexisNexis for the period January 1st, 2020 to July 31st, 2020. All columns use OLS with robust standard errors. *** p<0.01, ** p<0.05, * p<0.1

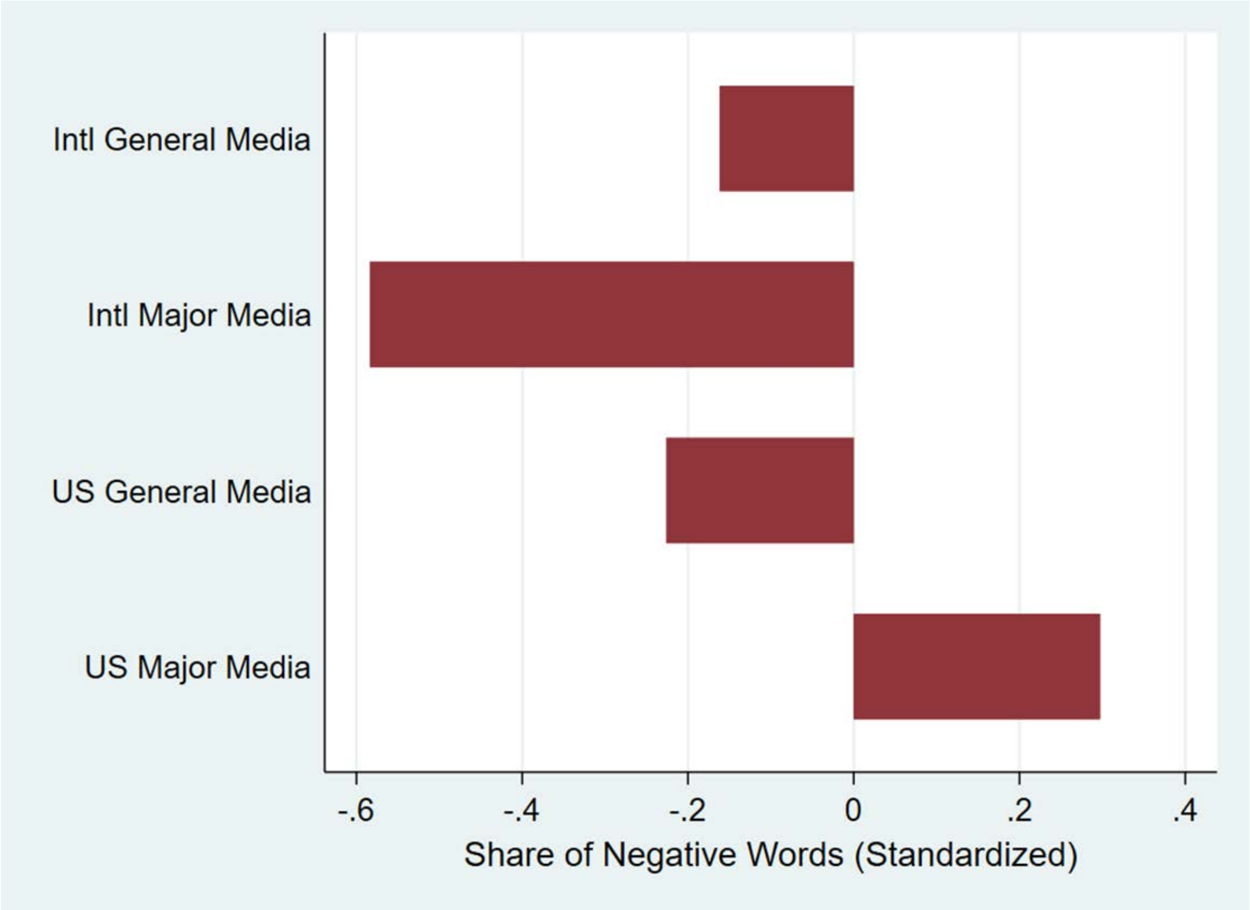
Appendix Figure 1:

Media Negativity by Source for COVID-19 Vaccine Articles



Appendix Figure 2:

Media Negativity by Source for School Reopening Articles



Appendix Table 1: Summary Statistics

	(1) N	(2) mean	(3) sd	(4) min	(5) max
Word Count of Article	42,757	1,512	2,102	7	57,166
Estimated P(Article is Negative)	42,757	0.646	0.387	0	1
Share of Words That Are Negative	42,757	0.0406	0.0192	0	0.164
Increase/Decrease in Cases Article	42,757	0.313	0.464	0	1
Reopenings Article	42,757	0.344	0.475	0	1
Vaccine Article	42,757	0.340	0.474	0	1
US Major Media	42,757	0.324	0.468	0	1
US General Media	42,757	0.165	0.372	0	1
International Major Media	42,757	0.296	0.456	0	1
International General Media	42,757	0.211	0.408	0	1
Fraction of Conservatives Who Trust This Source	16,088	0.243	0.126	0.116	0.742

Notes: We present summary statistics for our main variables. Each article is one observation. Probability of the article being negative is estimated using supervised machine learning on article phrases coupled with a training data set. Share of negative words is estimated as the fraction of negative words in the article and is standardized. The raw share of negative words is .043 with a standard deviation of .021. Negative words are defined by the Hui-Lu (1997) dictionary. Articles are manually downloaded from LexisNexis for the period January 1st, 2020 to July 31st, 2020.

Appendix Table 2: Total COVID-19-Related Media Articles by Topic: January 31st, 2020 to July 31st, 2020

Topic	U.S. Total	U.S. mainstream	U.S. non-mainstream	Non-U.S. Total	Non-U.S. mainstream	Non-U.S. non-mainstream
Coronavirus/COVID-19	2,594,510	90,600	2,503,910	6,823,410	453,900	6,369,510
Vaccines	33,980	2,375	31,605	69,600	3,257	66,343
Vaccines + Sarah Gilbert Etc.	28,740	1,371	27,369	54,860	2,299	52,561
Increases Whole Time Period	325,550	15,200	310,350	666,895	41,386	625,509
Decreases Whole Time Period	87,550	2,462	85,088	99,630	3,067	96,563
Increases 4/24-6/27 Period	103,700	3,581	100,119	314,548	16,660	297,888
Decreases 4/24-6/27 Period	33,000	676	32,324	53,850	1,297	52,553
Reopening	412,780	19,300	393,480	680,052	31,630	648,422
Masks	386,890	23,600	363,290	670,994	43,090	627,904
Masks and Trump	56,579	8,756	47,823	46,187	2,339	43,848
Benefits Masks	51,700	4,436	47,264	61,680	2,687	58,993
Social Distancing	378,940	19,600	359,340	811,503	55,610	755,893
Benefits Social Distancing	60,450	3,975	56,475	86,249	4,163	82,086
Hydroxychloroquine	21,440	2,273	19,167	33,005	2,746	30,259
Hydroxychloroquine and Trump	10,640	1,636	9,004	12,503	929	11,574

Notes: Article counts come from a LexisNexis for the period January 1st, 2020 to July 31st, 2020. The left most column indicates the search terms used (see methodology documents for exact searches). The article can be counted in multiple rows if the article contains both sets of terms.

Appendix Table 3: Negativity by Specific Media Source

	(1) Prob Article is Negative--US Sources	(2) Share of Negative Words (Standardized)-- US Sources
Fox	0.378***	0.611***
	(0.0138)	(0.0355)
MSNBC	0.301***	0.298**
	(0.0530)	(0.136)
ABC	0.324***	0.588***
	(0.0180)	(0.0465)
CBS	0.333***	0.574***
	(0.0227)	(0.0585)
CNN	0.377***	0.727***
	(0.00725)	(0.0187)
NBC	0.0684	0.288**
	(0.0483)	(0.124)
NPR	0.215***	0.386***
	(0.0110)	(0.0284)
LAtimes	0.385***	0.913***
	(0.0177)	(0.0455)
Newsweek	0.196***	0.130
	(0.0392)	(0.101)
Politico	0.308***	0.885***
	(0.0264)	(0.0680)
TheHill	0.549***	0.962***
	(0.143)	(0.368)
NYtimes	0.215***	0.970***
	(0.0114)	(0.0292)
NYPost	0.0901*	0.605***
	(0.0510)	(0.131)
USAToday	0.257***	0.728***
	(0.0254)	(0.0653)
Observations	13,888	13,888
R-squared	0.393	0.323

Omitted category consists of all U.S. sources not named above. Regressions are estimated using a linear probability model. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table 4 : Correlation Between School Reopenings and Media Negativity

	(1) Percent In Person Jan 2021	(2) Percent In Person Jan 2021 Full Controls	(3) Percent In Person Jan 2021	(4) Percent In Person Jan 2021 Full Controls	(5) Percent In Person (County-Week)	(6) Percent In Person (County-Week)
Nat'l News Google Trends Score	-1.126***	-0.397*				
	(0.241)	(0.230)				
CNN Google Trends Score			-1.061***	-0.501***		
			(0.150)	(0.186)		
Weekly Trend*National News Score						-0.00640
						(0.00659)
Weekly Trend*CNN Score					-0.00322	
					(0.00454)	
Vote Share Democratic 2016 Pres		-0.697***		-0.541***		
		(0.153)		(0.163)		
Ave Daily Covid Cases		0.00104		0.000983		
		(0.000869)		(0.000837)		
County Demographic Controls	No	Yes	No	Yes	No	No
County Fixed Effects	No	No	No	No	Yes	Yes
Week Fixed Effects	No	No	No	No	Yes	Yes
Constant	1.248***	3.828***	1.222***	2.722***	6.669	12.43
	(0.135)	(0.982)	(0.0872)	(0.837)	(8.544)	(12.18)
Observations	3,092	3,087	3,095	3,090	86,660	86,576
R-squared	0.071	0.290	0.135	0.299	0.761	0.761

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Appendix 1

Dataset Construction Details and Search Terms Used

Data Set Construction

Our dataset was assembled from Nexus Lexis articles. We utilized the following instructions:

1. Click on the link (links were derived from search terms at the bottom of this document)
(setting pages to display 50 at a time instead of 10)
2. Click the dropdown on the left that says location by publication
3. Click edit settings
4. In results display settings, switch it from 10 to 50.
5. Scroll to the bottom and hit save (you may have to do this every time for each link, not entirely sure how it “saves”
(downloading)
6. Before downloading, double check that you are sorting by relevance, and the slider is set to group duplicates
7. Click the little box beside the folder to select the whole page
8. Go to the next page and do the same
9. Click the download button which looks like it’s a downwards pointing arrow
10. In the dialog box, make sure the format is RTF and “save as individual files” these likely won’t be done already.
11. Download, and repeat until reaching 2500/link. In the final dataset this number may be less due to duplicates.

Lexus Nexus Article Search Process

vaccines	inc/dec	reopening
coronavirus or COVID-19 and ATLEAST5(vaccine)	coronavirus or COVID-19 and cases and increase or decrease	COVID-19 or coronavirus and reopening

American mainstream sources in our dataset consisted of:

US Mainstream Sources	International Mainstream Sources		
Fox	AFR	IndianExpress	Hindu
MSNBC	Analysis	MetroUK	Sun (England)
ABC	AsiaPacific	Newcastle	SunHerald
			Sydney Morning
CBS	AustralianFin	Northern Territory	Herald
CNN	BrisbaneTimes	SundayAge	Times of India
NBC	CTV	SundayHerald	TorontoStar
NPR	CanberraTimes	SydneyMorning	WestAZ
LAtimes	DailyMirror	Advertiser	WAToday
Newsweek	Geelong Advertiser	TheAge	Telegraph
Politico	HeraldSun	TheAustralian	Guardian
TheHill	HinduTimes	AustralianMag	
NYtimes	Hobart	Courier	
NYPost	IllawarraMercery	EveningStandard	
USAToday	IndiaToday	GlobeMail	