FAIR AND BALANCED?
QUANTIFYING MEDIA BIAS THROUGH CROWDSOURCED CONTENT ANALYSIS

CEREN BUDAK*
SHARAD GOEL
JUSTIN M. RAO

Abstract   It is widely thought that news organizations exhibit ideological bias, but rigorously quantifying such slant has proven methodologically challenging. Through a combination of machine-learning and crowdsourcing techniques, we investigate the selection and framing of political issues in fifteen major US news outlets. Starting with 803,146 news stories published over twelve months, we first used supervised learning algorithms to identify the 14 percent of articles pertaining to political events. We then recruited 749 online human judges to classify a random subset of 10,502 of these political articles according to topic and ideological position. Our analysis yields an ideological ordering of outlets consistent with prior work. However, news outlets are considerably more similar than generally believed. Specifically, with the exception of political scandals, major news organizations present topics in a largely nonpartisan manner, casting neither Democrats nor Republicans in a particularly favorable or unfavorable light. Moreover, again with the exception of political scandals, little evidence exists of systematic differences in story selection, with all major news outlets covering a wide variety of topics with frequency largely unrelated to the outlet's ideological position. Finally, news organizations express their ideological bias not by directly advocating for a preferred political party, but rather by disproportionately criticizing one side, a convention that further moderates overall differences.

CEREN BUDAK is an assistant professor in the School of Information at the University of Michigan, Ann Arbor, MI, USA. SHARAD GOEL is an assistant professor in the Management Science and Engineering Department at Stanford University, Palo Alto, CA, USA. JUSTIN M. RAO is a senior researcher at Microsoft Research, Redmond, WA, USA. The authors thank Seth Flaxman, Matthew Salganik, and Sid Suri for comments. *Address correspondence to Ceren Budak, University of Michigan, School of Information, 105 S. State St. Ann Arbor, MI 48109-1285, USA; e-mail: cbudak@umich.edu.

Introduction

To what extent are the US news media ideologically biased? Scholars and pundits have long worried that partisan media may distort one's political knowledge and in turn exacerbate polarization. Such bias is believed to operate via two mechanisms: selective coverage of issues, known as issue filtering (McCombs and Shaw 1972; Krosnick and Kinder 1990; Iyengar and Kinder 2010), and how issues are presented, known as issue framing (Gamson and Lasch 1981; Gamson and Modigliani 1989; Gamson 1992; Iyengar 1994; Nelson and Kinder 1996; Nelson, Clawson, and Oxley 1997; Nelson, Oxley, and Clawson 1997). Prior work has indeed shown that US news outlets differ ideologically (Patterson 1993; Sutter 2001) and can be reliably ordered on a liberal-to-conservative spectrum (Groseclose and Milyo 2005; Gentzkow and Shapiro 2010). There is, however, considerable disagreement over the magnitude of these differences (D'Alessio and Allen 2000), in large part due to the difficulty of quantitatively analyzing the hundreds of thousands of articles that major news outlets publish each year. In this paper, we tackle this challenge and measure issue filtering and framing at scale by applying a combination of machine-learning and crowdsourcing techniques.

Past empirical work on quantifying media bias can be broadly divided into two approaches: audience-based and content-based methods. Audience-based approaches are premised on the idea that consumers patronize the news outlet that is closest to their ideological ideal point (as in Mullainathan and Shleifer [2005]), implying that the political attitudes of an outlet's audience are indicative of the outlet's ideology. Though this approach has produced sensible ideological orderings of outlets (Gentzkow and Shapiro 2011; Zhou, Resnick, and Mei 2011; Bakshy, Messing, and Adamic 2015), it provides only relative, not absolute, measures of slant, since even small absolute differences between outlets could lead to substantial audience fragmentation along party lines.

Addressing this limitation, content-based methods, as the name implies, quantify media bias directly in terms of published content. For example, Gentzkow and Shapiro (2010) algorithmically parse congressional speeches to select phrases that are indicative of the speaker's political party (e.g., "death tax"), and then measure the frequency of these partisan phrases in a news outlet's corpus. Similarly, Groseclose and Milyo (2005) compare the number of times a news outlet cites various policy groups with the corresponding frequency among Congresspeople of known ideological leaning. Ho and Quinn (2008) use positions taken on Supreme Court cases in 1,500 editorials published by various news outlets to fit an ideal point model of outlet ideological position. Using automated keyword-based searches, Puglisi and Snyder (2011) find that an outlet's coverage of political scandals systematically varies with its endorsement of electoral candidates. Finally, Baum and Groeling (2008) investigate issue filtering by tracking the publication of stories from Reuters

and the Associated Press in various news outlets, where the topic and slant of the wire stories were manually annotated by forty undergraduate students.

Collectively, these content-based studies establish a quantitative difference between news outlets, but typically focus on a select subset of articles, which limits the scope of the findings. For example, highly partisan language from Congressional speeches appears in only a small minority of news stories, editorials on Supreme Court decisions are not necessarily representative of reporting generally, and political scandals are but one of many potential topics to cover. In response to these limitations, our approach synthesizes various elements of past content-based methods, combining statistical techniques with direct human judgments. This hybrid methodology allows us to directly and systematically investigate media bias at a scale and fidelity that were previously infeasible. As a result, we find that on both filtering and framing dimensions, US news outlets are substantially more similar—and less partisan—than generally believed.

Data and Methods

Our primary analysis is based on articles published in 2013 by the top thirteen US news outlets and two popular political blogs. This list includes outlets that are commonly believed to span the ideological spectrum, with the two blogs constituting the likely endpoints (Daily Kos on the left and Breitbart on the right), and national outlets such as *USA Today* and Yahoo News expected to occupy the center. See table 1 for a full list. To compile this set of articles, we first examined the complete web-browsing records for US-located users who installed the Bing Toolbar, an optional add-on application for the Internet Explorer web browser. For each of the fifteen news sites, we recorded all unique URLs that were visited by at least ten toolbar users, and we then crawled the news sites to obtain the full article title and text.[1] Finally, we estimated the popularity of an article by tallying the number of views by toolbar users. This process resulted in a corpus of 803,146 articles published on the fifteen news sites over the course of a year, with each article annotated with its relative popularity.

IDENTIFYING POLITICAL NEWS ARTICLES

With this corpus of 803,146 articles, our first step is to separate out politically relevant stories from those that ostensibly do not reflect ideological slant (e.g., articles on weather, sports, and celebrity gossip). To do so, we built two binary classifiers using large-scale logistic regression. The first classifier—which we

---

1. We estimate each article's publication date by the first time it was viewed by a user. To mitigate edge effects, we examined the set of articles viewed between December 15, 2012, and December 31, 2013, and limit to those first viewed in 2013.

Table 1. Average Number of Daily "News" and "Political News" Stories Identified in Our Sample for Each Outlet, with the Percent of News Stories That Are Political in Parentheses

Outlet	Average number of "news" stories per day	Average number of "political news" stories per day
BBC News	72.8	4.3 (6%)
Chicago Tribune	16.0	3.8 (24%)
CNN News	100.1	29.1 (29%)
Fox News	95.9	44.2 (46%)
Huffington Post	118.7	44.8 (38%)
Los Angeles Times	32.5	9.1 (28%)
NBC News	52.6	14.6 (28%)
New York Times	68.7	24.7 (36%)
Reuters	30.3	10.8 (36%)
Washington Post	65.9	37.9 (58%)
USA Today	33.7	11.8 (35%)
Wall Street Journal	11.7	4.6 (39%)
Yahoo News	173.0	53.9 (31%)
Breitbart News Network	15.1	11.2 (74%)
Daily Kos	14.0	9.8 (70%)

refer to as the news classifier —identifies "news" articles (i.e., articles that would typically appear in the front section of a traditional newspaper). The second classifier—the politics classifier —identifies political news from the subset of articles identified as news by the first classifier. This hierarchical approach shares similarities with active learning (Settles 2009), and is particularly useful when the target class (i.e., political news articles) comprises a small overall fraction of the articles. Given the scale of the classification tasks (described in detail below), we fit the logistic regression models with the stochastic gradient descent (SGD) algorithm (see, for example, Bottou [2010]) implemented in the open-source machine-learning package Vowpal Wabbit (Langford, Li, and Strehl 2007).[2]

To train the classifiers, we require both article features and labels. For features, we use a subset of the words in the article, as is common in the machine-learning literature. Given the standard inverted-pyramid model of journalism, we start by considering each article's title and first 100 words, which are strongly indicative of its topic. We then compute the 1,000 most frequently occurring words in these snippets of article text (across all articles in our sample), excluding stop words (e.g., "a," "the," and "of"). Finally, we

2. In the supplementary materials online, we compare this approach to the use of support vector machines (SVM) and find nearly identical performance.

represent each article as a 1,000-dimensional vector, where the *i*th component indicates the number of times the *i*th word in our list appears in the article's title and first 100 words.

The article labels for both the news and politics classifiers were collected through Amazon Mechanical Turk (http://mturk.amazon.com), a popular crowdsourcing platform. We required that workers reside in the United States, have good Mechanical Turk standing (i.e., have completed at least 1,000 tasks on the platform and have a 98 percent approval rate), and pass a test of political knowledge (described in the supplementary materials online). Although the answers to the test could be found using a web search, these types of screening mechanisms have nonetheless proven useful to ensure worker quality (Kittur, Chi, and Suh 2008).[3]

For the news-classification task, workers were presented with an article's title and first 100 words and asked to categorize it into one of the following nine topics, roughly corresponding to the sections of a newspaper: (1) world or national news; (2) finance/business; (3) science/technology/health; (4) entertainment/lifestyle; (5) sports; (6) travel; (7) auto; (8) incoherent text/foreign language; and (9) other. We then collapsed topics (2) though (9) into a single "non-news" category, producing a binary division of the articles into "news" and "non-news." For the training set, workers categorized 10,005 randomly selected articles stratified across the fifteen outlets (667 articles per outlet), with each article categorized by a single worker. Applying the trained news classifier to the full corpus of 803,146 articles, 340,191 (42 percent) were classified as news.

To evaluate the news classifier, we constructed a test set by first collecting labels for an additional random set of 1,005 articles (67 per outlet), where each article was now rated by two workers to ensure accurate ground-truth categories.[4] Of these 1,005 articles, 794 (79 percent) were identically labeled by the two workers. On this subset of articles, we find the classifier had 82 percent precision, 90 percent recall, and 87 percent overall accuracy. We also evaluated the classifier on the full set of 1,005 articles by randomly selecting one of the two labels as the "ground truth," again finding that it performed well, with 74 percent precision, 81 percent recall, and 79 percent overall accuracy.

Starting with the 340,191 articles classified as news, we next trained the politics classifier by again asking workers to label a random subset of 10,005 articles (667 per outlet), with each article classified by a single worker. In

---

3. These quality checks help address limitations of the Mechanical Turk labor force identified by related work, such as substandard performance by low-quality workers (Wais et al. 2010). We note that non-representativeness is also among the problems identified by past research (Berinsky, Huber, and Lenz 2012). However, this does not pose a problem for our particular study (details in the supplementary materials online).

4. For cost-effectiveness, only one label per article was collected for the training set, since the supervised learning techniques we used are robust to noise. However, to accurately evaluate the classifiers, it is important for the test set to be free of errors, and we thus collect two labels per article.

this case, we asked workers to "identify whether the following news article is about a US political issue," and we provided three options: (1) political; (2) not political; and (3) incoherent/corrupted text. We also provided a list of examples to help the workers in their decisions. On the set of 340,191 news articles, 114,814 (34 percent) were classified as political. Thus, 14 percent of the original set of 803,146 articles was identified as political news stories.

To evaluate performance of the politics classifier, 1,005 randomly selected news articles (67 per outlet) were classified as political or not by two workers, and of these, 777 (77 percent) had concordant labels. On this test set of 777 articles, the politics classifier had 91 percent precision, 81 percent recall, and 87 percent overall accuracy. As before, we further evaluated the classifier on the full set of 1,005 news articles by randomly selecting one of the two labels as the official ground truth; on this full set, the classifier had 84 percent precision, 69 percent recall, and 78 percent accuracy.

Overall, both the news and politics classifiers performed well, yielding results in line with recent work (Flaxman, Goel, and Rao 2015). Moreover, as shown in the supplementary materials online, we find almost identical results when we use support vector machines (SVMs) (Cortes and Vapnik 1995) to classify the articles instead of logistic regression. Finally, we note that there may be genuine disagreement about an article's classification. For example, some may consider a story about President Obama's vacation plans political news, while others may classify it as a travel or lifestyle piece. Thus, at least part of the differences we observe between the algorithmic and human labels can be attributed to the ambiguity inherent in the classification task. A similar pattern has been observed in related work on linguistic annotation (Plank, Hovy, and Søgaard 2014), where the authors show that disagreement between annotators reveals debatable linguistic cases and therefore should be embraced as opposed to eliminated.

Table 1 lists the average daily number of "news" and "political news" articles identified in our sample. Notably, there is substantial variation in the number of articles across outlets. In part, this variation is due to real differences in the number of published articles—Yahoo News and CNN, for example, do indeed publish more news stories than the niche blogs Daily Kos and the Breitbart News Network. Some of the variation, however, is due to the fact that we examine only articles that were visited by at least ten toolbar users. We thus obtain lower coverage for smaller outlets (e.g., the *Chicago Tribune*) and those with a paywall (e.g., the *Wall Street Journal*). As described below, we conduct our analysis on a popularity-weighted sample of articles, and since these popular articles are likely represented in our sample, we do not expect the differences in coverage to qualitatively affect our results.

IDENTIFYING ARTICLE TOPICS AND MEASURING IDEOLOGICAL SLANT

Having identified approximately 115,000 political news articles, we next seek to categorize the articles by topic (e.g., gay rights, healthcare, etc.), and to

quantify the political slant of the article. To do so, we again turn to human judges recruited via Mechanical Turk to analyze the articles. Even with crowd-sourcing, however, classifying over 100,000 articles is a daunting task. We thus limit ourselves to a readership-weighted sample of 10,502 political news articles. Specifically, for every day in 2013, we randomly selected two politi-cal articles, when available, from each of the 15 outlets we study, with sam-pling weights equal to the number of times the article was visited by our panel of toolbar users. We note that while we consider only a fraction of the political news articles in our corpus, our crowdsourcing approach allows us to analyze many more articles than would be feasible in a traditional laboratory setting.

To detect and control for possible preconceptions of an outlet's ideological slant, workers, upon first entering the experiment, were randomly assigned to either a blinded or unblinded condition. In the blinded condition, workers were presented with only the article's title and text, whereas in the unblinded condition, they were additionally shown the name of the outlet in which the article was published. Each article was then analyzed by two workers, one each from the sets of workers in the two conditions.

For each article, each worker completed the following three tasks. First, they provided primary and secondary article classifications from a list of fif-teen topics: (1) civil rights; (2) Democrat scandals; (3) drugs; (4) economy; (5) education; (6) elections; (7) environment; (8) gay rights; (9) gun-related crimes; (10) gun rights/regulation; (11) healthcare; (12) international news; (13) national security; (14) Republican scandals; and (15) other. We manually generated this list of topics with the aid of Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003), a popular technique for exploring large text corpuses.[5] Only 12 percent of all political articles received a primary clas-sification of "other," suggesting that our list of topics was sufficiently com-prehensive. Second, workers determined whether the article was descriptive news or opinion. Third, to measure ideological slant, workers were asked, "Is the article generally positive, neutral, or negative toward members of the Democratic Party?" and separately, "Is the article generally positive, neu-tral, or negative toward members of the Republican Party?" Choices for these last two questions were provided on a five-point scale: very positive, some-what positive, neutral, somewhat negative, and very negative. To mitigate question-ordering effects (Schuman and Presser 1996), workers were initially randomly assigned to being asked either the Democratic or Republican party

---

5. Though LDA was helpful for exploring the corpus and generating the list of topics, it did not produce article classifications that were sufficiently accurate for our purposes; we thus relied on human labels for the final article-level topic classification. In particular, manual inspection revealed that LDA resulted in poor classification for rare topics, such as gay rights. For example, various articles about California, including many on gun rights regulation, were incorrectly clas-sified as pertaining to gay rights, presumably due to the relatively large number of articles on gay rights that mentioned California.

question first; the question order remained the same for any subsequent articles the worker rated.

Finally, we assigned each article a partisanship score between –1 and 1, where a negative rating indicates that the article is net left-leaning and a positive rating indicates that it is net right-leaning. Specifically, for an article's depiction of the Democratic Party, the five-point scale from very positive to very negative is encoded as –1, –0.5, 0, 0.5, 1. Analogously, for an article's depiction of the Republican Party, the scale is encoded as 1, 0.5, 0, –.0.5, –1. The score for each article is defined as the average over these two ratings. Thus, an average score of –1, for example, indicates that the article is very positive toward Democrats and very negative toward Republicans. The result of this procedure is a large, representative sample of political news articles, with direct human judgments on partisanship and article topic.

Whereas past work has relied on undergraduate student judges to evaluate media bias (Baum and Groeling 2008; Ho and Quinn 2008), ours is the first to use crowdsourcing. This approach facilitates far greater scale and diversity of workers, but also raises concerns regarding data quality (Berinsky, Huber, and Lenz 2012). For instance, the small partisan differences we observe across outlets (discussed below) could simply reflect limited political awareness of workers. With these concerns in mind, we took several steps (described in more detail in the supplementary materials online), consistent with established best practices (Mason and Suri 2012), to ensure high quality ratings. First, we restricted participation to US-located workers with an exceptional track record on the crowdsourcing platform. Second, we required workers to pass a screening test. Third, in a preliminary analysis, multiple workers were assigned to the same article; we found that interrater reliability was on par with previous studies, even if we consider only those articles rated to have a political leaning (i.e., excluding "neutral" articles). Fourth, we limited the number of articles a single worker could rate to 100, ensuring a large pool of independent evaluations. Finally, as noted above, we presented only the name of the publication venue to a randomly selected subset of the workers, so as to check whether their perceptions of an outlet's ideological leaning affected their ratings. We found that article ratings were similar regardless of whether the outlet name was listed. Nevertheless, to be cautious, we limit our primary analysis to ratings generated by workers who did not see the outlet source.

We additionally conducted ex-post checks to validate the quality of the article slant ratings. The median subject spent over two minutes reading and rating each article, in line with expectations. The ratings were uncorrelated with a worker's stated political affiliation and only weakly related to a worker's intensity of news consumption. Interrater reliability—computed by comparing labels when the source was revealed versus blinded— is 81 percent, consistent with previous studies (Baum and Groeling 2008). We further note that this number should be considered a lower bound on agreement, since some differences could be due to the impact of revealing the source. Detailed statistics

on interrater reliability can be found in the supplementary materials online. The totality of evidence thus suggests that our workers produced high-quality article ratings. This finding is consistent with the growing literature demonstrating that crowd workers reliably replicate the behavior of undergraduate students across a wide variety of behavioral experiments (Paolacci, Chandler, and Ipeirotis 2010; Buhrmester, Kwang, and Gosling 2011; Berinsky, Huber, and Lenz 2012; Mason and Suri 2012; Goodman, Cryder, and Cheema 2013), and produce verifiably high-quality work in labeling tasks (Sorokin and Forsyth 2008; Callison-Burch 2009).

Results

OUTLET-LEVEL SLANT

We start by providing outlet-level estimates of ideological position. As described above, each article is first assigned a partisanship score between –1 and 1, with negative values indicating a net left-leaning article and positive values indicating a net right-leaning article. For each outlet, we then average the scores of the corresponding articles in our sample. Thus, since articles were randomly sampled in proportion to their popularity, an outlet's score is the average, popularity-weighted slant of articles in that outlet.

As can be seen in figure 1a, these ideological scores result in an ordering of outlets that is largely in line with past research.[6] For example, the Breitbart News Network is the most right-leaning outlet, and the Daily Kos is the most left-leaning outlet in our set. However, though the rank-ordering mirrors past work, the magnitude of the observed differences between the mainstream news outlets is remarkably small. For example, the *New York Times* and Fox News—which are the most ideologically distant mainstream outlets in our sample—have slant coefficients that differ by only 0.16 points (–0.05 versus 0.11). To put these numbers in perspective, we recall that the distance between each category in our five-point ideology scale (e.g., the distance between neutral and somewhat positive for Democrats) is 0.5. The two political blogs exhibit much larger differences (–0.24 versus 0.17), both from each other and from the mainstream media.

The outlet-level partisanship score is based on both descriptive news and opinion articles. Concerns over partisan media, however, largely stem from worries that descriptive news is ideologically biased, since readers do not necessarily interpret such coverage as representing only a single individual's perspective. To investigate this issue, we next examine outlet-level partisanship

---

6. One difference is that we identify the *Wall Street Journal* as a relatively conservative outlet—in accordance with convention wisdom—while past automated methods based on audience and co-citation measures have characterized it as left-leaning (Groseclose and Milyo 2005; Flaxman, Goel, and Rao 2015).
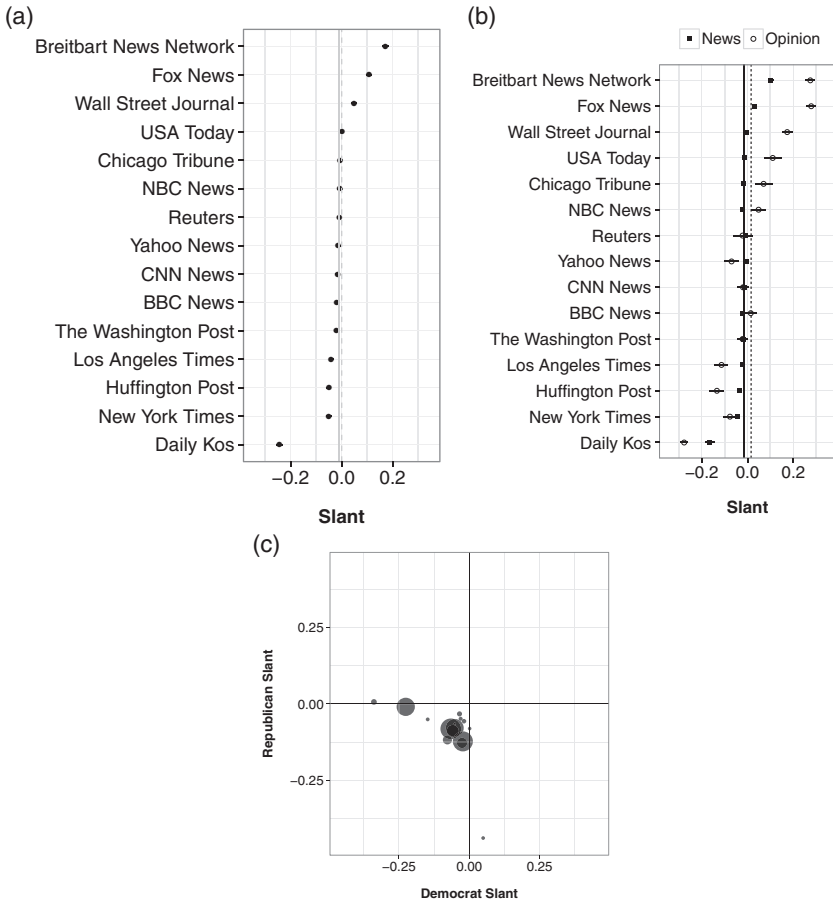
Figure 1a. Overall Outlet-Level Slant. Figure 1b. Outlet-Level Slant Estimated Separately for Opinion and Descriptive News. Figure 1c. Democrat and Republican Slant for Each News Outlet, Where Point Sizes Indicate the Relative Popularity of the Outlet.

scores separately for opinion pieces and descriptive reporting. As expected, figure 1b shows that partisanship is much more extreme for opinion than for descriptive news. For example, opinion stories on Fox News have a slant of 0.28, compared to 0.03 for descriptive news stories. Notably, even though descriptive news stories are largely neutral across the outlets, the differences still produce an ideological ranking of the outlets that is approximately the same as when we include opinion stories.[7] This finding indicates that ideological slant,

---

7. There are exceptions: for instance, the Huffington Post is more left-leaning on descriptive news than the *New York Times*.

while small in an absolute sense, is indeed present in descriptive reporting, and is directionally consistent with conventional wisdom.

Why is it that the partisan differences we find are so small? Figure 1c in part answers this question by splitting outlet-level slant into its two constituent pieces: Democratic and Republican slant. That is, we look at how, on average, outlets portray Democrats, and separately, how they portray Republicans. Strikingly, nearly all the outlets (with the exception of Daily Kos and Breitbart News Network) lie in the lower left quadrant, meaning that on average, they portray both Democrats and Republicans negatively. While one might have expected that net left-leaning or net right-leaning outlets would favorably portray one party while unfavorably characterizing the other, what we find is quite different. An outlet's net ideological leaning is identified by the extent of its criticism, rather than its support, of each party. In particular, net conservative outlets treat Republicans about the same way as centrist outlets, but are much more critical of Democrats. Analogously, net liberal outlets are more critical of Republicans but treat Democrats quite similarly compared to centrist outlets. This apparently widespread reporting practice of critical rather than supportive coverage in turn limits the ideological differences between outlets.[8]

Two additional contributing factors for the similar outlet-level slants that we observe are that (1) the vast majority of political articles in most outlets are neutral; and (2) among the partisan pieces, ideologically opposed articles are present in a single outlet. In figure 2, we plot the fraction of descriptive articles in each outlet that are net-left (score < 0) and net-right (score > 0), with the remaining fraction consisting of articles rated as neutral. We find that in most mainstream outlets, about one-third of descriptive news articles show a partisan leaning, and among these, slightly more than half are net-left, with the exact ratio varying with the outlet's overall ideology. For example, in the *New York Times*, about 20 percent of articles are net-left, 10 percent are net-right, and 70 percent are neutral. This similarity among outlets persists even when we restrict to more clearly partisan articles. In particular, about 1 percent of descriptive articles in most mainstream outlets have a slant score above 0.5 (indicating that the article is quite right-leaning), and about 1 percent have a score below –0.5 (indicating that the article is quite left-leaning). Among opinion articles, there are, as expected, typically far more partisan pieces; and among these partisan articles, the balance of net-left and net-right coverage generally reflects an outlet's overall ideology. For example, opinion pieces in Fox News are 63 percent net-right, 6 percent net-left, and 31 percent neutral.[9]

8. In the supplementary materials online, we look at treatment of Democrats and Republican by issue, and similarly find that on most issues, both parties are portrayed slightly negatively on average.

9. For more details about the fraction of articles that are highly partisan, please refer to the supplementary materials online.
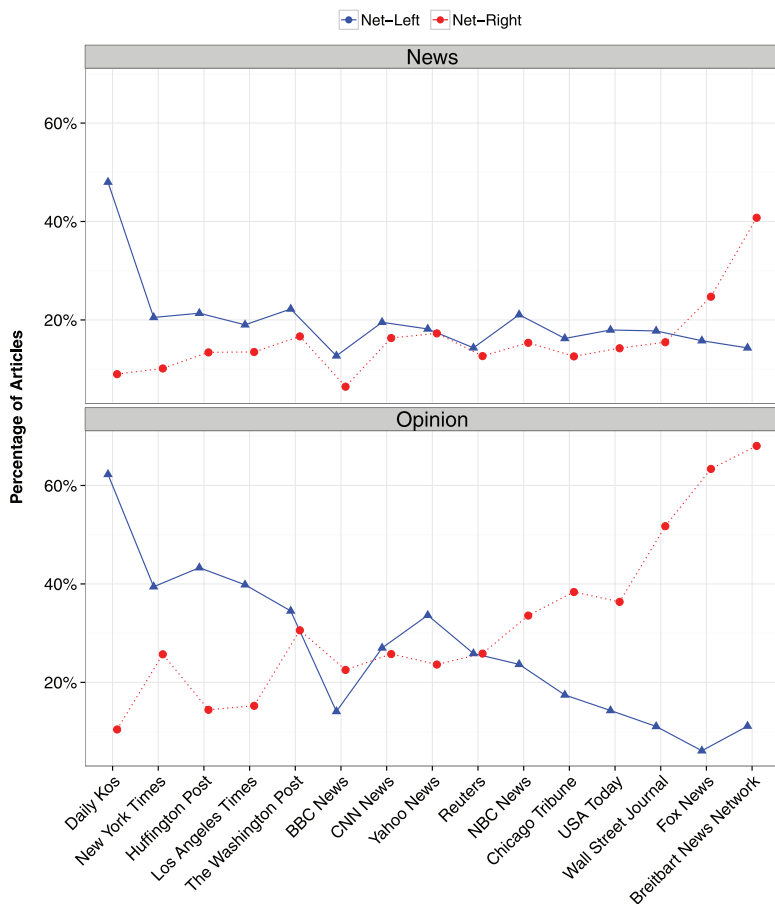
Figure 2. Article-Level News and Opinion Ideological Position, by Outlet.

Since the outlets in figure 2 are ordered left to right by ideological position, this relationship is revealed by the downward-sloping net-left line and upward-sloping net-right line.

ISSUE FRAMING

A key strength of our approach is that we not only can assess an outlet's overall slant, but we can also evaluate bias on an issue-by-issue basis. Figure 3 compares the ideological slant of the *New York Times* to Fox News for each of the fourteen topics we consider (excluding topic "other"). The issues are ordered top to bottom from largest to smallest differences in slant between the two outlets—thus, issues at the top of the list can be thought of as the most
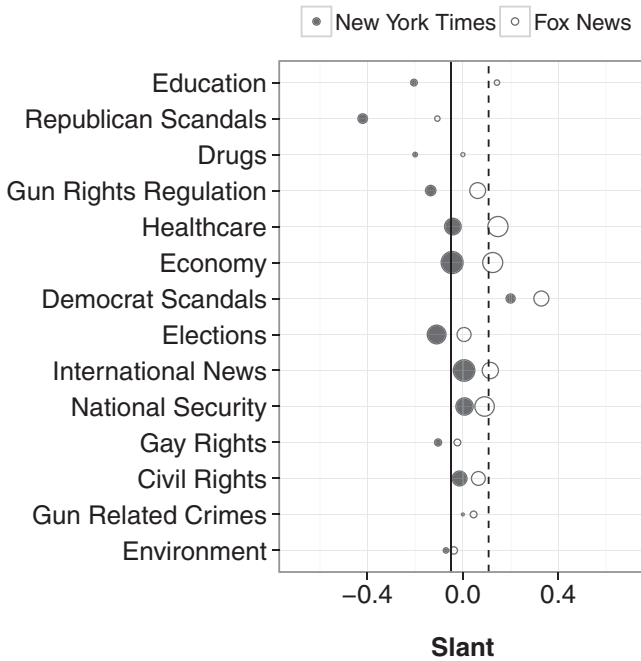
Figure 3. Comparison of Issue-Level Slant of the *New York Times* to Fox News. Point sizes indicate coverage intensity, and vertical lines give outlet averages.

polarizing. The point sizes reflect the coverage intensity in the corresponding outlet. The plot illustrates three high-level points. First, Fox News is consistently to the right of the *New York Times* on every issue we examined. Second, for many issues, the differences are remarkably small. For civil rights, for example, the net slant for the *New York Times* is –0.01, compared to 0.07 for Fox News. Finally, even for topics where there are relatively large differences between the two outlets, their slants are related. For example, in both outlets, Republican scandals have the most left-leaning coverage, and analogously, Democratic scandals have the most right-leaning coverage. This last observation further explains the relatively small overall differences between the outlets: many issues (e.g., scandals) are inherently left- or right-leaning, and thus mitigate the potential for bias; it would be difficult, for example, for the *New York Times* to write a net-left article about a scandal perpetrated by Democrats.

Figure 4 generalizes these findings to the fifteen outlets we study. Outlets are ordered on the *x*-axis from left to right based on overall outlet-level slant; each line corresponds to an issue, colored according to its mean slant across the outlets, and the *y*-axis indicates the average slant of articles on that topic in each outlet. As noted above, across outlets, Democrat and Republican scandals are among the few
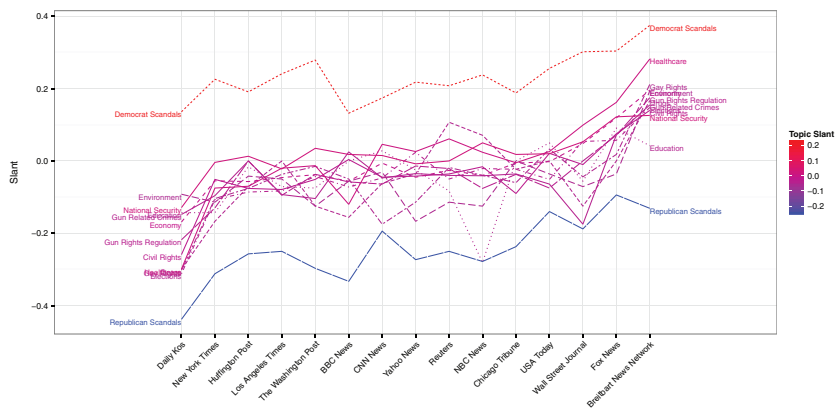
Figure 4. Issue-Specific Slant. Outlets are ordered left to right by their overall ideological position, and issues are colored blue to red according to their slant. The *y*-axis gives the average relative Republican slant for a particular domain on a specific issue.

issues exhibiting large partisan slant. Moreover, on all the issues, Fox News and the two political blogs—Daily Kos and the Breitbart News Network—and consistently more partisan than the other outlets. For the remaining issues and outlets, the ideological differences are quite small and do not appear to vary systematically.

ISSUE FILTERING

We next examine the extent to which news outlets selectively report on topics (i.e., issue filtering). Such potential issue filtering is consequential for at least two reasons. First, by selectively reporting on partisan topics (e.g., scandals), issue filtering can amplify an outlet's overall ideological slant. Second, even for issues that are reported in a largely nonpartisan manner, selective coverage may leave readers of different outlets with materially different exposure to political issues.

To gauge filtering effects, for each outlet we first estimate the proportion of articles that were categorized (by the human judges) under each topic.[10] Figure 5a compares the distribution of topics in Fox News to that in the *New York Times*, where points lying on the dashed diagonal line indicate equal coverage. Perhaps surprisingly, the plot shows that most topics receive similar coverage in the two outlets. Moreover, this similarity is not solely restricted to the popular topics—such as the economy, international news, and elections—but

---

10. We performed this analysis using both the primary and the secondary topics; as a robustness check, we used only the primary topic and did not find any qualitative differences. We used the labels gathered from both the blinded and unblinded groups, since seeing where the article was published should have little to no effect on the topic identified by participants. Moreover, we exclude articles labeled as "other."
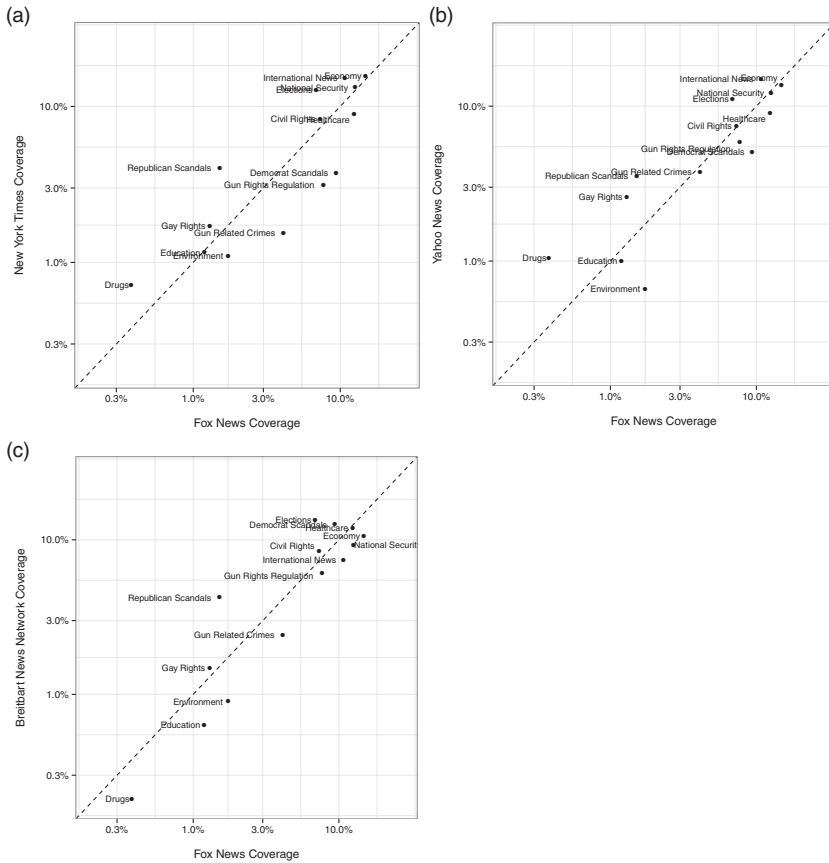
(a)

(b)

(c)

Figure 5a. Comparison of Fox News Topic Coverage with *New York Times*. Figure 5b. Comparison of Fox News Topic Coverage with Yahoo News. Figure 5c. Comparison of Fox News Topic Coverage with Breitbart News Network.

also carries over to more niche areas, including civil rights, gay rights, and education. Overall, the correlation in topic coverage between the *New York Times* and Fox News is 0.83. As another point of comparison, Figures 5b and 5c contrast Fox News to the centrist Yahoo News and to the right-wing blog Breitbart News Network. We again find, strikingly, that the distribution of topics is remarkably similar, despite their ideological differences. One exception is coverage of scandals. For example, Democrat scandals make up only 4 percent of political articles in the *New York Times*, while they account for almost 10 percent of those on Fox News. Similarly, Republican scandals make up 4 percent of all political articles in the *New York Times* and account for just 1.5 percent on Fox News.
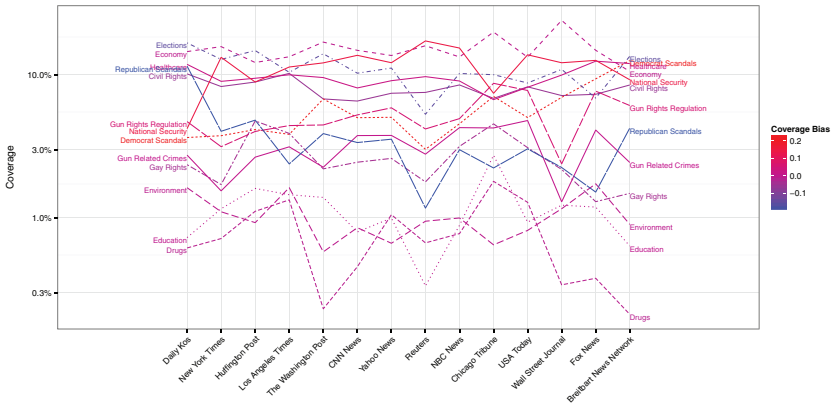
Figure 6. Issue Coverage by Outlet. Issue lines are colored blue to red to reflect the degree to which the issue-level average is relatively pro-Republican.

Figure 6 extends these results to the full set of outlets.[11] Outlets are ordered left to right based on their overall ideological slant. Each line corresponds to a particular topic, and is colored according to the average ideological slant of outlets that cover that topic: the more blue the line is, the more it is covered by left-leaning outlets, and the more red it is, the more it is covered by right-leaning outlets. Since the lines in figure 6 are largely flat across the outlets, there appears to be little systematic issue filtering in the US news media.

As a more quantitative test of this visually suggestive result, for each topic we separately fit the following regression model:

$$\log\left(\frac{C_i}{1-C_i}\right) = \beta_0 + \beta_1 I_i + \varepsilon_i, \tag{1}$$

where C_i is the coverage of the topic in outlet i (i.e., the percentage of articles in the outlet about the topic), and I_i is the outlet's overall ideological slant. The estimated coefficient β_1 thus captures how coverage relates to the position of the outlet on the political spectrum. We find that β_1 is small across the topics we study, and is in fact statistically significant for only two issues: Republican scandals ($p = 0.027$) and Democrat scandals ($p = 0.0016$). We note that previous work identified similar coverage differences for scandals (Baum and Groeling 2008; Gentzkow and Shapiro 2011); our results show that selective coverage of scandals, while consequential, is not representative of issue filtering more broadly.

---

11. For ease of viewing, we remove BBC News and the international news topic. That BBC News emphasizes international news is orthogonal to the question we address here, but its inclusion in the plot would have substantially changed the scale of the *y*-axis.

CONSUMPTION VS. PRODUCTION CHOICES

We have throughout considered a popularity-weighted sample of articles in order to study what a typical reader of each news outlet might consume. For example, our primary analysis aims to measure the differences—in terms of coverage and slant—between articles read by those who frequent the *New York Times* and those who frequent Fox News. These consumption choices are ostensibly driven by the availability and promotion of content by the news publishers: the headline story is likely to attract considerable attention, regardless of its topic or ideological slant. One might, however, reasonably seek to isolate the inherent production bias of an outlet, absent the consumption choices of its readers.

Given that online outlets frequently update which content to promote based on continuous popularity metrics, the distinction between production and consumption is not clear-cut. Nevertheless, here we report the results of two natural variations on our analysis. First, we estimate outlet-level slants for the full corpus of articles appearing in each outlet; second, we compute these metrics for articles appearing on an outlet's home page, the online equivalent of the traditional front page of a newspaper.[12] For each of the fifteen news sites, the three outlet-level estimates (popularity weighted, full corpus, and home-page articles) are nearly identical (figure 7). This pattern is indicative of the close relationship between production and consumption decisions, and illustrates the robustness of our results to the specifics of the measurement procedure. At the extremes of the ideological spectrum (e.g., Daily Kos and Breitbart News Network), popular articles do appear to be more partisan than those across an outlet's entire corpus or articles promoted on the home page, though the effect is small. In the supplementary materials online, we further investigate this relationship between article slant and popularity with article-level regression models, and confirm that there is a statistically significant but substantively small effect.

Discussions and Conclusion

In terms of both coverage and slant, we find that the major online news outlets—ranging from the *New York Times* on the left to Fox News on the right—have surprisingly similar, and largely neutral, descriptive reporting of US politics. This result stands in contrast to what one might reasonably conclude from past academic studies, and from the regular laments of popular commentators on the rise of partisan media. For example, by analyzing the think tanks that news outlets cite, Groseclose and Milyo (2005) conclude that there is a

---

12. Our analysis is specifically based on articles that garnered at least ten views by our sample of users. Due to data-retention policies, we could determine if an article appeared on an outlet's home page only for the last six months of our sample period, which accounts for the slight shuffling in the ordering of outlets. Details are provided in the supplementary materials online.
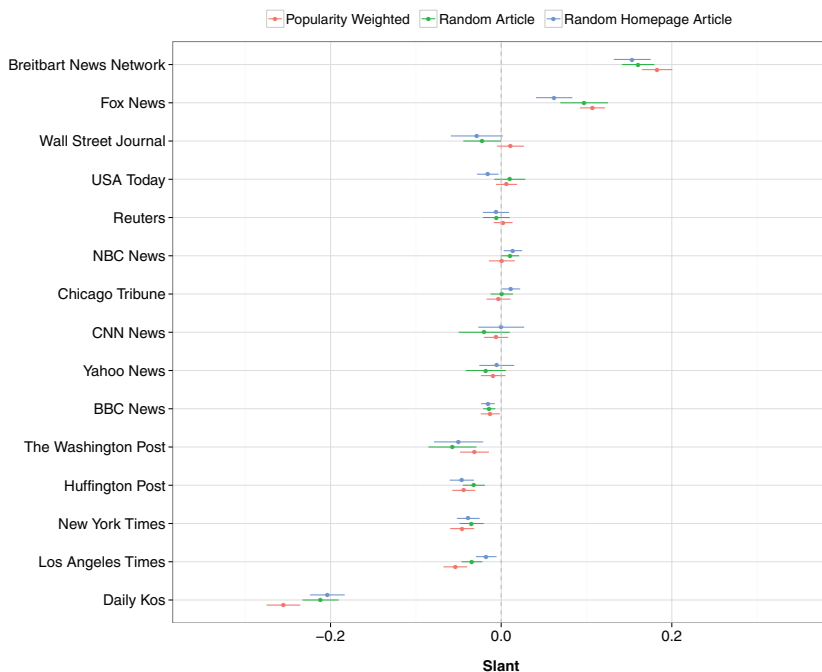
Figure 7. Overall Outlet-Level Slant Using Different Sample Weighting Procedures: Readership-Weighted, Uniform Weights, and Uniform Weights for Those Articles Appearing on the Outlet's Home Page. Whiskers give ±1 standard errors.

"strong liberal bias" in the US media. Further, given the ideological fragmentation of media audiences (Iyengar and Hahn 2009), theoretical and empirical work in economics predicts that news outlets would present biased perspectives in an effort to cater to their readers (Mullainathan and Shleifer 2005; Gentzkow and Shapiro 2006, 2010). Indeed, Gentzkow and Shapiro (2010) find that an outlet's use of highly partisan language (e.g., "death tax" instead of "estate tax") is strongly correlated with popular perceptions of its political leanings.

How, then, can we reconcile our results with the prevailing conventional wisdom? In part, the differences stem from the difficulty of directly measuring the ideological slant of articles at scale. For example, most articles do not cite policy groups, nor do they use highly partisan language, and those that do are not necessarily representative of political reporting more generally. In contrast, our combination of machine-learning and crowd-sourcing techniques does appear to yield accurate article-level assessments. Moreover, despite an increasingly polarized American public (Kohut et al. 2012), a substantial fraction of news consumers (64 percent) still prefer sources that do not have a systematic ideological leaning—and the proportion is particularly high among

those who get their news online (Purcell et al. 2010)—tempering partisan pressures. Finally, the traditional news-desk/editorial divide (Machin and Niblock 2006) may yet encourage publishers to maintain a degree of nonpartisanship in their descriptive reporting, reserving their opinion pages to show their ideological stripes and appeal to their audience base.

It may be tempting to cheer the relative lack of overt partisanship in the descriptive political reporting of most major American news outlets. Our analysis, however, also reveals some hurdles for robust political discourse. First, given the ideological distance between Democrats and Republicans— by some measures the largest of the modern political era—balanced coverage in the point/counterpoint paradigm (D'Alessio and Allen 2000) may not optimally inform voters about the issues. One might reasonably expect the facts to favor one party over the other—at least on some of the issues—and thus largely nonpartisan reporting may not accurately reflect the substantive differences between the political parties. (Coverage of political scandals is a notable exception to this convention of nonpartisanship; for example, scandals involving Democrats unsurprisingly portray Democrats more harshly than Republicans.) Second, for both Democrats and Republicans, we find that news outlets are almost universally critical rather than supportive, a practice some have called "gotcha journalism." For example, as many political commentators have observed, the failures of the Affordable Care Act received far more media attention than its successes. This tendency to print predominantly critical news may stem from publishers' desires to appear nonpartisan by avoiding apparent advocacy, or from readers' appetites for negative coverage, or from a combination of the two. Regardless of the rationale, predominantly critical coverage likely masks relevant facts and may hinder readers from developing informed opinions. Finally, though the relative uniformity we observe across news outlets provides readers with a common base of knowledge, it also limits the diversity of available perspectives.

We conclude by noting three important limitations of our study. First, we have characterized ideological slant by assessing whether an article is generally positive, neutral, or negative toward members of the Democratic and Republican parties. This codification has the advantage of side-stepping the tricky—and perhaps impossible—task of assessing bias against an objective "truth." However, it is certainly possible for an article not to explicitly favor a political party while still advocating a generally liberal or conservative position. Second, we considered a relatively short, twelve-month time span that did not coincide with a presidential or midterm election. While several hot-button issues attracted substantial media attention during this stretch—such as healthcare reform and marriage equality—other periods may exhibit more partisan dissent, which could in turn amplify differences between outlets. Third, we do not consider the effects of media bias on readers' attitudes or actions, such as voting, volunteering, and donating. It could be the case, for example, that even though we find that outlets have

ideologically similar coverage overall, a single partisan article could have an out-sized effect on readers. Despite these limitations, we believe our study is a natural starting point for investigating media bias at scale, and we hope the approach we have taken will benefit future exploration of such issues.

Supplementary Data

Supplementary data are freely available online at http://poq.oxfordjournals. org/.

References

Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. 2015. "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science* 348:1130–1132.

Baum, Matthew A, and Tim Groeling. 2008. "New Media and the Polarization of American Political Discourse." *Political Communication* 25:345–65.

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20:351–68.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3:993–1022.

Bottou, Léon. 2010. "Large-Scale Machine Learning with Stochastic Gradient Descent." Proceedings of the 19th International Conference on Computational Statistics (COMP- STAT), edited by Yves Lechevallier and Gilbert Saporta, 177–87. Paris: Springer.

Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011. "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?" *Perspectives on Psychological Science* 6:3–5.

Callison-Burch, Chris. 2009. "Fast, Cheap, and Creative: Evaluating Translation Quality Using Amazon's Mechanical Turk." Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, 1:286–95. Association for Computational Linguistics.

Cortes, Corinna, and Vladimir Vapnik. 1995. "Support-Vector Networks." *Machine Learning* 20:273–97.

D'Alessio, Dave, and Mike Allen. 2000. "Media Bias in Presidential Elections: A Meta- Analysis." *Journal of Communication* 50:133–56.

Flaxman, Seth, Sharad Goel, and Justin M. Rao. 2015. "Ideological Segregation and the Effects of Social Media on News Consumption." Social Science Research Network 2363701. Available at http://ssrn.com/abstract=2363701.

Gamson, William A. 1992. *Talking Politics*. Cambridge: Cambridge University Press.

Gamson, William A., and Kathryn E. Lasch. 1981. "The Political Culture of Social Welfare Policy." University of Michigan CRSO Working Paper No. 242.

Gamson, William A., and Andre Modigliani. 1989. "Media Discourse and Public Opinion on Nuclear Power: A Constructionist Approach." *American Journal of Sociology* 95:1–37.

Gentzkow, Matthew, and Jesse M. Shapiro. 2006. "Media Bias and Reputation." *Journal of Political Economy* 114:280–316.

———. 2010. "What Drives Media Slant? Evidence from US Daily Newspapers." *Econometrica* 78:35–71.

———. 2011. "Ideological Segregation Online and Offline." *Quarterly Journal of Economics* 126: 1799–1839.

Goodman, Joseph K., Cynthia E. Cryder, and Amar Cheema. 2013. "Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples." *Journal of Behavioral Decision Making* 26:213–24.

Groseclose, Tim, and Jeffrey Milyo. 2005. "A Measure of Media Bias." *Quarterly Journal of Economics* 120:1191–1237.

Ho, Daniel E., and Kevin M. Quinn. 2008. "Measuring Explicit Political Positions of Media." *Quarterly Journal of Political Science* 3:353–77.

Iyengar, Shanto. 1994. *Is Anyone Responsible? How Television Frames Political Issues*. Chicago: University of Chicago Press.

Iyengar, Shanto, and Kyu S. Hahn. 2009. "Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use." *Journal of Communication* 59:19–39.

Iyengar, Shanto, and Donald R. Kinder. 2010. *News That Matters: Television and American Opinion*. Chicago: University of Chicago Press.

Kittur, Aniket, Ed H. Chi, and Bongwon Suh. 2008. "Crowdsourcing User Studies with Mechanical Turk." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 453–56. ACM.

Kohut, Andrew, Carroll Doherty, Michael Dimock, and Scott Keeter. 2012. "In CNews Landscape, Even Television Is Vulnerable." Pew Center for the People and the Press. Available at http://www. people-press.org/2012/09/27/section-4-demographics-and-political-views-of-news-audiences/.

Krosnick, Jon A., and Donald R. Kinder. 1990. "Altering the Foundations of Support for the President Through Priming." *American Political Science Review* 84:497–512.

Langford, John, Lihong Li, and Alex Strehl. 2007. "Vowpal Wabbit Open Source Project." Available at https://github. com/JohnLangford/vowpal_wabbit/wiki.

Machin, David, and Sarah Niblock. 2006. *News Production: Theory and Practice*. London: Routledge.

Mason, Winter, and Siddharth Suri. 2012. "Conducting Behavioral Research on Amazon's Mechanical Turk." *Behavior Research Methods* 44:1–23.

McCombs, Maxwell E., and Donald L. Shaw. 1972. "The Agenda-Setting Function of Mass Media." *Public Opinion Quarterly* 36:176–87.

Mullainathan, Sendhil, and Andrei Shleifer. 2005. "The Market for News." *American Economic Review* 95:1031–1053.

Nelson, Thomas E., Rosalee A. Clawson, and Zoe M. Oxley. 1997. "Media Framing of a Civil Liberties Conflict and Its Effect on Tolerance." *American Political Science Review* 91:567–83.

Nelson, Thomas E., and Donald R. Kinder. 1996. "Issue Frames and Group-Centrism in American Public Opinion." *Journal of Politics* 58:1055–1078.

Nelson, Thomas E., Zoe M. Oxley, and Rosalee A. Clawson. 1997. "Toward a Psychology of Framing Effects." *Political Behavior* 19:221–46.

Paolacci, Gabriele, Jesse Chandler, and Panagiotis G. Ipeirotis. 2010. "Running Experiments on Amazon Mechanical Turk." *Judgment and Decision Making* 5:411–19.

Patterson, T. E. 1993. *Out of Order*. New York: Knopf.

Plank, Barbara, Dirk Hovy, and Anders Søgaard. 2014. "Linguistically Debatable or Just Plain Wrong?" Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL, 2:507–11.

Puglisi, Riccardo, and James M. Snyder. 2011. "Newspaper Coverage of Political Scandals." *Journal of Politics* 73:931–50.

Purcell, Kristen, Lee Rainie, Amy Mitchell, Tom Rosenstiel, and Kenny Olmstead. 2010. "Understanding the Participatory News Consumer: How Internet and Cell Phone Users Have Turned News into a Social Experience." Pew Internet & American Life Project. Available at http://www.pewinternet.org/files/old-media/Files/Reports/2010/PIP_Understanding_the_ Participatory_News_Consumer.pdf.

Schuman, Howard, and Stanley Presser. 1996. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. Thousand Oaks, CA: Sage.

Settles, Burr. 2009. "Active Learning Literature Survey." Technical Report, University of Wisconsin–Madison.

Sorokin, Alexander, and David Forsyth. 2008. "Utility Data Annotation with Amazon Mechanical Turk." *Urbana* 51:820.

Sutter, D. 2001. "Can the Media Be So Liberal? The Economics of Media Bias." *Cato Journal* 20:431–51.

Wais, Paul, Shivaram Lingamneni, Duncan Cook, Jason Fennell, Benjamin Goldenberg, Daniel Lubarov, David Marin, and Hari Simons. 2010. "Towards Building a High-Quality Workforce with Mechanical Turk." Proceedings of Computational Social Science and the Wisdom of Crowds, NIPS, 1–5.

Zhou, Daniel Xiaodan, Paul Resnick, and Qiaozhu Mei. 2011. "Classifying the Political Leaning of News Articles and Users from User Votes." Proceedings of the Fifth International Conference on Weblogs and Social Media.