Received: 21 January 2016 | Revised: 6 January 2017 | Accepted: 19 January 2017

DOI: 10.1111/jems.12207

ORIGINAL ARTICLE

WILEY Journal of Economics & Management Strategy

Content aggregation by platforms: The case of the news media

Lesley Chiou[1] | **Catherine Tucker**[2,3]

[1]Economics Department, Occidental College, CA (Email: lchiou@oxy.edu)

[2]MIT Sloan School of Management, MIT, Cambridge, MA (Email: cetucker@mit.edu)

[3]National Bureau of Economic Research

Abstract

The digitization of content has led to the emergence of platforms that draw information from multiple sources. This paper investigates whether aggregation of content by a single platform encourages users to "skim" content or to investigate it in depth. We study a contract dispute that led a major aggregator to remove information from a major content provider. After the removal, users were less likely to investigate additional, related content in depth, particularly sources that were horizontally or vertically differentiated.

1 | INTRODUCTION

In recent years, the digitization of content has led to the prominence of platforms as aggregators of content in many economically important industries, including media and Internet-based industries (Evans & Schmalensee, 2012). These new platforms consolidate content from multiple sources into one place, thereby lowering the transaction costs of obtaining content and introducing new information to consumers. Although an extensive literature focuses on pricing and piracy by platforms (Danaher, Dhanasobhon, Smith, & Telang, 2010; Oberholzer-Gee & Strumpf, 2007; Rob & Waldfogel, 2006), little is known about how the quantity and quality of content provided by a platform influences consumer search.

We examine how a change in content provided by a platform affects subsequent consumer search for different types of information and its consequences for content providers. For identification, we exploit a contract dispute as an exogenous shifter of content on a major news aggregator, Google News. In January 2010, after a breakdown in licensing negotiations, Google removed all news articles that were syndicated by a major content provider, The Associated Press (AP), from its news aggregator (Haddad, 2010). These articles were typically shortened versions of stories that appeared in a select number of newspapers associated with The AP.

Our setting presents an attractive opportunity to study these issues for several reasons. First, content aggregation is a prominent concern in the news media industry due to the rapid growth of digital news content, multiple content providers who have expressed fears about the viability of their business model in the presence of aggregators, and the prominence of platforms such as Google News. Second, our experiment studies a large shock as we focus on arguably two of the largest players in U.S. news media industry, Google News and The Associated Press. Google News is among the most-read news aggregators, automating the aggregation of news content from 25,000 news sources. The Associated Press is a prominent content provider formed in 1846 to fund news-gathering activities between its newspaper participants. The Associated Press has received 51 Pulitzer Prizes, reflecting its status in the news industry and its investments in journalism. Finally, due to the scale of The Associated Press, the removal of content most likely did not reflect the wishes of a particular newspaper or any self-selection. The AP formed historically with deep ties to the news media, so the content removal of The AP captures a broader effect of content change for news media.

Earlier versions of this paper were circulated under the titles "Copyright, Digitization, and Aggregation" and "News and Online Aggregators." We thank Robert Seamans for useful comments. We thank Christopher Hafer of Experian Hitwise. We also thank Shreya Bhaskaran, Cassandra Crosby, Sara Mcknight, Anthony Quach, and Yueh Yue for excellent research assistance. Financial support from the NBER Innovation Policy Group and NET Institute (www.NETinst.org) is gratefully acknowledged.

In particular, we empirically test for two theoretical effects. We first look for a "scanning effect" that may exist where consumers use platforms to scan the extracts of content without clicking through to pursue more in-depth material. To empirically test for this, we collect data on monthly visits and search intensity and examine whether overall traffic to Google News shifted after the contract dispute relative to Yahoo! News, which continued to provide The Associated Press content during this period. We do not find evidence of a scanning effect, as overall traffic to Google News and Yahoo! News remained relatively comparable during our time period. Our results suggest that the reduction in quality may not affect overall traffic whereas prior work suggests that other changes, such as increasing customization of content, can increase traffic by as much as 50% (Athey & Mobius, 2012).

Second, a "traffic effect" may exist where consumers use platforms to explore new material in more depth. To empirically test for a traffic effect, we compare subsequent website visits by users of Google News before and after the contract dispute relative to users of Yahoo! News, which continued to provide The Associated Press content during this period. Our results indicate that after The Associated Press content was removed from Google News, 28% fewer users subsequently visited news sites after navigating to Google News relative to users who used Yahoo! News. The pattern was driven by news websites with local content or news websites with national recognition as high quality. Thus, we find evidence that the traffic effect is large, as aggregators may guide users to new content. Our results are comparable in magnitudes to concurrent and recent work that finds that technological adoption can increase consumers' consumption of content by 26–30% (Athey & Mobius, 2012; Xu, Forman, Kim, & Van Ittersum, 2014). Back-of-the-envelope calculations suggest that the removal of content potentially led to a decrease of 110 million visits each month from Google News to news media websites hosted in the United States. We also explore the institutional relationship between news sites and The AP and find that websites with stronger ties to The AP suffered a drop in traffic after the dispute.

Our results inform legal and public policy debates. Recent regulation in the European Union attempts to make content on a platform an "opt-in" decision (Eddy, 2013; Pfanner, 2012), where a content provider has the right to decide whether or not their content appears on the aggregation platform. Our results suggest that the decision to opt-in to an aggregation platform should depend on whether the content provider is considered high quality or highly unusual. Both of these characteristics appear to encourage users to explore content more deeply instead of scanning content. One surprising development is that despite German publishers lobbying for an opt-in law, none have chosen to opt-out (Lomas, 2013). Our paper provides an explanation of such behavior—ultimately aggregators may benefit many newspapers, especially high-quality ones, and the purpose of the opt-in provision may be to increase bargaining power over payments to news providers rather than an actual desire by copyright holders to opt-out of the aggregator.

More broadly, our study is related to prior work that describes how digital technologies affect search costs and generate spillovers (Bakos, 1997; Ghose, Goldfarb, & Han, 2011; Greenstein, 2011; Shapiro & Varian, 1999). The novelty of our study is that we are the first to explore how access to different types of digital content affects information gathered by consumers. Note that the focus of our study is not how aggregators affect direct navigation to the content providers' websites—Sandoval (2009), Arrington (2010), and Athey and Mobius (2012) discuss this. Instead, we measure how a platform's expansion or contraction of content affects subsequent navigation by users.

Our results have implications for copyright policy regarding platforms that aggregate digital content. The digital revolution challenges various aspects of copyright protection (Greenstein, Lerner, & Stern, 2011), but much of the focus has been on peer-to-peer piracy rather than newer legal models of business that aggregate specific types of content. Online aggregators in media assert that their practice is protected by copyright law because they only display small extracts of information, and often this information is factual (Isbell, 2010). Our empirical distinction between a scanning effect where the aggregator substitutes for original content and a traffic effect where the aggregator is complementary, is useful for analyzing the potential policy implications of such business models. The fact that we find evidence of a traffic effect, even with a relatively large amount of content on an aggregator, is perhaps evidence that the "fair use" exemptions are less potentially damaging to the original copyright holder than is often thought.

2 | INSTITUTIONAL SETTING

Google News is ranked as the fifth most visited news website by Hitwise. Receiving 2.90% of all news site visits, Google News is the second most popular news aggregator service after Yahoo! News, which received 7.09% of all news site visits. Google News electronically aggregates different news sources based upon a proprietary algorithm. As of December 2009, Google News claimed to receive news content from 25,000 publishers across the world and to send one billion clicks to these publishers every month (Cohen, 2009). Figure 1 provides a screenshot of Google News. Google News has two noticeable features that distinguish

FIGURE 1 Screenshot of Google News [Color figure can be viewed at wileyonlinelibrary.com]
Notes: Wayback Machine (https://archive.org/web/) July 15, 2009.

TABLE I Timeline of negotiations between Google and The Associated Press

Date	Event
August 2006	Google and The Associated Press first sign contract to enable The Associated Press content to appear on Google News for 30 day window.
December 24, 2009	The Associated Press content no longer appears on Google. Industry press speculates that this is in preparation for the expiration of contract between The Associated Press and Google in one month's time.
End January 2010	The Associated Press and Google contract set to expire.
February 2010	The Associated Press content returns to Google News.

it from traditional news sites. First, a variety of sources is listed for each story. Second, the order of news is electronically determined based upon users' preferences, the recency of the story, and the interest it receives from other users.

The Associated Press (AP), founded in 1846, is one of the largest news agencies in the world. Since the demise of United Press International, The AP is the only national news service in the United States, and its major competitors are Reuters (based in the United Kingdom) and Agence-France Presse (based in France). The AP is a cooperative owned by various newspapers and radio and television stations in the United States. These stakeholders both contribute stories to The AP and use material written by The AP staff journalists. During the past decade, The AP has been at the forefront of efforts by copyright holders to circumscribe "fair use" for digital content and to protect copyholders' rights. For example, in June 2008, The AP invoked the Digital Millennium Copyright Act and insisted that various bloggers remove The AP content (Ardia, 2008).

The origins of The AP and its business model reveal that The AP enabled newspapers to pool content and stories in the old media world of physical newspapers and to hence enjoy economies of scale for news reporting in response to the new telegraph technology. Little evidence exists that The AP has tried to push its own website as an alternative "news-wire" service; instead, The AP website functions mainly as a corporate site which simply lists member newspapers. The AP's reluctance to perform a news-wire role may be due to its origins as a newspaper association; it may be reluctant to compete directly with a newspaper's business model. It is not clear how an organization founded under the traditional model where each newspaper provided full news coverage to individual print subscribers fits into a world where consumers consume news digitally. Table I, which summarizes the major events of The AP and Google relationship, makes clear that The AP is worried about the rise of search and aggregation technology for its business model. The Associated Press attempts to grapple with the shift to digital content. As described by Halberstam (2007), in 2005, the CEO of the Associated Press stated that

"Advertising is following the migrating eyeballs, and new distribution networks are requiring us to rethink how our content reaches consumers."

Because both The AP and Google News are key players in the distribution of news online, it is not surprising they forged a partnership. Their licensing agreement also protects Google News from allegations of copyright infringement over The AP content, given the current uncertainty over copyright law for aggregators. We study a discontinuity in this relationship, surrounding negotiations of contract renewal at the end of January 2010. As part of their existing contract, Google and The AP agreed that The AP content could be hosted by Google for a period of 30 days. Therefore, if the contract ended in January 2010 and was not renewed, Google would stop posting new content from The Associated Press 30 days prior to the end of the contract. Presumably to make this "clean break" a credible outside option, Google did indeed stop posting content for seven weeks during these contract negotiations (Krazit, 2010). We should emphasize that our discussion is necessarily based upon the observations of industry outsiders, because both Google and The AP signed binding non-disclosure agreements that prevent them from ever commenting on the course or outcome of negotiations (Sullivan, 2010).

The removal of The AP content represents a useful quasi-experiment. Because the removal of content was provoked by the intricacies of contract negotiations, its timing can be thought of as reasonably exogenous, as the removal was determined by the expiration of the contract rather than any considerations of the popularity (or lack thereof) of The AP content at that time. As detailed in Table I, Google removed The AP content from December 23, 2009 until sometime in February 2010. Fortunately for our purposes, Yahoo! News continued to host The AP content without interruption during this time, which enables us to use the behavior of Yahoo! News users as a control in our regressions. We compare which websites consumers navigated after visiting a news aggregator (either Google News or Yahoo! News) before and after the removal of content on Google News.

It is not clear whether the removal of content will lead aggregator users to seek more or less news after visiting the aggregator. In essence, do consumers use aggregators in lieu of content sites (a "scanning effect") or to go more in depth into content (a "traffic effect")? This depends upon whether consumers view a news aggregator as a complement or substitute to original news sources.

For example, The AP ran a news story about the economic depression in Michigan. Figure 2 depicts the screenshot of how the story appeared on Google News. The links related to The Associated Press story are also depicted in Figure 2. After reading The Associated Press summary of the story, readers are free to explore the issue further in local newspapers such as the Detroit News and Lansing State Journal. These papers are local affiliates of The Associated Press and typically expand in their newspapers on the summary content from The Associated Press content. We ask whether the presence of The Associated Press content on Google News makes it more or less likely that a news consumer would then visit Detroit News or the Lansing State Journal, both of which are members of The Associated Press Network.

Our preliminary analysis focuses on the period immediately prior to and during the removal of The AP articles from Google News from December 2009 to January 2010. Contract negotiations continued until August 30, 2010 when a long-term contract was signed between Google News and The AP (Krazit, 2010). We also examine the consequences of the reinstatement of the long-term relationship between Google News and The AP between January 2010 and October 2010.[1] October is the appropriate date of analysis for the reinstatement of The AP material. Because content is added daily and appears for 30 days on Google News, one month following August is the month of October 2010 when all content for the past 30 days was fully reinstated and available as hosted content on Google News.

Note that during this period, Yahoo! News hosted stories from The AP and did not withdraw any content. In fact, Yahoo! renewed their deal with The AP in 2010, and its readers did not observe any disruption in content from The AP (Parr, 2010). During this period, both Google News and Yahoo! News hosted news stories from The AP in a similar fashion; the news aggregator would have a headline or snippet of The AP article on the front page with a link to the full hosted article on the news aggregator's site.

3 | ANALYSIS

3.1 | Theoretical predictions of scanning and traffic

Two countervailing forces exist when users use aggregators to search for information. Consumers may use aggregators in lieu of content sites ("scanning effect") or to explore content more in depth ("traffic effect"). These effects have been captured in theoretical models of platforms and aggregators. For instance, Jeon and Esfahani (2012) describe "business-stealing" and "market expansion" effects of aggregators. Rutt (2011) examines a model with two types of consumers—"searchers" who use aggregators and those "loyal" to original content sites. More generally, the effects can be described as to what

Michigan voters search for economic savior

By KATHY BARKS HOFFMAN (AP) – 1 day ago

LANSING, Mich. — Michigan voters frustrated over lost jobs, home foreclosures and budget deficits will vote in Tuesday's primary election for leaders they hope can move the state out of its economic morass.

With seven men running for governor and nearly two dozen candidates running for three open congressional seats, the hardest task may be sorting through the barrage of names, campaign ads and economic rhetoric.

The candidates and voters agree that Michigan is at a crossroads. After a decade of malaise that has left the state with the nation's second-highest unemployment rate and one in every four residents relying on unemployment insurance, Medicaid, cash assistance or food stamps, creating more jobs is the overwhelming priority and topic of debate.

The gubernatorial candidates are competing to succeed outgoing Democratic Gov. Jennifer Granholm, who can't run again because of term limits and whose popularity sank with her struggles to revive the economy.

All seven gubernatorial candidates say they plan to cut business taxes to attract employers. Most of the five Republicans also say they would slash state regulations and cut state spending. One, Oakland County Sheriff Mike Bouchard, proposes getting rid of laws forcing workers to join unions to get certain jobs.

Among the Democrats, Lansing Mayor Virg Bernero is visiting factory gates and union halls to pledge he'll stand up for middle-class workers and jobs. His opponent, Andy Dillon, a business turnaround specialist who's now the House speaker, promises to bring in more alternative energy jobs to replace lost manufacturing work.

With platforms that are similar, the Republicans are using their job credentials to assure voters they would be the best at managing the economy.

National GOP interest in unseating freshmen Democratic Reps. Mark Schauer in mid-Michigan's 7th District and Gary Peters in the Detroit suburbs in Oakland County has Republicans vying in both districts for the chance at a November matchup.

Related articles

Gov candidates woo undecided voters during final weekend before primary
The Detroit News - 7 hours ago

Race to get names on November ballot in Michigan governor's race is wide open
Lansing State Journal - 18 hours ago

Editorial: Focus on jobs, not abortion in governor race
The Detroit News - 21 hours ago

More coverage (1) »

Associated Press

President Barack Obama addresses employees at the Chrysler's Jefferson North Assembly Plant in Detroit, Friday, July 30, 2010. (AP Photo/Carlos Osorio)

Map

FIGURE 2 Example screenshot of The Associated Press article hosted on Google News [Color figure can be viewed at wileyonlinelibrary.com]
Notes: Google News, August 2010. Text of article has been slightly edited to fit on page.

extent content on platforms and aggregators are "substitutes" or "complements" (Athey, Calvano, & Gans, 2011; George & Hogendorn, 2012).

Under the "scanning effect," consumers scan all articles at an aggregator and then finish their search. The interpretation of "fair use" and other facets of copyright law shape the scanning effect. If fair use is more permissive, then more content is featured on the aggregator, which makes scanning more valuable to consumers. Our study examines a change in content at one aggregator, Google News. If consumers use the aggregator for scanning, then the removal of The AP content will reduce the quality of scanning, and subsequently, total visits to Google News may decline as consumers seek other aggregators.

The "traffic effect" suggests that consumers are interested in pursuing content more in depth. When consumers read a headline or excerpt on the aggregator, they will be prompted to click on the links to the content provider for further details of the story. If consumers use aggregators to "dig" for more content, then the removal of The AP content will lead to less traffic to news sites.

We empirically test for the two effects of scanning and traffic in our analyses below.

3.2 | Testing for scanning effect: Overall visits to an aggregator

To test for the scanning effect, we investigate whether the removal of The AP content from Google News leads to a shift away from Google News. If consumers use aggregators to merely scan headlines and excerpts of articles, then the removal of content from Google News will lower the quality of scanning on Google News. Consequently consumers will shift away from Google News toward other aggregators. This test assumes that content from non-AP sources on Google News is not a close substitute for content from The AP. This assumption may be reasonable in light of the earlier discussion of the role of The AP. Founded in 1846 with an intent to create scale economies associated with telegraph transmission of the news, The AP is one of the largest

TABLE II Monthly visits, page views, and visit time to Google News and Yahoo! News

Metric	Date	Google News	Yahoo! News
Visits	November 2009	75,667,000	224,160,000
Visits	December 2009	78,160,000	267,570,000
Visits	January 2010	79,373,000	262,700,000
Page views	December 2009	3%	7%
Page views	January 2010	3%	7%
Visit time	December 2009	22 seconds	5 seconds
Visit time	January 2010	22 seconds	5 seconds

Notes: ComScore and Hitwise.

news agencies in the world, and it is the only national news service in the United States. Given that other major competitors are based abroad, it seems plausible that The AP is distinct from other non-AP sources.[2]

We collect data from comScore on total visits to Google News and Yahoo! News. ComScore tracks the online activity of a panel of more than 2 million users based in the United States and subsequently aggregates their search patterns for resale to commercial clients. ComScore recruits its panel members through affiliate programs and partnering with third party application providers. ComScore emphasizes and discusses the representativeness of their sample to the general population in their Marketer User Guide. ComScore data have also been used in several academic studies and noted as a "highly regarded proprietary [source] for information on the size and composition of media audiences" (Chiou & Tucker, 2010; De Los Santos, Hortacsu, & Wildenbeest, 2012; Gentzkow & Shapiro, 2011; Montgomery, Li, Srinivasan, & Liechty, 2004).

Table II reports the number of monthly visits, page views, and visit time to each aggregator during our period of study. Because the content removal occurs at the very end of the month on December 23, 2009, we consider November and December 2009 to be the period before content removal and January 2010 to be the period after content removal. If the scanning effect persists, we would expect to observe a decline in visits and in reader attention through page views and visit time to Google News relative to Yahoo! News, because Google News would have a lower quality of scanning with the removal of The AP content whereas Yahoo! News would not. When we compare the metrics, we do not find evidence of a precipitous drop in monthly visits to Google News relative to Yahoo! News in the wake of the dispute from December 2009 to January 2010.

We also collect data from Experian Hitwise to check for any further evidence of changes in behavior. Hitwise "develops proprietary software that Internet Service Providers (ISPs) use to analyze website logs created on their network." Once the ISP aggregates the anonymous data, the data are provided to Hitwise. According to their website, Hitwise collects these usage data from a "geographically diverse range of ISP networks and opt-in panels, representing all types of Internet usage, including home, work, education and public access." Currently, Hitwise has usage data from a sample of 25 million people worldwide. We include further details on Hitwise's data collection in the Appendix.

When we checked the monthly data from Hitwise, we found no other evidence of changes in behavior in the months before (November and December 2009) and after (January 2010) the content removal. Indeed, throughout the period we study, Google News remained solidly ranked as fifth for unique visits among news websites whereas Yahoo! remained ranked as first. As seen in the table, no change occurred in alternative metrics such as "average visit time" or the number of pages navigated within a website. The share of page views among all news and media sites were 7% and 3% for Yahoo! News and Google News. The average visit time for Google was 22 seconds, and the average visit time for Yahoo was 5 seconds. Overall, the metrics suggest that scanning behavior did not change for Google News relative to Yahoo! News after the removal of content from The AP.

In our tests for the scanning effect, a potential concern may be whether Google News or Yahoo! News experienced any pre-existing trends prior to the dispute. For instance, although we do not find a relative shift in metrics for Google News compared to Yahoo! News during the period of study, we consider whether we would have expected Google News to fare better than it did in the absence of the content removal. For instance, did Google News trend positively in the months leading to the removal of The AP content and then drop precipitously to the levels of Yahoo! News after the policy change?

To explore this alternative explanation, we collect additional data from Google Trends on weekly search activity for Google News and Yahoo! News for the year leading up to the dispute. In order to document whether any dramatic changes occur over the course of the year prior to December 2009, Figure 3 graphs the search indices for Google News and Yahoo! News from January to December 2009 where the search indices are normalized between 0 and 1. Search activity varies for both news aggregators over the course of the year. Reassuringly, we do not observe declining popularity for Google News in absolute levels or relative to Yahoo! News during the period prior to the dispute.

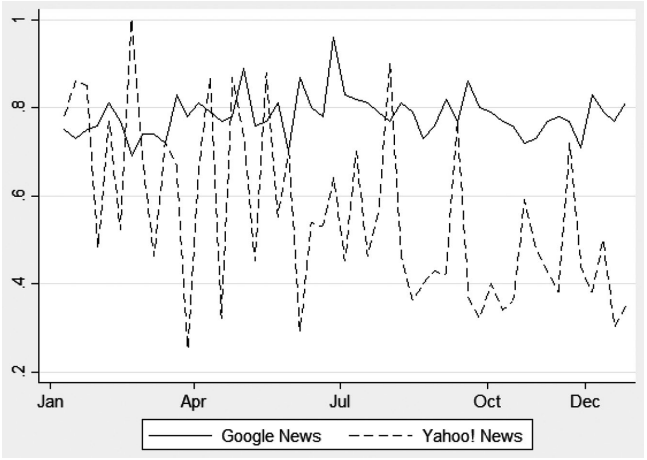

FIGURE 3 Search activity for Google News and Yahoo! News prior to the dispute

Notes: This figure graphs the search index of Google News and Yahoo! News from Google Trends, which is normalized between 0 and 1. The vertical axis is the index of search intensity, and the horizontal axis is time from January to December 2009.

TABLE III No evidence of scanning effect during removal and reinstatement of content from The Associated Press

	(1)	(2)	(3)	(4)
APContentRemoval × Google	0.0485			
	(0.0763)			
APContentRestored × Google		−0.0720	−0.0235	
		(0.0707)	(0.0532)	
December × Google				0.153
				(0.0875)
APContentRemoval	−0.0143			
	(0.0664)			
Google	0.417***	0.466***	0.417***	0.265**
	(0.0419)	(0.0629)	(0.0419)	(0.0764)
APContentRestored		0.0260	0.167***	
		(0.0841)	(0.0360)	
December				−0.0963
				(0.0663)
Observations	18	20	18	16

Notes: Robust standard errors. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. The outcome variable is the Google Trends index of search intensity for Google News or Yahoo! News, which is normalized between 0 and 1. Column (1) covers the period from December 2009 to January 2010, before and after the removal of The AP content from Google News on December 23, 2009. Column (2) compares January 2010 (when no content from The AP was available) to October 2010 (the restoration of hosted articles by The Associated Press in Google News). Column (3) compares the weeks in December 2009 and October 2010 when The AP content was available in Google News as a falsification check. Column (4) compares the weeks in November 2009 and December 2009 when The AP content was available to test for a pre-trend.

For a more formal test, we use the data from Google Trends to estimate a difference-in-difference analysis of search intensity for news aggregator j in week t:

$$searches_{jt} = \alpha_0 + \alpha_1 Google_j \times APContentRemoval_t + \alpha_2 Google_j + week_t + \nu_{jt}, \quad (1)$$

where $Google$ is an indicator variable equal to one for searches on Google News, and $APContentRemoval$ is an indicator variable equal to one for the weeks after the removal of The AP content from Google News on December 23, 2009. The vector $week$ contains weekly fixed effects to capture national variation in the volume and interest generated by news stories in that week.

We also collect additional data after the dispute was resolved. Around August 30, 2010, Google News and The AP formally signed a long-term contract to continue their relationship (Krazit, 2010). Because content is added each day and appears for 30 days on Google News, we collect data for the weeks in October 2010 when all content for the past 30 days was fully reinstated and

available. In Table III, Column (1) tests the period from December 2009 to January 2010 before and after the content removal of The AP. If a scanning effect exists, we would expect search intensity for Google News to decline relative to Yahoo! News after the content removal, as the removal of content leads to a lower quality of scanning for Google News. The estimated coefficient on the interaction term is not statistically significant; we do not find evidence of any change in search intensity for Google News relative to Yahoo! News. Column (2) tests the period of content reinstatement between January 2010 (when no content from The AP was available) to October (when The AP content was fully reinstated). If a scanning effect exists, we would expect to observe a positive interaction term, as the reinstatement of content from The AP improves the quality of scanning for Google News. The coefficient is not statistically significant; we do not find evidence for a scanning effect. As additional falsification checks, Column (3) compares December 2009 and October 2010 when The AP content was available in Google News in both months, and Column (4) compares November 2009 and December 2009 to test for a pre-trend. For both of these falsification checks, we would not expect to observe any shift in search intensity for Google News relative to Yahoo! News as the quality of scanning for Google News does not change between the periods of comparison. In fact, the coefficients are not statistically significant, and we do not find any evidence of a shift in search intensity for Google News relative to Yahoo! News during these periods.

An additional concern is whether any interaction occurs between the scanning and traffic effects. Potentially the two effects are separate, as each relies on a different distribution of users. For instance, we may expect users who scan and skim content to differ from users who seek further information and drive traffic. In the aggregate, we find that the overall set of customer characteristics across Google News and Yahoo! News are similar according to Table A1 in the Appendix.[3] Therefore, as long as those customer characteristics adequately predict the way in which consumers use news aggregators, it is likely that the scanning and traffic effects do not interact differently across the two aggregators.

In theory, removing The AP content reduces the quality of scanning on Google News and could potentially reduce total visits to Google News as it loses competition to other aggregators. We do not find evidence that competing aggregators act as substitute platforms for one another. Perhaps if the reduction in quality affected the customization of the website, we would observe an effect. For instance, Athey and Mobius (2012) find that the ability to customize stories to local content on Google News increased adoption of the news aggregator.

Our results should be interpreted with two caveats. First, it is possible that the length of the period in which The AP content was unavailable may have been too short to have had a noticeable impact, as users did not have sufficient time to notice the change in quality and switch to alternative aggregators. This is one explanation for the lack of effect we measure. Of course, unlike other products that are consumed less frequently or where quality is not readily ascertainable, the timeliness of news and the regularity of its consumption suggests that it may be reasonable to assume that enough consumers did visit the website and observe changes in quality. Nevertheless, our results should be interpreted with the caveat that consumers' awareness of the policy change may play a role. Second, even if readers are aware of the change, it is also possible that switching costs prevent readers from changing their behavior much. For instance, a reader may prefer to remain in the Google "environment" for other related services such as mail, search, etc.—all of which are easily accessible by embedded icons on the Google web page. Typing in an alternative web address to navigate to a different news aggregator presents a higher cost than clicking on an icon, so such switching costs, although small, are probably present.

3.3 | Testing for traffic effect: Downstream visits after an aggregator

In this section, we investigate the magnitude of the "traffic effect." To explore whether content on a news aggregator prompts users to seek further news, we examine whether traffic to news sites from Google News falls after the removal of The AP content from Google News.

Hitwise provides aggregate information on the sites that users visit immediately after navigating to Google News or Yahoo! News. We use weekly data on the top 2,000 sites navigated by consumers after visiting Google News or Yahoo! News during the week ending December 5, 2009 to the week ending January 30, 2010. Hitwise reports the fraction of total traffic that arrives at these "downstream" sites immediately after a visit to Google News and Yahoo! News. When Hitwise reports data for downstream sites, it tracks Web-surfing behavior by recording where people navigate after visiting a particular site at an aggregate level.

We construct a panel of the percentage of weekly visits a downstream website received from either Google News or Yahoo! News. For instance, we observe the weekly share of visits that nytimes.com receives out of all visits to websites by users immediately after using Google News. In our sample, 26% of websites received incoming traffic from both Google News and Yahoo! News. The remainder of websites were only visited after navigating to one particular aggregator. This pattern may reflect internal complementarities for these companies. For instance, someone using Google News is unlikely to navigate to Yahoo! Mail, and similarly, someone using Yahoo! News is unlikely to navigate to Gmail.

TABLE IV Summary statistics for downstream websites from Google News and Yahoo! News

	Mean	Std Dev	Min	Max
% visits	0.00016	0.0019	0	0.18
Google News	0.50	0.50	0	1
Yahoo! News	0.50	0.50	0	1
APContentRemoval	0.67	0.47	0	1
News Site	0.15	0.36	0	1
Non-news Site	0.85	0.36	0	1
Aggregator Site	0.0013	0.036	0	1
International Site	0.048	0.21	0	1
Weather Site	0.0067	0.081	0	1
Observations	98,730			

Notes: This table reports weekly statistics for websites visited immediately after Google News and Yahoo! News during December 2009 and January 2010. The variable $\%visits$ refers to the percentage of visits from each search engine that navigated to a particular site; this variable is measured from 0 to 1. The dispute between The Associated Press and Google News occurred after December 23, 2009. The variable $APContentRemoval$ is an indicator variable for whether the week occurred during the period of the dispute. News sites refer to print media and broadcast media sites as defined by Hitwise, excluding weather sites, international news sites, and top news aggregators.

To identify which sites are related to The AP, we categorized the websites into two main classes: "news" (e.g., newyorktimes.com, bostonherald.com) and "non-news" (e.g., Yahoo! Mail, Youtube.com). Our news category consists of a strict definition of sites that fall under Hitwise's categories of print media and broadcast media. We made sure that these news sites reflected The AP network of member news organizations as well as news outlets that subscribe to The AP service and provide The AP content. As we are interested in traffic to websites of primary news sources, we exclude weather sites and the top aggregators—e.g., Yahoo! News, Google News, AOL News, Bing News, Ask News, Huffington Post—from the "news" category. In addition, we use Hitwise's identification of non-US domains to exclude international sites (e.g., bbc.com/news, hindustantimes.com) from the "news" category, because we do not expect the removal of The AP content to affect international sites that tend to either generate their own content or rely on non-American news agencies for their content. We use data on international sites in our robustness checks. Given the set of "news" sites, we refer to all other sites within our sample as "non-news."

Table IV reports the summary statistics for our data. $Google$ and $Yahoo$ are indicator variables equal to one if the traffic originated after viewing Google News or Yahoo! News. $APContentRemoval$ is an indicator variable equal to one for the weeks after the removal of The AP content from Google News on December 23, 2009. $News$ is an indicator variable equal to one if the website is a news site, and $Non\text{-}news$, $Aggregator$, $International$, and $Weather$ are indicator variables for other types of websites such as non-news, aggregators, international, and weather. A site received on average 0.016% of downstream visits. News sites represent 15% of all sites where we observe subsequent visits within our sample, and non-news sites account for 85%. Aggregator, international, and weather sites account for a smaller fraction of sites compared to news sites.

Table V displays the top 40 news websites in our dataset and the average percentage of downstream visits they received from either Google News or Yahoo! News. Downstream visits refer to the number of visits to a website immediately after navigating to the news aggregator. Table VI displays the top 40 non-news websites in our dataset, excluding international and aggregator sites, and the average percentage of downstream visits they receive. As shown in Table VI, the top non-news websites reflect the top website brands on the Internet.

To verify that Yahoo! News could be considered an appropriate control group for Google News, we check that the users shared similar observable demographics. Hitwise reports the fraction of users within each demographic category for a particular site. As seen in Table A1 in the Appendix, the users of Yahoo! News and Google News do indeed look reasonably similar; the users are skewed toward being older, predominantly male, and wealthier than the general U.S. population. For comparison, we also report demographics for users of the New York Times website. The users of the New York Times site are similar, though significantly older than the average users of a news aggregator. Table A1 also provides suggestive evidence of why the debate over ad revenues from news content is so contentious. These readers are a remarkably attractive demographic group from an advertiser's perspective.

Our preliminary analysis examines visits to news sites after navigating to an aggregator. Figure 4 illustrates the aggregate mean percentage of downstream traffic to news and non-news sites for users that visited Google News and Yahoo! News during this period. As seen in the graph, little change occurs in downstream site navigation for Yahoo! However, news sites experience a decline in visits from Google News during the period of the removal of The AP content, relative to the change in traffic from

TABLE V Top 40 news websites visited after Google News and Yahoo! News

	Avg Visit Pct
nytimes.com	0.029
abcnews.com	0.021
cnn.com	0.019
washingtonpost.com	0.017
wsj.com	0.017
nydailynews.com	0.016
reuters.com	0.014
examiner.com	0.013
time.com	0.012
foxnews.com	0.011
politico.com	0.011
msnbc.com	0.0083
people.com	0.0078
usatoday.com	0.0072
bloomberg.com	0.0051
nypost.com	0.0051
boston.com	0.0048
latimes.com	0.0048
usmagazine.com	0.0046
mercurynews.com	0.0044
edition.cnn.com	0.0040
bostonherald.com	0.0038
cbsnews.com	0.0037
pcworld.com	0.0037
sfgate.com	0.0033
npr.org	0.0032
businessweek.com	0.0031
csmonitor.com	0.0030
miamiherald.com	0.0030
philly.com	0.0030
theweek.com	0.0029
chron.com	0.0027
voanews.com	0.0026
freep.com	0.0025
seattletimes.nwsource.com	0.0022
dallasnews.com	0.0021
mcclatchydc.com	0.0019
startribune.com	0.0017
wired.com	0.0017

Notes: All visit percentages are measured between 0 and 1. News sites are defined as sites for print and broadcast media. International and aggregator sites are excluded from the table above. Please see text for further details.

Yahoo! News. To investigate whether this pattern could be due to underlying seasonality in news consumption, we examine the change in visits in the prior year during the same calendar months. Figure 5 illustrates that no such change in visits occurred between December 2008 and January 2009.

To formalize the insights provided by Figure 4, we run a difference-in-differences regression for the policy change and estimate the percentage of visits to website i after visiting news aggregator j in week t. We use a Generalized Linear Model (GLM)

TABLE VI Top 40 non-news websites visited after Google News and Yahoo! News

	Avg Visit Pct
google.com	0.12
mail.yahoo.com	0.099
yahoo.com	0.072
facebook.com	0.062
search.yahoo.com	0.044
youtube.com	0.025
gmail.com	0.015
myspace.com	0.015
my.yahoo.com	0.013
mail.live.com	0.013
finance.yahoo.com	0.012
howlifeworks.com	0.010
msn.com	0.010
images.google.com	0.010
ebay.com	0.0100
cosmos.bcst.yahoo.com	0.0095
weather.yahoo.com	0.0079
blogsearch.google.com	0.0077
livescience.com	0.0075
finance.google.com	0.0072
weather.com	0.0067
omg.yahoo.com	0.0064
bing.com	0.0062
amazon.com	0.0059
members.yahoo.com	0.0058
espn.com	0.0056
video.google.com	0.0054
sports.yahoo.com	0.0052
wikipedia.org	0.0050
aol.com	0.0046
maps.google.com	0.0046
tmz.aol.com	0.0039
ask.com	0.0039
bankofamerica.com	0.0036
bleacherreport.com	0.0035
thedailybeast.com	0.0034
gather.com	0.0034
education.yahoo.net	0.0034
jaamla.com	0.0017

framework to estimate the following equation:

$$g(E(\%visits_{ijt})) = \beta_0 + \beta_1 News_i \times Google_j \times APContentRemoval_t + \beta_2 News_i \times APContentRemoval_t + \beta_3 News_i \times Google_j + \beta_4 Google_j + \theta_i + week_t + \epsilon_{ijt}, \quad (2)$$

where $g(.)$ is the generalized function.[4] $E(\%visits)$ is the expected value of the percentage of visits. The control θ contains downstream-website fixed effects.[5] The vector $week_t$ contains weekly fixed effects.

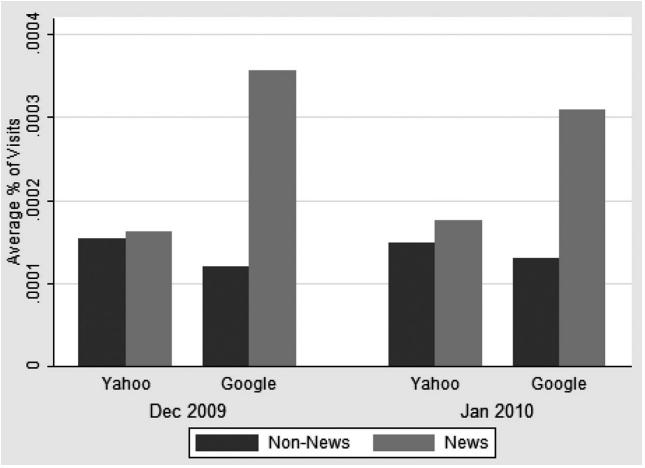

FIGURE 4 Downstream sites visited after Google News and Yahoo! News before and after The Associated Press content was removed

Notes: This figure shows the average percentage of visits to news and non-news sites after users visited Google News and Yahoo! News before and after the removal of The Associated Press from Google News in December 2009 and January 2010.

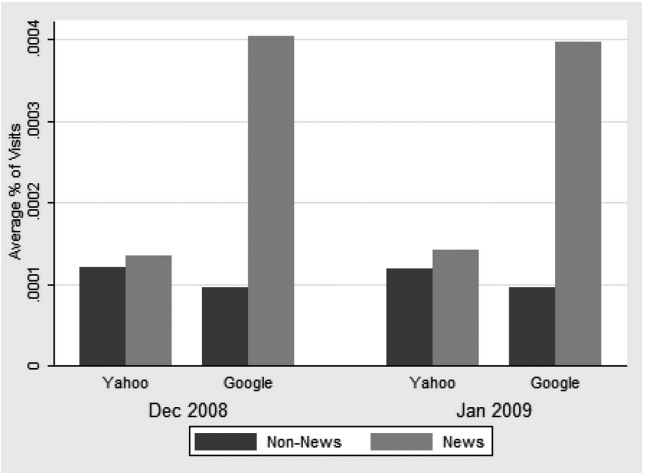

FIGURE 5 Downstream sites visited after Google News and Yahoo! News in prior year when no content was removed (December 2008 and January 2009)

Notes: This figure shows the average percentage of visits to news and non-news sites after users visited Google News and Yahoo! News in December 2008 and January 2009 for the year prior to the removal of The Associated Press content from Google News.

The coefficient β_1 on the interaction term $News \times Google \times APContentRemoval$ captures the effect of The AP content removal on visits to news sites compared to non-news sites from Google News with Yahoo! News as a control. In other words, the coefficient compares consumer visit behavior on Google News before and after the removal of content from The AP with consumer visit behavior on Yahoo! News. We control for difference in the levels of visits across websites and news aggregators through fixed effects for websites and news aggregators. We also control for general trends in news consumption over the year by including weekly dummies that capture seasonality or the timing of breaking news stories.

Our outcome variable is the probability or share for each website as measured relative to a particular aggregator. The specification captures the probability or percentage of visits to website i after visiting news aggregator j in week t where the percentage is calculated from all outgoing visits from aggregator j. For instance, if V_{jt} is the number of all outgoing visits from search engine j in week t to any site and v_{ijt} is the number of outgoing visits from search engine j in week t to site i, then we observe the percentage of visits where $\%visits_{ijt} = v_{ijt}/V_{jt}$. If consumers were equally likely to visit news sites before and after the policy change, then the measure of the share of outgoing traffic to news sites from each aggregator will remain the same. The shares of each site outgoing from Google News are measured as a fraction or probability conditional on traffic to Google News.

In other words, the market shares are calculated separately for Google News and Yahoo! News. Because the dependent variable is measured relative to traffic from each news aggregator separately, the share of outbound traffic to news sites would remain the same even with shifts in total traffic to each news site.

We estimate this specification using a Generalized Linear Model with a fractional response variable (Papke & Wooldridge, 1996). Following Papke and Wooldridge (2008), a GLM with link logit function $g(.)$ and family binomial takes into account that the dependent variable (percentage of visits) lies between 0 and 1. We cluster our standard errors at the website level to avoid the downward bias reported by Bertrand, Duflo, and Mullainathan (2004).

Our dependent data are proportional data and bounded between zero and one, so panel data methods propose the GLM framework to keep predicted values within the unit interval (Papke & Wooldridge, 2008). The GLM framework also maintains the advantage that no *ad hoc* transformations are required to handle the data at the extreme values near zero and one (Papke & Wooldridge, 1996). Given that the share of visits to a particular site may be small, this is attractive in our empirical setting. Furthermore, we interpret the results as an odds ratio. Because we use the binomial family with logit link, the coefficients may be used to compute an odds ratio, that is, ratio of probabilities of success (visiting a news site) and of failure (visiting a non-news site) between Google News and Yahoo! News.

As Lechner (2010) points out, an arithmetic difference between the expected value of the dependent variable before and after treatment will not difference out the common trend for the group. Instead, because we work with fractional response data, we are able to take the ratio of the probabilities to compute the odds for the treatment and control group, and the common trend for each group will proportionally difference out.[6] We generalize to the difference-in-differences framework by computing the ratio of the odds for visiting news and non-news sites for the treatment and control groups of Google News and Yahoo! News:

$$\frac{\left\{\frac{[Odds|Google=1,Post=1]}{[Odds|Google=1,Post=0]}\right\}}{\left\{\frac{[Odds|Google=0,Post=1]}{[Odds|Google=0,Post=0]}\right\}} = exp(\beta_3), \tag{3}$$

where $G(E(Y)) = \beta_0 + \beta_1 News * Google * Post + \beta_2 News * Post + \beta_3 News * Google + \beta_4 Google + \epsilon$, and $Post$ is a dummy variable equal to one after the policy change.[7] To facilitate interpretation, we report the exponentiated coefficients or odds ratio for our results under the GLM logit link.

Table VII reports the results in Column (1) for the full specification as described by equation (2).[8] The time period covers December 2009 to January 2010, which encompasses the weeks before and after the dispute on December 23, 2009 between The AP and Google News. In our setting, the odds are the probability of visiting a news site compared to the probability of not visiting a non-news site, or the share of visits to news site compared to the share of visits to all non-news sites. Note that because the coefficients are exponentiated, we interpret them and test for statistical significance relative to the value of one, which represents no effect ($exp(0) = 1$). In other words, if the policy change has a negative effect (coefficient is less than zero), we would expect an odds ratio or exponentiated coefficient to be less than one; if the policy change has a positive effect (coefficient is greater than zero), we would expect an odds ratio or exponentiated coefficient to be greater than one.

If news aggregators complement the news sources that they feature, we would expect the aggregators to direct further traffic to news sites. Accordingly, if the hosting of content from The AP by Google News prompts readers to seek further information, we would expect the removal of content from The AP to lead to a decline in referrals from Google News to other news sites. Because Table VII reports the exponentiated coefficients, we would expect β_1 to be negative, and therefore the odds ratio or the exponentiated coefficient of β_1 to be less than one.

We find that the exponentiated coefficient of 0.72 on $APContentRemoval \times Google \times News$ indicates that the odds ratio is 0.72 or 72% of its level prior to the policy change. In other words, the odds ratio fell by 28%. Consequently, the odds of visiting a news site on Google News relative to a non-news site on Google News decreased by 28% compared to the odds of visiting a news site on Yahoo! News relative to a non-news site on Yahoo! News. This suggests that the presence of The Associated Press articles in Google News prompted users to seek further information at news sites. More generally, our results suggest that news aggregators may complement the news sources that they feature by directing traffic to these news sites.

We also collect additional data after the dispute was resolved. Around August 30, 2010, Google News and The AP formally signed a long-term contract to continue their relationship (Krazit, 2010). Because content is added each day and appears for 30 days on Google News, we collect data for the weeks in October 2010 when all content for the past 30 days was fully reinstated and available. In Table VII, Column (2) compares January 2010 (when no content from The AP was available) to October (when The AP content was fully reinstated). If the complementary relationship between Google News and The AP exists, then we would

TABLE VII Downstream traffic from Google News and Yahoo! News during removal and reinstatement of content from The Associated Press

	(1) Removal	(2) Reinstatement	(3) Falsification	(4) Falsification
APContentRemoval × Google × News	0.718*			
	(0.127)			
APContentRestored × Google × News		1.955**	1.347	
		(0.553)	(0.259)	
December × Google × News				0.964
				(0.115)
APContentRemoval × Google	1.120			
	(0.181)			
Google	1.092	1.187	1.115	1.291***
	(0.125)	(0.142)	(0.0862)	(0.0768)
APContentRemoval × News	1.114*			
	(0.0703)			
News × Google	0.811	0.575***	0.718***	0.667***
	(0.109)	(0.0837)	(0.0849)	(0.0563)
APContentRestored × Google		0.902	1.016	
		(0.216)	(0.131)	
APContentRestored × News		0.601***	0.680***	
		(0.0899)	(0.0869)	
December × Google				0.998
				(0.105)
December × News				1.114
				(0.121)
Week Fixed Effects	Yes	Yes	Yes	Yes
Website Fixed Effects	Yes	Yes	Yes	Yes
Observations	98,730	119,640	103,113	84,048

Notes: Robust standard errors clustered at website level. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. The outcome variable is the fraction of traffic to websites after visiting Google News or Yahoo! News. The exponentiated coefficients are reported with the corresponding standard errors for interpretation as odds ratios. Note that because the coefficients are exponentiated, we interpret them and test for statistical significance relative to the value of one, which represents no effect ($exp(0) = 1$). In other words, if the policy change has a negative effect (coefficient is less than zero), we would expect an odds ratio or exponentiated coefficient to be less than one; if the policy change has a positive effect (coefficient is greater than zero), we would expect an odds ratio or exponentiated coefficient to be greater than one. Column (1) covers the period from December 2009 to January 2010, before and after the removal of The AP content from Google News on December 23, 2009. Column (2) compares January 2010 (when no content from The AP was available) to October 2010 (the restoration of hosted articles by The Associated Press in Google News). Column (3) compares the weeks in December 2009 and October 2010 when The AP content was available in Google News as a falsification check. Column (4) compares the weeks in November 2009 and December 2009 when The AP content was available to test for a pre-trend.

expect an increase in visits to other news sites when Google News restores content from The AP. As expected, we observe a positive effect; as the odds ratio of visiting a news sites increased by 96% after the reinstatement of content through a long-term contract.

So far our analyses in this section of the removal and subsequent reinstatement of The AP content suggest that a "traffic effect" does exist and that the relationship is complementary. Consumers do appear to use platforms to seek new and further content. One striking feature of how The AP content was featured on Google News is that in general, as shown in Figure 2, quite a large amount of news content was displayed rather than merely a snippet. In light of this, our evidence of a traffic effect rather than merely a scanning effect is striking.

News sites on Google experience a 28% decrease in visits after the removal of The Associated Press articles. The magnitude of this estimate is comparable to concurrent and prior studies of technological adoption on consumers' consumption of content. Athey and Mobius (2012) study how the adoption of Google News toolbar affects consumption of local news and finds that the additional content increases local news consumption by more than 26%, and over time (beyond an 8-week period) the increase persists at more than 14%. In our study, The AP represents both local and national news, so we would expect our results for

traffic to be comparable. Furthermore, Xu et al. (2014) study how the adoption of a news app increases the probability of visiting a news site by nearly 30%.

If the claim in Cohen (2009) is true that Google sends a billion clicks each month to its partner news providers, then this percentage translates into a very large change in the number of clicks that news websites receive from Google News. Although we do not know precisely the international breakdown, our data from Hitwise suggest that before the policy change, news media websites hosted in the United States account for 40% of all clicks for the subset of users who use Google News. Therefore, this 28% decrease could imply an approximately 110 million decrease in visits each month from Google News users to news media websites hosted in the United States.[9]

As seen in Figure 1, The AP holds the topmost position under "Top Stories" on Google News. The large effects of the policy may be expected if Google News places The AP in a prominent position on its website. Although the exact algorithm is unknown for how Google News decides which stories to include and where to place them on its site, we do know that recency and popularity of the story play key factors. As The AP is the only remaining national newswire agency in the United States, it is likely that The AP would provide breaking news on recent events and be very popular among readers. Consequently, The AP may often hold a prominent position on the page of Google News, and therefore we would expect the removal of The AP content to have a relatively large effect on traffic from Google News.

When we consider the reinstatement, an even larger measured increase exists with the point estimate suggesting that clicks increase by 95%. One potential reason for the larger size of the reinstatement effect is that when Google restored all the AP articles at once in fall 2010, the articles dominated the most valuable position on the web page in Google News.

4 | ROBUSTNESS CHECKS

4.1 | The relationship between the Associated Press and news organizations

As a sharper test of our theory and as a robustness check, we examine whether the traffic effect was strongest among sites more embedded within The AP ecosystem. To clarify further institutional details of The AP, recall that news organizations may subscribe to and disseminate from The AP on their own site. We collect data on the number of stories featured from The AP on each news site.[10] We create a measure $APstories$ of the fraction of stories from The AP featured on the homepage of the news site. Our measure is intended to identify which news sites are more likely to be featured alongside the hosted article and therefore removed from Google News when The AP hosted content was removed. The measure $APstories$ reflects the strength of the relationship between the news site and The AP.

These data were collected from a cross-section of news sites in 2015. Our assumption is that the relationship between the news sites and The AP stories had a common trend over the years. This seems reasonable given that The Associated Press has a long history of relationships with newspapers, which seem unlikely to change quickly. The biggest concern is that newspapers who faced financial turmoil could have severed links to The AP to save money, potentially influencing our results. However, conversations with industry insiders suggest to us that newspapers facing financial constraints are more likely to stop producing their own content rather than severing ties with The Associated Press. Therefore, it seems likely that The Associated Press relationships with newspapers have not changed dramatically in the time between 2009 and 2016.

Table VIII reports the results of equation (2) with our continuous measure $APstories$ as our variable of interest instead of the indicator variable $News$. Once again, we report the exponentiated coefficients or odds ratios. Consistent with our results on the traffic effect, we find that sites that are more embedded in The AP ecosystem experienced a larger drop in traffic relative to Yahoo! News. Columns (2)–(4) check robustness of the results to alternative definitions of the control group. As described previously, users navigated to a variety of "non-news" sites after visiting a news aggregator. In Columns (2) and (3), our robustness checks omit the top international and news aggregators websites as part of the control group. These alternative definitions of the control group could be warranted if the removal of The Associated Press content also affected navigation to these sites directly (e.g., if The Associated Press content had previously encouraged people to visit international websites) or if the removal of The Associated Press content on Google altered people's perceptions of news aggregators. In Column (4), we check robustness to removing both aggregators and international sites from our control group. In general, the results are robust in sign and similar in magnitude.

Note that $APstories$ is a continuous measure, so the reported results represent a percentage change in the odds of visiting compared to not visiting a news site. Using the conservative estimate of 0.52, for every one percentage point increase in the fraction of stories from The AP, the odds of visiting a site fall by 0.48%.[11]

TABLE VIII The dispute had a larger effect for sites that featured more stories from The Associated Press.

	(1)	(2)	(3)	(4)
	All	**No international**	**No aggregators**	**No international & no aggregators**
PeriodDispute × Google × APstories	0.328**	0.319**	0.507***	0.511***
	(0.160)	(0.164)	(0.128)	(0.130)
PeriodDispute × Google	1.052	1.061	0.929	0.928
	(0.133)	(0.144)	(0.0442)	(0.0457)
Google	1.068	1.046	1.158***	1.143***
	(0.0947)	(0.0994)	(0.0472)	(0.0481)
APstories	3.987***	4.080***	3.986***	4.079***
	(0.768)	(0.789)	(0.770)	(0.791)
PeriodDispute × APstories	1.291	1.292	1.294	1.296
	(0.205)	(0.207)	(0.208)	(0.209)
APstories × Google	0.776	0.827	0.590*	0.615*
	(0.306)	(0.337)	(0.173)	(0.181)
Website Fixed Effects	Yes	Yes	Yes	Yes
Week Fixed Effects	Yes	Yes	Yes	Yes
Observations	97,668	92,889	97,542	92,763

Notes: Robust standard errors clustered at website level. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. The outcome variable is the fraction of traffic to websites after visiting Google News or Yahoo! News. The table covers the period from December 2009 to January 2010. The variable $APstories$ measures the percentage of stories that were from The Associated Press. The exponentiated coefficients are reported with the corresponding standard errors for interpretation as odds ratios.

4.2 | The availability of content from the Associated Press

We also perform an additional falsification check by comparing downstream traffic to Google News during two periods when The AP content was available. We expect no difference in traffic between these two periods, because The AP content was available in both time periods. We collect additional data on traffic during October 2010 when the dispute was resolved and compare this to traffic in December 2009; in both months, The AP content was featured on Google News. Table VII reports the results of the estimation in Column (3). As expected, no distinguishable effect exists between the two time periods, as The AP content was available in both instances for the past 30 days. The odds ratio is not significantly different from one; in other words, the odds of visiting a news site did not change between these two periods.

4.3 | Checking for a pre-trend

As a final set of falsification checks, we test for a pre-trend in the data prior to the dispute. The concern may be that the policy change coincides with a pre-existing trend in the data. We collected weekly data from November until the dispute in December 23, 2009, and we re-ran our analysis with a dummy variable for the weeks in December instead of a post variable. As shown in Table VII Column (4), we do not find evidence of a pre-trend in the months preceding the contract dispute. The odds ratio is not significantly different from one; in other words, the odds of visiting a news site did not change in our falsification test.

Furthermore, we include additional data from comScore on the number of visits received by each news aggregator to directly control for trends in visits to Yahoo! News and Google News during the period. In the Appendix, Table A4 reports our results with direct controls for visits to Yahoo! News and Google News during this time period. The goal is to observe how directly controlling for trends affects the estimates of our analysis—that is, how trends may affect our results. Across all four specifications during the period of removal, the period of reinstatement, and the two falsification checks, the estimated coefficients are similar in magnitude, direction, and statistical significance to our main results.

Despite these checks, it is important to remember that broader limits exist to the generalizability of the experiment we study, because we focus on a particular natural experiment. However, as a single natural experiment, it is reasonably attractive, as it captures the dispute between the second largest news aggregator and one of the largest provider of news content in the United States. Even if our results do not generalize across the industry, the results may themselves be inherently interesting. This is especially the case as news readership can be highly concentrated (Gentzkow & Shapiro, 2011), so it is instructive to focus on larger players in the industry.

TABLE IX The dispute harmed sites that were either local or national

	(1) All	(2) No international	(3) No aggregators	(4) No international & no aggregators
PeriodDispute × Google × Local	0.817**	0.817**	0.817**	0.817**
	(0.0709)	(0.0709)	(0.0709)	(0.0709)
PeriodDispute × Google × National	0.742***	0.742***	0.742***	0.742***
	(0.0844)	(0.0844)	(0.0844)	(0.0844)
PeriodDispute × Google × News	0.917	0.900	1.081	1.081
	(0.156)	(0.167)	(0.0812)	(0.0829)
Google	1.092	1.064	1.219***	1.203***
	(0.125)	(0.134)	(0.0540)	(0.0554)
PeriodDispute × Google	1.120	1.141	0.950	0.950
	(0.181)	(0.203)	(0.0490)	(0.0510)
PeriodDispute × Local	1.122**	1.122**	1.122**	1.122**
	(0.0524)	(0.0524)	(0.0524)	(0.0524)
Google × Local	1.203	1.203	1.203	1.203
	(0.161)	(0.161)	(0.161)	(0.161)
PeriodDispute × National	1.129*	1.129*	1.129*	1.129*
	(0.0771)	(0.0771)	(0.0771)	(0.0771)
Google × National	1.546**	1.546**	1.546**	1.546**
	(0.297)	(0.297)	(0.297)	(0.297)
PeriodDispute × News	0.986	0.986	0.987	0.987
	(0.0534)	(0.0540)	(0.0542)	(0.0548)
News × Google	0.616***	0.633***	0.552***	0.560***
	(0.0967)	(0.105)	(0.0643)	(0.0655)
Website Fixed Effects	Yes	Yes	Yes	Yes
Week Fixed Effects	Yes	Yes	Yes	Yes
Observations	97,704	92,925	97,578	92,799

Notes: Robust standard errors clustered at website level. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. The outcome variable is the fraction of traffic to websites after visiting Google News or Yahoo! News. The table covers the period from December 2009 to January 2010. The variable *Local* is a dummy variable that equals one if the majority of stories on the news site were local news; the variable *National* is a dummy variable that equals one if the majority of stories on the news site were national news. The exponentiated coefficients are reported with the corresponding standard errors for interpretation as odds ratios. Note that because the coefficients are exponentiated, we interpret them and test for statistical significance relative to the value of one, which represents no effect ($exp(0) = 1$). In other words, if the policy change has a negative effect (coefficient is less than zero), we would expect an odds ratio or exponentiated coefficient to be less than one; if the policy change has a positive effect (coefficient is greater than zero), we would expect an odds ratio or exponentiated coefficient to be greater than one.

5 | LOCAL AND NATIONAL SITES

In the previous section, we found that users employ technological advances, such as aggregation, to seek further, more specific information. That is, our results imply that the "traffic effect" dominates the "scanning effect." Given the expansion in users' information set, we next consider what information users seek and which types of content benefit from aggregation. Depending upon its content, a site may be horizontally differentiated with a very local audience, or vertically differentiated with a national audience and acclaimed standards of quality.

Given our finding that downstream traffic to news sites from Google News declined after the removal of The AP content, we explore which sites were most affected by the removal of the news content from the aggregator and consequently which sites benefit the most from aggregation. Specifically, we explore whether the extent of the decline varied by a site's type of content differentiation. News sites may be local in news coverage with a readership that is geographically local, or may be national with a broad readership. Tastes for local news sites vary horizontally, depending upon the consumer's interest in regional news, whereas tastes for national news can be vertically differentiated with readers seeking sites, such as The New York Times, with acclaimed standards of quality.

To identify the extent to which a site's content is very local or very national, we examine the top 20 stories featured on the homepage of each news site in our sample, and we compute the fraction of these stories that are local or national news. We classify a story as "local news" if the news is specific to the site's state of headquarters. We classify a story as "national news" if the news focuses on an area outside of the local state headquarters. Note that some stories may not fall into either category, such as movie reviews, culture and opinion pieces, etc. To ensure that we properly identify local news stories, we focus on sites affiliated with newspapers or local TV or radio stations in order to exclude travel or leisure sites that feature local articles, often in the guise of advertising. We classify a news site as "local" if a majority of the stories featured on the site are local news. We classify a news site as "national" if a majority of stories featured on the site are national news.

We estimate a GLM equation similar to equation (2) where we include additional interactions between our measures of local and national news sites. We include the main effect of the policy change on the news category, and we allow an incremental effect if a news site is local or national. For ease of interpretation, we report the exponentiated coefficients or odds ratios of our results, so we interpret and test our estimated coefficients for statistical significance relative to the value of one, which represents no effect ($exp(0) = 1$). As seen in Column (1) of Table IX , our results indicate that visits to sites decrease the most for sites that are either local or national. Columns (2)-(4) check the robustness of the results to alternative definitions of the control group. In general, the results are robust in sign and magnitudes.

The odds of visiting a local news site on Google News declined by an extra 18% relative to other news sites and compared to Yahoo! News, because the odds ratio is 0.82. The odds of visiting a national news site on Google News declined by an additional 26% relative to other news sites and compared to Yahoo! News, because the odds ratio is 0.74.

Our results are consistent with news aggregators reducing consumers' search costs and allowing readers to easily find sites that specialize in local news. Local news sites may not otherwise find an audience outside of their local region. Our results have an important public policy implication as policymakers enact legislation to encourage the growth of local media, which is often viewed as necessary to encourage civic engagement.

Our findings also suggest that aggregators encourage visits to vertically differentiated sites such as national newspapers with acclaimed standards of quality. As Gentzkow and Shapiro (2011) note, news is vertically differentiated, with a small number of sites capturing a large fraction of readers. We examine two pieces of evidence that suggest that these sites are of higher "quality." First, such sites account for a disproportionate number of visits. For instance, 25% of national sites account for over half of all visits to news sites. Second, we obtain a list of Pulitzer Prize winners and finalists and confirm that a disproportionate number fall among national sites. Even though national sites comprise 20% of all sites, they account for the majority (60%) of winners and finalists for the 13 categories of news reporting in 2009 and 2010.[12]

6 | CONCLUSION

In the first wave of digitization of content, peer-to-peer networks disseminated copyrighted content, sometimes illegally. Now the second wave of the digitization of content has led to the rise of platforms which aggregate content in a convenient and unified form for consumers. The new practice of digital aggregation has led to both lawsuits and uncertainty over the economic consequences.

To investigate the consequences of these disputes, we examine a breakdown in contract negotiations between The Associated Press (AP) and Google—which prompted Google to stop hosting The AP content. We find little evidence of a "scanning effect" where customers use news aggregators as substitutes for traditional news media sites. Instead we find evidence of a "traffic effect" where news aggregators complement the consumption of other related news sites. Specifically, we find that when Google News no longer hosted content from The AP, Google News users were less likely to visit other news websites with content related to The AP after visiting Google News relative to Yahoo! News users who experienced no such removal of The AP content. This pattern was driven by a reduction in visits to either local or national websites. Consequently, the relaxation of intellectual property rights may benefit content that is either horizontally differentiated, such as local sites, or vertically differentiated, such as top news websites with acclaimed standards of quality.

This is important both for content producers trying to understand the likely consequences of the rise of these platforms for their business and also for regulators trying to understand whether regulation of such platforms is advisable. In the European Union, for example, legislators now require aggregators to offer an explicit opt-in for all their content. Our paper offers an explanation for why content producers have not "opted-out" of aggregators despite lobbying for laws that require such consent. One possible reason is that ultimately aggregators may benefit platforms, and the purpose of the opt-in provision may be to increase bargaining power over payments to news providers.

Several limitations of this paper exist. First, our data are at the aggregate level, so we focus on uncovering heterogeneity in responses at the website level rather than at the consumer level. Thus we may not uncover other moderating factors that could explain the propensity to use new aggregators for scanning or traffic purposes. Second, as with any attempt at analyzing a quasi-experiment, limitations may exist both because of the potential endogeneity of actions of agents surrounding the experiment and also because of questions over its generalizability. For example, the relative improvement in traffic from Google News to The AP websites may reflect improved terms in the deal they struck, which we do not observe. Third, we focus on the aggregation of news content which has attracted a lot of attention, but may have different search and consumption patterns from other content such as music and movies. Consequently our results may not generalize broadly to other content platforms. Notwithstanding these limitations, we believe this paper provides useful first evidence about the effects of digital aggregation on the consumption of copyrighted content.

NOTES

[1] In the interim, on February 2010, a temporary short-term deal was agreed upon which gave rise to the content that can be seen in Figure 2. It is not clear at which point in February the relationship was resumed between Google News and The AP. It is also not apparent whether the short-term deal during this time consisted of the older, missing content or new content or whether Google changed the presentation of The AP articles afterwards. For example, it would be problematic if Google decided to highlight The AP content after the contract negotiations were concluded, perhaps as a "sweetener" although on-going negotiations continued until September. For these reasons, we collect data on the full reinstatement of the long-term relationship between Google News and The AP in the fall of 2010.

[2] Mark Twain said in 1906, "There are only two forces that can carry light to all corners of the globe, only two, the sun in the heavens and the Associated Press down here" (Halberstam, 2007).

[3] Hitwise reports the fraction of users within each demographic category for a particular site.

[4] We also re-estimate our model using a linear specification of OLS on the dependent variable of percentage of visits. Table A2 in the Appendix reports the results of the linear model. Reassuringly, the results are qualitatively similar.

[5] Due to computational limitations, we include fixed effects for the top 500 websites.

[6] For a simple example, suppose that $E(Y_i) = \beta_0 + \beta_1 X_i + \epsilon_i$ where Y_i is the fraction of successes from n_i binomial trials for observation i, and X is a dummy variable equal to 0 or 1. Using the logit link for our link function $G(.)$, the odds are $E(Y)/(1 - E(Y)) = exp(\beta_0 + \beta_1 X_1 + \epsilon)$. Consequently, the odds for $X = 1$ are $(Odds|X = 1) = E(Y|X = 1)/(1 - E(Y|X = 1)) = exp(\beta_0 + \beta_1 + \epsilon)$, and the odds for $X = 0$ are $(Odds|X = 0) = E(Y|X = 0)/(1 - E(Y|X = 0)) = exp(\beta_0 + \epsilon)$. We simplify the ratio of the two odds as $(Odds|X = 1)/(Odds|X = 0) = exp(\beta_1)$.

[7] This is a simplified version of Equation (2) for ease of notation. Note that if we had taken the arithmetic difference for news and non-news sites, $[E(Y|Google = 1, Post = 1) - E(Y|Google = 1, Post = 0)] - [E(Y|Google = 0, Post = 1) - E(Y|Google = 0, Post = 0)]$, instead of the ratio of the odds, we would not difference out the trends for each group.

[8] We also re-estimate our model using top websites with traffic levels among the 100 highest in our sample. Table A3 in the Appendix reports the results, which are qualitatively similar.

[9] Forty percent of 1 billion clicks is 40 million clicks. A 28% reduction of 40 million clicks is a decrease of 112 million clicks.

[10] Specifically, for each news site, we counted the number of articles from The AP that are featured in the top 20 stories on its home page.

[11] We also attempt instrumenting for $APstories$ by using the age of the media organization. The estimated coefficients for GLM with instrumental variables is of similar magnitude and sign, though the first stage is weak.

[12] We obtained the list of Pulitzer Prize winners and finalists from the official website www.pulitzer.org. The categories include breaking news reporting, breaking news photography, commentary, correspondence, criticism, editorial cartooning, editorial writing, explanatory reporting, feature photography, feature writing, international reporting, local reporting, and national reporting.

REFERENCES

Ardia, D. (2008, June 16). Associated press sends DMCA takedown to drudge report, backpedals, and now seeks to define fair use for bloggers. *Citizen Media Law Project*.

Arrington, M. (2010, Feb 2). Everybody forgets the readers when they bash news aggregators. *Techcrunch*.

Athey, S., Calvano, E., & Gans, J. (2011). The impact of the internet on advertising markets for news media. working paper.

Athey, S., & Mobius, M. (2012). *The impact of news aggregators on internet news consumption: The case of localization*. Working paper.

Bakos, J. Y. (1997). Reducing buyer search costs: Implications for electronic marketplaces. *Management Science*, *43*(12), 1676–1692.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, *119*(1), 249–275.

Chiou, L., & Tucker, C. (2010). *How does pharmaceutical advertising affect consumer search*? Mimeo, MIT.

Cohen, J. (2009, December 2). Same protocol, more options for news publishers. *Posting on Google News's Blog*.

Danaher, B., Dhanasobhon, S., Smith, M. D., & Telang, R. (November/December 2010). Converting pirates without cannibalizing purchasers: The impact of digital distribution on physical sales and internet piracy. *Marketing Science*, *29*(6), 1138–1151.

De Los Santos, B., Hortacsu, A., & Wildenbeest, M. R. (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review*, *102*(6), 2955–2980.

Eddy, M. (2013). German copyright law targets Google links. *The New York Times*.

Evans, D., & Schmalensee, R. (2012). *The antitrust analysis of multi-sided platform businesses*. Coase-Sandor Working Paper Series in Law and Economics.

Gentzkow, M., & Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics*, *126*, 1799–1839.

George, L., & Hogendorn, C. (2012). Aggregators, search and the economics of new media institutions. *Information Economics and Policy*, *24*(1), 40–51.

Ghose, A., Goldfarb, A., & Han, S. (2011). *How is the mobile internet different? Search costs and local activities*. Mimeo, New York University.

Greenstein, S. (2011, April). The direction of broadband spillovers. *Micro Economics*, 103–104.

Greenstein, S., Lerner, J., & Stern, S. (2011). *The economics of digitization: An agenda for NSF*. NSF Report.

Haddad, M. (2010). Associated press Google news partnership ends. Business 2.0 Press, January 12.

Halberstam, D. (2007). *Breaking news: How the Associated Press has covered war, peace, and everything else*. New York, NY: Princeton Architectural Press.

Isbell, K. (2010). The rise of the news aggregator: Legal implications and best practices. *Berkman Center, Research Publication No. 2010-10*.

Jeon, D., & Esfahani, N. N. (2012). *News aggregators and competition among newspapers in the internet*. Working paper.

Krazit, T. (2010). Google, AP reach deal for Google news content. *CNET*.

Lechner, M. (2010). The estimation of casual effects by difference-in-difference methods. *Foundations and Trends in Econometrics*, *4*, 165–224.

Lomas, N. (2013, August 13). German publishers can't wean themselves off Google News, despite winning copyright law change. *TechCrunch*.

Montgomery, A. L., Li, S., Srinivasan, K., & Liechty, J. C. (2004). Modeling online browsing and path analysis using clickstream data. *Marketing Science*, *23*(4), 579–595.

Oberholzer-Gee, F., & Strumpf, K. (2007). The effect of file sharing on record sales: An empirical analysis. *Journal of Political Economy*, *115*, 1–42.

Papke, L., & Wooldridge, J. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, *11*, 619–632.

Papke, L., & Wooldridge, J. (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics*, *145*, 121–133.

Parr, B. (2010). AP and Yahoo Ink Content Deal, Leave Google in Limbo. *Mashable*.

Pfanner, E. (2012). A clash across Europe over the value of a click. *The New York Times*.

Rob, R., & Waldfogel, J. (2006). Piracy on the high C's: Music downloading, sales displacement, and social welfare in a sample of college students. *Journal of Law & Economics*, *49*(1), 29–62.

Rutt, J. (2011). *Aggregators and the news industry: Charging for access to content.* Working paper.

Sandoval, G. (2009). Google May Lose WSJ, Other News Corp. Sites. CNET News, November 9.

Shapiro, C., & Varian, H. (1999). *Information rules: A strategic guide to the network economy*. Boston: Harvard Business School Press.

Sullivan, D. (2010, January 8). Where is AP In Google News? Apparently in Limbo, As contract running out. *Search Engine Land*.

Xu, J., Forman, C., Kim, J., & Van Ittersum, K. (2014). News media channels: Complements or substitutes? Evidence form mobile phone usage. *Journal of Marketing*, *78*, 97–112.

How to cite this article: Chiou L, Tucker C. Content aggregation by platforms: The case of the news media. *J Econ Manage Strat*. 2017;26:782–805. https://doi.org/10.1111/jems.12207.

APPENDIX A

The following contains excerpts from Experian Hitwise "How We Do It" description on its official website.

> *Hitwise has developed proprietary software that Internet Service Providers (ISPs) use to analyze website logs created on their network. This anonymous data is aggregated and provided to Hitwise, where it is analyzed to provide a range of industry standard metrics relating to the viewing of websites including page requests, visits, average visit length, search terms and behavior.*
>
> *Hitwise is able to combine this rich ISP data with data from opt-in panel partners and with region specific consumer demographic and lifestyle information.*
>
> *Hitwise collects aggregate usage data from a geographically diverse range of ISP networks and opt-in panels, representing all types of Internet usage, including home, work, educational and public access. To ensure this data is accurate and representative, it is weighted to universe estimates in each market. Because of the extensive sample size (25 million people worldwide, including 10 million in the United States), Hitwise can provide detailed insights into the search terms used to find thousands of sites as well as a range of clickstream reports, analyzing the movement of visitors between sites.*
>
> *Hitwise only extracts aggregate information. No personal information is seen or captured by Hitwise in according with local and international privacy guidelines. Hitwise's methodology is audited by PricewaterhouseCoopers on an annual basis.*

TABLE A1 Demographic description of users

Measure	Yahoo! News	Google News	New York Times
Male	59.95	63.8	61.21
Age 18–24	12.12	13.89	6.17
Age 25–34	18.05	14.72	13.93
Age 35–44	19.03	17.08	12.98
Age 45–54	21.41	22.24	19.45
Age 55+	29.38	32.06	47.47
Income <30k	22.33	20.77	20.76
Income 30–60k	28.82	27.53	26.36
Income 60–100k	24.95	24.6	24.82
Income 100–150k	14.61	17.5	17.29
Income >150k	9.29	9.6	10.77

Source: Hitwise

Notes: This table reports the percentage of users within each demographic category. Statistics are reported for users of Yahoo! News, Google News, and the New York Times website.

TABLE A2 Linear model of downstream traffic from Google News and Yahoo! News during removal and reinstatement of content from The Associated Press

	(1) Removal	(2) Reinstatement	(3) Falsification	(4) Falsification
APContentRemoval × Google × News	−0.00742**			
	(0.00316)			
APContentRestored × Google × News		0.0102**	0.00247	
		(0.00441)	(0.00434)	
December × Google × News				0.000126
				(0.00246)
APContentRemoval × Google	0.00147			
	(0.00218)			
Google	−0.0102*	−0.00715	−0.0105**	−0.0117**
	(0.00569)	(0.00543)	(0.00475)	(0.00569)
APContentRemoval × News	0.00176			
	(0.00114)			
News × Google	0.0360***	0.0245***	0.0365***	0.0379***
	(0.00805)	(0.00728)	(0.00736)	(0.00818)
APContentRestored × Google		−0.00121	0.000197	
		(0.00279)	(0.00177)	
APContentRestored × News		−0.00697***	−0.00505**	
		(0.00263)	(0.00213)	
December × Google				0.0000501
				(0.00180)
December × News				0.00168
				(0.00186)
Week Fixed Effects	Yes	Yes	Yes	Yes
Website Fixed Effects	Yes	Yes	Yes	Yes
Observations	98730	119640	103113	84048

Notes: Robust standard errors clustered at website level. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. The outcome variable is the percentage of traffic (measured between 0 and 100) to a website after visiting Google News or Yahoo! News. The specifications are estimated with OLS. In Column (1), the removal of The AP content is the removal of hosted articles by The Associated Press from Google News in January 2010. In Column (2), the restoration of The AP content is the restoration of hosted articles by The Associated Press in Google News in October 2010. In Column (3), the falsification check compares December 2009 and October 2010 when The AP content was available in Google News in both months. In Column (4), the falsification check compares November 2009 and December 2009 to test for a pre-trend.

TABLE A3 Downstream traffic for top sites from Google News and Yahoo! News

	(1) Removal	(2) Reinstatement	(3) Falsification	(4) Falsification
APContentRemoval × Google × News	0.569*			
	(0.175)			
APContentRestored × Google × News		2.666*	1.451	
		(1.369)	(0.477)	
December × Google × News				0.689
				(0.227)
APContentRemoval × Google	1.304			
	(0.374)			
Google	0.700	0.914	0.817	0.809**
	(0.261)	(0.177)	(0.104)	(0.0869)
APContentRemoval × News	1.193*			
	(0.120)			
APContentRestored × Google		0.749	0.990	
		(0.333)	(0.233)	
APContentRestored × News		0.524**	0.635**	
		(0.137)	(0.139)	
December × Google				1.010
				(0.238)
December × News				1.575**
				(0.346)
Week Fixed Effects	Yes	Yes	Yes	Yes
Website Fixed Effects	Yes	Yes	Yes	Yes
Observations	1746	1960	1755	1755

Notes: Robust standard errors clustered at website level. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. The outcome variable is the fraction of traffic to a website after visiting Google News or Yahoo! News. The specifications are estimated using the sample of websites with traffic levels among the top 100. The exponentiated coefficients are reported with the corresponding standard errors for interpretation as odds ratios. Note that because the coefficients are exponentiated, we interpret them relative to the value of one, which represents no effect ($exp(0) = 1$). In other words, if the policy change has a negative effect (coefficient is less than zero), we would expect an odds ratio or exponentiated coefficient to be less than one; if the policy change has a positive effect (coefficient is greater than zero), we would expect an odds ratio or exponentiated coefficient to be greater than one. In Column (1), the removal of The AP content is the removal of hosted articles by The Associated Press from Google News in January 2010. In Column (2), the restoration of The AP content is the restoration of hosted articles by The Associated Press in Google News in October 2010. In Column (3), the falsification check compares December 2009 and October 2010 when The AP content was available in Google News in both months. In Column (4), the falsification check compares November 2009 and December 2009 to test for a pre-trend.

TABLE A4 Robustness check of downstream traffic from Google News and Yahoo! News during removal and reinstatement of content from The Associated Press

	(1) Removal	(2) Reinstatement	(3) Falsification	(4) Falsification
APContentRemoval × Google × News	0.718*			
	(0.127)			
APContentRestored × Google × News		1.955**	1.347	
		(0.553)	(0.259)	
December × Google × News				0.964
				(0.115)
APContentRemoval × Google	1.118			
	(0.172)			
Google	1.690	1.978	1.643	1.596
	(3.518)	(2.908)	(1.546)	(1.830)
APContentRemoval × News	1.114*			
	(0.0703)			
News × Google	0.811	0.575***	0.718***	0.667***
	(0.109)	(0.0837)	(0.0849)	(0.0563)
APContentRestored × Google		1.375	1.386	
		(1.604)	(1.139)	
APContentRestored × News		0.601***	0.680***	
		(0.0899)	(0.0869)	
December × Google				1.045
				(0.161)
December × News				1.114
				(0.121)
Week Fixed Effects	Yes	Yes	Yes	Yes
Website Fixed Effects	Yes	Yes	Yes	Yes
Trends	Yes	Yes	Yes	Yes
Observations	98,730	119,640	103,113	84,048

Notes: Robust standard errors clustered at website level. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. The outcome variable is the fraction of traffic to websites after visiting Google News or Yahoo! News. All regressions include direct controls for trends in visits to Google News and Yahoo! News. The exponentiated coefficients are reported with the corresponding standard errors for interpretation as odds ratios. Note that because the coefficients are exponentiated, we interpret them relative to the value of one, which represents no effect ($exp(0) = 1$). In other words, if the policy change has a negative effect (coefficient is less than zero), we would expect an odds ratio or exponentiated coefficient to be less than one; if the policy change has a positive effect (coefficient is greater than zero), we would expect an odds ratio or exponentiated coefficient to be greater than one. In Column (1), the removal of The AP content is the removal of hosted articles by The Associated Press from Google News in January 2010. In Column (2), the restoration of The AP content is the restoration of hosted articles by The Associated Press in Google News in October 2010. In Column (3), the falsification check compares December 2009 and October 2010 when The AP content was available in Google News in both months. In Column (4), the falsification check compares November 2009 and December 2009 to test for a pre-trend.