Volume 13 Number 8 October 2025

DIGITAL JOURNALISM

Digital Journalism

Routledge Taylor & Francis Group

ISSN: 2167-0811 (Print) 2167-082X (Online) Journal homepage: www.tandfonline.com/journals/rdij20

The Role of Syndicated Content in Retaining Digital Newspaper Subscribers: Evidence from Clickstream and Subscription Data

Jaewon Royce Choi, Su Jung Kim, Yayu Zhou & Edward C. Malthouse

To cite this article: Jaewon Royce Choi, Su Jung Kim, Yayu Zhou & Edward C. Malthouse (17 Feb 2025): The Role of Syndicated Content in Retaining Digital Newspaper Subscribers: Evidence from Clickstream and Subscription Data, Digital Journalism, DOI: 10.1080/21670811.2025.2466029

To link to this article: https://doi.org/10.1080/21670811.2025.2466029

DIGITAL JOURNALISM
https://doi.org/10.1080/21670811.2025.2466029

Routledge
Taylor & Francis Group

ORIGINAL ARTICLE

The Role of Syndicated Content in Retaining Digital Newspaper Subscribers: Evidence from Clickstream and Subscription Data

Jaewon Royce Choi[a], Su Jung Kim[b], Yayu Zhou[c] and Edward C. Malthouse[d]

[a]Manship School of Mass Communication, Louisiana State University, Baton Rouge, LA, USA; [b]Annenberg School for Communication and Journalism, University of Southern California, Los Angeles, CA, USA; [c]Department of Industrial Engineering and Management Sciences, McCormick School of Engineering, Northwestern University, Evanston, IL, USA; [d]Spiegel Research Center, Medill School of Journalism, Media, Integrated Marketing Communications, Northwestern University, Evanston, IL, USA

ABSTRACT

Syndicated content from news agencies remains a pivotal component of newspaper coverage. Especially for smaller local/regional news organizations with limited resources, syndicated content helps them deliver important news that is beyond their capabilities. However, research on syndicated content has heavily focused on its potentially negative impact such as content homogenization or extensive news recycling (i.e., churnalism) from the production side, whereas investigations on the effects of syndicated content on news audience engagement have been an uncharted territory. Considering the importance of syndicated content for complementing quality local/regional journalism, understanding how it contributes to building readership is especially important for local/regional news organizations. We aim to fill this gap by investigating the impact of syndicated content on news reading behaviors. Using news subscriber behavioral data from two local/regional news sites in the US, we identify subscriber segments and study how syndicated content differentially influences the regularity of news reading. Results indicate that syndicated content can help develop news reading habits of relatively less engaged subscribers and imply that local/regional news organizations should strive to understand their subscriber segments deeper, and syndicated content can be a cost-efficient way to achieve balanced news portfolios when developing content strategies.

KEYWORDS

Local newspapers; syndicated content; news agency; subscription; clickstream data; subscriber retention; news reader segments

Introduction

The advent of digital and social media has brought changes to the production, distribution, and consumption of news. For instance, production has become more diversified (e.g., citizen journalism, Documenters Network[1]) (Wall 2015). Furthermore,

CONTACT Jaewon Royce Choi roycechoi@lsu.edu

digital information sources have become increasingly central to journalists' news sourcing practices (Van Leuven et al. 2018). The triopoly of Alphabet (Google), Meta, and Amazon, reaped 64% of all US digital advertising spending in 2021 (Lebow 2021) and the total estimated circulation revenue for news organizations surpassed advertising revenue for the first time in 2020 (Pew Research Center 2021). Facing such challenges, news organizations had to reassess their business models, implementing various strategies to make readers pay for news, including revamping digital subscriptions (Goyanes 2014; Newman et al. 2021; Sjøvaag 2016).

Despite the rise of digital subscriptions, the financial struggles of news organizations remain, and maintaining newsroom resources is a big challenge. The number of newsroom employees has continuously declined, harming the resource-limited local and regional newspapers more severely (Abernathy 2022). One of the breakthroughs of dealing with such financial and human resource scarcity is economizing the cost that is associated with the news-gathering process, for example, using news content provided by news agencies, news wire services, or other syndicated sources (hereafter called syndicated content). By using news content from these news wholesalers, news organizations can reduce news production costs.

Against the backdrop of the role and influence of syndicated content, this study revisits the value of news agency content from the perspective of sustaining local/regional news organizations. Previous studies have mostly viewed syndicated content in the context of structural consolidation and the reduction of content diversity as a result of extensive syndication, potentially leading to a long-term negative effect on sustaining subscribers (Adams and Baldasty 2001). On the other hand, a study by Sjøvaag (2014) finds that extensive use of syndicated content does not lead to homogenization. McPherson (2002) notes a positive side of using syndicated content, pointing out that reliance on syndicated content allows smaller news outlets to spare the resources they can use otherwise for local or regional coverage. However, the general sentiment surrounding syndicated content has been negative. This study offers a unique perspective on the implications of syndicated content by examining its impact on audiences, more specifically how reading it differentially influences news reading habits and subscription decisions by news subscriber segments. Using a unique dataset that links subscribers' clickstream data and payment records at the individual level, this study directly examines news reading and subscription behaviors.

Specifically, we identify several audience groups based on their news reading patterns and investigate how news content provided by the Associated Press (AP) affects the regularity of their online news reading and decision to renew or cancel their subscription. We choose the AP because it is one of the largest international news agencies. The AP operated 243 locations in 96 countries worldwide and created 440,000 stories in 2021 (Associated Press 2022). This study is, to our best knowledge, one of the few using clickstream and subscription data at the individual subscriber level to estimate the effects of syndicated content on news reading behavior and subscriber retention. Our investigation adds to the burgeoning literature on audience engagement by focusing on reception-oriented engagement (Nelson 2021). We consider exposure-based engagement as the "behavioral manifestation of audience engagement (Kim, Wang, and Malthouse 2022, 1017)" that

reflects the value that digital news audiences seek when they consume news (Olsen 2021). This study also makes a contribution by providing an understanding of the effects of news on audiences, specifically the paying subscribers who are of high value to news organizations. Moreover, we provide practical implications for news organizations on how to strategically incorporate syndicated content and use clickstream data to optimize the allocation of newsroom resources and selection of diverse content.

Literature Review

Existing Research on Syndicated Content and News Agencies in the News Industry

News agencies have played a central role in providing stories for news organizations. They are news wholesalers providing news contents to clients and are indispensable in the news industry. News stories provided by news agencies relieve news organizations from the burden of having journalistic resources in every corner. News agencies operate at both international and national levels, and news agencies have redefined their key features in recent years, especially after the advent of advanced communication technologies (e.g., shift of focus from text-based to more picture- or video-based news service) (Boyd-Barrett 2008, 2015; Lorenz 2017; Rantanen 2019). The Internet has brought greater opportunities for news agencies as their resources and brand reputation offered added-value to news organizations facing a fierce competition in the online space (Boyd-Barrett 2008). While the Internet also allowed news agencies to be news "retailers" (i.e., providing news content directly to consumers), online news audiences still access news agency's content through intermediaries such as news aggregators or individual newspaper websites (Boyd-Barrett 2015).

In academic research, the influence of syndicated content has been discussed with a critical tone. For instance, news flow studies have shown disproportionate amounts of syndicated news content originating from the West and ignoring large parts of the world (Rantanen 2019). Another line of research on news agencies points out the issues of "churnalism," which refers to the journalistic practice involving the inflow of both PR material and news wire copies being recycled extensively without journalists' critical assessment of the content (Davies 2009; Johnston and Forde 2017). Critique from the churnalism perspective extends to the discussion of gatekeeping as news agencies often become a convenient and accessible source of information for journalists. Studies find that news agencies play an important gatekeeping role, especially for online newspapers. For instance, researchers find that the influence of the news agency is substantial on online newspapers, whereas it is only marginal on print newspapers in the Dutch news industry; thus, news agencies are a gatekeeper among gatekeepers (Welbers et al. 2018). Furthermore, when news agencies are one of few correspondents with large international institutions such as the EU for small news organizations, they are often closely involved in the agenda building process (Lorenz 2017). Such growing dependence on syndicated content from elite agency sources has generally conjured concerns about the potential decrease of quality independent reporting and journalistic norms (Saridou, Spyridou, and Veglis 2017).

Syndicated content has also been examined in the context of market consolidation and news homogenization. An early empirical investigation of a national media corporation concluded that the company's heavy dependence on syndicated materials and the lack of local coverage had a substantial long-term impact on the company's subordinated market position (Adams and Baldasty 2001). Another study in the context of Norwegian regional media examines influences of consolidation strategies, including content syndication, on content homogenization. The comparative content analysis concludes that there is not enough evidence of content overlap to the level of significant homogenization, and content more suitable for syndication (e.g., sports and lifestyle) could rather be seen to reflect the differentiation strategies of regional newspapers (Sjøvaag 2014). In the context of the Dutch newspaper industry, a large-scale textual analysis finds that prominent online newspapers depend substantially on the only national news agency (Boumans et al. 2018).

Overall, there is a relative lack of attention to syndicated content despite its importance in the media industry. Existing literature has discussed the potential impact of syndicated content from a more critical viewpoint. However, findings of empirical studies remain equivocal on whether high dependence on news agencies leads to detrimental effects on the general public by failing to deliver diverse news stories. Investigations of relationships between news agencies catering to wider audience segments and local/regional newspapers remain scarce.

Furthermore, previous research on syndicated content has been mostly historical archival studies, newsroom ethnographies, or quantitative content analysis from the production side of news (Rantanen 2019), neglecting the audience point of view. How news audiences consume syndicated content, especially those who are paying for news, and how reading syndicated content subsequently impacts reading habits and audience loyalty has rarely been investigated.

While engagement can be defined differently by contexts at various scopes (Choi, Hong, and Kim 2023; Malthouse and Peck 2011; Nelson 2021), content characteristics including both formal characteristics and news factors can be influential to audience engagement with news. Formal characteristics such as original news outlet, article length, news section, and article visualization impact the degree to which the audience shares them on social media (Karnowski et al. 2021; Trilling, Tolochko, and Burscher 2017). News topics or news values also affect what the audience reads. For instance, contrary to journalists' expectations, news readers read less public affairs articles (Boczkowski and Mitchelstein 2013). News values such as geographical proximity, cultural proximity, and presence of conflict also influences individuals' tendency to share news stories on social media (Karnowski et al. 2021; Trilling, Tolochko, and Burscher 2017). Besides sharing news on social media, content characteristics are related to how much the audience visits news publishers directly. Audiences visit the news site more frequently when they read more news about local issues, entertainment, or timely matters (e.g., COVID-19 news during the pandemic), whereas national/world news or general business can drive them away (Kim, Wang, and Malthouse 2022).

While news topics matter, whether news content is syndicated or not can be another significant content factor. Studies have found that syndicated content is less likely to be shared on social media and has a negative impact on subscriber retention

and regularity (Kim et al. 2024; Trilling, Tolochko, and Burscher 2017). However, there is yet an investigation that focuses on how syndicated content is related to the reading and subscription behaviors of local/regional news audiences. Audiences could engage differently with original content from local news organizations than syndicated content. In addition to audiences' perceived credibility or reputation of the news agency that appear in the byline, the extent and depth of coverage could differ between contents from resource-constrained local news publishers and news agencies. For instance, for breaking news such as crime or accidents, local news content may cover these incidents in the vicinity of the organization's community, whereas syndicated content may cover them in larger areas (e.g., regional, national, international). These differences reflect how syndicated content can interact with news values (Karnowski et al. 2021), potentially influencing how audiences consume news. Understanding the ways in which syndicated content may help local/regional newspapers attract user traffic, build readership, and increase subscriber retention are equally critical inquiries that have not been studied. Addressing the knowledge gaps is important especially because it will not only add knowledge about the significance of syndicated content from the audience perspective, but also because it can provide practical implications regarding sustainability and content strategies for local/regional newspapers.

The Dual Role of Syndicated Content as Traffic Driver and Habit Builder

Despite the mostly negative implications of syndicated content discussed in previous studies, scholars also note the benefits of syndicated content. In addition to supplying news to small local newspapers with scarce resources (McPherson 2002), syndicated content can drive user traffic to news sites for interested readers. Previous studies on paywalls have found that newspapers place syndicated content before paywalls so that they can increase site traffic (Myllylahti 2017; Sjøvaag 2016). Syndicated content covers broader national or international issues and is mostly available elsewhere. Some raise concerns that such a wide availability paired with hard paywalls may jeopardize newly incoming traffic, potentially hurting long-term advertising revenues (Myllylahti 2016; Pickard and Williams 2014), while others show evidence of the positive effect of paywalls on advertisers by offering more loyal eyeballs (Olsen and Solvoll 2018). However, to the best of our knowledge, there has not yet been discussions about the role of syndicated content in maintaining and attracting traffic from already *subscribed* readers. Given that syndicated content carries timely news on national/international issues as well as broader subject categories that cater to wider user interests, this can be an effective strategy especially for those readers who have a "loose connection" to the news sites at the top of the digital subscription funnel (Benton 2019). For these casual or light news readers, syndicated content can function as a gateway to repeated visits to the news site, and eventually a continuation of their paid subscriptions.

This also implies that syndicated content may not work for all types of news readers. Previous research on post-paywall user behaviors shows that long-time subscribers tend to exhibit more active usage patterns in terms of site visits, usage time, platforms used, and categories consumed, demonstrating stronger reading habits that have

been established over time (Wadbring and Bergström 2021). These long-term subscribers are more committed readers who tend to seek out unique content that they cannot find elsewhere, compared to short-term subscribers. Less committed readers may be more interested in news stories that are of immediate and disruptive nature such as crime, accidents, or traffic and weather (Sjøvaag 2016) or that are more shareworthy (Trilling, Tolochko, and Burscher 2017). Furthermore, common gratifications sought in news include surveillance and staying informed about the broader social and political environment (Ruggiero 2000). Given that stories from wire services mostly deliver topics that appeal to a mass audience rather than catering to specific locales, we expect that syndicated content will have a more positive impact on building reading habits for less committed readers than more committed readers who have already formed their distinctive reading habits.

Based on our discussion of previous literature on the role and nature of syndicated content and online news reading behavior, we address the following research questions and hypothesis. We first ask how reading different categories of news content impacts people's reading habits (i.e., regularity) across news subscriber segments identified in our sample. We then hypothesize that syndicated content will positively contribute to habit building of less committed subscribers because of its wide coverage on national/international news stories as well as breaking news including wars, crimes, or disasters. Then, we expand our inquiry of the role of different news content, syndicated content specifically, in subscriber retention. Given the nascent nature of connecting the readership of different content types and subscriber retention of different subscriber segments, we pose a research question:

RQ1: How does reading different content types influence the regularity of reading across subscriber segments?

H1: Syndicated content will differentially influence the regularity of reading across subscriber segments. More specifically, syndicated content will increase the regularity of reading among subscriber segments that exhibit a minimal or moderate level of news reading.

RQ2: How does reading different content types influence the subscriber retention across subscriber segments?

Methods

We present two complementary analyses that address our hypothesis and research questions. Before conducting the analyses, we use factor analysis to reduce the dimension of raw news content tags (e.g., politics, sports, entertainment, etc.) and create high-level factors that represent the aggregated page views based on those tags, which will then be used as predictors in later modeling steps to avoid multicollinearity. Then we perform a *k*-means cluster analysis to identify groups of subscribers with similar reading preferences. Subscribers are grouped based on the (logged and standardized) number of page views of each high-level factor they consume in

each month. Having the news content factors and subscriber segments, the first analysis investigates how different news content categories, including syndicated content affect subscribers' reading habit, i.e., regularity (how often a subscriber reads per month), and test whether the effects differ by subscriber segments identified from the *k*-means cluster analysis (H1 & RQ1). In the second analysis, we build a multivariate survival model to explore how different content categories influence subscriber retention (RQ2).

Data

We collect clickstream data from two local/regional news sites in the US that can be matched with subscriber payment history. The two sites (labeled Sites A and B) cover one larger and one midsize city, respectively. Our samples consist of 111,449 and 3994 digital-only subscribers from Sites A and B. Our study spans 24 months (2020-01-01 to 2021-12-31) and 19 months (2020-05-01 to 2021-12-31) for the respective sites. We merge the subscriber payment history with the raw clickstream data recording every click that a subscriber makes, and aggregate the joined data into monthly periods. As a result, each row of our data contains information about a subscriber's digital reading behavior and their subscription status of the given month. Table 1 presents the descriptive statistics of the variables used in our models.

Measures

Dependent Variables

We study two dependent variables. One is regularity, defined as the number of days the subscriber reads the digital newspaper in one month (H1, RQ1). Here we assume subscribers read something every time they visit, therefore "reading the digital newspaper" means simply visiting the news website. The second is the cancellation of a digital news subscription, i.e., churn (RQ2), a binary variable indicating whether a digital news subscriber canceled payment during the current month. To avoid issues of simultaneity, dependent variables regularity and churn refer to the next month and we use independent variables from the current month. Table 1 shows the average percentage of subscription cancellation and regularity in two sites.

Table 1. Descriptive statistics of dependent variables and selected independent variables.

	Site A	Site B
Number of subscribers	111,449	3,994
Number of observations	1,443,425	25,203
Dependent variables		
Cancellation of digital subscription	0.028 (0.165)	0.006 (0.077)
Regularity	1.719 (1.289)	1.161 (1.161)
Independent variables		
News content – News	1.735 (1.571)	1.140 (1.352)
News content – AP	0.246 (0.540)	0.291 (0.695)
News content – Sports	0.482 (0.940)	0.165 (0.482)
News content – Opinion	0.207 (0.542)	
Customer tenure	103.674 (299.028)	52.259 (92.000)

Standard deviation is provided in the parenthesis.

News Content Factors

We obtain news content categories used by each local newspaper and page views for each category. These category tags are classified as high-level factors using principal component analysis (PCA) with a varimax rotation (Appendix A). Among the content categories is a tag provided by the news organization whether the content originates from the AP. The page view counts of articles are right-skewed, so we log-transformed them. We also standardize the data to make sure all the predictors are on a comparable scale. We create four content category factors from Site A: *AP* factor represents all articles published by the AP, including business, entertainment, national news, etc. *Sports* factor combines high school, college, professional sports, and AP sports. The third factor is *opinion* and the last factor is *news* including the remaining general news. Similarly, the PCA generated three categories (*news*, *AP*, *sports*) from Site B. These content factors will be used as independent variables.

Subscriber Segments

Subscriber segments are generated using *k*-means cluster analysis and are used as moderating variables. We conduct analyses generating up to eight clusters, and conclude that the four-cluster solution represents data from both sites the best. The four-cluster solution offers better interpretability against models with larger numbers of clusters. The resulting subscriber segment variable is a four-level categorical variable with each level indicating a segment of subscribers with similar reading preference (i.e., subscribers belonging to the same segment have a similar amount of page views of high-level content factors).

Control Variables

Across all models, we use a spline to capture the baseline hazard function, which describes the risk of cancellation conditional on current survival and represents a lifetime distribution. We often find that new subscribers are more likely to cancel and the probability of churn diminishes the longer a subscriber has been around. We include a four-level factor variable representing seasonality (fall is the baseline) to control for seasonality. Regularity, page views and customers' tenure in months are numerical with right-skewed distribution and outliers, so we log transformed them after adding an offset of 1.

Analysis

We first fit multiple linear regression models to understand how reading behaviors influence regularity. To answer RQ1 and test H1, we include time-dependent covariates tracking the logged number of page views for each content factor, the four-level cluster variable indicating to which cluster one subscriber belongs, and the interaction between different content page views and cluster. To address RQ2, we next build multivariate, discrete-time survival models (Allison 2010). Discrete-time survival models are logistic regression models that have one observation for every month where the subscriber was active. The dependent variable, churn, was described above with the other measures. The model uses the same control variables and independent variables as in the linear regression models.

Results

Identifying News Subscriber Segments

Table 2 shows descriptive statistics of the four subscriber segments of each Site. Note that the page views of content factors are standardized; thus 0 is the overall average. Generally, subscriber segments of Sites A and B show fairly similar content preference patterns except for one segment. For Site A, segment 1 represents 48% of the sample and corresponds to those subscribers whose reading page views (log transformed) of the four content categories are below average. Segment 2's page views of reading *sports* and *news* contents are above average while that of *AP* content is only slightly above the average. The size of this segment is much smaller than segment 1, representing 10% of the sample. In contrast, subscribers in segment 3 have the highest average reading frequency of all four content categories (large number of *opinion* and *AP* articles and moderate number of *news* and *sports*). However, this is the smallest segment, representing only 8% of the Site A sample. Finally, segment 4 represents the remaining 33% of the sample, and only shows slightly above average readership in general news content with no clear preference for other content types. According to subscribers' content preferences, we call segment 1 *inactive readers* who read very few articles of any type. In other words, they show below average readership for all types of articles. Segment 2 (*sports & news readers*) is a group of subscribers with an interest in sports and news articles but less interest in AP and opinion content. *Heavy readers* in segment 3 read a large amount of all types of articles. We refer to segment 4 as *news light readers* because they read slightly above average number of news articles and very few of the other three types.

As presented in Table 2, the four segments from Site B show similar patterns to Site A and are labeled as: *inactive readers*, *AP & news readers*, *heavy readers*, and *news light readers,* respectively. Note that in both sites, the *inactive readers* segment comprises the largest proportion of subscribers (48% in Site A and 60% in Site B), while the smallest group is always the *heavy readers* (8% in Site A and 6% in Site B). This reveals that only a small portion of subscribers actively read a lot of news, while a large number of customers pay for subscriptions but rarely read the paper.

Table 2. Summary of the four-cluster solution from cluster analysis for Site A and B.

Clusters	Size (%)	Cluster Means				RMSE	Total page views (SD)	Regularity (SD)
		News	Ap	Sports	Opinion			
A1: *Inactive readers*	696,045 (48%)	−0.90	−0.44	−0.46	−0.36	0.239	0.788 (1.077)	0.598 (0.803)
A2: *Sports + News*	147,320 (10%)	0.84	0.46	2.11	−0.08	0.862	4.663 (0.901)	3.025 (0.438)
A3: *Heavy readers*	118,852 (8%)	1.53	2.09	1.10	2.62	1.239	5.370 (0.842)	3.212 (0.337)
A4: *News Light*	481,208 (33%)	0.66	−0.01	−0.26	−0.09	0.612	3.647 (0.931)	2.573 (0.585)
B1: *Inactive readers*	15,055 (60%)	−0.71	−0.39	−0.31	—	0.232	0.215 (0.395)	0.404 (0.670)
B2: *AP + News*	2,025 (8%)	1.48	1.97	0.08	—	0.782	3.403 (0.770)	2.726 (0.611)
B3: *Heavy readers*	1,512 (6%)	1.75	2.02	3.33	—	1.375	3.909 (0.l926)	2.974 (0.514)
B4: *News Light*	6,611 (26%)	0.76	−0.18	−0.07	—	0.530	2.235 (0.649)	1.994 (0.666)

The Influence of Reading Different Content Types on Regularity by Subscriber Segments

Table 3 presents the parameter estimates of the linear regression models. To answer RQ1, we include interaction terms between the page views of different content types and the cluster of subscribers to see how the level of individuals' page views of different types affect regularity and whether the effects differ by subscriber segments. For the cluster variable, *inactive readers* are used as the reference group. In Site A, the main effects of segments are positive and significant, which indicates that compared to the reference group of *inactive readers*, all other segments have significantly higher regularity on average when page view is zero. This is consistent with the average regularity across segments presented in Table 2. More specifically,

Table 3. Panel regression models Predicting regularity using subscriber segments, content variables and the interaction between segments and content variables.

	Dependent Variable: Regularity in next month			
	Site A		Site B	
	Slope	*SE*	*Slope*	*SE*
bs(custtime)1	0.6594***	0.0134	0.9716***	0.0593
bs(custtime)2	−1.1146***	0.0430	−0.2254**	0.0755
bs(custtime)3	0.0866***	0.0034	0.5512***	0.0587
seasonSpring	0.0308***	0.0017	−0.0671***	0.0152
seasonSummer	0.0220***	0.0017	−0.0237	0.0140
seasonWinter	0.1449***	0.0018	−0.0437**	0.0158
seg[a]Sports + News[b]	1.4439***	0.0079		
segHeavy	1.7274***	0.0115	0.9392***	0.0922
segNewsL	0.5823***	0.0041	−0.2364***	0.0353
segAP + News			0.1776*	0.0714
type[c]News	0.8274***	0.0017	0.5191***	0.0192
typeAP	0.6343***	0.0130	0.4278***	0.0514
typeSports	0.6537***	0.0037	0.3291***	0.0594
typeOpinion	0.4662***	0.0089		
segSports + News:typeNews	−0.5904***	0.0026		
segHeavy:typeNews	−0.5678***	0.0036	−0.1281***	0.0317
segNewsL:typeNews	−0.3196***	0.0022	0.1509***	0.0245
segSports + News:typeAP	−0.6067***	0.0135		
segAP + News:typeNews			0.0479	0.0288
segHeavy:typeAP	−0.6435***	0.0134	−0.4355***	0.0568
segNewsL:typeAP	−0.5465***	0.0133	−0.1287*	0.0612
segSports + News:typeSports	−0.5092***	0.0043		
segAP + News:typeAP			−0.3644***	0.0582
segHeavy:typeSports	−0.6190***	0.0041	−0.2682***	0.0683
segNewsL:typeSports	−0.4878***	0.0045	−0.0579	0.0679
segSports + News:typeOpinion	−0.4293***	0.0105		
segHeavy:typeOpinion	−0.5156***	0.0093		
segNewsL:typeOpinion	−0.3724***	0.0094		
segAP + News:typeSports			−0.1389	0.0835
Constant	0.2888***	0.0016	0.2536***	0.0132
Observations	1,331,976		21,209	
R^2	0.7112		0.5813	
Adjusted R^2	0.7112		0.5808	
Residual Std. Error	0.6967 (df = 1,331,950)		0.7673 (df = 21,187)	
F Statistic	131,229.8000*** (df = 25; 1,331,950)		1400.4870*** (df = 21; 21,187)	

[a]seg refers to segment variables; [b]Reference group is *inactive reader* segment; [c]type refers to content type variables.

$p < 0.05$*; $p < 0.01$**; $p < 0.001$***; NewsL refers to *news light* segment.

heavy readers have the highest regularity level (3.212), followed by *sports & news* readers (3.025), *news light readers* (2.573), and *inactive readers* (0.598).

When it comes to the main effects of content variables, all news content variables' effects are statistically significant. *News* (0.827), *AP* (0.634), *sports* (0.654), and *opinion* (0.466) are positively associated with regularity, which suggests that reading different types of content helps subscribers form habits to read more often. RQ1 asks whether such effects differ across subscriber segments. Furthermore, H1 posits the existence of greater effects of AP content on the regularity of subscribers with minimal or moderate level of news reading. This requires us to further investigate the interactions. Figures 1 and 2 show interaction plots of Site A and Site B.

Results show significant interaction effects, indicating that while reading different content categories all increase regularity, such effects vary across subscriber segments. To test H1, we first focus on AP content. The slope of AP for *inactive readers* is positive (0.634) and significantly different from 0, which implies that *inactive readers* who read more AP articles tend to visit the site more often. The interaction terms between segment and AP represent the difference in the slope of AP for the corresponding segment to the reference group. Thus, the AP slopes for *sports & news*, *heavy* and *news light readers* are computed as $0.6343 - 0.6067 = 0.0276$; $0.6343 - 0.6435 = -0.0092$; $0.6343 - 0.5465 = 0.0878$, respectively. In summary, AP content has a positive association with regularity for all segments except *heavy readers* and its effect is the strongest for *inactive readers*. In other words, AP content can develop reading habits and it works effectively among *inactive readers* and *news light readers*. For *heavy readers,* who

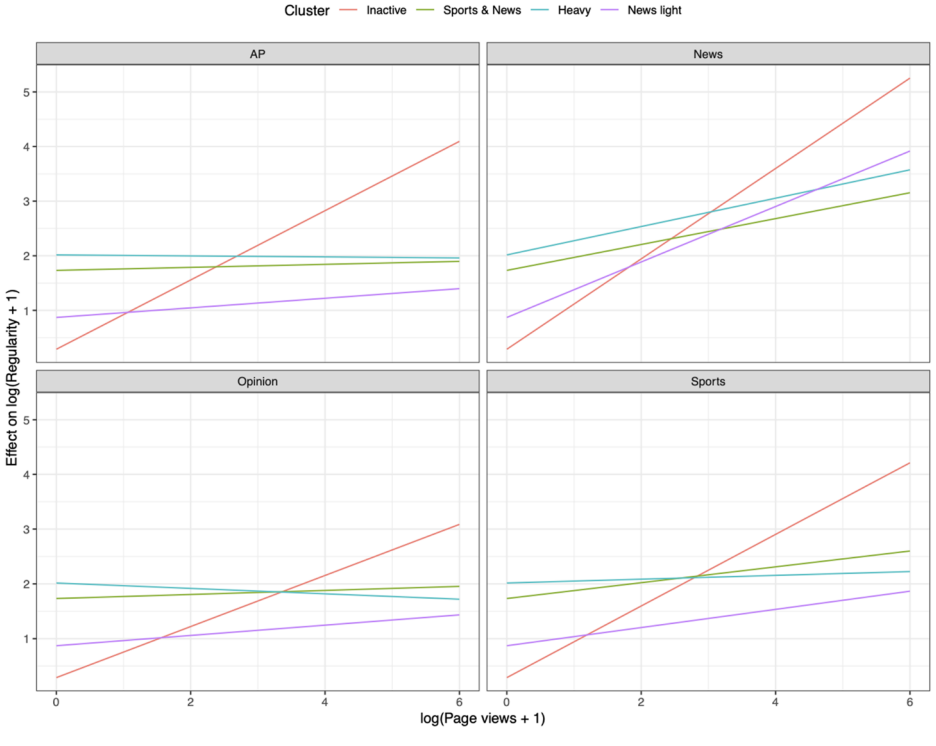

Figure 1. Interaction Plot for four-cluster × four content factors on regularity in Site A.

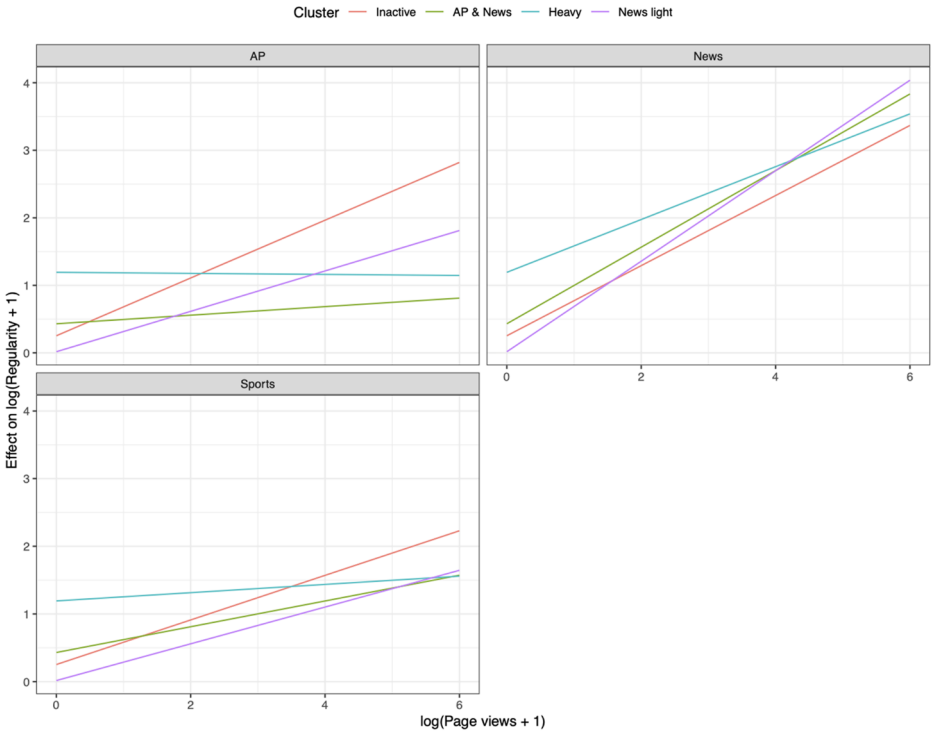

Figure 2. Interaction Plot for four-cluster × three content factors on regularity in Site B.

have already formed a habit or preference, AP content only has a neutral effect. The magnitude of the interaction effect is near zero. This confirms H1 for Site A sample. For the remaining three contents, general news content shows overall positive interaction effects on all subscriber segments, including *heavy readers*. Interaction effects of sports content on *news light readers* and *sports & news* are slightly positive, whereas for *heavy readers* it is neutral. Opinion content's effects on the remaining three subscriber segments are marginal. For *inactive readers*, any type of article facilitates developing a reading habit.

In the case of Site B, we observe results similar to Site A. For the subscriber segment variable, the main effects of segment groups are positive and statistically significant except for *news light readers*. This indicates that compared to *inactive readers*, *AP & news* readers, and *news heavy readers* have significantly higher regularity. In terms of main effects of content variables, all news content types show significant positive influence on regularity. An increase in reading general news (0.519), AP (0.428), and sports (0.329) content all contribute to building news reading habits. Examination of interaction terms finds that they are generally statistically significant, which indicates that effects of reading different news categories vary among subscriber segments. Focusing on the AP content effects, the AP slopes for *AP & news*, *heavy*, and *news light readers* are calculated as $0.4278 - 0.3644 = 0.0634$; $0.4278 - 0.4355 = -0.0077$; $0.4278 - 0.1287 = 0.2991$, respectively. Therefore, AP content has a positive association with regularity for all segments except *heavy readers* and the effect is

most prominent for *inactive readers* followed by *news light readers*. This echoes the findings from Site A, confirming H1 for Site B.

The Influence of Reading Different Content Types on Churn by Subscriber Segments

Our second research question addresses how reading different content types influences subscriber retention across the four subscriber segments. We fit a discrete-time survival model to Site A. For Site B, the sport content type variable shows little variation, compromising the validity of the result. Therefore, we decide to focus on the results from Site A. Table 4 presents the results of the discrete-time survival model using the sample in Site A and Figure 3 shows the interaction effects between content types and subscriber segments on churn. In terms of the subscriber segments, our results indicate that *news light readers* are significantly more likely to churn (0.315), whereas *heavy readers* (−0.423) are less likely to churn than *inactive readers*.

The direct effects of different content types on subscriber retention are statistically significant for general news (−0.269) and sports (−0.095) content. The interaction terms between content types and subscriber segments imply that reading general news significantly lowers the probability of canceling subscription next month for all subscriber segments. Interestingly, the results indicate that AP and opinion content may have adverse effects on subscriber retention for *heavy readers*. That is, reading more AP or opinion content could increase the risk of *heavy readers* canceling their subscription. The nuanced difference is also illustrated in Figure 3. While directionally AP content reduces the churn probability of *inactive readers*, *sports & news readers*, and *news light readers*, it has the opposite effect among *heavy readers*. Overall, the direct influence of content types on subscriber retention among subscriber segments is not as pronounced as in the case of regularity.

Discussion and Conclusion

Syndicated content has been criticized for promoting "churnalism" (Johnston and Forde 2017) by journalism scholars and professionals yet has not received much attention on its contribution to smaller local/regional news organizations or its role in attracting news readers and/or building news readership. Aligning with the burgeoning literature on audience engagement, especially exposure-based engagement behaviors, this study examines how reading syndicated content influences digital news reading habits across different subscriber segments. More specifically, this study reshifts its focus to an understudied subject in news audience behaviors—the role of syndicated content in news engagement behavior among less engaged audience segments. We do this by investigating whether and how AP news stories help digital news subscribers with a low to moderate level of news reading build a habit of reading digital news to deepen our understanding of how we can promote audience engagement of those with mediocre consumption levels by finding which content types they consider valuable. This study also examines whether the influence of syndicated news content readership extends to subscriber retention, substantially impacting audience revenue. In doing so, we use unique data sets—individual-level

Table 4. Discrete-time survival models predicting cancellation of digital news subscription using segments, content variables and the interaction between segments and content variables.

	Dependent Variable: Churn in next month	
	Site A	
	Slope	*SE*
bs(custtime)1	−18.8799***	0.1956
bs(custtime)2	34.9733***	0.4674
bs(custtime)3	−2.9731***	0.0512
seasonSpring	0.2247***	0.0141
seasonSummer	0.0409**	0.0144
seasonWinter	−0.0744***	0.0157
seg[a]Sports + News[b]	0.0729	0.1106
segHeavy	−0.4230*	0.1673
segNewsL	0.3147***	0.0409
segAP + News		
type[c]News	−0.2689***	0.0122
typeAP	−0.1594	0.0956
typeSports	−0.0953**	0.0299
typeOpinion	0.0099	0.0638
segSports + News:typeNews	−0.0801**	0.0291
segHeavy:typeNews	−0.1379**	0.0479
segNewsL:typeNews	−0.1928***	0.0198
segSports + News:typeAP	0.0282	0.1110
segAP + News:typeNews		
segHeavy:typeAP	0.2394*	0.1111
segNewsL:typeAP	0.0285	0.0998
segSports + News:typeSports	−0.0045	0.0466
segAP + News:typeAP		
segHeavy:typeSports	0.0464	0.0451
segNewsL:typeSports	−0.0066	0.0408
segSports + News:typeOpinion	−0.1709	0.1108
segHeavy:typeOpinion	0.2429**	0.0798
segNewsL:typeOpinion	−0.0940	0.0715
segAP + News:typeSports		
Constant	−2.3206***	0.0127
Observations	1,331,976	
Log Likelihood	−166,371.6000	
Akaike Inf. Crit.	332,795.2000	

[a]seg refers to segment variables; [b]Reference group is *inactive reader* segment; [c]type refers to content type variables.
$p < 0.05$*; $p < 0.01$**; $p < 0.001$***; NewsL refers to *news light* segment.

clickstream data and payment records—from two news sites that allow us to estimate the differential effects of syndicated content on reading regularity and customer churn. Below, we summarize and discuss our major findings around how syndicated content can help build reading regularity for readers with different engagement levels, and further delineate the theoretical and practical contributions of this study.

Syndicated Content Building Regularity of Lighter News Readers

We first identified distinct subscriber segments with different levels of news readership and found that both Site A and Site B had a similar composition of subscriber groups—*inactive*, *light* and *heavy* news readers with the exception of *sports & news* readers in Site A vs. *AP news* readers in Site B. The results of our study indicate that reading syndicated content differentially influences reading habits across subscriber segments. More specifically, syndicated content positively affects the regularity of reading in all subscriber segments except for the *heavy reader* group.

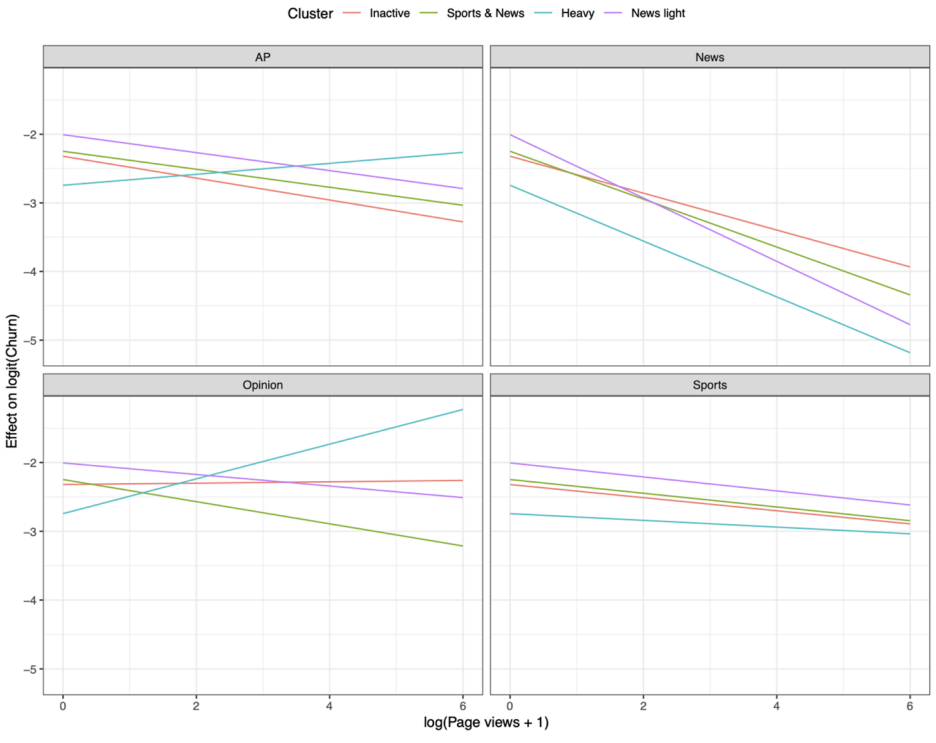

Figure 3. Interaction Plot for four-cluster × four content factors on cancellation in Site A.

The impact of syndicated content is greater for *inactive* and *light reader* groups. While our data do not show specific mechanisms of how syndicated content contributes to these less committed news readers' reading habits, the positive role of syndicated content in establishing reading habits may be due to the nature of syndicated content and the unique characteristics of *inactive* and *light reader* segments. First, syndicated content brings news stories covering a wide variety of subject categories and user interests such as national or international news or breaking news including crime, wars, or disasters that is often accompanied by high-quality visual elements, which are difficult to acquire for local/regional news organizations. This can engage loosely connected news readers. Second, *inactive* or *light news readers* are less likely to be newshounds who hold multiple news subscriptions or actively search for news stories online. That is, the local/regional news outlets function as the one-stop destination for all news for these subscriber segments and syndicated content helps complete news coverage beyond their region, fulfilling their need for environmental surveillance that has evolved in the human alarm system (Ng and Zhao 2020). On the other hand, *heavy readers* are more likely to be newshounds who may already have satisfied their wider surveillance needs from other sources and value the local paper predominantly for local news.

Our results challenge the conventional understanding that syndicated content would hurt journalism and news readership by examining the impact of AP content on segments with different news reading preferences, essentially revealing Simpson's

paradox. Simpson's paradox is when a relationship between two variables disappears or reverses when it is examined by subpopulation groups (Freedman, Pisani, and Purves 2007). Syndicated content might have appeared to have adverse effects on readership, but when we look beneath the surface, we find differential effects by reader segments. Our finding that AP content increases reading regularity of *inactive* and *light news readers* implies that including syndicated content may increase the revenue potential for local/regional news organizations. Considering the ways syndicated content can reduce costs for smaller news organizations, this study calls for a reassessment of syndicated content for local/regional news organizations based on audience-based evidence. Specifically, local/regional news organizations can benefit from strategically providing syndicated content to their less-engaged subscribers. This strategy has the potential to help less committed subscribers build a habit of visiting their digital news site more often and read more content, eventually transforming them into more loyal customers (Benton 2019).

The relationship between syndicated content and willingness to pay (i.e., retention) is not as prominent. The results indicate that AP content—alongside other content types—is negatively related to subscription cancellation for most reader segments, albeit statistically not significant. However, for *heavy readers*, syndicated content and opinion content significantly increase risks of subscription cancellation. One possible interpretation is that *heavy readers* have multiple news sources in addition to the local/regional news outlet where they encounter news alike to syndicated content (Ksiazek, Malthouse, and Webster 2010). Such redundancy can lead to canceling subscription. Similarly, it is conceivable that *heavy readers* may already have their news repertoire including sources that publish opinionated pieces that are congruent to their political beliefs. In such cases, their encounter with opinion contents from local/regional news sources can cause psychological incongruity leading to subscription cancellation. Furthermore, it is plausible that an alternative indirect relationship exists, in which regularity is associated with subscription retention (Kim et al. 2024).

Contribution to Theory and Practice

Nonetheless, our investigation contributes to journalism literature by examining the role of syndicated content in local/regional news organizations from the audience perspective, which has been uncharted territory. Our findings imply that local/regional news organizations can benefit from implementing syndicated content wisely to different groups of readers. It is critical for news organizations to deliver a portfolio of news including not only the specialized original content on issues of local communities or subjects of interest, but also content that readers "need to know" as informed democratic citizens (Abdollahpouri et al. 2021). Syndicated content can be a cost-efficient way to complete this portfolio for local/regional news with limited resources. Furthermore, this study offers evidence that such syndicated content can help less engaged subscribers to build more robust reading habits. Research on audience engagement has highlighted content characteristics related to formal characteristics (e.g., article length, original source), topics, and news values as influential factors for audience engagement (Boczkowski and Mitchelstein 2013; Karnowski et al. 2021; Trilling, Tolochko, and Burscher 2017). This study adds to the literature an

in-depth investigation of syndicated content as one type of content characteristics that can influence audience engagement behaviors (see Appendix B for additional analysis using original content tags). Our results shed a new insight into the generally negative influence of syndicated content found in previous studies (Kim et al. 2024; Trilling, Tolochko, and Burscher 2017).

Our work offers several implications for practitioners in local/regional news organizations. First and foremost, there is a greater need for local/regional news organizations to understand their readers beyond simple attention measures such as page views. News organizations, large or small, have been heavily dependent on advertisements as their revenue source. Therefore, their business has been centered on the overall amount of traffic and page views as key performance indicators implemented in systems that help editors' decision-making in newsrooms (Møller 2022). However, our investigation implies that local/regional newsrooms should invest more in understanding their audience behavioral patterns and start considering measures such as regularity as a critical performance indicator. This also puts local/regional news organizations in a better place to navigate the industrial shift towards a paid subscription business model. Second, syndicated content should be considered a key part of the news portfolio, especially for less engaged subscribers, to drive engagement as in regular news reading. One way to implement this in practice would be to identify segments of subscribers and develop segment-customized content distribution strategies. For instance, feeding a balanced diet of local and general news to *inactive* and *light* news readers, and more local, differentiated content to *heavy* readers. Furthermore, local/regional news organizations may systemize such news recommendations as content bundle products (e.g., newsletters) curated for different subscriber segments. To do that, news organizations also need to understand the specific needs and values of each segment when it comes to digital news readership (Abdollahpouri et al. 2021).

Limitations

This study is not without its limitations. First, there are inherent limitations that naturally come from the type of data we used. For instance, we do not know about subscribers' external reasons for cancellation (e.g., change of address). The data also relies on the content tags and do not include any additional contextual information such as authorship. Secondly, although we find generally similar subscriber segments and effects of syndicated content on regularity in the two local/regional news sites, one should be wary of generalizing these findings. Interests of constituents should differ according to community characteristics, which may result in different subscriber segments. Similarly, we only examine AP content as the representative of syndicated content when there are other types of syndicated content that need further investigation. In this vein, future research can replicate our approach in different local/regional news markets and other types of syndicated content. An interesting question can be whether there are community characteristics associated with higher-level subscriber segments. Additionally, we are limited to the content tags provided by the publishers, which limits more in-depth investigations regarding content characteristics such as the news value or formal characteristics. Future studies can dig

deeper into building automated methods of identifying and classifying these common content characteristics to examine how these may interact with syndicated content to affect audience engagement.

Furthermore, one can also explore how distinct types of syndicated content may have differential effects on regularity. Our findings are also limited in terms of showing direct effect on the financial outcome of subscriber retention. Future studies should explore the hidden mechanisms between subscriber segments, content types, regularity and subscriber retention to come up with a more generalizable conceptual framework. Such efforts could include qualitative work from the audience's point of view like interviews or an online experiment testing the difference between syndicated news and original local news articles regarding audience perceptions of quality, news style, or credibility. Adding audience perceptions will allow us to better understand the mechanisms behind news reading behaviors. Some final suggestions include more practical implementations of the current study's findings to advance emerging practices in digital journalism such as algorithmic recommender systems. What is the place of syndicated content when developing such algorithmic systems for news organizations? How much of it should we include, to whom and when? These questions are integral not only for the business side of local/regional news organizations, but also in the context of achieving adequate and appropriate level of diversity (Loecherbach et al. 2020) in algorithmic journalism tools by design.

Note

1. See https://www.citybureau.org/documenters for more details.

Disclosure Statement

The data used in this study came from the Northwestern Local News Initiative (LNI) and the Spiegel Research Center (SRC). Both LNI and SRC are supported by the Lilly Endowment, Myrta Pulliam Charitable Trust, John Mutz, and several anonymous donors. No potential conflict of interest was reported by the author(s).

ORCID

Jaewon Royce Choi http://orcid.org/0000-0003-1847-4709
Su Jung Kim http://orcid.org/0000-0003-2025-4019
Yayu Zhou http://orcid.org/0000-0002-3246-4278
Edward C. Malthouse http://orcid.org/0000-0001-7077-0172

References

Abdollahpouri, Himan, Edward C. Malthouse, Joseph A. Konstan, Bamshad Mobasher, and Jeremy Gilbert. 2021. "Toward the Next Generation of News Recommender Systems." In Companion Proceedings of the Web Conference 2021, 402–406. Ljubljana Slovenia: ACM. https://doi.org/10.1145/3442442.3452327

Abernathy, Penny. 2022. "The State of Local News: The 2022 Report." *Medill Local News Initiative*. https://localnewsinitiative.northwestern.edu/research/state-of-local-news/report/

Adams, Edward E., and Gerald J. Baldasty. 2001. "Syndicated Service Dependence and a Lack of Commitment to Localism: Scripps Newspapers and Market Subordination." *Journalism & Mass Communication Quarterly* 78 (3): 519–532. https://doi.org/10.1177/107769900107800308

Allison, Paul D. 2010. *Survival Analysis Using SAS: A Practical Guide*. 2nd ed. Cary, North Carolina, USA: SAS Press.

Associated Press. 2022. "AP by the Numbers." Associated Press. https://www.ap.org/about/annual-report/2021/ap-by-the-numbers

Benton, Joshua. 2019. "The News Revenue Hub Is Launching a Pilot Project to Help News Orgs Increase Their Readers' Loyalty." Nieman Lab (Blog). October 16, 2019. https://www.niemanlab.org/2019/10/the-news-revenue-hub-is-launching-a-pilot-project-to-help-news-orgs-increase-their-readers-loyalty/

Boczkowski, Pablo J., and Eugenia Mitchelstein. 2013. *The News Gap: When the Information Preferences of the Media and the Public Diverge*. Cambridge, MA, USA: The MIT Press.

Boumans, Jelle, Damian Trilling, Rens Vliegenthart, and Hajo Boomgaarden. 2018. "The Agency Makes the (Online) News World Go Round: The Impact of News Agency Content on Print and Online News." *International Journal of Communication* 12: 1768–1789.

Boyd-Barrett, Oliver. 2008. "News Agencies in the Era of the Internet." *Media Asia* 35 (3): 163–169. https://doi.org/10.1080/01296612.2008.11726880

Boyd-Barrett, Oliver. 2015. "News Agencies, History Of." In *The International Encyclopedia of Communication*, edited by Wolfgang Donsbach. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781405186407.wbiecn016.pub2

Choi, Jaewon Royce, Sooyeon Hong, and Junghwan Kim. 2023. "Does Social Capital Matter to the Millennials? Social Capital and User Engagements in Online Video Platforms." *Telematics and Informatics* 80: 101967. https://doi.org/10.1016/j.tele.2023.101967

Davies, Nick. 2009. *Flat Earth News: An Award-Winning Reporter Exposes Falsehood, Distortion and Propaganda in the Global Media*. London: Vintage.

Freedman, David, Robert Pisani, and Roger Purves. 2007. *Statistics*. 4th ed. New York, NY: W.W. Norton & Company, Inc.

Goyanes, Manuel. 2014. "An Empirical Study of Factors That Influence the Willingness to Pay for Online News." *Journalism Practice* 8 (6): 742–757. https://doi.org/10.1080/17512786.2014.882056

Johnston, Jane, and Susan Forde. 2017. "Churnalism: Revised and Revisited." *Digital Journalism* 5 (8): 943–946. https://doi.org/10.1080/21670811.2017.1355026

Karnowski, Veronika, Dominik J. Leiner, Anna Sophie Kümpel, and Larissa Leonhard. 2021. "Worth to Share? How Content Characteristics and Article Competitiveness Influence News Sharing on Social Network Sites." *Journalism & Mass Communication Quarterly* 98 (1): 59–82. https://doi.org/10.1177/1077699020940340

Kim, Su Jung, Xiaohan Wang, and Edward C. Malthouse. 2022. "Digital News Readership and Subscription in the United States during COVID-19: A Longitudinal Analysis of Clickstream and Subscription Data from a Local News Site." *Digital Journalism* 10 (6): 1015–1036. https://doi.org/10.1080/21670811.2021.1984972

Kim, Su Jung, Yayu Zhou, Edward C. Malthouse, and Yasaman Kamyab Hessary. 2024. "In Search for an Audience-Supported Business Model for Local Newspapers: Findings from Clickstream and Subscriber Data." *Digital Journalism* 12 (9): 1283–1303. https://doi.org/10.1080/21670811.2021.1948347

Ksiazek, Thomas B., Edward C. Malthouse, and James G. Webster. 2010. "News-Seekers and Avoiders: Exploring Patterns of Total News Consumption across Media and the Relationship to Civic Participation." *Journal of Broadcasting & Electronic Media* 54 (4): 551–568. https://doi.org/10.1080/08838151.2010.519808

Lebow, Sara. 2021. "Google, Facebook, and Amazon to Account for 64% of US Digital Ad Spending This Year." EMarketer. https://www.emarketer.com/content/google-facebook-amazon-account-over-70-of-us-digital-ad-spending

Loecherbach, Felicia, Judith Moeller, Damian Trilling, and Wouter van Atteveldt. 2020. "The Unified Framework of Media Diversity: A Systematic Literature Review." *Digital Journalism* 8 (5): 605–642. https://doi.org/10.1080/21670811.2020.1764374

Lorenz, Hannah. 2017. "News Wholesalers as Churnalists?: The Relationship between Brussels-Based News Agency Journalists and Their Sources." *Digital Journalism* 5 (8): 947–964. https://doi.org/10.1080/21670811.2017.1343649

Malthouse, Edward C., and Abe Peck, eds. 2011. *Medill on Media Engagement*. Cresskill, New Jersey: Hampton Press Inc.

McPherson, Jim. 2002. "Mergers, Chains, Monopoly, and Competition." In *American Journalism: History, Principles, Practices*, edited by W. David Sloan and Lisa Mullikin Parcell, 116–124. Jefferson, N.C: McFarland & Co.

Møller, Lynge Asbjørn. 2022. "Recommended for You: How Newspapers Normalise Algorithmic News Recommendation to Fit Their Gatekeeping Role." *Journalism Studies* 23 (7): 800–817. https://doi.org/10.1080/1461670X.2022.2034522

Myllylahti, Merja. 2016. "Newspaper Paywalls and Corporate Revenues: A Comparative Study." In *The Routledge Companion to Digital Journalism Studies*. edited by Bob Franklin and Scott A. Eldridge II. London: Routledge.

Myllylahti, Merja. 2017. "What Content Is Worth Locking behind a Paywall?: Digital News Commodification in Leading Australasian Financial Newspapers." *Digital Journalism* 5 (4): 460–471. https://doi.org/10.1080/21670811.2016.1178074

Nelson, Jacob L. 2021. "The Next Media Regime: The Pursuit of 'Audience Engagement' in Journalism." *Journalism* 22 (9): 2350–2367. https://doi.org/10.1177/1464884919862375

Newman, Nic, Richard Fletcher, Anne Schulz, Simge Andi, Craig T. Robertson, and Rasmus Kleis Nielsen. 2021. "Reuters Institute Digital News Report 2021." Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2021-06/Digital_News_Report_2021_FINAL.pdf

Ng, Yu-Leung, and Xinshu Zhao. 2020. "The Human Alarm System for Sensational News, Online News Headlines, and Associated Generic Digital Footprints: A Uses and Gratifications Approach." *Communication Research* 47 (2): 251–275. https://doi.org/10.1177/0093650218793739

Olsen, Ragnhild Kristine. 2021. "The Value of Local News in the Digital Realm – Introducing the Integrated Value Creation Model." *Digital Journalism* 9 (6): 810–834. https://doi.org/10.1080/21670811.2021.1912623

Olsen, Ragnhild Kristine, and Mona Kristin Solvoll. 2018. "Reinventing the Business Model for Local Newspapers by Building Walls." *Journal of Media Business Studies* 15 (1): 24–41. https://doi.org/10.1080/16522354.2018.1445160

Pew Research Center. 2021. "Estimated Advertising and Circulation Revenue of the Newspaper Industry." Pew Research Center's Journalism Project (Blog). June 29, 2021. https://www.pewresearch.org/journalism/chart/sotnm-newspapers-newspaper-industry-estimated-advertising-and-circulation-revenue/

Pickard, Victor, and Alex T. Williams. 2014. "Salvation or Folly?: The Promises and Perils of Digital Paywalls." *Digital Journalism* 2 (2): 195–213. https://doi.org/10.1080/21670811.2013.865967

Rantanen, Terhi. 2019. "News Agencies from Telegraph Bureaus to Cyberfactories." In *Oxford Research Encyclopedia of Communication*, edited by Terhi Rantanen. Oxford: Oxford University Press. https://doi.org/10.1093/acrefore/9780190228613.013.843

Ruggiero, Thomas E. 2000. "Uses and Gratifications Theory in the 21st Century." *Mass Communication and Society* 3 (1): 3–37. https://doi.org/10.1207/S15327825MCS0301_02

Saridou, Theodora, Lia-Paschalia Spyridou, and Andreas Veglis. 2017. "Churnalism on the Rise?: Assessing Convergence Effects on Editorial Practices." *Digital Journalism* 5 (8): 1006–1024. https://doi.org/10.1080/21670811.2017.1342209

Sjøvaag, Helle. 2014. "Homogenisation or Differentiation?: The Effects of Consolidation in the Regional Newspaper Market." *Journalism Studies* 15 (5): 511–521. https://doi.org/10.1080/1461670X.2014.885275

Sjøvaag, Helle. 2016. "Introducing the Paywall: A Case Study of Content Changes in Three Online Newspapers." *Journalism Practice* 10 (3): 304–322. https://doi.org/10.1080/17512786.2015.1017595

Trilling, Damian, Petro Tolochko, and Björn Burscher. 2017. "From Newsworthiness to Shareworthiness: How to Predict News Sharing Based on Article Characteristics." *Journalism & Mass Communication Quarterly* 94 (1): 38–60. https://doi.org/10.1177/1077699016654682

Van Leuven, Sarah, Sanne Kruikemeier, Sophie Lecheler, and Liesbeth Hermans. 2018. "Online and Newsworthy: Have Online Sources Changed Journalism?" *Digital Journalism* 6 (7): 798–806. https://doi.org/10.1080/21670811.2018.1498747

Wadbring, Ingela, and Lovisa Bergström. 2021. "Audiences behind the Paywall: News Navigation among Established versus Newly Added Subscribers." *Digital Journalism* 9 (3): 319–335. https://doi.org/10.1080/21670811.2021.1878919

Wall, Melissa. 2015. "Citizen Journalism: A Retrospective on What We Know, an Agenda for What We Don't." *Digital Journalism* 3 (6): 797–813. https://doi.org/10.1080/21670811.2014.1002513

Welbers, Kasper, Wouter van Atteveldt, Jan Kleinnijenhuis, and Nel Ruigrok. 2018. "A Gatekeeper among Gatekeepers: News Agency Influence in Print and Online Newspapers in the Netherlands." *Journalism Studies* 19 (3): 315–333. https://doi.org/10.1080/1461670X.2016.1190663

Appendices

Appendix A

Table A1. Loadings of the four principal components of content categories in Site A.

	Loadings			
Content Categories	Factor 1	Factor 2	Factor 3	Factor 4
Other	0.826			
Local govt or state	0.799			
National or election	0.664			
Local or state	0.692			
Education	0.690			
CFAD	0.716			
Health	0.937			
Food or dining	0.754			
Business	0.865			
AP – Business		0.741		
AP – Entertainment		0.720		
AP – Health		0.603		
AP – Nation/World		0.621		
AP – Local		0.402		
AP – Other		0.494		
High school Sports			0.745	
College Sports			0.863	
Pro Sports			0.760	
AP – Sports		0.315	0.538	
Letters				0.830
Staff opinions				0.674
Signed opinions				0.639
Syndicated				0.711
Celebrity or TV				
Event or art	0.487			
Obituary or Announcement				
SS loadings	6.071	2.560	2.290	2.181
Proportion Variance	0.234	0.098	0.088	0.084
Cumulative Variance	0.234	0.332	0.420	0.504

CFAD: Crime, Fire, Accident, Disaster.

Table A2. Loadings of the four principal components of content categories in Site B.

	Loadings			
Content Categories	Factor 1	Factor 2	Factor 3	Factor 4
News	0.821			
CFAD	0.659			
Event or art	0.736			
Local news	0.896			
Sports (other)	0.770			
Health	0.863			
Local govt or politics	0.698			
Business	0.781			
Travel	0.501			
Food or dining	0.575			
Life	0.662			
AP – Nation/World		0.689		
AP – Other		0.588		
AP – Business		0.763		
AP – Politics		0.817		
AP – Health		0.505		0.331
AP – Entertainment		0.698		
AP – Local news	0.328	0.454		−0.414
AP – Lifestyle		0.450		0.380
AP – Sports		0.445	0.461	
High school sports			0.611	
College sports			0.798	
Pro sports			0.569	
WE				0.683
Nation/World				0.702
Opinion		0.493		
National govt or election	0.457	0.341		
Education	0.418			
Other	0.466			
SS loadings	6.851	4.057	1.666	1.601
Proportion Variance	0.236	0.140	0.057	0.055
Cumulative Variance	0.236	0.376	0.434	0.489

CFAD: Crime, Fire, Accident, Disaster; WE: Washington Examiner is a particular type of syndicated content provided by the publication at Site B.

Appendix B

Below we report results of additional analyses to show relationships between content characteristics (without AP) and dependent variables (i.e., regularity, churn), and relationships between content characteristics (with raw AP tags) and dependent variables. The results show significant relationships between different content types and our DVs. However, AP content categories also add additional effects when brought in, implying that whether the content is syndicated or not matters even after considering traditional content type tags.

The overall effects of syndicated content is negative (consistent with previous studies). However, we find positive effects of syndicated content to *inactive* and *light news reader* segments, uncovering the Simpson's Paradox. This further highlights the contribution of our examination of how reader segments interact with syndicated content.

Table B1. Discrete-time survival models predicting cancellation of digital news subscription using content variables (without and with AP content).

	Dependent Variable: Churn in next month (Site A)			
	Without AP		With AP	
	Slope	*SE*	*Slope*	*SE*
bs(custtime)1	−18.6748***	0.1947	−18.6894***	0.1949
bs(custtime)2	34.6793***	0.4667	34.7054***	0.4668
bs(custtime)3	−2.9507***	0.0512	−2.9511***	0.0512
seasonSpring	0.2258***	0.0143	0.2283***	0.0144
seasonSummer	0.0211	0.0146	0.0229	0.0147
seasonWinter	−0.0802***	0.0157	−0.0792***	0.0158
Other	−0.2189***	0.0106	−0.2204***	0.0107
High-school Sport	−0.1305*	0.0560	−0.1326*	0.0561
College Sport	−0.0331	0.0191	−0.0408	0.0193
Professional Sport	−0.0737***	0.0135	−0.0752***	0.0136
Local Government/State	−0.0590***	0.0152	−0.0602***	0.0152
National/Election	−0.1602***	0.0137	−0.1620***	0.0138
Local/State	−0.0906***	0.0194	−0.0947***	0.0194
Education	−0.0447*	0.0201	−0.0468*	0.0201
CFAD[a]	0.0197	0.0134	0.0141	0.0136
Health	−0.2439***	0.0094	−0.2433***	0.0094
Food/Dine	−0.1953***	0.0172	−0.1950***	0.0172
Celebrity/TV	−0.0232	0.0220	−0.0250	0.0221
Event/Art	−0.0776***	0.0225	−0.0804***	0.0226
Obituary/Announce	0.1531*	0.0713	0.1517*	0.0714
Business	0.0012	0.0132	0.0004	0.0132
Letters	0.3176***	0.0645	0.3117***	0.0646
Staff Opinion	0.0755*	0.0302	0.0697*	0.0303
Signed Opinion	−0.0797*	0.0341	−0.0868*	0.0342
Syndicated	0.1353***	0.0300	0.1283***	0.0302
AP Local			0.2936***	0.0652
AP Business			0.0623	0.0604
AP Entertainment			−0.0696	0.0634
AP Health			−0.0564	0.0589
AP Sport			0.1152	0.0820
AP National/World			−0.0011	0.0314
AP Other			0.0570	0.0339
Constant	−2.3771***	0.0121	−2.3770***	0.0121
Observations	1,331,976		1,331,976	
Log Likelihood	−166,359.7000		−166,345.9000	
Akaike Inf. Crit.	332,771.3000		332,757.8000	

[a]CFAD refers to crime, fire, accidents, and disasters.
$p < 0.05$*; $p < 0.01$**; $p < 0.001$***.

Table B2. Panel regression models predicting regularity using content variables in Site A (without and with AP content).

	Dependent Variable: Regularity in Next Month (Site A)			
	Without AP		With AP	
	Slope	*SE*	*Slope*	*SE*
bs(custtime)1	0.3824***	0.0161	0.4898***	0.0160
bs(custtime)2	−0.9060***	0.0515	−1.1561***	0.0511
bs(custtime)3	0.0792***	0.0040	0.0867***	0.0040
seasonSpring	0.0637***	0.0021	0.0597***	0.0021
seasonSummer	0.0308***	0.0020	0.0245***	0.0020
seasonWinter	0.1964***	0.0022	0.1857***	0.0022
Other	0.5119***	0.0013	0.5178***	0.0013
High-school Sport	−0.0735***	0.0043	−0.0578*	0.0043
College Sport	0.0728***	0.0017	0.0942***	0.0017
Professional Sport	0.1645***	0.0013	0.1690***	0.0013
Local Government/State	0.0517***	0.0017	0.0552***	0.0016
National/Election	0.1106***	0.0014	0.1301***	0.0014
Local/State	−0.0971***	0.0019	−0.0743***	0.0019
Education	−0.0949***	0.0019	−0.0902***	0.0019
CFAD[a]	0.0508***	0.0015	0.0834***	0.0015
Health	0.3214***	0.0011	0.3216***	0.0011
Food/Dine	0.0201***	0.0016	0.0255***	0.0016
Celebrity/TV	−0.0454***	0.0019	−0.0154***	0.0019
Event/Art	−0.0800***	0.0020	0.0427***	0.0020
Obituary/Announce	−0.2986***	0.0056	−0.2645***	0.0056
Business	0.0653***	0.0015	0.0716***	0.0015
Letters	−0.1351***	0.0060	−0.1068***	0.0059
Staff Opinion	−0.0756***	0.0028	−0.0573***	0.0028
Signed Opinion	0.0050	0.0027	0.0389***	0.0027
Syndicated	−0.1572***	0.0026	−0.1247***	0.0026
AP Local			−0.2361***	0.0050
AP Business			−0.2142***	0.0047
AP Entertainment			−0.1602***	0.0047
AP Health			−0.1524***	0.0045
AP Sport			−0.2031***	0.0057
AP National/World			−0.1510***	0.0026
AP Other			−0.1294***	0.0029
Constant	0.7320***	0.0018	0.7158***	0.0018
Observations	1,331,976		1,331,976	
R^2	0.5848		0.5916	
Adjusted R^2	0.5848		0.5916	
Residual Std. Error	0.8354 (df = 1,331,950)		0.8285 (df = 1,331,943)	
F Statistic	75,036.5300*** (df = 25; 1,331,950)		60,289.1900*** (df = 32; 1,331,943)	

[a]CFAD refers to crime, fire, accidents, and disasters.
$p < 0.05$*; $p < 0.01$**; $p < 0.001$***.

Table B3. Panel regression models predicting regularity using content variables in Site B (without and with AP content).

	Dependent Variable: Regularity in next month (Site B)			
	Without AP		With AP	
	Slope	*SE*	*Slope*	*SE*
bs(custtime)1	0.9013***	0.0627	0.9155***	0.0625
bs(custtime)2	−0.1416	0.0799	−0.1480	0.0797
bs(custtime)3	0.5136***	0.0622	0.5197***	0.0620
seasonSpring	−0.0924***	0.0163	−0.0840***	0.0164
seasonSummer	−0.0420**	0.0149	−0.0337*	0.0149
seasonWinter	−0.0692***	0.0169	−0.0667***	0.0169
News	0.1693***	0.0152	0.1849***	0.0153
High-school Sport	0.0868***	0.0213	0.1004***	0.0213
College Sport	−0.1016*	0.0506	−0.0785	0.0506
Professional Sport	0.1354***	0.0182	0.1580***	0.0188
Other Sport	0.2364***	0.0146	0.2398***	0.0146
Local Government/Politics	0.0501**	0.0164	0.0671***	0.0166
National/Election	−0.0847***	0.0243	−0.0543*	0.0246
Local News	0.4726***	0.0115	0.4664***	0.0115
National/World	0.0830	0.1496	0.0283	0.1494
Education	−0.1214***	0.0266	−0.0983***	0.0266
CFAD[a]	−0.0551**	0.0169	−0.0284	0.0171
Health	0.2121***	0.0123	0.2167***	0.0123
Food/Dine	−0.0952**	0.0319	−0.0872**	0.0319
Event/Art	0.1027***	0.0164	0.1066***	0.0164
Business	0.1361***	0.0155	0.1402***	0.0155
Opinion	−0.0123	0.0139	0.0303*	0.0145
Travel	−0.1211***	0.0302	−0.0943**	0.0302
Life	−0.0297	0.0231	−0.0175	0.0231
WE[b]	−0.1445**	0.0471	−0.0375	0.0492
AP Local			−0.1322***	0.0358
AP Business			−0.0950**	0.0304
AP Entertainment			−0.1814**	0.0679
AP Health			−0.1425***	0.0423
AP Sport			0.0036	0.0484
AP National/World			−0.0359	0.0239
AP Politics			−0.0801**	0.0289
AP Lifestyles			−0.2187***	0.0624
AP Other			−0.0557	0.0403
Constant	0.4512***	0.0130	0.4359***	0.0130
Observations	21,209		21,209	
R^2	0.5305		0.5337	
Adjusted R^2	0.5299		0.5330	
Residual Std. Error	0.8127 ($df = 21,182$)		0.8100 ($df = 21,173$)	
F Statistic	920.3887*** ($df = 26$; 21,182)		692.4757*** ($df = 35$; 21,173)	

[a]CFAD refers to crime, fire, accidents, and disasters; [b]WE refers to syndicated content.
$p < 0.05$*; $p < 0.01$**; $p < 0.001$***.