What Drives Demand for Media Slant?*

Marcel Garz† Gaurav Sood‡ Daniel F. Stone§ Justin Wallace¶

February 21, 2018

Abstract

We study the determinants of demand for political news. We examine stories on the 2016 US presidential debates and on poll results throughout the 2012 and 2016 campaigns (horse race news). We use incentivized surveys and web data on "most viewed" stories to conduct within-outlet-topic analyses, holding fixed two main components of the information value of news. Our results support both the cognitive dissonance and the credibility theories of demand for media slant, and fail to support the instrumental value theory.

*Our title alludes to that of Gentzkow and Shapiro (2010) ("What Drives Media Slant?"). We thank Matthew Gentzkow, Jesse Shapiro, Greg Martin, Matt Botsch, and participants at the 2018 Economics of Media Bias Workshop for very helpful comments, and Isaiah West for excellent research assistance.

†Marcel is a postdoctoral researcher at Hamburg Media School, Germany, and a research fellow at the University of Hamburg's Institute for Media Economics. He can be reached at: m.garz@hamburgmediaschool.com.

‡Gaurav can be reached at: gsood07@gmail.com.

§Dan is Assistant Professor of Economics at Bowdoin College. He can be reached at dstone@bowdoin.edu

¶Justin is a Ph.D. student at the University of Washington's Department of Economics. He can be reached at: jtw95@uw.edu.

1 Introduction

People tend to get news from like-minded sources. This behavior raises now very well-known concerns about echo chambers and filter bubbles exacerbating political polarization, gridlock, political hostility, and related issues.[1] The taste for like-minded news (news from an ideologically aligned source), a.k.a. *partisan selective exposure*, is often assumed to be rooted in psychology—loosely speaking, the desire to read news that "feels good" and/or avoid news that is unpleasant, perhaps due to cognitive dissonance.[2]

However, a number of economics papers have shown that rational consumers can choose like-minded sources because they provide greater instrumental value. That is, news from a like-minded source leads to a better decision, as compared to news from other sources (e.g., Burke, 2008; Chan and Suen, 2008; Oliveros and Várdy, 2015; Fang, 2016). A third possibility is that consumers perceive news to be more accurate when it confirms existing beliefs. Thus, like-minded sources have reputations as being more credible (Gentzkow and Shapiro, 2006), although this perception could be mistaken (Vallone et al., 1985; Stone, 2011).

The distinction between these theories—which we refer to, for short, as the psychology, instrumental, and reputation mechanisms—has important welfare implications. If consumers choose like-minded news because it actually has greater instrumental value, then proliferation of diverse media has greater social benefits (Gentkzow et al., 2015), and concerns about selective exposure are more likely to be overblown. If consumers choose like-minded sources because they are more trusted, then consumers may at least be willing to see a variety of stories from these outlets, and would be more informed than if consumers were only willing to view "good news" from like-minded outlets.

The empirical literature addressing the distinction between mechanisms underlying news

[1]See, e.g., Pariser (2011), Bakshy et al. (2015), Flaxman et al. (2016), Halberstam and Knight (2016), Sood and Lelkes (2016), Peterson et al. (2017).

[2]See, for example, Iyengar and Hahn (2009), and from economics, Mullainathan and Shleifer (2005) and Bernhardt et al. (2008). Stroud (2011) discusses three psychological factors that could cause such demand: cognitive dissonance, motivated reasoning, and "cognitive misers" (that processing conflicting information takes more resources than consistent information).

demand is limited.[3] Our paper seeks to contribute to this literature. We examine the relationship between "congeniality" of a story's headline—whether it conveys good news or bad news, for a reader with particular political preferences—and demand for the full story. We do this for two related issues: assessments of performance in US presidential campaign debates and "horse-race" news on the presidential candidates' chances of winning the election. Both of these issues have the advantage of being reported on repeatedly by various major news outlets, with substantial variation in congeniality of reported stories both across and within outlets.

We discuss relevant theory in Section 2. A key prediction is that only the psychology mechanism could plausibly explain greater demand for stories with more congenial headlines from a like-minded outlet, on a given topic. These stories could certainly be more enjoyable due to various psychological factors, and it is unlikely that these stories would be perceived as offering more useful or more accurate information. We also argue that the psychology mechanism would be the most plausible cause of outlets producing a greater number of stories on days when news is more congenial for their readers. The instrumental and reputation mechanisms could explain other patterns in within-outlet effects. Specifically, it is plausible that for debate news at least, demand could be higher for stories with *less* congenial headlines due to the instrumental mechanism. These stories are more likely to persuade a reader to change her vote, and thus perhaps more worth spending the time to read in full. And for both debate and horse race stories, the reputation channel could cause demand to be higher for stories with "atypical" headlines for the outlet—headlines that lean left from outlets considered

[3] The most closely related paper to ours is Garz et al. (2018), who use observational data to show that users engage more with congenial Facebook posts than with uncongenial ones. Other literature in economics on demand for congenial information more generally (beyond political media) is better developed. For instance, Karlsson et al. (2009) find that investors exhibit "ostrich effect," checking their portfolios more often when markets are rising rather than falling. See Golman et al. (2017) for a review of the literature. See Masatlioglu et al. (2016) for experimental work on preferences over the distribution of non-instrumental information. There is a large literature outside of economics on selective exposure but most of it uses laboratory data. A paper that is especially similar to ours in that it tries to directly analyze the determinants of the demand for like-minded slant is Metzger et al. (2015); our papers complement one another, using different types of data but reaching similar conclusions.

to typically lean right, and vice versa—these headlines could be perceived as signaling articles with particularly credible information.[4]

Looking across outlets, all three mechanisms could explain a general preference for getting debate news from a like-minded outlet. Such an outlet might offer the most voting-decision relevant information, be considered most credible, or just be most enjoyed. For our horse race data, we do not directly observe individuals' choices of which outlet to get news from, but can compare average headline "slants" across outlets, which likely reflect consumer preferences. We argue that the instrumental mechanism cannot explain differences in these slants, since horse race news is of little instrumental value in general, and skewed horse race news is especially unlikely to be of particular instrumental value. Differences in mean slant across outlets for horse race stories could be explained by the psychology mechanism, as consumers could enjoy news favoring their preferred candidates' chances, and also by the reputation mechanism, if outlets pandered to consumer priors to appear more credible. This reputation effect might dissipate as the election approached and "truth" was sooner to be revealed.

Independent of these arguments, our work offers two more general contributions. First, the analysis of the relation between within-outlet story demand and headline congeniality can be conservatively viewed as a descriptive analysis of within-outlet selective exposure. This topic is important in its own right, and has received much less attention than that of selective exposure across outlets. Second, the analysis of variation in mean slant in horse race reporting across outlets contributes to the broader literature on estimating media bias. Most studies in this literature analyze text and word choice, and relatively few study reporting on a relatively objective topic like the horse race (Puglisi and Snyder Jr, 2015). This aspect of our paper complements two related studies from other disciplines that also study bias in reporting on poll results using different data sets, Tremayne (2015) and Searles et al. (2016).

[4]Such atypical headlines would be uncongenial for a story from a like-minded outlet. So this prediction varies sharply from the first prediction stated above in this case. However, this reputation effect and the psychology effect point in the same direction for atypical headlines from outlets that are not ideologically aligned with the reader, so we cannot separate these mechanisms as easily for this case. We consider this possibility, and discuss accordingly, as appropriate throughout the paper.

We study both debate and horse race stories so that we can examine both complementary topics, with similar but distinct theoretical predictions, and complementary types of data. We use incentivized surveys for debate stories, and use (observational) web data for horse race stories. The survey data have the advantages of being micro-level, and giving us control over the timing and menu of news options offered, which was ideal given the known timing of debates. The web data have the advantage of reflecting real-world choices from the entire relevant population for the outlets studied, for a larger number of events (poll results). We describe the survey data and analysis in Section 3, and the web data and analysis in Section 4, and summarize each of these here before proceeding.

The surveys were conducted on Amazon's Mechanical Turk (MTurk). A survey was conducted in the morning following each of the first three presidential election debates in 2016. We offered subjects a choice of real headlines for timely stories on the debate "winner" from the New York Times (NYT) and Fox News, and two other headlines on issues unrelated to presidential politics. Subjects were given an incentive to click on and read the story they most wanted to read. Both the NYT and Fox deemed the Democrat the winner of the first debate, the Republican the winner of the second, and there were mixed verdicts for the third. We then estimate whether subjects were more likely to click on a debate story when the headline was congenial, and whether this effect varies across the outlets.

Survey respondents on both sides of the political spectrum were indeed more likely to choose news about debates, as opposed to one of the non-debate options, when the debate news was more favorable toward their preferred candidate, supporting the psychology theory. But for Clinton supporters, demand for NYT debate news was not significantly sensitive to the congeniality of headline, and the overall variation in their demand was driven by variation in demand for Fox News stories. Since headlines congenial to Democrats from Fox were unusual for Fox, this result could also plausibly be explained by the reputation mechanism, as noted above. For Trump supporters, there was also evidence of increased demand for more congenial news, especially for Fox. It is less plausible that Fox's pro-Trump stories were perceived as Fox's most credible stories, and so this result provides cleaner evidence in support of the psychology mechanism.

We obtained the web data by scraping homepage and politics section headlines, and "most viewed" headlines from the NYT, Fox, and four other prominent outlets with a variety of ideological reputations in both 2012 and 2016. The other outlets for 2016 were: `washingtonpost.com`, `news.google.com`, `news.yahoo.com`, and `wsj.com`. We used the web archive to obtain the 2012 data; the *Post* and Google News sites were not available for that year so we replaced them with `huffingtonpost.com` and `usatoday.com`. We focus the discussion on the NYT and Fox results for consistency and because of the outlets' especially high importance. We used MTurk coders to rate the "slant" (favorability to one party or the other) of each headline. We provide various types of validation of our slant measures, including Krippendorff alpha calculations, showing that the mean slants are highly correlated with daily poll results, and that the slant of each outlet in each year is predictive of daily polls, even conditional on the slants of news from the other outlets that day. Our final samples include 425 headlines for 2012, and 696 for 2016.

We find that the slants of horse race stories for the different outlets varied over the course of a campaign in similar ways. Both left-wing and right-wing outlets tended to report bumps for each candidate at the same time, consistent with Barberá and Sood (2014)'s finding that intra-media content is perhaps surprisingly diverse. However, the mean slants vary across the outlets substantially, in ways mostly consistent with their reputations. Fox's headlines were slanted in favor of the Republicans' chances in both years, versus those of each of the other outlets that year. There is some evidence that the NYT's were more favorable to Democrats than Yahoo's in both years, and solid evidence of this relative slant as compared to Google in 2016. These results are consistent with those reported by Tremayne (2015) and Searles et al. (2016), but differ in substance from the findings of Budak et al. (2016), who conclude that the political slants of "news outlets are considerably more similar than generally believed." The variation we find in mean slant across outlets on an issue which, in theory, is an objective fact likely reflects psychology, but could also reflect reputation, as noted above. If this were the case, we would expect slant to decline as the election approached, since there would be reputational costs of slant when the electoral outcome was realized. We only find evidence of such a decline in slant over time for Fox in 2016, indicating that slants for most cases were unlikely driven by reputation concerns.

Next, we analyzed the relationship between headline congeniality and whether the story made the outlet's most viewed list. We consider this analysis our paper's most significant contribution, and, frankly, expected a priori to find numerous positive relationships. Instead, we found only one—for Fox News in 2012, and even that result was not robust. Moreover, there is more robust evidence of a negative relationship between congeniality and popularity for both Fox and the NYT in 2016. These results may have been due to the reputation mechanism (readers may have found these stories particularly credible due to their non-standard slants as noted above) or perhaps because the outlets became too extreme for their readers' tastes in 2016 for supply-side reasons. Fox readers may have preferred more right-slanted news in 2012, as compared to 2016, because a higher fraction of readers supported the Republican candidate in 2012. It is also possible that non-typical readers of those outlets (readers with non-standard ideologies for the outlets) flock to these stories because of the psychological value. The bottom line, however, is that the within-outlet popularity of a horse race story was not particularly driven by the headline indicating congenial news for the outlet's typical reader.[5]

Last, we analyze whether outlets reported more horse race stories on days in which the news was more congenial for those outlets' typical readers, and find modest evidence of this for Fox in 2012, and for the NYT in 2016. We also present informal evidence that the NYT and Fox produced more debate news stories following debates in which their preferred candidate was considered to have "won" in 2016, but did not find this was the case in 2012. These results provide some additional support for the psychology mechanism.

In summary, we obtain some results best explained by psychology (especially for right-of-center consumers), some potentially explained by psychology or reputation, and some that seem best explained by reputation alone (the Fox and NYT most viewed effects for 2016 in

[5]If total website traffic was affected by the congeniality of horse race news at a given point in time, the relative popularity of a story (for a given outlet) at that time could be a misleading measure of its level of popularity. We present some data indicating that this should not be a concern for the interpretation of our results.

particular). Our results corroborate the literature on the importance of credibility of source, and that perhaps this is the key factor driving political information-seeking, as shown by, e.g., Miller and Krosnick (2000), and more recently supported by Metzger et al. (2015). Our results consistently fail to support the instrumental/welfare-improving theory of selective exposure. In the final section we provide concluding remarks, including further discussion of the relationship between the psychology and reputation mechanisms.

2 Theory

We discuss how ideas from the media bias theory literature imply various hypotheses for our empirical setting. As discussed above, the theory literature has noted that demand for news from like-minded sources could be due to psychological forces (the desire for belief confirmation and to avoid dissonance), and this demand can also be rationalized in two primary ways: instrumental value and reputation. The empirical implications of the psychology mechanism are fairly clear: psychological factors would likely cause readers to choose like-minded sources for political news on most, perhaps all, topics (including debate verdicts and the horse race), and to be more likely to choose stories with more congenial headlines for most or all topics.

Instrumental value theories imply that slanted news can be more useful for decision-making for consumers with strong priors or preferences, as compared to more neutral news. Such consumers can only be persuaded to change their decision by strong contrary evidence, which can only be produced by an outlet whose reporting strategy favors the consumer's ex ante optimal choice. For example, a consumer who leans toward voting Democrat might not be persuaded to vote Republican by endorsements from Republican-leaning outlets, or even more centrist ones. The consumer would know that these outlets might report an endorsement of the Republican even if the underlying supporting information was not very strong. But this consumer may be persuaded to vote Republican by an endorsement by an outlet known to support Democrats, as this endorsement would only occur if the outlet observed very compelling evidence supporting the Republican candidate, which is necessary for the voter to wish to change from her strong prior favoring the Democrat. Thus, only a Democratic outlet provides instrumental information value to a Democratic reader.

In general, models making this point (such as Chan and Suen, 2008, and others noted above) can be interpreted as examples of the broader literature on delegation and advice. Just as a decision-maker is better off being advised by, or delegating a decision to, an agent who shares the decision-maker's preferences, a news media consumer may be better off getting news from a media outlet with a similar political viewpoint (Gentkzow et al., 2015).

Instrumental value is potentially relevant to the consumer's choice of which, if any, outlet to get debate news from, since interpretation of debate performance could depend on a subjective viewpoint. For example, if the Democratic candidate focused on security in a debate and the Republican focused on growth, it would be reasonable for security-concerned outlets to assess the Democrat's performance more generously than other outlets, and for security-concerned citizens to get news from these outlets.

But since an outlet's viewpoint is likely stable, at least within the time period in which the debates occur in a given year, instrumental value is unlikely to explain greater demand for debate news stories from a given outlet just because their headlines are more congenial. It is plausible that consumers who read debate news for instrumental value are unaffected by the congeniality of the headline (and simply use it as a cue for the story's topic). It is also plausible that the instrumental value mechanism could explain greater demand for debate stories with *uncongenial* headlines, since these are stories more likely to provide decision-relevant information (i.e., to cause the reader to switch her vote).

Horse race stories are about objective, albeit unobserved, facts: which candidate is more likely to win the election at a given point in time. Thus, these are not stories on an issue that consumers with different political viewpoints could reasonably interpret in different ways. Consequently, instrumental information seeking seems unlikely to explain the choice of which outlet to get horse race news from. Instrumental value seems even more unlikely to explain variation in demand for horse race news driven by variation in headline congeniality from the same outlet. Perhaps in the few days before the election, stories on the tightness of the race could affect the consumer's decision of whether to vote. But for most of the campaign, horse race stories do not seem to provide much instrumental information.[6]

The reputation theory of demand for like-minded outlets is as follows. In a nutshell: if my prior is that X is likely true, and I hear you say independently that X is likely true, I rationally infer that you are pretty smart. Similarly, news consumers with a political prior rationally infer that a media outlet that reports news consistent with this prior is relatively credible and accurate (Gentzkow and Shapiro, 2006). Thus, consumers who simply wish to be informed (whether for intrinsic or instrumental reasons) may prefer congenial news because they find it more trustworthy, and consumers with different priors will trust different outlets.[7]

The reputation mechanism can plausibly explain variation in demand for both debate news and horse race news across outlets. For example, media consumers who had an inflated belief about the chance of Clinton winning the 2016 election may have inferred (correctly or not) that outlets that reported that Clinton was very likely to win were relatively credible outlets. Gentzkow and Shapiro (2006) discuss how such reputation-driven slant is likely greater when objective feedback is less likely. For our context, this means that reputation-driven slant would likely decline late in the campaign. If an outlet reported inaccurate horse race news early in the campaign this could easily be forgotten; if the outlet did so the day before the election, while the consumer might draw a positive inference about the outlet that day, the inference would be more likely to decline quickly when the electoral results come in.

It is also possible that the reputation mechanism explains a within-outlet relationship between headline slant and interest in the story, for both debate and horse race stories. If

[6] It is conceivable that there are voters who are unsure of which candidate to support and draw inferences on this from horse race stories via social learning (e.g., if the crowd favors candidate X, she must be a good candidate). This is most plausible for news consumer-voters with weak priors, and for these people, no headline would be particularly congenial and so the value of this instrumental information would not be correlated with congeniality of the headline.

[7] Note that in instrumental value models, consumers also "trust" that like-minded outlets report news in a particular way (e.g., the leftist outlet only endorses the Republican when the underlying case for this candidate is very strong). The key difference between instrumental value and reputation models is that in the former, news consumers have shared, correct beliefs about outlet reporting strategy (and consumers with different utility functions simply find some strategies more useful than others), while in the latter, consumers disagree on the accuracy of an outlet's reporting, and thus disagree on the usefulness of an outlet for a consumer with a given utility function.

the perceived accuracy of articles varies within an outlet, and the slant of a headline affects an article's perceived accuracy, then this could make consumers more or less likely to click on slanted articles. We think it is useful here to draw a distinction between headlines whose slants are consistent and inconsistent with the outlet's typical slant. It seems reasonable to think that if an outlet is known for usually reporting news favoring party A, and then this outlet reports news favoring party B, the facts supporting this unusual story are likely to be especially compelling. Thus, the reputation mechanism could increase demand for such "atypical" news—uncongenial news from a typically congenial outlet and congenial news from an uncongenial one.[8] In the case of more typical news, or especially congenial news from a like-minded outlet, it seems less likely the news would be seen as more accurate and credible. Therefore, if we were to find greater demand for stories with more congenial headlines from like-minded outlets, we think this would likely be due to psychological factors.

Since firms likely report more stories on a topic when the topic is in higher demand, the quantities of stories reported could also shed light on the mechanism underlying news demand. It is very plausible that psychological factors lead to demand for a greater quantity of stories on more congenial news days. But it seems unlikely that either of the other mechanisms would lead to this outcome. Even if more or less congenial stories from a given outlet are considered more credible, this would likely not make readers want to read multiple stories making the same point for the sake of being better informed. But if the stories are being read because they "feel good" then we can see why readers would want to read more stories when they are more congenial.

We summarize this discussion in Table 1.[9] The table shows that the three mechanisms have distinct sets of empirical implications. These predictions are not unambiguous, and are simplified (it is particularly worth noting that we ignore supply-side factors). But we think this set of predictions is still very reasonable and useful for organizing the analysis and

[8]Another reason demand could be higher for atypical news is that it is more surprising (Ely et al., 2015).

[9]These three mechanisms are of course not the only factors explaining news demand. An additional cause of demand worth noting is the desire to be informed in social interactions: to try to persuade others, or to signal to others. Social factors would have implications very similar to the psychological factors that we discuss.

interpretation of results. And as noted in the introduction, independent of these predictions, the analysis can conservatively be interpreted as documentation of within-outlet selective exposure for news on the two topics.

Table 1: Summary of theoretical predictions

	News demand mechanisms		
	Instrumental information seeking	Reputation and credibility	Psych. value
Debate news			
Increase in demand for ...			
1. ... more like-minded outlets?	✓	✓	✓
2. ... stories w/ more congenial headlines, within outlet?			✓
3. ... stories w/ less congenial headlines, within outlet?	✓		
4. ... stories w/ atypical headlines, within outlet?		✓	
5. ... greater # congenial stories, within outlet?			✓
Horse race news			
Increase in demand for ...			
1. ... more like-minded outlets?		✓	✓
2. ... stories w/ more congenial headlines, within outlet?			✓
3. ... stories w/ less congenial headlines, within outlet?			
4. ... stories w/ atypical headlines, within outlet?		✓	
5. ... greater # congenial stories, within outlet?			✓

3 Survey data

3.1 Description

In the morning (between 9:00 AM and 10:00 AM) following each of the first three (of four) 2016 US presidential election debates, we conducted a survey on MTurk.[10] The first and third

[10]MTurk is a market for trading small services. "Workers" are paid for finishing small tasks like filling out surveys, tagging an image, etc. Research suggests that participants recruited on MTurk are "slightly more demographically diverse than are standard Internet samples and are significantly more diverse than typical American college samples" (Buhrmester et al., 2011). See Appendix Table A1 for sample statistics for key variables. Compared to the population, survey respondents recruited on MTurk tend to be younger, better educated, and more likely to identify with the Democratic party (Berinsky et al., 2012), though sectoral breakdown of employment is similar to more representative online surveys—the sectoral differences are no more than 7% (Huff and Tingley, 2015). Pleasingly, a broad variety of experiments done on MTurk tend to reach similar conclusions as those done on more representative samples (e.g., Mullinix et al., 2015). The two major advantages of MTurk over a survey firm are: a) MTurk

debates were between the presidential nominees, and the second one was between the vice president candidates. In each survey, we asked a small number of demographic and party affiliation questions, and asked respondents to pick the article they were most interested in reading from four headline options—two articles on which candidate won the debate, one from the NYT and one from Fox, and two articles on other topics from news.yahoo.com. The headlines are provided in Table A2. We use articles from the NYT and Fox because they are leading outlets with reputations and readers skewing left- and right-of-center, respectively (e.g., Pew, 2014). Respondents were also told that after choosing the article, they would be asked a question on the article's content. We noted that the difficulty of the question would be the same regardless of what article they picked, and that they would receive an additional payment if they answered the question correctly (respondents were paid $0.50 per survey and a bonus of $0.25 for answering the reading question correctly). The exact wording of the instructions can be found in the supplementary appendix (see Table A2 notes).

Since respondents were presented with a choice of articles that were both timely and real, and given an incentive to actually read the article they selected, respondent choices should reflect the articles that they would be most likely to read in similar real-world situations. The most unnatural element of our design was that respondents were presented with news choices from diverse outlets at the same time. This would be unrealistic for news consumers who, for example, go straight to the NYT website to decide which story to read, or have a Facebook feed consisting of ideologically similar sources. We consider this issue in the interpretation of our results. Another issue was priming: In the first two surveys, we asked which candidate the respondents were planning to vote for last, after choosing and reading their article. We chose this order to avoid priming respondents about the election, which could influence article selection in an unnatural way. In survey 3, we randomly asked some respondents about their preferred candidate before presenting the article options to assess if the order mattered. We find (in unreported tests) that the order did not have significant effects on article choice. Finally, the respondents might have already gotten their debate news prior to the survey.

is much more cost effective, allowing us to obtain a larger sample, 2) MTurk gives us control over the timing of surveys, which, as we explain in the next section, is crucial for their validity.

We try to minimize this possibility by conducting the surveys early in the morning after the debates.

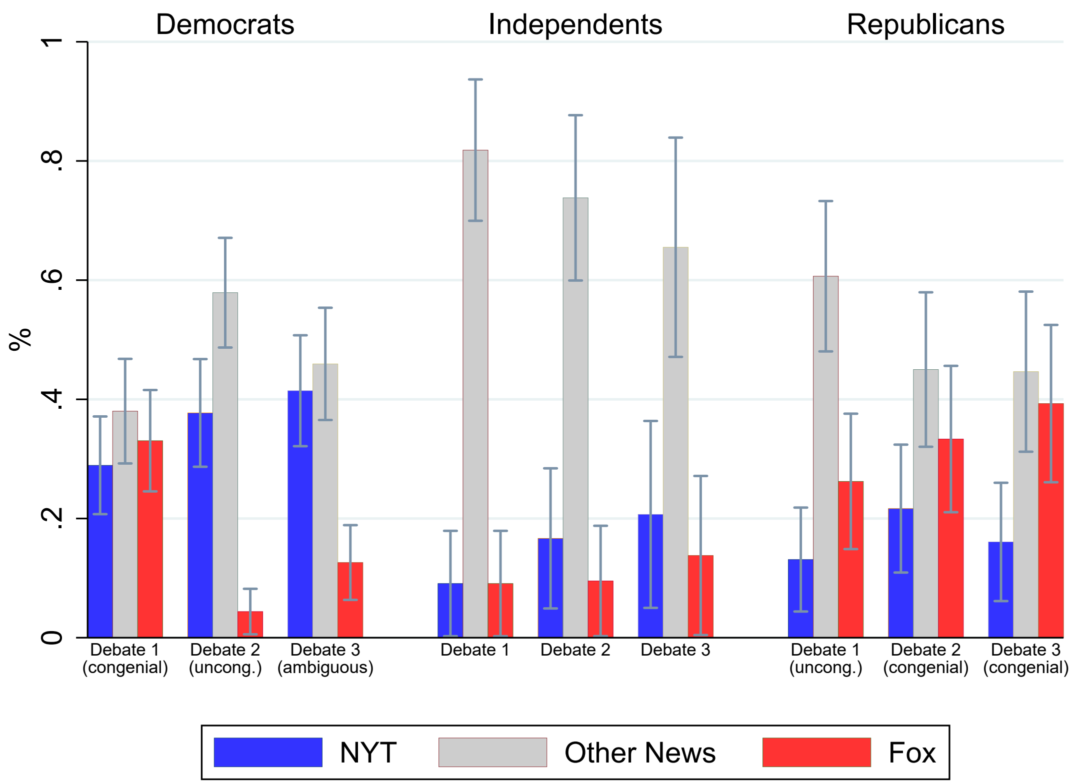
To be clear, our surveys do not randomize congeniality across groups of respondents labeled treatment and control. We do, however, examine variation in congeniality of news across surveys for partisan respondents. Assuming congeniality of news remained neutral for non-partisan respondents, they constitute a control group of sorts, allowing us to account for general changes in the importance of debate news over time. Moreover, we examine behavior (news choice) in response to an incentive. Consequently, we think it would be misleading to refer to this data as plain vanilla survey data.

We ran the surveys on debate news, and not news on other topics (such as the horse race), because the timing of these stories was known in advance. This allowed us to prepare in advance to select the stories, and post the surveys shortly after the stories became available. Debate news stories also worked well for our surveys because we expected the congeniality of these articles to vary over time and across outlets in an unambiguous way—most outlets usually proclaim one candidate the winner of each debate. Indeed, both the Fox and NYT headlines that we used in the first survey stated that Clinton won the first debate, and both said that Pence won the second. The outlets only disagreed on the third debate; Fox's headline said that Trump won, while the NYT's headline did not declare a winner.

We surveyed 250 MTurkers (US residents aged 18 or older) after each debate. We discarded observations in which the respondent did not answer the reading comprehension question correctly. We also discarded 16 observations in which the respondent chose "other" in response to the partisan self-identification question as we suspected many of these respondents were partisan, potentially influencing their article choices, but did not want to reveal this. Our final sample had 637 observations, with 345 identifying as Democrats or leaning toward Democrats, 177 identifying as Republican or leaning toward Republicans, and 115 as independent.

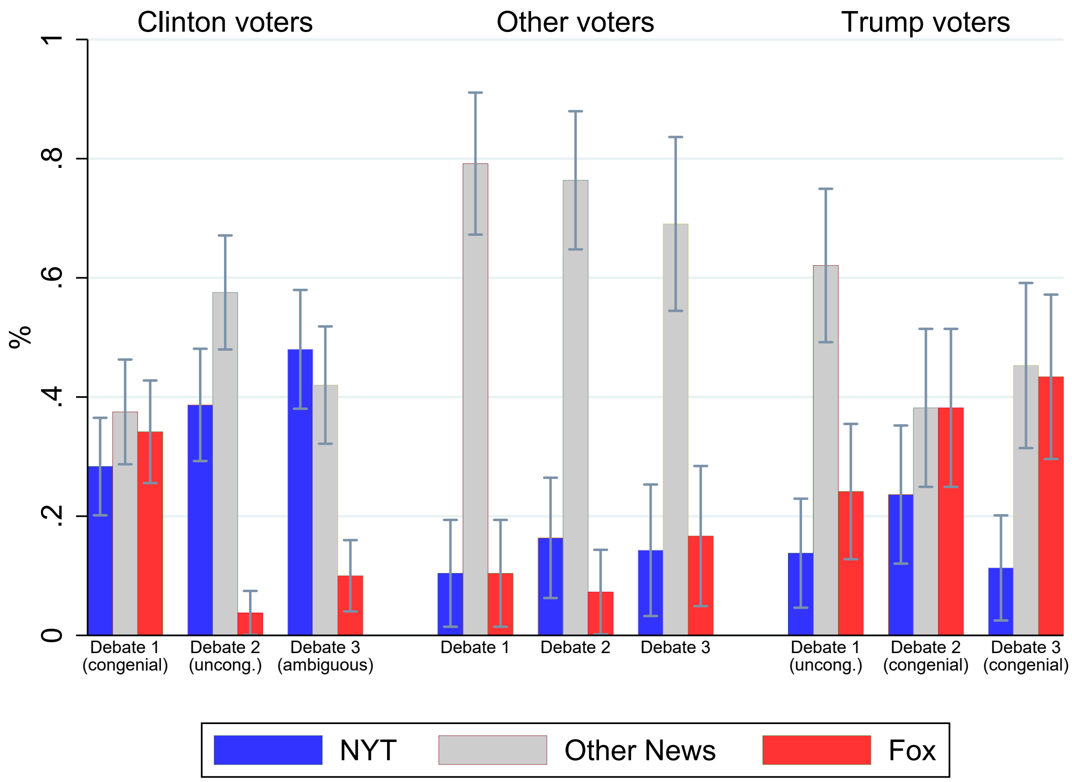
Figures 1 and 2 foreshadow the econometric results from the next section. Across the three surveys, Democrats were least likely, and Republicans most likely, to pick an "other [non-debate] news" story after the first debate, which was the only debate that both the NYT and Fox headlines said Clinton won. However, Democrats were slightly more likely to choose the NYT story after the second and third debates, as compared to the first debate. Democrats were significantly less likely to choose the Fox News story after the second and third debates compared to the first debate. Republicans' demand for NYT news was more stable across the debates, while their demand for Fox news grew as the congeniality (and congeniality relative to NYT) grew. Independents were much less likely than partisans to get political news in general, but became more likely to get political news in later debates, suggesting a secular increase in interest in political news. Results are similar but somewhat sharper when respondents are split out by preferred candidate rather than party.

Figure 1: News choices by debate and party



Note: 1) Both NYT and Fox survey 1 headlines said Clinton (Democrat) won first debate, 2) Both NYT and Fox survey 2 headlines said Pence (Republican, Trump's VP) won the second debate, and 3) Fox survey 3 headline said Trump won third debate while NYT survey 3 headline was ambiguous. The error bars denote 95% confidence intervals.

Figure 2: News choices by debate and preferred candidate



Note: 1) Both NYT and Fox survey 1 headlines said Clinton (Democrat) won first debate, 2) Both NYT and Fox survey 2 headlines said Pence (Republican, Trump's VP) won the second debate, and 3) Fox survey 3 headline said Trump won third debate while NYT survey 3 headline was ambiguous. The error bars denote 95% confidence intervals.

3.2 Analysis

We focus on the debate news predictions of Table 1 for which the theories vary, and which we have the data to address: #2 (#3) (within-outlet increase (decrease) in headline congeniality increases demand for psychological reasons) and #4 (atypical congeniality increases demand due to the reputation mechanism). At the end of the section, we present and discuss a limited amount of data relevant to prediction #5 (that outlets supplied more stories when the news was more congenial).

To analyze survey respondent choice, since respondents chose among four unordered alternatives, a multinomial model is ideal for analysis. However, in the interest of simplicity and transparency, we relegate the multinomial analysis to the appendix (see Table A3), and present here results of linear probability models predicting a binary outcome equal to: 1) Y^{NYT} (= 1 if the respondent chose the NYT article); 2) Y^{Fox} (defined analogously); 3) whether either type of debate news story is chosen ($Y^{NYT} + Y^{Fox}$). We run two sets of regressions, one using party identity as a measure of the respondent's politics and one using the respondent's preferred candidate for this, each with the following structure:

$$Y_i = \alpha + \beta_L D_i^L + \beta_R D_i^R + \beta_{S^2} S_i^2 + \beta_{S^3} S_i^3 + \beta_{L,2} D_i^L S_i^2 + \beta_{L,3} D_i^L S_i^3 + \beta_{R,2} D_i^R S_i^2 + \beta_{R,3} D_i^R S_i^3 + \beta_X X_i + \epsilon_i. \quad (1)$$

D_i^L is a dummy for respondent i being "type L" (a Democrat in one set of regressions, or Clinton-voter in the other set), D_i^R is analogous (independents are the omitted group); S_i^t is a dummy for respondent i taking survey t (the survey 1 dummy is omitted); X_i is a vector of demographic and other controls. Some respondents took more than one survey, but including fixed effects for these respondents has very little effect on the results.

The parameters of interest are the politics-survey interactions: $\beta_{L,2}$ and $\beta_{L,3}$ can be interpreted as mean changes in type L demand for surveys 2 and 3, respectively, as compared to survey 1, and $\beta_{R,2}$ and $\beta_{R,3}$, have analogous interpretations for type R respondents. Recall that the NYT's first headline was most (least) congenial, and the second headline least (most) congenial, to L (R) types. Fox's first headline was most (least) congenial to L (R) types. The psychology mechanism thus implies that for Y^{NYT}, $\beta_{L,3} < \beta_{L,2} < 0$. For Y^{Fox}, $\beta_{L,3} < 0$ and $\beta_{L,2} < 0$, but $\beta_{L,3}$ and $\beta_{L,2}$ cannot be compared since the headlines for those surveys are both uncongenial. For Republicans and Trump supporters, the signs of these predictions are reversed. The reputation mechanism implies that for Y^{NYT}, $\beta_{n,2} > \beta_{n,3} > 0$, and for Y^{Fox}, $\beta_{n,3} < 0$ and $\beta_{n,2} < 0$ (these predictions are the same for respondents of type n = L or R). Note that the effects of the psychology and reputation mechanisms go in opposite directions for Y^{NYT}, and the same direction for Y^{Fox} for L types, and vice versa for R types.

Table 2 reports the results. Congeniality drives Democrat and Clinton supporter demand for Fox news, but not for NYT news. Democrats are nearly 30 percentage points less likely to get Fox news when it is uncongenial than when it is congenial ($\beta_{L,2} < 0$). Congeniality also drives both Democrats' and Clinton supporters' demand for debate news from either source. There are no significant survey effects for Republicans. For Trump supporters, there is also evidence of a congeniality effect but it is primarily for Fox: Trump supporters are around 20 percentage points more likely to get Fox news in survey 2 as compared to survey 1 ($\beta_{R,2} > 0$).

Table 2: Analysis of survey data

	Y^{NYT}	Y^{Fox}	$Y^{NYT} + Y^{Fox}$
Panel A (Party Identity)			
Debate 2 × Democrat ($\beta_{L,2}$, uncongenial)	0.004	-0.293***	-0.289***
	(0.095)	(0.079)	(0.111)
Debate 3 × Democrat ($\beta_{L,3}$, uncongenial for Fox only)	-0.023	-0.221**	-0.243*
	(0.110)	(0.100)	(0.126)
Debate 2 × Republican ($\beta_{R,2}$, congenial)	0.009	0.071	0.079
	(0.101)	(0.104)	(0.128)
Debate 3 × Republican ($\beta_{R,3}$, congenial)	-0.095	0.112	0.016
	(0.113)	(0.121)	(0.140)
Adj. R^2	0.059	0.085	0.062
N	637	637	637
Panel B (Candidate supported)			
Debate 2 × Clinton supporter ($\beta_{L,2}$, uncongenial)	0.039	-0.274***	-0.235**
	(0.093)	(0.075)	(0.106)
Debate 3 × Clinton supporter ($\beta_{L,3}$, uncongenial Fox only)	0.122	-0.270***	-0.148
	(0.098)	(0.094)	(0.116)
Debate 2 × Trump supporter ($\beta_{R,2}$, congenial)	0.056	0.194*	0.250**
	(0.100)	(0.103)	(0.122)
Debate 3 × Trump supporter ($\beta_{R,3}$, congenial)	-0.069	0.186	0.117
	(0.099)	(0.117)	(0.132)
Adj. R^2	0.078	0.123	0.091
N	637	637	637

Note: All models are estimated using OLS with robust standard errors and include survey, education, gender, age, and party identity (Democrat, lean Democrat, Republican, lean Republican, independent) fixed effects. The reference category is debate 1. The reports on this debate were congenial for Democrats/Clinton supporters and uncongenial for Republicans/Trump supporters. *, **, *** denote 10%, 5%, 1% significance.

For L types, these results are consistent with off-setting reputation and psychology mechanisms for NYT demand, or neither mechanism affecting NYT demand, and for either reputation or psychology effects for Fox demand. For Trump supporters, the results are only consistent with the psychology mechanism for Fox demand, and do not indicate any variation in NYT demand. Thus, the results support the existence of psychology and/or reputation mechanisms for L types, and of the psychology mechanism for R types. The magnitudes are substantial (news demand can approximately double due to changes in these effects) but not completely dominant (these effects cannot be the only explanation for observed news demand).

To assess prediction #5, we manually collected the number of debate-related links (both articles and videos) from web.archive on both the Fox and NYT websites in the morning following each of the four debates of both 2012 and 2016. These numbers are presented in Table 3. The table shows that the number of stories was fairly constant for both outlets in 2012, but this was not the case in 2016. Fox had the most links after the third debate that year, and the fewest links after the fourth debate. The NYT had the fewest links after the second debate. Recall that Fox claimed Trump won the third debate, and the NYT said the Republican (Pence) won the second. Fox's headlines favored Clinton after the fourth debate (e.g., "Trump winning on points until terrible mistake"). Thus, both Fox's and the NYT's numbers of debate links are correlated with the congeniality of the debate outcome for their readers, and so the table provides further support for the psychology mechanism. The table also raises the question of how the slant of the outlets may have changed over time.

Table 3: Counts of links to reports on presidential debates the morning following each debate

	2012				2016			
Debate:	#1	#2 (VP)	#3	#4	#1	#2 (VP)	#3	#4
Fox	12	7	12	13	24	25	31	17
NYT	20	19	21	19	23	14	23	25

Note: The counts are based on web.archive.org snapshots of nytimes.com and foxnews.com at approximately 10:00 AM the morning following each debate.

4 Web data

4.1 Description

The survey analysis is intended to complement our analysis of real-world demand for horse race stories—stories about which candidate, if any, is winning the "race" at any given time, and by how much. For the 2016 presidential election, we started scraping news articles on July 27, 2016, well after the presumptive nominee for each party had been decided. The websites of *Fox News* (Fox), *Google News* (Google), *New York Times* (NYT), *Washington Post* (WashPost), *Wall Street Journal* (WSJ), and *Yahoo! News* (Yahoo) were scraped three times daily until election day (November 8, 2016). We downloaded outlets' landing pages, politics sections, and most viewed lists. We chose these outlets because of their prominence, ideological diversity, and because each of them publicly reports "top," "trending," "most popular," or "most viewed" stories; we use the last term ("most viewed") as short-hand to refer to all of these categories. The selection of Google's "top stories" is based on an algorithm, whereas the other five websites each use terms that explicitly or implicitly refer to stories being most frequently clicked on within a recent period of time. Given the inclusion of Fox News, it would be natural to include the two other major cable news outlets, CNN and MSNBC. However, neither of these outlets report most viewed stories.

News data for the 2012 presidential election were collected by scraping snapshots of the outlets' homepages stored by web.archive.org, also dating between July 27 and election day of that year. Since web.archive.org respects `robots.txt` advisory files, it did not collect data from Google and WashPost, and partly for Yahoo, which mandate `no follow`. We substituted for these losses with *USA Today* (USAT) and the *Huffington Post* (HuffPost).[11] We also downloaded data from snapshots in 2016 to complement the live data, and collected some of the archived data manually (rather than by scraping), when necessary.

Both the archive and the live data include article date, time, URL, source, headline, text, author, and keywords, and, where applicable, the current rank in the most viewed list. Next, we processed the text data so that it could be used for statistical analysis. To do this, we first had to identify a set of articles that were relatively likely to be horse race stories. We used a fairly broad set of keywords to make the initial set inclusive and then narrowed this set down with human coding and additional restrictions. The initial set included any article with one of the following keyword combinations: at least one of the terms from the set {Obama (Clinton in 2016), Romney (Trump in 2016), president, white house, electoral} in the headline or URL

[11] We scraped the text for USAT and the HuffPost for 2016 as well but exclude them from the analysis to maintain consistency across years (i.e., to maintain an estimation sample with two outlets considered left-of-center, two considered right-of-center, and two relatively neutral). Moreover, USAT stopped reporting most viewed articles relatively early in the 2016 campaign.

and at least one term from {win, winning, momentum, lead (and not "leader"), bounce, bump, tied, gallup} in the headline. Through manual checks of random subsamples, these keywords were determined to lead to a very small fraction of false negative classifications (actual horse race stories that were not classified as horse race stories) at the expense of having a high number of false positives. We erred in this direction because the cost of cutting false positives with additional steps was relatively low. We then dropped articles that included a clear indicator of being an opinion piece in the headline, due to our focus on hard news stories.[12] This should reduce the prevalence of within-outlet variation in reputation. For example, certain readers might perceive certain authors to be particularly credible. With the exceptions of opinion pieces, readers do not see the author (if any) before clicking on the headline, in most cases.

We then had three "master" MTurkers rate every headline on a five point scale—very good news or good news for either the Democratic or Republican candidate's chances of winning, or neutral— with two additional options, ambiguous or unclear, or not relevant to the candidates' chances of winning the election. The exact instructions provided to MTurkers are in Table A4. We did not allow the Mturkers to see the names of the outlets, so the ratings should be based just on headline content and comparable across outlets. We faced a trade-off between incentivizing effort and accuracy of the Mturkers' work, and excessive monitoring and potential "demand effects." Thus, we kept instructions intentionally vague, not specifying additional payment for particular results, but simply offering the incentive of generously paid additional work (workers were paid $3 for each batch of 40 headlines) if the work was done "carefully and reasonably." We monitored the coding done by these MTurkers by choosing four headlines with relatively unambiguous slants, and spot-checked each worker's ratings for these headlines, for each batch of 40 headlines. These spot-checked ratings were consistent with our expectations in all but one batch (out of dozens of batches); we manually checked the other headlines in that batch and they seemed reasonable so we continued to invite the worker to do additional work. We used the same three MTurkers for the vast majority of the

[12]We used the following headline keywords, determine by inspection, to identify articles as opinion pieces: {opinion, schoen, goodwin, rove, strassel, power play, juan williams, bias alert, gainor, reich, douthat, dan rather, whalen, starnes}.

coding for consistency and because presumably, as they gained experience, their ratings likely became more accurate, but we used three other MTurkers for a small number of batches that the original three workers were not available for.

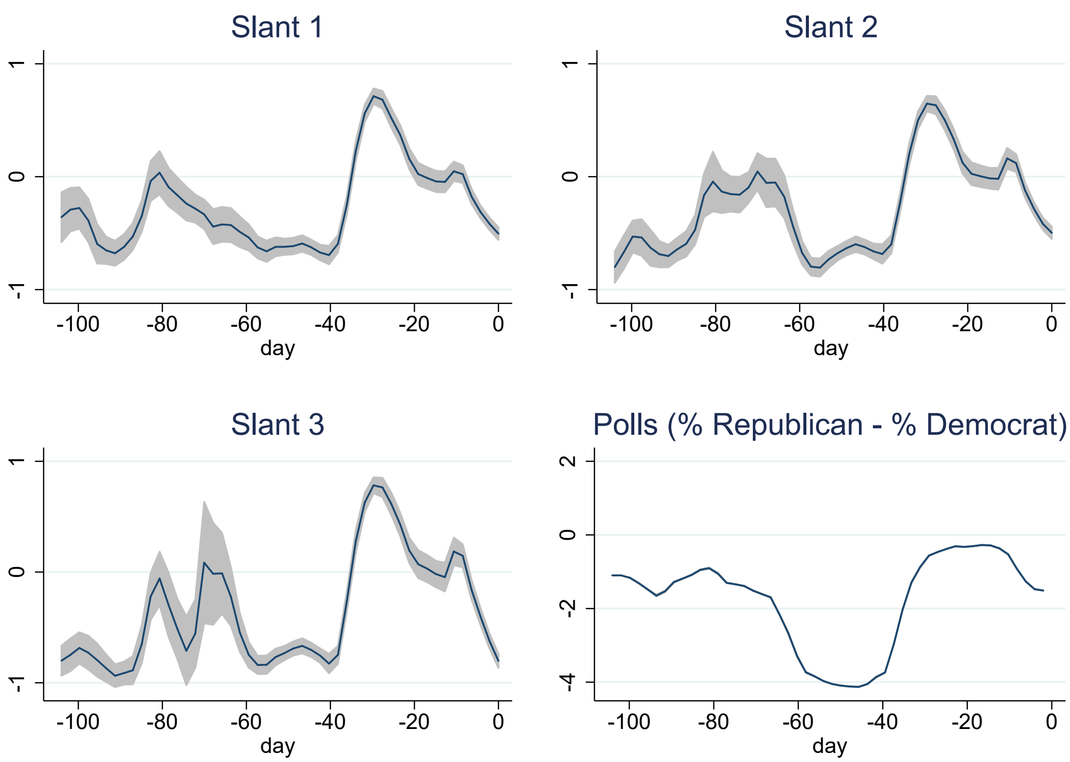
There were 2,025 headlines coded in total. The Krippendorff's α, a standard measure of inter-coder reliability, for all the coded headlines is 0.313. Condensing to an ordinal three-point scale—good news for the Democrats, good news for the Republicans, or neutral or ambiguous—increases α to 0.816, exceeding the standard threshold of 0.80.[13] Of the 1,177 headlines that were not rated as irrelevant by any of the coders, the α for the five-point scale is 0.405 and for the three-point scale, it is 0.859. Of the 871 headlines that were not rated as irrelevant or ambiguous by any of the coders, the α for the five-point scale is 0.440 and for the three-point scale, it is 0.900. Thus, the three-point scale appears much more valid than the five-point scale, and so we only use the three-point scale going forward, with very good or good news for Democrats coded as -1, very good or good news for Republicans coded as 1, and neutral and ambiguous news coded as 0. Restricting the sample to headlines that all three raters agreed were relevant and unambiguous increases the validity, but results in a substantial loss of observations. Thus, for transparency and to examine robustness, we consider three variations of the measure for all of our analyses: $Slant_1$ = mean slant of headlines rated as irrelevant or ambiguous by at most one coder; $Slant_2$ = mean slant of headlines not rated as irrelevant by any coder; $Slant_3$ = mean slant of headlines not rated as irrelevant or ambiguous by any coder. Headlines used for $Slant_3$ are a strict subset of those used for $Slant_1$. Most, but not all, $Slant_2$ headlines are also $Slant_1$ headlines.

Figures 3 and 4 present smoothed plots of daily means of each slant measure versus daily poll averages (we use R's "pollstR" library to obtain poll data from HuffPost Pollster). The plots are quite similar to one another, and to the polling average. Furthermore, Tables A5 and A6 show that the slant of each outlet's stories is predictive of that day's polls, for both years. Most of the outlets' slants are also significant predictors of polls even controlling for a date

[13]For the three-point scale, we used the ordinal method to calculate the Krippendorff α and thus coded "not relevant" as missing values. For the five-point scale, Krippendorff α values were very similar whether we used ordinal or nominal methods.

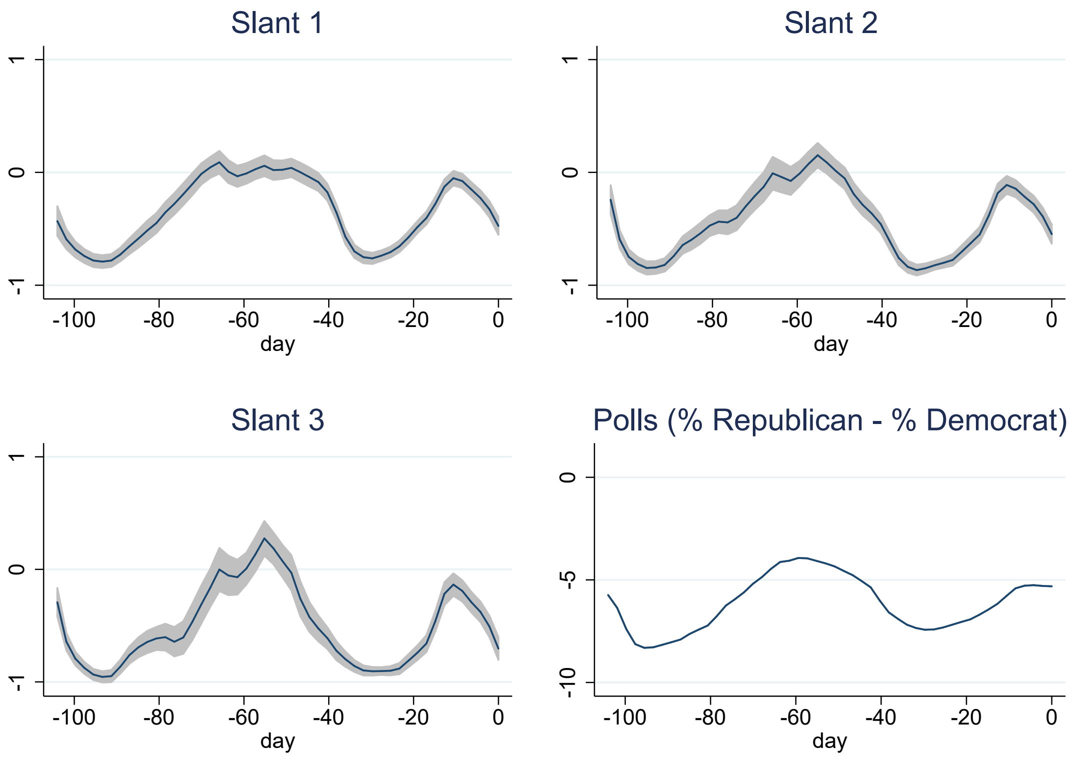
polynomial, and some slants are also significant predictors of changes in polls (sometimes with date polynomials). These results and figures provide further support for the validity of each slant measure. However, a number of results from the analyses (reported in the next section) are significant for one measure of slant but not for others, and so it could be misleading to restrict the analysis to just one of these measures. To illustrate the coding, Table A7 shows the three slant ratings for all headlines that contained the appropriate keywords for the day before the 2016 election (November 7, 2016). Most of the ratings seem very reasonable; sometimes $Slant_2$ and/or $Slant_3$ seem to appropriately drop a non-horse race headline (e.g., "Trump urges voters to deliver justice at polls"); sometimes these more restrictive versions seem to mistakenly drop a horse race headline (e.g., "polls Trump and Clinton virtually tied in key swing states"). Thus, the table also supports the use of the various *Slant* measures in the analysis.

Figure 3: Mean $Slant_i$ and poll average versus day relative to election day (day 0) in 2012.



Note: Curves are kernel-weighted smoothed local polynomials with 95% confidence interval bands. Positive values of *Slant* denote better chances of winning for the Republican candidate, whereas negative values indicate better chances for the Democratic candidate.

Figure 4: Mean $Slant_i$ and poll average versus day relative to election day (day 0) in 2016.



Note: Curves are kernel-weighted smoothed local polynomials with 95% confidence interval bands. Positive values of *Slant* denote better chances of winning for the Republican candidate, whereas negative values indicate better chances for the Democratic candidate.

Table 4 reports the number of unique most viewed and other articles per outlet, and their means for various slant measures.[14] The average slants vary across outlets substantially, largely consistently with the outlets' reputations. Panel A of the table, on the 2012 data, also suggests that the most viewed stories were slanted to the left of other stories for all outlets

[14]The number of headlines reported in this table is less than the corresponding number referred to in the Krippendorff alpha analysis (2,025) because the sample used in this table, and for most of the subsequent analysis, differs for two reasons. First, even the broadest slant definition that we use for the main analysis, $Slant_1$, is restricted to headlines coded as irrelevant or ambiguous by at most one coder, and therefore excludes many of the original 2,025 headlines. Second, the story-level data set collapses headlines with slight variants in wording to a unique observation, whereas the MTurkers coded multiple variants of headlines, with wording that slightly differed, for some stories (such as "FOX NEWS POLL Clinton leads Trump by 10 points both seen as flawed presidential candidates" and "Fox News Poll Clinton Leads Trump by 10 Pts Yet Both Flawed Say Voters"). Including these variants in the α calculations should not bias results up or down since the coders are as likely to disagree on variants of headlines for a given story as they are on a single version of a headline.

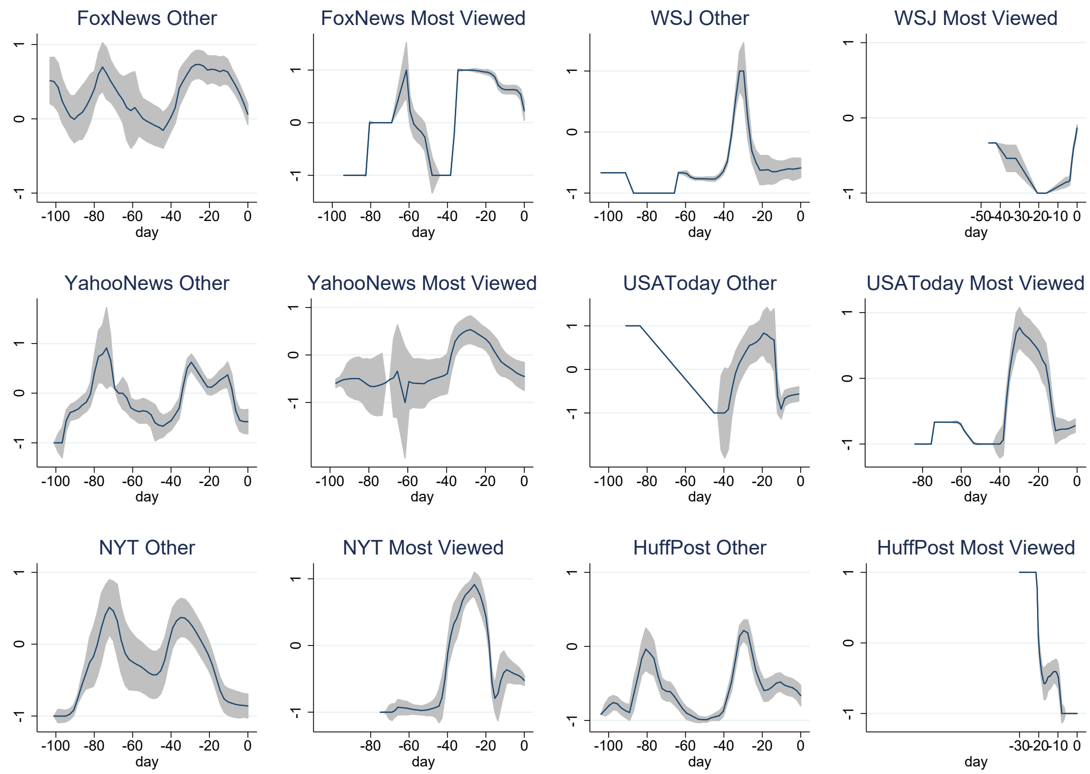
except for Fox. Panel B (on the 2016 data) suggests that both Fox's and the NYT's most viewed stories were more neutral than the outlet's other stories.

Table 4: Mean slants and article counts by outlet

Outlet	Type	$Slant_1$	N	$Slant_2$	N	$Slant_3$	N
Panel A: 2012							
Fox	Other	0.27	38	0.33	31	0.38	23
	Most viewed	0.37	25	0.34	27	0.42	22
WSJ	Other	-0.37	6	-0.37	6	-0.28	6
	Most viewed	-0.53	7	-0.52	7	-0.53	7
USAToday	Other	0.10	13	0.00	9	0.00	8
	Most viewed	-0.31	19	-0.24	20	-0.29	16
Yahoo	Other	-0.12	76	-0.23	50	-0.31	31
	Most viewed	-0.23	30	-0.21	32	-0.36	22
NYT	Other	-0.17	59	-0.15	58	-0.21	42
	Most viewed	-0.48	21	-0.31	24	-0.48	14
HuffPost	Other	-0.58	119	-0.50	102	-0.59	76
	Most viewed	-0.56	12	-0.37	16	-0.50	10
Panel B: 2016							
Fox	Other	0.35	23	0.50	12	0.67	8
	Most viewed	0.00	47	-0.11	33	-0.16	25
WSJ	Other	-0.65	12	-0.90	7	-1.00	6
	Most viewed	-0.67	11	-0.93	9	-0.93	9
Google	Other	-0.23	270	-0.28	205	-0.33	161
	Most viewed	-0.16	46	-0.36	34	-0.42	26
Yahoo	Other	-0.27	85	-0.42	60	-0.68	39
	Most viewed	-0.50	8	-0.33	6	-0.33	6
NYT	Other	-0.65	33	-0.79	24	-0.90	21
	Most viewed	-0.37	22	-0.35	19	-0.61	11
WashPost	Other	-0.38	96	-0.46	60	-0.61	45
	Most viewed	-0.36	43	-0.40	32	-0.51	24

To further examine the data before proceeding to the formal analysis, Figures 5 and 6 present smoothed polynomials with 95% confidence bands of the relationship of $Slant_1$ and days to the election for stories that made the most viewed list that day, and for all other stories, for each outlet and year. We restrict these figures to just $Slant_1$ for two reasons: 1) presenting these figures for the other slant measures would be unwieldy and 2) the figures are not formal statistical tests, and just illustrative. The figures use scraped headline-level data, not story-level data, i.e., the figures are created using a data set with a separate observation for each headline scraped for a given story, because here we want to include separate observations for the same story made available on different days. The confidence bands reflect variation in $Slant_1$ within an outlet, both within and across contiguous days, but the bands can be misleading as they can be small (or non-existent) due to limited data. Still, the trends and confidence bands are useful for illustrating broad trends in, and the availability of, the data.

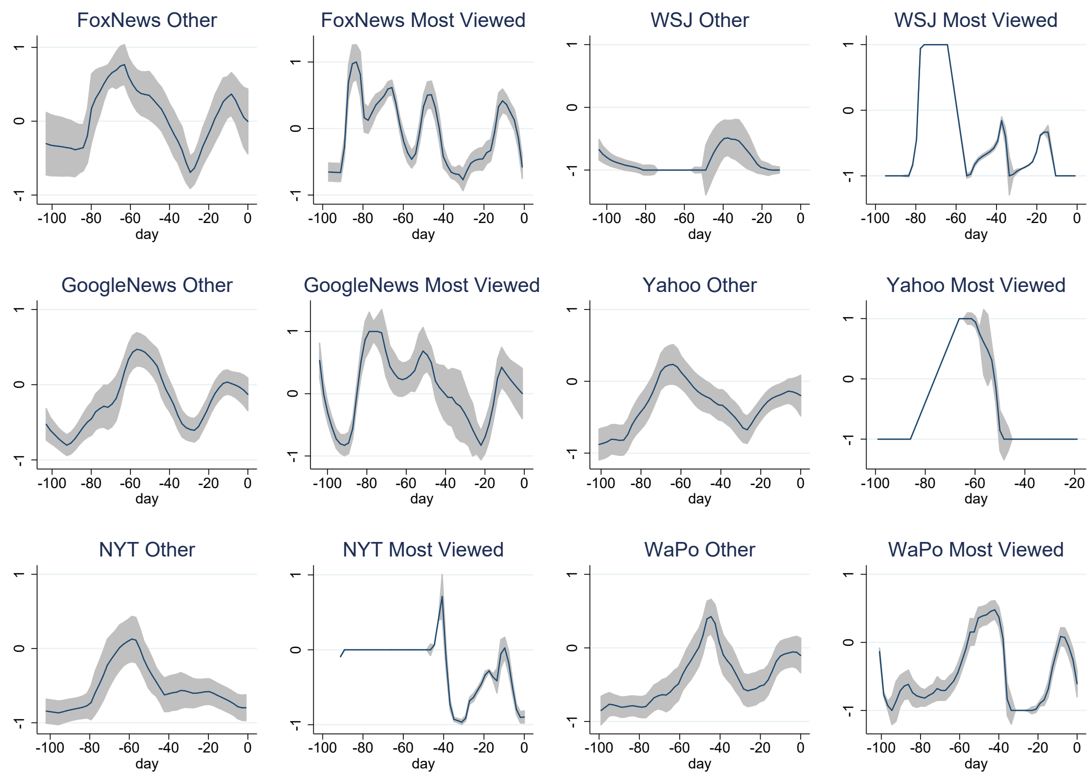
Figure 5: Mean $Slant_1$ versus day relative to election day (day 0) by outlet in 2012



Note: Curves are kernel-weighted smoothed local polynomials with 95% confidence interval bands. Positive values of *Slant* denote better chances of winning for the Republican candidate, whereas negative values indicate better chances for the Democratic candidate.

For 2012, the four outlets for which we have the most data—Fox, NYT, Yahoo, and HuffPost—all show pro-Republican bumps for "other" stories around 80 and 30 days prior to the election, consistent with Figure 3. The most viewed stories show the latter bump as well, while only Fox's most viewed stories reflects the earlier bump. The "most viewed" stories for HuffPost and WSJ all come from the last 30–40 days of the sample time-frame. For 2016, almost all outlets' most viewed and "other" news show pro-Republican bumps around 60 days prior to the election (50 days for WaPo), and also around 10 days prior to the election, with the NYT "other" stories being a notable exception. The pattern of slant in Fox's most viewed

Figure 6: Mean $Slant_1$ versus day relative to election day (day 0) by outlet in 2016



Note: Curves are kernel-weighted smoothed local polynomials with 95% confidence interval bands. Positive values of *Slant* denote better chances of winning for the Republican candidate, whereas negative values indicate better chances for the Democratic candidate.

stories was particularly volatile. The NYT most viewed graph also reveals a lack of data in general and in particular, early in our time-frame; this is because their most viewed list often included a link with variants on simple headlines along the lines of "2016 election polls". These headlines were not classified as horse race stories since they did not refer to the presidential race in particular and would almost surely be coded as neutral regardless.

4.2 Analysis

We first address prediction #1 of the horse race section of Table 1: that outlets may increase their supply of congenial news to satisfy consumer demand, either due to the reputation or the psychology mechanism. We do this by analyzing how the mean slant of stories varied across outlets. Table 5 presents results for simple regressions of $Slant_i$ (for each $i \in 1, 2, 3$) on outlet and day fixed effects, using the story-level data set. Here, we use Yahoo as the reference category, given the outlet's relative neutrality. The day fixed effects account for general trends affecting horse race story demand and supply—the content of these stories, such as the results of recent poll changes; the importance of those stories, which may be influenced by the timing relative to the election; and the importance of competing stories to fill the "news hole." The outlet coefficients thus represent the mean difference in slant for that outlet as compared to Yahoo given the mean slant of stories reported on that day across outlets. In a second version of each regression we include the difference in polls for the Republican and Democratic candidate for the day the story was first reported and the average change in this difference over the previous week as additional controls. Results are very similar with various alternative controls, such as a dummy for whether or not the story made the most popular list, or interactions of this dummy with day fixed effects.

As expected, in both years, there is evidence that Fox's stories were slanted to the right of the other outlets. The leftmost stories on average in 2012 were from the HuffPost. In 2016, the NYT was significantly to the left of Yahoo at the 5% level for two specifications, and WSJ was significantly to the left at 10% for one specification, while Google was significantly to the right at 5% for both $Slant_3$ models. However, these Google results (and lack of significant NYT results for these models) seem to be due to Yahoo's slant skewing to the left for the $Slant_3$ measure, as all of the other outlets' estimated slants are more conservative as compared to the other specifications.[15] The magnitudes of the effects are large: On a day on which Yahoo's expected $Slant_3$ is 0 in 2012, the expected $Slant_3$ for a Fox story is expected to be 0.795, i.e., there is a 79.5% chance the headline would be about good news for Republicans. In 2016, the expected value of $Slant_3$ for Fox for a neutral Yahoo story is 0.685.

As noted in the theory section, these results are consistent with both the psychology and reputation mechanisms, and we can shed some light on distinguishing them by examining

[15]When we use Google as the omitted outlet in 2016, both Fox's and the NYT's estimated slants are more stable and significant across specifications (see Table A8). In addition, we obtain results indicating that each of the other outlets aside from Fox are slanted to the left of Google. These results are precise and consistent for the WashPost, and imply a slant that is approximately half of that of the NYT. Of the six estimates for each outlet, three are significant (at at least 10%) for the WSJ, four are significant for WashPost, and two are significant for Yahoo, and all of the point estimates are negative.

whether slant declines as the election approaches, which is theoretically more likely for the case of reputation. In unreported results, we only find consistent evidence (across slant measures) that this is the case for Fox's slant in 2016, but not for any of the other slants (and in fact the NYT's and WashPost's increases over time in 2016).

Next, to address horse race predictions #2-4 from Table 1, that psychology would cause greater demand for stories with more congenial headlines within an outlet, and that reputation would cause greater demand for more atypical headlines within an outlet, respectively, we estimate the effect of within-outlet variation in slant on the probability of making the outlet's most viewed list. We do this by using linear probability models with day fixed effects, outlet fixed effects, and outlet-$Slant_i$ interactions. The coefficients on these interactions are the estimates of interest, as each represents the marginal effect of slant on the probability of being most viewed, for a given outlet. The outlet fixed effects account for differences in both the number of most viewed articles reported per outlet, and the total number of articles reported per outlet per day. We present results both with and without additional controls for the number of competing horse race headlines at the same time and from the same outlet.

Results are presented in Table 6. There is some evidence of a congeniality effect for Fox in 2012, but this is significant (at 5%) for $Slant_1$ only. The magnitudes of the $Slant_1$ estimates imply a one unit increase in slant caused a nearly 20 percentage point increase in being most viewed. There are no other significant results for that year. For 2016, there are a few marginally significant results for Fox and NYT, but each indicates that *less* congenial stories were the ones more likely to be most viewed. The significant effects for Fox are 13-19 percentage points per unit decrease in congenial slant, and for the NYT, 19-28 percentage points. The 2016 results only become stronger when we replace the day fixed effects with a date polynomial, while the 2012 Fox effects disappear in this case (these results are unreported). Of the explanations presented in Table 1, the 2016 results can be best explained by the reputation mechanism; i.e., stories with headlines with atypical slants appear in the most viewed lists because readers expect the stories to be particularly credible. The 2016 results could also be explained by the psychology mechanism, if the stories with headline slants that are atypical for the outlet are most viewed due to inordinate clicks by readers who are in the ideological minority for each outlet. By contrast, the 2012 Fox effects are only consistent with the psychology mechanism. A supply-side explanation for the 2016 results would be that Fox and the NYT mean horse race story slant "overshot" the partisanship of their readers' preferences in that year, perhaps unintentionally, or perhaps in an attempt to influence readers. Note that the standard errors for the insignificant estimates for the HuffPost, WashPost, and NYT in 2012 are all less than 0.1 (10 percentage points), implying reasonably good power. In unreported results, we examine specifications in which we split out slant into congenial and uncongenial effects for each outlet, and find no systematic patterns.

The most viewed data could be misleading if total website traffic changed depending on the congeniality of horse race news at the time. For example, suppose Republicans were less likely to visit foxnews.com on days when HR news for Republicans was less congenial, and Fox HR stories were indeed less congenial those days. Even if these stories were more likely to make the most viewed list than an HR story on a more congenial news day, it is possible these uncongenial stories received fewer total clicks than the more congenial stories. We do not have daily total website click data to directly address this issue, but can use other publicly available data to shed some light on it. Figure A.1 presents smoothed polynomials of Google Trends data on Google searches for "fox news" and "new york times" for the election seasons of 2012 and 2016. The curves all generally trend up over time, more sharply in the final two weeks before the election. There is no evidence of the trends differing between the two outlets in a given year, or of any correlation between these trends and the trends presented in earlier figures using poll and $Slant_i$ data. Thus, there is no evidence that Google searches are correlated with the congeniality of HR news on a given day. Google searches likely only account for a small fraction of total website traffic. Still, if total traffic was substantially correlated with congeniality we would expect to see some sign of this correlation in the search data as well.

Last, we address the final prediction of Table 1: that the quantity of horse race stories reported on particular days depends on the nature of horse race news that day. This prediction can be tested by using an outlet-level daily time series version of the data set. We use this to run a separate regression for each outlet with left-hand-side variable of number of horse race stories reported per day, and a measure of "true slant" for the day on the right-hand-side (the true nature of horse race news that day). Two such measures are used: 1) the average slant of stories reported by other outlets that day; 2) the Pollster average difference in polls (percent planning to vote Republican minus percent planning to vote Democrat) that day. Results are largely similar when we also include recent poll changes in the second specification, but cleaner to report when these are omitted. For the first measure, we also include a 4th order date polynomial to further control for general trends in horse race news interest. We do not do this for the second measure because the date polynomial is much more collinear with the poll average variable. For both measures, we control for the importance of horse race news on a given day using a 4th order polynomial of the total number of horse race stories reported by other outlets (we cannot use day fixed effects because there is only one observation per day). We use Poisson regressions because the left-hand side is a count variable; results are similar when we use OLS with Newey-West standard errors.

Results are reported in Table 7. For the Fox 2012 sample, one estimate is significant at 10%, and all of the estimates have signs consistent with the prediction. There are also two significant estimates consistent with the prediction for the NYT in 2016 (at 5% and 1%). These results provide some very modest additional support for the psychology mechanism. The strongest results, however, are for the WSJ: The majority of estimates in both years are significant at least at the 5%-level and indicate that the WSJ reported more horse race stories on days when news is *less* favorable to Republicans. This is inconsistent with the idea that the WSJ provides news that feels good to a predominantly conservative readership, and perhaps can be explained by the WSJ's news section being surprisingly liberal, consistent with a finding of Groseclose and Milyo (2005).

In summary, we find that Fox horse race news was slanted right (of the other outlets we study) in both 2012 and 2016, and that HuffPost stories in 2012, and NYT stories in 2016, were slanted left, and some evidence of the other outlets we study being slanted to the left of both

Google and Fox in 2016. These effects apply to both the content and quantity of headlines, though the evidence is stronger regarding the content. The content effects can be explained by the psychology or reputation mechanism, and the quantity effects are best explained by psychological factors. We find non-robust evidence that particularly right-slanted stories were more popular for Fox in 2012. We find somewhat more robust evidence that stories with slants that went against the grain for their outlets were more popular for Fox and the NYT in 2016. The 2012 effect is best explained by the psychology mechanism, and the 2016 effects by the reputation mechanism. We also find that the WSJ reported more horse race stories on days when news was more favorable to Democrats in both years. This result cannot be explained by our theoretical framework; a factor outside our framework that could explain it is that the WSJ's news staff is more liberal than the opinion staff, and the conservative reputation of the paper is based more on its opinion content.

5 Concluding remarks

Why do people demand congenially slanted news? We look for evidence to distinguish between three plausible mechanisms: instrumental value, reputation, and psychological value. Our results are complex, and yet consistent in some ways. We find both direct and indirect evidence that consumers prefer news for psychological reasons and because they perceive it as more informative (reputation). The direct evidence for the psychology mechanism is perhaps surprisingly weak. We fail to find evidence supporting the instrumental mechanism.

It is worth noting that the reputation mechanism may also relate to psychological factors like cognitive dissonance. Reputation is a function of trust, which is driven by psychological processes, which could be influenced by cognitive dissonance and confirmation bias affecting reactions to information observed in the past. These psychological factors are likely relevant given that reputations for a given outlet vary across the population. Thus, our results provide consistent support for the importance of psychology interpreted more broadly, and consistently fail to support theories of congenial news demand being driven by optimal instrumental information-seeking behavior. However, to be clear, the psychology and reputation mechanisms are still distinct. In particular, only the reputation mechanism can explain the reasonably robust demand for "bad news" that we observe.

The results indicating that mean slants were skewed for several outlets, both for all stories and for the most viewed stories of those outlets, imply that those outlets' readers were likely not exposed to accurate information about the candidates' chances throughout the campaigns. Although horse race reporting may seem innocuous, this distorted information may still have been harmful to social welfare, whether the distortions were driven directly by cognitive dissonance or motivated reasoning, or by reputational factors.[16] For example, distorted horse race reporting and news consumption could lead to distrust of election results and conspiracy theories.

Future work building on our paper could consider more detailed individual click-level web data, or other types of media data, such as social media data. And the complexity of our results demonstrates (or reminds us) that both media and reader behavior can vary substantially over time and across outlets. In 2016, Fox News may have moved to the right of its readers, and the New York Times may have further left. The Washington Post and New York Times have similar ideological reputations, but there seemed to be significant differences in their reporting. The Wall Street Journal's reporting appears quite distinct from its reputation (as opposed to that of Fox News and the New York Times). These results confirm the importance of considering context in interpretation, and the limitations for extrapolating results. While this is always true in empirical work, this point may be especially relevant to the analysis of media in recent years due to the fast-changing environment and uniqueness of the various outlets.

[16]Reputational factors could also be indirectly related to cognitive dissonance. If a partisan has a skewed prior about her preferred candidate's chance of winning, this is likely due to biased prior exposure and processing of information about the candidates' chance.

Table 5: Estimated mean differences in slant across outlets

Outlet	$Slant_1$	$Slant_2$	$Slant_3$	$Slant_1$	$Slant_2$	$Slant_3$
Panel A: 2012						
Fox	0.405***	0.515***	0.795***	0.327**	0.440***	0.671***
	(0.127)	(0.126)	(0.165)	(0.134)	(0.132)	(0.189)
WSJ	-0.150	-0.020	0.305	-0.392*	-0.276	0.016
	(0.227)	(0.210)	(0.303)	(0.223)	(0.207)	(0.330)
USAToday	-0.151	-0.023	0.062	-0.185	-0.067	0.039
	(0.176)	(0.179)	(0.222)	(0.192)	(0.198)	(0.236)
NYT	-0.187	-0.031	0.043	-0.258*	-0.141	-0.098
	(0.131)	(0.140)	(0.211)	(0.134)	(0.143)	(0.221)
HuffPost	-0.400***	-0.262**	-0.112	-0.444***	-0.262**	-0.146
	(0.103)	(0.112)	(0.167)	(0.110)	(0.120)	(0.182)
Poll controls?				✓	✓	✓
Adj. R^2	0.352	0.330	0.431	0.393	0.372	0.472
N	425	382	277	400	363	267
Panel B: 2016						
Fox	0.313*	0.342	0.685***	0.328*	0.389*	0.712***
	(0.172)	(0.222)	(0.245)	(0.173)	(0.219)	(0.248)
WSJ	-0.276	-0.353	0.152	-0.265	-0.379*	0.083
	(0.208)	(0.231)	(0.269)	(0.202)	(0.228)	(0.280)
Google	0.018	0.050	0.398**	0.055	0.119	0.431**
	(0.136)	(0.161)	(0.157)	(0.134)	(0.160)	(0.178)
NYT	-0.329**	-0.156	-0.024	-0.322**	-0.131	0.048
	(0.158)	(0.201)	(0.187)	(0.157)	(0.200)	(0.204)
WashPost	-0.159	-0.080	0.148	-0.109	0.001	0.194
	(0.149)	(0.174)	(0.171)	(0.148)	(0.170)	(0.193)
Poll controls?				✓	✓	✓
Adj. R^2	0.214	0.270	0.367	0.236	0.308	0.405
N	696	501	381	696	501	381

Note: OLS estimates, using story-level data. Left-hand side variable: $Slant_i$. The reference category for the outlet dummies is Yahoo. All models include day fixed effects. The poll controls include the Pollster daily average poll difference in the support for the Republican and Democratic candidate, as well as the average change in this difference over the previous week. The 2016 models also include dummies for Yahoo stories' first date occurring during one of two time-frames in which Yahoo data collection changed (see appendix). Standard errors are clustered by the first date the story was available. *, **, *** denote 10%, 5%, 1% significance.

Table 6: Probability of story making most viewed list

	$Slant_1$	$Slant_2$	$Slant_3$	$Slant_1$	$Slant_2$	$Slant_3$
Panel A: 2012						
Fox $\times Slant$	0.170**	0.096	0.116	0.172**	0.079	0.100
	(0.083)	(0.103)	(0.135)	(0.085)	(0.104)	(0.143)
WSJ $\times Slant$	-0.010	0.024	0.001	-0.042	-0.003	-0.055
	(0.202)	(0.210)	(0.186)	(0.195)	(0.201)	(0.180)
USAT $\times Slant$	-0.106	-0.087	-0.088	-0.110	-0.096	-0.118
	(0.088)	(0.115)	(0.111)	(0.091)	(0.120)	(0.114)
Yahoo $\times Slant$	-0.043	-0.038	-0.027	-0.027	-0.023	-0.027
	(0.043)	(0.053)	(0.064)	(0.040)	(0.047)	(0.056)
NYT $\times Slant$	-0.013	0.024	0.026	0.001	0.037	0.032
	(0.072)	(0.076)	(0.087)	(0.071)	(0.076)	(0.085)
HuffPost $\times Slant$	0.010	0.029	0.021	0.020	0.035	0.009
	(0.048)	(0.054)	(0.068)	(0.052)	(0.058)	(0.077)
Competing headline controls?				✓	✓	✓
	0.379	0.408	0.353	0.380	0.410	0.340
N	425	382	277	425	382	277
Panel B: 2016						
Fox $\times Slant$	-0.131*	-0.189**	-0.179*	-0.128*	-0.171*	-0.176
	(0.069)	(0.087)	(0.094)	(0.075)	(0.099)	(0.108)
WSJ $\times Slant$	-0.043	-0.187	0.389	-0.029	-0.091	0.463
	(0.244)	(0.523)	(0.293)	(0.247)	(0.546)	(0.285)
Yahoo $\times Slant$	-0.021	-0.009	0.089	-0.034	-0.061	0.022
	(0.087)	(0.116)	(0.141)	(0.093)	(0.137)	(0.209)
Google $\times Slant$	0.016	-0.026	0.009	0.023	-0.018	0.018
	(0.029)	(0.029)	(0.032)	(0.030)	(0.032)	(0.032)
NYT $\times Slant$	0.193*	0.284**	0.511	0.194*	0.314**	0.539
	(0.111)	(0.130)	(0.324)	(0.110)	(0.129)	(0.330)
WashPost $\times Slant$	-0.043	-0.013	0.019	-0.048	-0.046	-0.003
	(0.056)	(0.072)	(0.082)	(0.057)	(0.073)	(0.093)
Competing headline controls?				✓	✓	✓
Adj. R^2	0.286	0.378	0.371	0.286	0.392	0.371
N	696	501	381	696	501	381

Note: OLS estimates, using story-level data. Left-hand side variable: most viewed (yes/no). All models include day and outlet fixed effects. The 2016 models also include dummies for Yahoo stories' first date occurring during one of two time-frames in which Yahoo data collection changed (see appendix). Competing headlines controls are fixed effects for number of headlines with $Slant_1$ value for the outlet on first day that story was reported. Standard errors are clustered by the first date the story was available. *, **, *** denote 10%, 5%, 1% significance.

Table 7: Effects of slant and poll means on the daily number of horse race stories

	Fox	WSJ	NYT	Huff/WashPost
Panel A1: RHS = mean slant of other outlets (2012)				
$Slant_1$	0.251	-0.038	0.284	-0.075
	(0.195)	(0.359)	(0.207)	(0.203)
$Slant_2$	0.309	-0.740*	0.182	0.014
	(0.234)	(0.444)	(0.217)	(0.187)
$Slant_3$	0.202	-0.624*	0.224	-0.335**
	(0.207)	(0.349)	(0.195)	(0.169)
Panel A2: RHS = mean Republican poll advantage (2012)				
$Slant_1$	0.077	-0.584***	-0.067	0.051
	(0.094)	(0.161)	(0.092)	(0.076)
$Slant_2$	0.135	-0.437***	-0.027	0.037
	(0.089)	(0.164)	(0.085)	(0.075)
$Slant_3$	0.186*	-0.447***	-0.040	-0.006
	(0.103)	(0.156)	(0.083)	(0.090)
Panel B1: RHS = mean slant of other outlets (2016)				
$Slant_1$	-0.091	-0.417	-0.367	0.224
	(0.217)	(0.559)	(0.281)	(0.222)
$Slant_2$	-0.198	-1.078**	-0.787***	0.216
	(0.240)	(0.432)	(0.279)	(0.235)
$Slant_3$	-0.146	-1.206**	-0.609**	0.233
	(0.237)	(0.492)	(0.255)	(0.194)
Panel B2: RHS = mean Republican poll advantage (2016)				
$Slant_1$	0.204**	-0.335**	-0.155	0.107*
	(0.081)	(0.135)	(0.100)	(0.062)
$Slant_2$	0.134	-0.363**	-0.156	0.122*
	(0.085)	(0.179)	(0.096)	(0.071)
$Slant_3$	0.091	-0.341**	-0.128	0.076
	(0.102)	(0.162)	(0.136)	(0.090)

Note: Poisson regressions, using daily outlet-level time series. Left-hand side variable: number of horse stories of the outlet listed in the column header. Mean $Slant_i$ of other outlets equals 0 if no horse race stories of type i are available on a given day. All models include the 4th order polynomial of the number of horse race stories reported by all other outlets; models in Panels A2 and B2 also a 4th order date polynomial. N=103 and 101 for Panels A1 and A2, and N=105 for Panels B1 and B2. *, **, *** denote 10%, 5%, 1% significance.

References

E. Bakshy, S. Messing, and L. Adamic. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348:1130–1132, 2015.

P. Barberá and G. Sood. Follow your ideology: A measure of ideological location of media sources. In *MPSA Conference*, 2014.

A. J. Berinsky, G. A. Huber, and G. S. Lenz. Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, 20(3):351–368, 2012.

D. Bernhardt, S. Krasa, and M. Polborn. Political polarization and the electoral effects of media bias. *Journal of Public Economics*, 92(5):1092–1104, 2008.

C. Budak, S. Goel, and J. M. Rao. Fair and balanced? Quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly*, 80(S1):250–271, 2016.

M. Buhrmester, T. Kwang, and S. Gosling. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1):3–5, 2011.

J. Burke. Primetime spin: Media bias and belief confirming information. *Journal of Economics & Management Strategy*, 17(3):633–665, 2008.

J. Chan and W. Suen. A spatial theory of news consumption and electoral competition. *Review of Economic Studies*, 75(3):699–728, 2008.

J. Ely, A. Frankel, and E. Kamenica. Suspense and surprise. *Journal of Political Economy*, 123(1):215–260, 2015.

R. Y. Fang. Profit-maximizing media bias. *SSRN Working Paper*, 2016.

S. Flaxman, S. Goel, and J. M. Rao. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1):298–320, 2016.

M. Garz, J. Sörensen, and D. F. Stone. Congeniality and news consumption: Evidence from Facebook. Working paper, 2018.

M. Gentkzow, J. Shapiro, and D. Stone. Media bias in the marketplace: Theory. In S. Anderson, J. Waldfogel, and D. Stromberg, editors, *Handbook of media economics*, pages 623–645. Elsevier, 2015.

M. Gentzkow and J. Shapiro. Media bias and reputation. *Journal of Political Economy*, 114 (2):280–316, 2006.

M. Gentzkow and J. Shapiro. What drives media slant? Evidence from US daily newspapers. *Econometrica*, 78(1):35–71, 2010.

R. Golman, D. Hagmann, and G. Loewenstein. Information avoidance. *Journal of Economic Literature*, 55(1):96–135, 2017.

T. Groseclose and J. Milyo. A measure of media bias. *The Quarterly Journal of Economics*, 120(4):1191–1237, 2005.

Y. Halberstam and B. Knight. Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter. *Journal of Public Economics*, 143:73–88, 2016.

C. Huff and D. Tingley. "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics*, 2(3):1–12, 2015.

S. Iyengar and K. S. Hahn. Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, 59(1):19–39, 2009.

N. Karlsson, G. Loewenstein, and D. Seppi. The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty*, 38(2):95–115, 2009.

Y. Masatlioglu, A. Y. Orhun, and C. Raymond. Preferences for non-instrumental information and skewness. Working paper, 2016.

M. J. Metzger, E. H. Hartsell, and A. J. Flanagin. Cognitive dissonance or credibility? A comparison of two theoretical explanations for selective exposure to partisan news. *Communication Research*, forthcoming, 2015.

J. M. Miller and J. A. Krosnick. News media impact on the ingredients of presidential evaluations: Politically knowledgeable citizens are guided by a trusted source. *American Journal of Political Science*, 44(2):301–315, 2000.

S. Mullainathan and A. Shleifer. The market for news. *The American Economic Review*, 95 (4):1031–1053, 2005.

K. Mullinix, T. Leeper, J. Druckman, and J. Freese. The generalizability of survey experiments. *Journal of Experimental Political Science*, 2(2):109–138, 2015.

S. Oliveros and F. Várdy. Demand for slant: How abstention shapes voters' choice of news media. *The Economic Journal*, 125(587):1327–1368, 2015.

E. Pariser. *The filter bubble: What the Internet is hiding from you*. Penguin, 2011.

E. Peterson, S. Goel, and S. Iyengar. Echo chambers and partisan polarization: Evidence from the 2016 presidential campaign. Working paper, 2017.

Pew. Political polarization & media habits. Pew Research Center, October 2014.

R. Puglisi and J. M. Snyder Jr. Empirical studies of media bias. In *Handbook of media economics*, volume 1, pages 647–667. Elsevier, 2015.

K. Searles, M. H. Ginn, and J. Nickens. For whom the poll airs: Comparing poll results to television poll coverage. *Public Opinion Quarterly*, 80(4):943–963, 2016.

G. Sood and Y. Lelkes. Dont expose yourself: Discretionary exposure to political information. Working paper, 2016.

D. Stone. Ideological media bias. *Journal of Economic Behavior & Organization*, 78(3): 256–271, 2011.

N. J. Stroud. *Niche news: The politics of news choice*. Oxford University Press, 2011.

M. Tremayne. Partisan media and political poll coverage. *Journal of Information Technology & Politics*, 12(3):270–284, 2015.

R. P. Vallone, L. Ross, and M. R. Lepper. The hostile media phenomenon: Biased perception and perceptions of media bias in coverage of the Beirut massacre. *Journal of personality and social psychology*, 49(3):577–585, 1985.

A Supplemental materials

Table A1: Summary statistics for survey full sample

Category	Variable (all 0/1)	Mean
Party	Democratic	0.384
	Lean Dem.	0.124
	Independent (no lean)	0.158
	Lean Rep.	0.180
	Republican	0.154
Preferred candidate	Trump	0.261
	Clinton	0.512
	Not voting/other	0.228
Education	Some HS	0.006
	HS degree	0.121
	Some college	0.227
	2 yr degree	0.113
	4 yr coll. degree	0.415
	$>$ college	0.118
Gender	Female	0.458
	Male	0.542
Age	18-29	0.326
	30-39	0.382
	40-49	0.160
	50-64	0.113
	$\geq$ 65	0.019

Note: N=638 for all variables (N=226 from survey 1, 216 from survey 2, 196 from survey 3) except preferred candidate and gender (N=637). Respondents are master MTurkers who answered reading check question correctly.

Table A2: Headline options for each survey

Survey 1:	New York Times: Commentators Give Hillary Clinton Edge in Debate
	Fox News: Hillary won the first debate (it helps to be prepared)
	Yahoo News: Long dog-gone trip: Florida pooch travels to Boston and back
	Yahoo News: Houston gunman had two weapons, thousands of rounds at scene.
Survey 2:	New York Times: Who Won the Debate? Commentators Give Edge to Mike Pence
	Fox News: Pence triumphs in VP debate. And then there was the night's biggest loser...
	Yahoo News: 2 Vermont teachers accused of vandalizing sidewalk
	Yahoo News: Two young girls shot in Cleveland drive-by shooting
Survey 3:	New York Times: Who Won the Debate? Donald Trump Avoids Annihilation
	Fox News: Trump comes out swinging and wins second debate
	Yahoo News: Three police officers shot in Palm Springs, California
	Yahoo News: Record 1,201 couples renew wedding vows in Kalamazoo

Note: The instructions given to survey respondents before being asked to choose a headline were: "Choose one of the following articles to read. You should choose the article that you are more interested in − the one you would be more likely to read if you came across these links simply surfing the web, or on a social network, etc. After making your choice, you will have access to the article and a simple question on the article's content. Your payment will be $0.25 higher if you answer the question correctly. The question's difficulty is the same for each article. Thus, you might as well choose the article you are truly more interested in, as it will be more enjoyable to read, and you will be just as likely (or more likely) to get the extra payment." The median work time was approximately four minutes and so our payment, as an hourly rate, was relatively high for an MTurk task, which typically pay at rates less than $5 per hour.

Table A3: Multinomial logit survey analysis results

	(1)	(2)
Panel A: Party identity		
Dem. congeniality	0.676**	
	(0.269)	
Rep. congeniality	0.229	
	(0.291)	
NYT × Dem. congeniality		0.388
		(0.351)
Fox × Dem. congeniality		0.983***
		(0.366)
NYT × Rep. congeniality		0.043
		(0.407)
Fox × Rep. congeniality		0.209
		(0.378)
Panel B: Candidate preferences		
Clinton congeniality	0.656**	
	(0.256)	
Trump congeniality	0.579**	
	(0.282)	
NYT × Clinton congeniality		0.306
		(0.339)
Fox × Clinton congeniality		1.035***
		(0.345)
NYT × Trump congeniality		0.323
		(0.404)
Fox × Trump congeniality		0.591*
		(0.356)

Note: All models estimated on respondent-news alternative level data set (three alternatives, Fox, NYT, and other, per respondent, and so 1,911 total observations). "Dem. congeniality" = one for Democrats for both the NYT and Fox alternatives for survey 1, equals -1 for both NYT and Fox stories for survey 2, and equals -1 for the Fox story for survey 3, and 0 otherwise (for NYT survey 3 and for all "other" observations). "Clinton congeniality" takes the same values for those alternatives for Clinton supporters, and "Rep./Trump congeniality" take opposite signs for Republicans and Trump supporters, respectively. Models estimated with alternative-specific conditional logit with left-hand side variable of news choice (Fox, NYT or other), and right-hand side respondent (case)-level variables of party/party-strength, education, age, gender and survey fixed effects. *, **, *** denote 10%, 5%, 1% significance.

Table A4: Instructions for rating slant of horse race headlines

Category	Definition
Very good news for Clinton	"Very good news" about Clinton's chances of winning (and very bad news about Trump's chances) – that Clinton is substantially more likely to win, or that her chances have greatly improved recently
Good news for Clinton	"Good news" about Clinton's chances of winning (and bad news about Trump's chances) – that Clinton is more likely to win, or that her chances have improved recently
No change	No change in either candidate's chance
Good news for Trump	"Good news" about Trump's chances of winning (and bad news about Clinton's chances) – that Trump is more likely to win, or that his chances have improved recently
Very good news for Trump	"Very good news" about Trump's chances of winning (and very bad news about Clinton's chances) – that Trump is substantially more likely to win, or that his chances have greatly improved recently
Ambiguous or unclear	Relevant to the chances of one of the candidates winning, but unable to determine which candidate is being favored (if at all)
Not relevant	This headline does not seem to be about the candidates' chances of winning the election

Note: MTurkers were given the following instructions: "The following items are real headlines of reports of major US news outlets from July-November 2016 about the upcoming presidential election that year between Hillary Clinton and Donald Trump. Most, but not all, of these articles are about information about which candidate is more likely to win the election. How do you think a typical reader would perceive these headlines? [using the categories above] We would like for you to evaluate these 40 headlines. We will look over your evaluations and if we believe you have done them carefully and reasonably, we will invite you, by email, to do additional similar work (potentially quite a lot)." For 2012 headlines the references were to Obama and Romney.

Table A5: Outlet-slant interactions as predictors of 2012 polls and poll changes

	(1)	(2)	(3)	(4)
Fox $\times Slant_1$	0.835***	0.321**	0.664***	0.227*
	(0.206)	(0.152)	(0.141)	(0.128)
WSJ $\times Slant_1$	0.838*	0.744**	0.578*	0.340
	(0.477)	(0.336)	(0.320)	(0.266)
USA $\times Slant_1$	0.446**	0.544***	0.615***	0.050
	(0.208)	(0.180)	(0.178)	(0.117)
Yahoo $\times Slant_1$	0.752***	0.293**	0.543***	0.174*
	(0.176)	(0.133)	(0.112)	(0.096)
NYT $\times Slant_1$	0.402*	0.812***	0.520***	0.595***
	(0.210)	(0.178)	(0.131)	(0.116)
HuffPost $\times Slant_1$	0.334**	0.531***	0.316***	0.268***
	(0.148)	(0.148)	(0.118)	(0.097)
Adj. R^2	0.106	0.186	0.582	0.556
N	400	400	400	400
Fox $\times Slant_2$	0.777***	0.301*	0.593***	0.137
	(0.196)	(0.160)	(0.144)	(0.130)
WSJ $\times Slant_2$	0.904*	0.771**	0.547*	0.337
	(0.484)	(0.325)	(0.326)	(0.262)
USA $\times Slant_2$	0.516**	0.798***	0.757***	0.141
	(0.227)	(0.157)	(0.180)	(0.101)
Yahoo $\times Slant_2$	0.775***	0.273**	0.546***	0.123
	(0.206)	(0.126)	(0.132)	(0.095)
NYT $\times Slant_2$	0.494**	0.837***	0.564***	0.581***
	(0.217)	(0.170)	(0.135)	(0.105)
HuffPost $\times Slant_2$	0.451***	0.505***	0.387***	0.279***
	(0.161)	(0.165)	(0.113)	(0.104)
Adj. R^2	0.109	0.190	0.607	0.577
N	363	363	363	363
Fox $\times Slant_3$	0.754***	0.233	0.507***	0.074
	(0.196)	(0.164)	(0.160)	(0.141)
WSJ $\times Slant_3$	0.935**	0.703**	0.536*	0.306
	(0.443)	(0.320)	(0.285)	(0.264)
USA $\times Slant_3$	0.516**	0.772***	0.766***	0.203**
	(0.214)	(0.167)	(0.181)	(0.096)
Yahoo $\times Slant_3$	0.853***	0.310**	0.575***	0.151
	(0.209)	(0.127)	(0.140)	(0.098)
NYT $\times Slant_3$	0.489**	0.833***	0.539***	0.600***
	(0.230)	(0.163)	(0.147)	(0.109)
HuffPost $\times Slant_3$	0.466***	0.549***	0.413***	0.309***
	(0.169)	(0.174)	(0.118)	(0.116)
Adj. R^2	0.147	0.256	0.632	0.592
N	267	267	267	267

Note: OLS estimates, using story-level data. Dependent variables: Pollster daily average poll difference in the support for the Republican and Democratic candidate (columns 1 and 3); change in this difference over the past week (columns 2 and 4). The models in columns (3) and (4) include a third-order date polynomial. Standard errors are clustered by the first date the story was available. *, **, *** denote 10%, 5%, 1% significance.

Table A6: Outlet-slant interactions as predictors of 2016 polls and poll changes

	(1)	(2)	(3)	(4)
Fox $\times Slant_1$	0.467**	0.469***	0.183	0.172
	(0.192)	(0.149)	(0.137)	(0.117)
WSJ $\times Slant_1$	0.613*	0.350	0.333	0.017
	(0.327)	(0.241)	(0.273)	(0.183)
Yahoo $\times Slant_1$	0.735***	0.419**	0.355**	0.013
	(0.159)	(0.183)	(0.146)	(0.106)
Google $\times Slant_1$	0.776***	0.361**	0.451***	-0.002
	(0.145)	(0.144)	(0.101)	(0.068)
NYT $\times Slant_1$	0.638**	0.362	0.434***	0.214
	(0.276)	(0.261)	(0.139)	(0.173)
WashPost $\times Slant_1$	0.531***	0.023	0.436***	-0.086
	(0.128)	(0.133)	(0.106)	(0.077)
Adj. R^2	0.160	0.045	0.499	0.591
N	696	696	696	696
Fox $\times Slant_2$	0.641***	0.343**	0.376**	0.106
	(0.231)	(0.172)	(0.152)	(0.138)
WSJ $\times Slant_2$	0.732**	0.336*	0.563**	0.099
	(0.364)	(0.198)	(0.263)	(0.183)
Yahoo $\times Slant_2$	0.888***	0.443**	0.399**	-0.024
	(0.175)	(0.219)	(0.160)	(0.117)
Google $\times Slant_2$	0.866***	0.360**	0.453***	-0.053
	(0.156)	(0.161)	(0.114)	(0.073)
NYT $\times Slant_2$	0.717**	0.388	0.464***	0.211
	(0.286)	(0.311)	(0.142)	(0.178)
WashPost $\times Slant_2$	0.679***	0.064	0.494***	-0.103
	(0.173)	(0.129)	(0.123)	(0.107)
Adj. R^2	0.210	0.044	0.580	0.574
N	501	501	501	501
Fox $\times Slant_3$	0.694***	0.313*	0.453***	0.056
	(0.232)	(0.170)	(0.160)	(0.139)
WSJ $\times Slant_3$	0.633*	0.283	0.501*	0.094
	(0.356)	(0.196)	(0.265)	(0.186)
Yahoo $\times Slant_3$	0.834***	0.355	0.429**	-0.081
	(0.180)	(0.235)	(0.169)	(0.128)
Google $\times Slant_3$	0.807***	0.326**	0.460***	-0.061
	(0.152)	(0.153)	(0.109)	(0.073)
NYT $\times Slant_3$	0.673**	0.365	0.492***	0.210
	(0.288)	(0.312)	(0.138)	(0.180)
WashPost $\times Slant_3$	0.641***	0.025	0.498***	-0.120
	(0.176)	(0.130)	(0.127)	(0.112)
Adj. R^2	0.229	0.036	0.568	0.591
N	381	381	381	381

Note: OLS estimates, using story-level data. Dependent variables: Pollster daily average poll difference in the support for the Republican and Democratic candidate (columns 1 and 3); change in this difference over the past week (columns 2 and 4). The models in columns (3) and (4) include a third-order date polynomial. Standard errors are clustered by the first date the story was available. *, **, *** denote 10%, 5%, 1% significance.

Table A7: *Slant* values for headlines from November 7, 2016

Headline	$Slant_1$	$Slant_2$	$Slant_3$	Outlet	MV
polls trump and clinton virtually tied in key swing states	0.00	0.00		Fox	1
momentum buster? fbi's comey tells congress email review completed decision not to prosecute clinton stands	-0.50			Fox	0
trump supporters say they feel michigan momentum	0.67	0.67		Google	0
president obama makes closing argument if we win florida its a wrap	-0.50			Google	0
poll on eve of election day clinton maintains her edge over trump	-1.00	-1.00	-1.00	Google	0
iowa poll trump opens 7point lead over clinton	1.00	1.00	1.00	Google	0
us presidential election live countdown to the polls	0.00			Google	1
live blog last updated 7.30am aest us election live trump clinton in final pitch to voters latest polls	0.00			Google	0
obama if clinton wins florida she will win the election	-0.50			Google	0
our final map has clinton winning with 352 electoral votes. compare your picks with ours.	-1.00	-1.00	-1.00	Google	0
clinton has solid lead in electoral college trumps winning map is unclear	-1.00	-1.00	-1.00	NYT	1
trump and clinton tied in final upshot poll of north carolina	0.00	0.00		NYT	1
clinton cleared on new emails keeps small lead in polls	-1.00	-1.00	-1.00	NYT	0
clinton leads trump by 4 points in lates poll	-1.00	-1.00	-1.00	WSJ	1
1 hillary clinton has enough electoral votes to win the white house in final fix map	-1.00			WashPost	1
postabc tracking poll clinton 47 trump 43 on election eve	-1.00	-1.00	-1.00	WashPost	1
amid lastminute push in va. clinton holds 6point lead in latest poll	-1.00	-1.00	-1.00	WashPost	0
trump urges voters to deliver justice at polls	0.50			WashPost	0

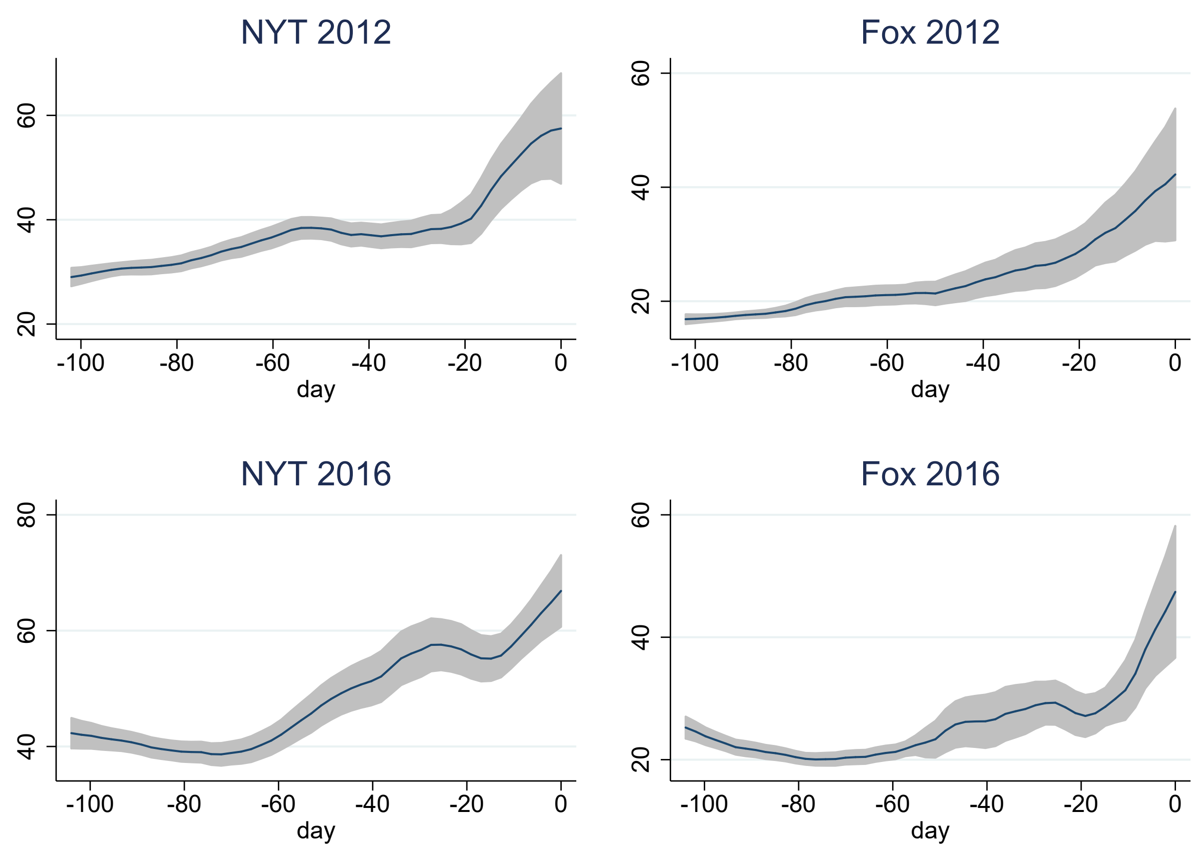
Note: MV = most viewed.

Table A8: Estimated mean differences in slant across outlets in 2016 with Google as reference outlet

Outlet	$Slant_1$	$Slant_2$	$Slant_3$	$Slant_1$	$Slant_2$	$Slant_3$
Fox	0.295**	0.292*	0.288	0.274**	0.270*	0.280
	(0.117)	(0.157)	(0.183)	(0.116)	(0.154)	(0.174)
WSJ	-0.293	-0.402**	-0.245	-0.320*	-0.498***	-0.348
	(0.190)	(0.191)	(0.240)	(0.187)	(0.184)	(0.233)
Yahoo	-0.018	-0.050	-0.398**	-0.055	-0.119	-0.431**
	(0.136)	(0.161)	(0.157)	(0.134)	(0.160)	(0.178)
NYT	-0.347***	-0.206	-0.422***	-0.377***	-0.250**	-0.383***
	(0.108)	(0.134)	(0.136)	(0.100)	(0.122)	(0.134)
WashPost	-0.177**	-0.129	-0.249**	-0.164*	-0.118	-0.237*
	(0.084)	(0.094)	(0.125)	(0.083)	(0.094)	(0.123)
Poll controls?				✓	✓	✓
Adj. R^2	0.214	0.270	0.367	0.236	0.308	0.405
N	696	501	381	696	501	381

Note: OLS estimates, using story-level data. Left-hand side variable: $Slant_i$. The reference category for the outlet dummies is Google. All models include day fixed effects. The poll controls include the Pollster daily average poll difference in the support for the Republican and Democratic candidate, as well as the average change in this difference over the previous week. The models also include dummies for Yahoo stories' first date occurring during one of two time-frames in which Yahoo data collection changed. Standard errors are clustered by the first date the story was available. *, **, *** denote 10%, 5%, 1% significance.

Figure A.1: Google Trends data (day 0 = election day)



Note: Curves are kernel-weighted smoothed local polynomials with 95% confidence interval bands. “NYT” = Google searches for “new york times”; “Fox” = Google searches for “fox news”.