CESifo WORKING PAPERS

9832 2022

July 2022

Competing for Attention on Information Platforms: The Case of News Outlets

Tim Meyer, Anna Kerkhof, Carmelo Cennamo, Tobias Kretschmer

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9832*

Competing for Attention on Information Platforms: The Case of News Outlets

Abstract

Mainstream logic supports the idea that platforms bring large benefits to firms, especially smaller ones, by opening up access to a broader set of consumers and making firms' products easier to find. However, this argument mostly applies to transaction platforms that match consumer preferences to products. On information platforms such as social media or news aggregators, firms compete for consumer attention, not matches. We argue that consumer attention and choice in contexts such as news content are driven by the size and focus of content providers. Providers sufficiently large to be recognized by consumers and sufficiently broad in their focus to cover multiple content categories of interest to consumers are better positioned to capture a significant share of consumer attention, and thus demand, compared to smaller and more narrow competitors. We develop a simple formalization of our reasoning and find empirical support for it by exploiting a legal dispute leading to the removal of a group of German news outlets from news aggregators.

Keywords: information platforms, competition for attention, consumer attention, news aggregators, news content.

Tim Meyer
LMU Munich School of Management
Munich / Germany
tim.meyer@lmu.de

Anna Kerkhof
ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Munich / Germany
kerkhof@ifo.de

Carmelo Cennamo
Copenhagen Business School / Denmark
cce.si@cbs.dk

Tobias Kretschmer
LMU Munich School of Management
Munich / Germany
t.kretschmer@lmu.de

Working Paper, June 2022

1 Introduction

The promise of many platforms is to provide firms with access to a wider set of potential consumers than they could reach on their own, which is particularly true for smaller firms that struggle to extend their reach (Brynjolfsson, Hu and Smith 2003, Brynjolfsson, Hu and Simester 2011, Kumar, Smith and Telang 2014). Since both consumer preferences and complements (i.e., products made available on the platform by firms) are typically heterogeneous, platforms can benefit smaller firms by acting as intermediaries that match consumer preferences and complements (Bakos 1997, Cennamo 2021, Tajedin, Madhok and Keyhani 2019). Platforms can expand the overall market and thus demand for these smaller firms.

This logic assumes that complementors (i.e., firms making complements available via the platform) compete for the best match between consumer preferences and complements, and that complement heterogeneity drives (and sways) consumer choices. However, on information platforms such as news aggregators, search engines or social media platforms, "sellers" produce information content and compete for consumer attention (Cennamo 2021). Consumers make choices based on the attention that complementor content can attract rather than on its quality and price (Bordalo, Gennaioli and Shleifer 2013), especially for experience goods (Nelson 1970) that are perceived as largely homogenous by consumers prior to consumption.[1] Platforms act as "attention brokers" (Prat and Valletti 2021) that attract consumer attention and channel it to content (Evans 2019, Rui and Whinston 2012). Competition on these platforms then turns into a "battle for consumers' [limited] attention" (Kumar et al. 2014, The Economist 2017). While smaller content providers may gain exposure to a broader set of consumers, larger ones may capture a disproportionate share of consumer attention. Hence, we ask *if smaller content providers indeed stand to gain from being on a platform in attention markets*.

We argue that attention spillovers between content providers on information platforms have a dual - incoming and outgoing - effect: part of a provider's consumer base will be attracted to other

[1] For example, the quality of a search result or of a post on social media is not known prior to consumption; they might thus be perceived as substitute consumption options ex ante.

content providers, but in turn, the focal provider will attract consumers of other content providers. The net effect depends on the provider's relative capacity to attract consumer attention, which is affected, as we explain in greater detail below, by its scale and focus. Providers that are better known to consumers due to their large scale benefit from attention spillovers as they are more likely to attract attention, while smaller content providers may face an "attention discount" compared to larger providers. Further, by being less focused on a subset of categories (or broadly covering multiple content categories), content providers benefit from spillovers within multiple categories and thus increase the overall number of consumers they attract.

We develop a simple formalization of our reasoning and find empirical support for it in the context of local news outlets listed on online news aggregators. News aggregators such as *Google News*, *Apple News* or *Yahoo! News* display the headlines and small excerpts of news articles produced by online news outlets (i.e., content providers). While local news outlets usually only operate in a limited geographic region, their content is also of interest outside their traditional market and could thus benefit from being listed on news aggregators. Making use of a legal dispute following a policy change in Germany that led to the removal of a group of German news outlets from several news aggregators, we compare the performance of local news outlets that were removed to those that were not, before and after the legal dispute. Using web traffic data on 140 local news outlets, we find evidence consistent with the dual attention spillover effect: Larger and less focused news outlets suffer more from being delisted than smaller and more narrow ones. A 10% larger scale results in ~18,000 fewer monthly visits for an average-sized news outlet after removal from news aggregators. Similarly, a one standard deviation reduction in focus results in ~360,000 fewer monthly visits post-removal.

We add to platform research by highlighting how competing for attention can be a double-edged sword for content providers *on* the platform and expose providers small in scale and narrow in focus to its negative effects. The starting point for most prior work on the effect of news aggregators on news outlets was competition *for* the market for attention – that is, competition between aggregator and news outlets as potential substitutes (Athey, Mobius and Pal 2021, Calzada and Gil 2020, Dellarocas, Sutanto, Calin and Palme 2016, Peitz and Reisinger 2014). Instead, we focus on competition *in* the market, i.e.

smaller and larger content providers competing for attention on the same platform (Evans 2013). Perhaps counterintuitively, smaller content providers suffer from the attention distribution dynamics on a platform because larger competitors will attract some of their consumers. We expect this effect to carry over to transaction platforms, albeit mitigated by product heterogeneity and market expansion through improved match quality.

2 Related Literature

2.1.1 Platforms and Attention

The relationship between platforms and their complementors has been studied widely. Typically, platforms are considered interfaces that "mediate transactions between two or more sides, such as […] complementors and users" (McIntyre and Srinivasan 2017), which implies indirect network effects. The more complementors are on the platform, the more attractive the platform becomes to users[2] and vice versa (Parker and van Alstyne 2005). The coordination and aggregation of product offerings by the platform allows for better value creation and capture than platform and complementors could achieve on their own (Jacobides, Cennamo and Gawer 2018). Indeed, the prospect of creating and capturing value through interactions with consumers is a key incentive for complementors to join the platform (Kretschmer, Leiponen, Schilling and Vasudeva 2022), particularly for smaller firms, which only have access to a limited market and face challenges in attracting potential consumers. If these firms make some of their products available as complements on a platform, they gain exposure to a wider set of consumers than they could access on their own.

The role of platforms as mediators between smaller firms and consumers is particularly important, as both complements and consumer preferences are typically heterogeneous (Panico and Cennamo 2022, Sun, Rajiv and Chu 2016, Rietveld and Eggers 2018). Platforms facilitate matches between consumers and the respective complement (Cennamo 2021, Tajedin et al. 2019) by reducing search costs compared to conventional markets (Bakos 1997). This logic has led to the assumption that

[2] While we acknowledge that there can be meaningful distinctions between "users" and "consumers" in some settings, we use the two terms interchangeably in this paper.

products at the lower end of the sales distribution (the so-called long tail) gain the most from joining a platform (Anderson 2004). Platforms make discovering these products easier for the relevant consumers, and ultimately allow producers to benefit from increased demand. Indeed, consumers are more likely to buy long-tail products when they move from physical stores to online channels (Zentner, Smith and Kaya 2013). Recommender systems facilitate the discovery of long-tail products in these settings (Brynjolfsson et al. 2011) and the sales distribution becomes less skewed as more information about products becomes available to potential consumers, improving match quality (Kumar et al. 2014, Tucker and Zhang 2011).

A key (implicit) assumption of this work is that consumers and products are heterogeneous and that it is mainly match quality driving consumer choices. While this assumption seems accurate for transaction platforms connecting sellers and buyers (e.g., *Amazon*) or technology platforms matching software developers and users (e.g., video game platforms), it applies less readily to information platforms such as social networks, search engines, and news aggregators. There, complementors largely provide informational content rather than goods or services and compete mainly for consumer attention, rather than competing on quality or price. Platforms act as "attention brokers" (Prat and Valletti 2021), that attract attention to their services, and then channel it to providers' content. Content providers compete and may attract more or less attention depending on their characteristics. Platform users choose from a wide set of options whose characteristics, and thus potential match with individual preferences, are hard to assess prior to consumption (Nelson 1970). Moreover, consumers only have limited attention to evaluate all available content (Evans 2019, Boik, Greenstein and Prince 2016).

We therefore posit two countervailing - incoming and outgoing – spillover effects that platforms can have. On the one hand, by attracting, aggregating and channeling consumer attention, platforms can help consumers discover new content they would otherwise miss. Firms competing for attention with similar competitors thus benefit from "*incoming attention spillover" effects*. On the other hand, processing of information requires cognitive resources (Kahneman 1973), making human attention a scarce and rivalrous good in the attention economy (Lanham 2006, Evans 2013, Calvano and Polo 2021).

Firms competing on a platform for consumer attention may thus face intense competition and be harmed by "*outgoing attention spillover" effects*.

The ability of content providers to attract attention depends on how recognizable they are. When content by multiple providers is shown, well-known providers will draw the attention of potential consumers more than less known ones. This effect might be even stronger if more content is available on the platform as consumers turn to well-known sources of information on crowded platforms (Piezunka and Dahlander 2015). Here, consumers base their decisions not predominantly on the content provided, but rather on the salience of the content source compared to other providers. Typically, content providers larger in scale are better known to potential consumers due to previous exposure through consumption or advertising (Alba and Hutchinson 1987). Content by well-known providers hence becomes the "default option" (Macdonald and Sharp 2000), while choosing other options requires additional cognitive effort.

Hence, despite reduced search costs on platforms, larger content providers capture the lion's share of consumer attention, similar to the case of so-called "hit" or "superstar" products (Elberse 2008, Kumar et al. 2014). Prior work has confirmed that demand remains highly concentrated at the top of the market, even if product variety increases (Tan, Netessine and Hitt 2017) and consumers increasingly transact through platforms (Elberse and Oberholzer-Gee 2006). Conversely, products in the tail of the distribution are likely to be consumed mostly by heavy consumers who seek variety after extensively consuming mainstream content, or by a limited number of connoisseurs with very specific product knowledge and tastes (Elberse 2008). Thus, consumers may end up choosing from a relatively narrow set of content, and it is not obvious whether information platforms deliver on the promise of promoting smaller firms or whether the dynamics lead to winner-take-all outcomes favoring large, well-known firms.

Prior work on competing for attention has focused mostly on competition *between* different information platforms (Evans 2013, Boik et al. 2016, Peitz and Reisinger 2014) or on the role of advertising in the business model of these platforms (Evans 2019, Prat and Valletti 2021). Work on competition *within* information platforms has focused on the motivations for content providers to contribute (Rui and Whinston 2012, Loh and Kretschmer 2022) and on how contribution behavior relates to competition among content providers, for instance on social media (Rossi and Rubera 2021) or intra-organizational knowledge platforms (Hansen and Haas 2001). We take a demand-side perspective to study how differences in scale and focus of content providers affect the demand they can attract. Previous research on the "long-tail" and "superstar" effects has mostly focused on product-level analyses (Brynjolfsson et al. 2011, Elberse and Oberholzer-Gee 2006) to explain market category expansion and success therein. However, this research does not consider the role of firm-level characteristics in content success, and how such characteristics drive the competitive dynamics at the platform market level. We add to this research by bringing these intertwined effects to the core of the analysis.

2.1.2 News Aggregators

One example of information platforms acting as attention brokers are news aggregators such as *Google News*, which combine content by online news outlets and make short excerpts (so-called snippets) of this content available to consumers who search for specific news topics.

Prior work has often focused on potential substitution effects between news aggregator and online news outlets, i.e. whether the headlines and content excerpts on news aggregators provide sufficient information for consumers and ultimately keep them from reading the full article on the news outlet's website (Athey et al. 2021, Dellarocas et al. 2016). Thus, the focus was on competition *for* the market for attention – that is, competition between the news aggregator and news outlets. Some types of news outlets may indeed benefit from news aggregators, for instance horizontally or vertically differentiated outlets (Chiou and Tucker 2017), smaller publishers whose content is hard to find (Athey et al. 2021) or lower-performing websites and local news outlets (Calzada and Gil 2020), in line with the idea that news aggregators facilitate discovery of unknown content. Conversely, larger outlets such as *Axel Springer* members that were temporarily excluded from *Google News* in Germany benefitted significantly from being on news aggregators (Calzada and Gil 2020), and aggregators may not always steer consumers towards new content (George and Hogendorn 2020).

Most work that specifically describes competitive mechanisms *in* the market, i.e., competition among news outlets for the same readers on news aggregators has been theoretical. As potential readers have limited time and attention to process information, they will only consume a (small) subset of all news available on the aggregator (Alaoui and Germano 2020). This may lead to more concentrated demand for popular news outlets due to herding behavior, as consumers have incomplete information on the true quality of news articles and infer quality from the behavior of others (Hong 2011). However, competition among news outlets for the same readers on news aggregators can also drive content quality, as news outlets may increase the quality of their content to increase their chances of being featured on the aggregator (Dellarocas, Katona and Rand 2013) or if the web traffic's increase for news outlets is sufficiently large (Jeon and Nasr 2016).

We extend prior work by conducting a nuanced empirical analysis of competition among news outlets that lets us focus on competition *in* the market for attention rather than *for* the market. Next, we build a simple theoretical model to develop our hypotheses.

3 Theory and Hypotheses

3.1 Theoretical Framework

We propose a simple model with two (news) outlets and two (news) categories[3] in which outlets compete *online* for the attention of a fixed set of readers. We introduce the off-platform setting as benchmark: off-platform, outlets operate as local monopolists in an online market of limited size.[4] We contrast this case with the on-platform case, where outlets gain access to a broader set of potential readers, but compete with other outlets for reader attention.[5]

[3] For instance, Politics (P) and Sports (S).

[4] This reflects the situation of many outlets that have traditionally not had access to the entire market but held a very strong position in the part of the market they were active in.

[5] A simple way of thinking about this situation is that off-platform, readers access each paper separately, creating a local quasi-monopoly for each outlet and user if there are search/switching costs. On-platform, readers access the platform and multiple papers are available at equal cost, hence they compete "for the same eyeballs" by readers.

3.1.1 Off-Platform

Consider two outlets i = A,B. Each outlet i faces a readership $\tilde{\imath} \geq 0$ and is a monopolist in its market, i.e., a consumer who considers reading i does not consider reading $k \neq i$ (e.g., because search costs for another outlet are prohibitively high).[6] Assume w.l.o.g. that the total mass of readers $\sum_i \tilde{\imath} = 1$. Each outlet can offer up to two content categories $j \in \{P, S\}$ and each reader will consume content in either category P or S.[7] The mass of off-platform readers of outlet i consuming content category j is denoted by $\widetilde{\imath_j}$ (see *Table 1*). Prices are normalized to zero and we abstract from fixed or marginal costs of production.

Insert *Table 1* about here

3.1.2 On-Platform

In the on-platform case, outlets A and B are listed on a (news) aggregator. Each outlet now has access to the entire mass of readers $\tilde{A} + \tilde{B} = 1$ (as readers face no on-platform search costs), but must compete with the other outlet for reader attention, as all readers can now potentially read content from either outlet.[8] When outlets A and B both appear on an aggregator, readers choose which one to consume. Readers that have been consuming content by outlet i in category j *off*-platform can now choose from a wider set of options available *on*-platform.

When facing this extended set of potential options, readers can decide to continue consuming the content consumed *off*-platform (i.e. outlet i in category j) or they can switch to a different option. If they decide to switch to a different option, the most immediate alternatives are likely either switching

---

[6] This assumption provides a strong contrast to the on-platform case and lets us focus on our main mechanisms.
[7] We assume that readers only read a single article (unit demand), but that they are indifferent between articles in categories P and S from the same outlet and articles in the same category across different outlets.
[8] Note that, both in the off-platform case and the on-platform case, we are interested in the readership consuming content on a given news outlet's *own* web page (i.e., the web traffic on the news outlet's web page). While a share of each news outlet's readership may decide to consume only the news excerpts (i.e., snippets) available on the news aggregator in the on-platform case, our model considers the *net* effect on each news outlet's readership. This allows us to focus on competition between news outlets rather than competition between news outlets and news aggregators, and it accounts for the fact that news outlets only care about visits on their own web page as only these visits yield advertising revenues.

to another outlet (and consuming the same content category *j*) or switching to another content category (and consuming the same outlet *i*). The content of outlet *i* in category *j* hence competes with alternatives within the *same category* and with alternatives within the *same outlet*, but is unlikely to compete with alternatives in other categories *and* other outlets.

Readers' choices among these alternatives are driven by the *attention* each alternative draws compared to the attention drawn by all potential alternatives. Specifically, content that attracts a large readership *off*-platform is likely to attract more attention *on*-platform. Hence, the attention received by a specific type of content *on*-platform depends on the size of its own *off*-platform readership relative to the *off*-platform readership of all relevant *on*-platform options.

Insert *Figure 1* about here

To illustrate this logic, consider a Hotelling-like setting where each off-platform readership $\widetilde{\iota_j}$ is uniformly distributed on a line of length 1 (*Figure 1*).[9] For concreteness, consider $\tilde{A}_P$, i.e., the off-platform readership that consumes content of outlet *A* in category *P*. The relevant on-platform options for this readership – $\hat{A}$ and $\widehat{B_P}$ – are located at the extremes of the Hotelling line. The *attraction* of option $\hat{A}$ for a reader located at x_{A_P} can be described as

$$\frac{\tilde{A}}{\tilde{A} + \widetilde{B_P}} - t \cdot x_{A_P}, \tag{1}$$

where $\tilde{A} = \widetilde{A_P} + \widetilde{A_S}$ and $\widetilde{B_P}$ indicate off-platform readership, *t* denotes the readers' transport costs, and the term $\frac{\tilde{A}}{\tilde{A}+\widetilde{B_P}}$ corresponds to the attention that $\hat{A}$ receives.[10] Analogously, the *attraction* of $\widehat{B_P}$ from a reader located at x_{A_P} is given by

$$\frac{\widetilde{B_P}}{\tilde{A} + \widetilde{B_P}} - t \cdot \left(1 - x_{A_P}\right). \tag{2}$$

[9] We can think of each off-platform readership $\widetilde{\iota_j}$ as a submarket in which the available alternatives compete.
[10] In our setting, transport costs refer to the strength of preference when moving away from a reader's off-platform content. *Attraction* of an on-platform option corresponds to attention minus transport costs.

Finding the indifferent reader for whom the options $\hat{A}$ and $\widehat{B_P}$ are equally attractive yields the respective shares of "remaining" readers (who stay with news outlet *A*) and "churning" readers (who switch to news outlet *B*).[11] We can repeat this procedure for all off-platform readerships $\widetilde{\iota_J}$ to find the overall number of remaining and churning readers for *A* and *B*.

While this simple framework abstracts away from many complexities underlying media choice, it is a useful workhorse to generate testable hypotheses in the following sections. In particular, we assume that on-platform market shares are affected by two drivers of reader attention: *scale* and *focus*. While these dimensions may be correlated, e.g., an outlet that is relatively large in scale may be less focused, we study the impact of each channel on readers' attention separately.

3.2 Hypotheses

We use our model to develop two hypotheses on the effect of being *delisted* from an aggregator that we will test in our empirical setting.[12] Consider the illustrative baseline setting in *Figure 2*, where each off-platform readership $\widetilde{\iota_J}$ is uniformly distributed on a Hotelling line of length 1, with the relevant on-platform options located at the extremes of the line. The shaded area gives the mass of $\widetilde{\iota_J}$,[13] and x_{i_j} is the reader who is indifferent between the options. The dark arrows give the share of readers of outlet *A* on-platform, the light arrows indicate the share of readers of outlet *B* on-platform. Suppose that off-platform readerships across outlets and categories are evenly distributed in the baseline case, i.e. $\tilde{A}_P = \tilde{A}_S = \tilde{B}_P = \tilde{B}_S = \frac{1}{4}$. In this symmetric case, the number of "churning" readers is equal to the number of "incoming" readers for each outlet.[14] The size of the on-platform readership is thus equal to the size of the off-platform readership.

[11] With $t = 1$, equating (I) and (II), and solving for x_{A_P} yields $x_{A_P} = \frac{1}{2}\left(\frac{\tilde{A}-\widetilde{B_P}}{\tilde{A}+\widetilde{B_P}} + 1\right)$

[12] Our empirical setting lets us test the effect of being "struck off" (delisted from) a news aggregator, and we formulate our hypotheses accordingly. However, the analogous logic (with inverse sign) applies (by definition) for complementors joining an information platform.

[13] This is the density of the uniform distribution.

[14] "Incoming" readers are those who churn away from the other news outlet.

Insert *Figure 2* about here

3.2.1 Outlet Scale

Consider now the case where the off-platform readership of outlet A is larger than the off-platform readership of outlet B ($\tilde{A} > \tilde{B}$). To isolate the impact of outlet scale, suppose that the readerships of A and B are equally distributed across categories P and S ($\tilde{A}_P = \tilde{A}_S$ and $\tilde{B}_P = \tilde{B}_S$). Since the overall number of readers is constant, the off-platform readerships $\tilde{A}_P$ and $\tilde{A}_S$ are larger, while $\tilde{B}_P$ and $\tilde{B}_S$ are smaller than in the baseline case.

Figure 2 helps illustrate how changes in the off-platform readerships $\tilde{\iota_J}$ affect $\hat{A}$ and $\hat{B}$. First, the on-platform options provided by outlet A receive relatively more attention than the options provided by B. In particular, when $\tilde{A}_P$ and $\tilde{A}_S$ increase while $\tilde{B}_P$ and $\tilde{B}_S$ decrease, the indifferent consumers x_{A_P} and x_{A_S} move to the right (increasing A's share of remaining readers), and x_{B_P} and x_{B_S} move to the left (increasing A's share of incoming / B's share of churning readers). Second, the off-platform markets $\tilde{A}_P$ and $\tilde{A}_S$ become relatively larger than $\tilde{B}_P$ and $\tilde{B}_S$ (in terms of *Figure 2,* the shaded areas above the Hotelling-lines grow or shrink, respectively). However, since A can increase its on-platform share of readers in every market, the attention effect dominates when considering the overall effect of outlet scale.

In sum, larger outlets can capture more demand and attract readers from smaller competitors in the on-platform case, so we expect larger outlets to be more negatively affected in their web traffic when they are removed from aggregators.

Hypothesis 1: Web traffic to outlets will be affected more negatively by their removal from aggregators the larger in scale the outlet.

3.2.2 Outlet Focus

Consider now the case where outlets *A* and *B* are equally large in scale ($\tilde{A} = \tilde{B}$), but outlet *A* is more focused ($\tilde{A}_P > \tilde{A}_S$ and $\tilde{B}_P = \tilde{B}_S$), i.e. readers of outlet *A* are not equally distributed over both categories as there are more readers in category *P* than in category *S*.

Again, changes in the off-platform readership affect on-platform readership in two ways. First, outlet *A* attracts relatively more attention in category *P* and relatively less in *S*. This implies that the indifferent consumer x_{B_P} moves to the left (increasing *A*'s share of incoming readers), while x_{B_S} moves to the right (diminishing *A*'s share of incoming readers). Since $x_{B_P} = \frac{1}{2}\left(\frac{\tilde{B}-\tilde{A}_P}{\tilde{B}+\tilde{A}_P} + 1\right)$ and $x_{B_S} = \frac{1}{2}\left(\frac{\tilde{B}-\tilde{A}_S}{\tilde{B}+\tilde{A}_S} + 1\right)$ are convex in $\tilde{A}_P$ and $\tilde{A}_S$ respectively, a marginal increase in $\tilde{A}_P$ affects x_{B_P} to a smaller extent than a marginal decrease in $\tilde{A}_S$ affects x_{B_S}. Thus, *A*'s gain in incoming readers from $\tilde{B}_P$ cannot offset the loss from $\tilde{B}_S$, and *A* loses more readers on-platform than it can gain. Note that since $\tilde{A}$, $\tilde{B}_P$ and $\tilde{B}_S$ do not change relative to the baseline case, the indifferent readers x_{A_P} and x_{A_S} are unaffected. Further, the relative importance of the off-platform readership $\tilde{A}_P$ ($\tilde{A}_S$) grows (shrinks), but since x_{A_P} and x_{A_S} do not change, these effects cancel each other out. Hence, we expect:

> *Hypothesis 2: Web traffic to outlets will be affected more negatively by their removal from aggregators the less focused the outlet.*

4 Data and Methods

4.1 Empirical Setting

Our empirical setting is the German newspaper industry. Similar to developments in many other countries, the German newspaper industry has undergone increasing digitization over the past years. Most print media outlets do not just produce physical newspapers but make (some) of their content available on online websites as well. This content is often combined by news aggregators, which are considered a potential threat for the original content producers. For this reason, the German government introduced the so-called "ancillary copyright for press publishers" (*Leistungsschutzrecht für*

Presseverleger), which allows print media companies to charge royalty fees if other companies reuse their content. The bill was passed by the German parliament on March 22, 2013 (Bundesrat 2013) and came into force on August 1, 2013 (Bundesanzeiger 2013). Importantly, the German government exempted "short excerpts of text" from the regulation the week before the bill was passed, arguing that it would constrain the public's fundamental right for information (Klaiber 2013).

Exempting "short excerpts of text" led to opposing views on whether the copyright bill applied to the text snippets that news aggregators typically provide along with a news article's title and URL. While news aggregators were reluctant to pay, news outlets insisted. In particular, *VGM* (*VG Media, Gesellschaft zur Verwertung der Urheber- und Leistungsschutzrechte von Sendeunternehmen und Presseverlegern mbH*),[15] a German copyright collecting society of privately owned broadcasters and press publishers, urged aggregators to pay royalty fees for the reuse of content that its members, a subset of German newspapers, produce. Consequently, *VGM* compiled a pricing schedule that would allow licensees to reuse its members' original content and threatened to file lawsuits against news aggregators that refused to do so (Kuri 2014). At the time of these events, *VGM* members included the majority of private German television and radio broadcasters, and several press publishers with their online news outlets. Prominent examples include *Axel Springer*, *Funke Mediengruppe*, and *ProSiebenSat.1*. *Table 2* provides an overview of all *VGM* members considered in this study.

To avoid further dispute with *VGM*, several German news aggregators, including *gmx.de*, *web.de*, and *t-online.de*, removed all content of *VGM* members from their platforms in August 2014, but continued to display news articles of non-members (Kruse 2014). We exploit this removal of *VGM* news articles from several news aggregators to study how the web traffic of *VGM* news outlets is affected compared to the web traffic of news outlets whose articles remained on the aggregators.[16]

---

[15] Note that *VG Media* has changed its name to *Corint Media* in 2021.

[16] A similar setting has been used by Calzada and Gil (2020) to study the effect of news aggregators on news outlets. The empirical setting of their study however differs in two main points. First, Calzada and Gil (2020) study a more limited two-week time period in 2014, during which *VGM* members where temporarily removed from *Google News*, which ultimately led *VGM* to allow *Google News* to use excerpts of their members' content for free. Second, while Calzada and Gil (2020) focus on a limited sample of domains that include local, national, business and sports outlets, our study specifically covers a much larger sample of local news outlets.

We focus on *local* news outlets. These news outlets have a particularly strong potential to increase their access to consumers when joining a news aggregator, as they would have difficulties in attracting readers outside their limited market in the absence of news aggregators. Further, local news outlets have traditionally played a much bigger role in the German newspaper industry than in many other countries, both in terms of the sheer number of outlets and in terms of the extent to which readers appreciate their content (Media Landscapes 2021, Newman 2020). At the same time, the number of subscribers to local news outlets is declining, which raises the question of whether digitization of the news industry and the increasing availability of free content play a role in this process (Media Landscapes 2021). All this makes the German newspaper industry an excellent setting to explore whether smaller firms benefit from being on a platform.

4.2 Data

To analyze the effect of the legal dispute on news outlet performance, we collected data from the website of the *German Audit Bureau of Circulation IVW* (*Informationsgemeinschaft zur Feststellung der Verbreitung von Werbeträgern e.V.*). The *IVW* is an independent, non-commercial organization that collects circulation data on print media outlets (e.g. newspapers) as well as traffic data on digital advertising media (e.g. online news outlets) and makes it available to advertisers and advertising agencies. The aim is to create transparency in the market for advertising and to provide advertisers with reliable data to monitor the performance of the media in which they advertise. Data quality is ensured through standardized measurement and continuous auditing under the supervision of both advertising media (e.g. publishers) and advertiser (e.g. advertising agencies) representatives. The *IVW* data includes information on most German media outlets and can be publicly accessed through the *IVW* website. Compared to other web traffic data from sources like *SimilarWeb* or *Alexa*, this data is not based on estimates but represents the actual traffic on a given website. We complement this data with additional information gathered from *Onlineatlas der Zeitungen (Online Newspaper Atlas)*, a website providing information on a news outlet's (offline) distribution area.

4.2.1 Dependent Variables

We use a news outlet's number of visits per month to measure our dependent variable, web traffic, where a visit is defined as an entire user session.[17] If a user accesses three articles of one news outlet, *IVW* counts one visit. If a user accesses two articles, leaves the news outlet's domain for at least thirty minutes, and returns to access another article, *IVW* counts two visits.[18] As few news outlets attract the lion's share of user attention, we use the logarithm of visits as dependent variable.

4.2.2 Independent Variables

Treatment indicator

To generate our treatment indicator, we retrieve a list of all *VGM* members from the association's website *vg-media.de*. Our analysis includes 57 members and 83 nonmembers (see *Table 2* and the following section for details on the selection of observations).

Insert *Table 2* about here

Scale

We generate the average number of visits per month *before* March 2013 as our measure for news outlet scale. Like the dependent variable, our measure for news outlet scale is heavily skewed, so we use its logarithm as independent variable in the empirical analysis. Our results are robust to using alternative measures for news outlet scale, such as average circulation per month before March 2013, the number of counties that make up a news outlet's (offline) distribution area or the relative scale compared to its competitors (see Appendix).

[17] Note that these are visits of the news outlet's *own* web page, not of the outlet's content on the news aggregator. Our measure thus captures the *net* effect on the web traffic that news outlets attract to their web page.
[18] We use page impressions per month as an alternative dependent variable in our robustness checks. For instance, if a user accesses three articles of one news outlet, *IVW* would count this as three page impressions.

Focus

To measure the news outlets' focus, i.e. the extent to which news outlets are active in different categories, we use their average monthly number of page impressions before March 2013 by categories (see *Table 3*).

Insert *Table 3* about here

Although these category page impressions do not directly measure a news outlet's content category composition, they are a valid proxy because a large number of page impressions within a particular category indicates that the news outlet puts more focus on that category.

For each news outlet *i*, we compute the relative number of page impressions for each of the eight main categories listed in *Table 3*. Denote these fractions as $c_{i,j}$, with $j = 1, \ldots, 8$ and $\sum_j c_{i,j} = 1$. Based on the fractions $c_{i,j}$, we then compute the Herfindahl diversity measure, a well-known measure of within-firm diversity (e.g. Zahavi and Lavie 2013):[19]

$$Herfindahl_i = 1 - \frac{\sum_j c_{i,j}^2}{\left(\sum_j c_{i,j}\right)^2} \tag{3}$$

As an alternative measure of focus, we compute a news outlet's scope by considering the number of categories a news outlet covers. An outlet covering many different topics ranging from, say, political news to celebrities is broader in scope than a news outlet covering a single topic (e.g., sports).[20] We use a simple count of the number of a news outlet's active categories:

$$Count_i = Count(c_{i,j} \neq 0) \tag{4}$$

We again use the logarithm as independent variable in our analysis.

4.2.3 Observations

Our main analysis comprises 18 months before the copyright bill was passed in March 2013 and 18 months after the German news aggregators removed *VGM* members from their platforms in August

[19] We use the entropy measure (Jacquemin and Berry 1979, Palepu 1985) as a robustness check.
[20] This can be thought of as an extreme case of focus where a given news outlet focuses on a subset of news categories.

2014 (i.e., we discard the 17 months in between).[21] By focusing on local news outlets, we avoid (potentially unobserved) news outlet heterogeneity beyond scale and focus (e.g., in terms of quality or news content). *Table 4* summarizes all variables used in our study.

Insert *Table 4* about here

4.3 Empirical Strategy

4.3.1 Baseline Specification

To isolate the causal effect of platform removal on local news outlets' web traffic, we use the passage of the copyright bill in March 2013 and the subsequent unexpected dispute between *VGM* and German news aggregators that eventually led to the removal of *VGM* members from the aggregators' platforms. We compare the change in web traffic of *VGM* members before March 2013 and after August 2014 to the change in web traffic of non-members in a difference-in-differences framework. In our main specification, we omit the period between these two dates to exclude potential confounding effects from unobserved actions of news aggregators or the short-term removal of *VGM* members from *Google News*. The baseline regression is:

$$\log(Visits)_{it} = \beta(VGM_i * Post_t) + \varphi_i + \lambda_t + \varepsilon_{it} \tag{5}$$

where the dependent variable is the web traffic (in visits) of news outlet *i* in month *t*, VGM_i is a dummy equal to one for all *VGM* members, $Post_t$ is a dummy equal to one for all time periods after the passage of the copyright bill, φ_i and λ_t are outlet and monthly fixed effects, respectively. The parameter of interest in (5), β, gives the average change in web traffic of *VGM* members after August 2014 relative to the change of non-members.

Note that it is crucial to include monthly and outlet fixed effects in our regression analysis. The monthly fixed effects control for general changes in all news outlets' web traffic, including seasonality and a growing online audience. The outlet fixed effects capture (time-invariant) unobserved

[21] We obtain similar results when using different time windows (see Appendix).

heterogeneity among our observations, including differences in quality and specific features of the local market they operate in. Omitting such fixed effects could lead to over- or underestimation of the impact of platform removal on outlets' web traffic. For example, high-quality, well-organized, or visually attractive news outlets could tend to join *VGM.* As the web traffic of such outlets is likely to be larger, too, omitting outlet fixed effects would lead to overestimating the impact of platform removal. Including outlet fixed effects prevents omitted variable bias from such unobserved time-invariant heterogeneity. Moreover, outlet fixed effects align our empirics more closely with the theoretical framework in Section 3. Our model considers the *ceteris paribus* effects of scale and focus: we isolate the impact of both moderators while holding everything else fixed. Using outlet fixed effects to capture time-invariant heterogeneity among outlets brings our empirical analysis as close to this idea as possible.

4.3.2 Scale and Focus as Moderators

As argued in *Section 3*, scale and focus of a news outlet can moderate the effect of being delisted from an aggregator. Hence, we augment *(5)* to a triple-difference-in-differences equation:

$$\log(Visits)_{it} = \beta_1(VGM_i * Post_t) + \beta_2(S_i * Post_t) + \beta_3(VGM_i * S_i * Post_t) + \varphi_i + \lambda_t + \varepsilon_{it} \qquad (6)$$

where S_i refers to the measures of outlet scale and focus as defined in *Section 4.2*. The parameter of interest in *(6)* is β_3, as it corresponds to the moderating effect of outlet scale and focus on the effect of being removed by the aggregator platforms in August 2014.

5 Results

5.1 Baseline Specification

The first column of *Table 5* shows OLS estimates for baseline regression *(5)*. All specifications in *Table 5* include outlet and monthly time fixed effects. Standard errors are robust to heteroskedasticity and clustered at the outlet level; *p*-values are in parentheses. The estimate for β is close to zero and statistically insignificant. Thus, being delisted from aggregators in August 2014 left outlets' web traffic unchanged on average.

Insert *Table 5* about here

5.2 Scale as a Moderator

Column 2 of *Table 5* displays the OLS estimate of the triple difference-in-differences (*6)* when we consider the moderating effect of outlet scale. Consistent with Hypothesis 1, we find that the estimate for β_3 is negative and statistically significant at the 5%- level.

The point estimate can be interpreted as follows. According to $\hat{\beta}_3$, a 1% increase in outlet scale is associated with a 0.09% decrease in web traffic after August 2014. In contrast to β_3, our estimate for β_1 is positive and statistically significant at the 1% level. To interpret our estimates, note that β_1 captures the baseline effect of being removed from an aggregator for an outlet of zero scale. Outlets of zero scale, however, do not exist and examining $\hat{\beta}_1$ in isolation is not meaningful. Thus, we jointly evaluate $\hat{\beta}_1$and $\hat{\beta}_3$ at meaningful margins of outlet scale. As before, consider the median *VGM* member (in terms of scale) in our sample with $\log(avgVisits)_i \approx 13.58$. For this outlet, the overall treatment effect is negative and equal to -0.008 $(1.199 - 0.0877 \times 13.58)$. In other words, the median *VGM* member's web traffic decreases by about 0.8% after August 2014. While this effect grows (i.e., becomes more negative) if outlet scale increases, it approaches zero and eventually switches its sign if outlet scale shrinks. The reversal of the treatment effect occurs around the 45$^{\text{th}}$ percentile of the distribution of outlet scale, hence, the overall effect of being removed from aggregator platforms is negative for the majority of *VGM* members in our sample.

Thus, in addition to supporting Hypothesis 1, our analysis on outlet scale reveals that the sign of the overall treatment effect differs between small- and large-scale outlets, which explains the zero average effect that we document in column 1 of *Table 5*. Moreover, the effect heterogeneity supports our theory from *Section 3.2.1*, where we argue that large-scale outlets are more likely to benefit from aggregators as they are better able to capture a disproportionate share of demand from the platforms than small-scale outlets. Small-scale outlets are unlikely to reap additional demand from aggregators and may even suffer from intensified competition for readers on platforms. Consequently, small-scale outlets may benefit from platform removal. Our estimates provide empirical support for this reasoning.

5.3 Focus as Moderator

Columns 3 and 4 of *Table 5* show the OLS estimates of the triple difference-in-differences *(6)* when we consider the moderating effects of outlet focus (column 3) and scope (column 4). Analogous to column 2, we find that both estimates for β_3 are negative and statistically significant at the 1%- (column 3), or at the 5%-level (column 4), thus supporting Hypothesis 2.

To interpret the point estimates, take column 3, where we consider outlet focus in terms of the Herfindahl diversity measure, as an example. According to $\hat{\beta}_3$, a one standard deviation decrease in outlet focus (sd = 0.16) corresponds to a 17.6% increase in *VGM* members' web traffic after August 2014. The median *VGM* member in terms of scale attracted an average of 917,000 visits per month before March 2013. If the focus of this median outlet decreases by one standard deviation, its web traffic would increase by about 141,000 visits per month.

Our estimate for β_1 is positive in column 3 and statistically significant at the 1%-level. Again, we jointly evaluate $\hat{\beta}_1$ and $\hat{\beta}_3$ at meaningful margins of outlet focus. To this end, consider the median *VGM* member in our sample in terms of the Herfindahl diversity measure ($Herfindahl_i = 0.48$). For this outlet, the full treatment effect is equal to -0.038 $(0.490 - 1.100 \times 0.48)$ and therefore negative. While this effect grows (i.e., becomes more negative) if outlet focus decreases, it approaches zero and eventually switches its sign if outlet focus increases. The reversal of the treatment effect again occurs around the 45th percentile of the distribution of outlet focus. Hence, the overall effect of being removed from aggregators is negative for the majority of *VGM* members in our sample.

5.4 Scale and Focus as Moderators

It is possible that large-scale outlets also have a broad focus. In other words, it could be that outlet scale and focus do not operate as independent moderators, but simultaneously characterize the same outlets. The unconditional correlation between scale and focus is .1687 (significant at $p < 0.000$, indicating moderate collinearity. To evaluate the relationship between our two main moderators, we include both scale and focus in our regression in column 5 of *Table 5*. Our estimates for the moderating effects of scale and focus remain negative, albeit at reduced statistical significance (p=.011 → p=.090 and p=.004 → p=.028 for scale and focus, respectively), and their magnitudes decrease by 25% and 37%, respectively. Thus, despite some moderate collinearity, outlet scale and focus capture two different dimensions of heterogeneity.

6 Discussion and Conclusion

We explore on-platform effects of competition for consumer attention for content providers. While platforms can grant smaller providers access to larger potential demand, they also expose them to more intense competition with other, potentially larger providers, all battling to capture the same limited consumer attention. Studying local news outlets (i.e., news outlets which have traditionally only had access to a limited market) in Germany, we analyze how the performance of a subset of these outlets that were removed from several news aggregators in Germany after a legal dispute evolved compared to local news outlets that were not removed.

We argue (and show empirically) that news outlets need to be relatively large in scale and/or broad in their focus to benefit from being listed on news aggregators. When displayed side-by-side, articles of larger outlets are likely to attract a larger share of consumer attention compared to those of smaller outlets. We posit that content provided by larger-scale outlets is the default option for potential readers whereas directing attention to smaller-scale outlets would require additional cognitive effort from readers to move past the first option in their mind. Put differently, outlet size and the resulting recognizability becomes a salient attribute commanding attention, which helps larger outlets capture more of the demand on news aggregators. Outlets also capture a larger share of on-platform demand if they cover different types of content broadly. This helps them attract the attention of readers across multiple content categories.

Indeed, very small or focused news outlets may be better off not being on news aggregators at all. By making their articles available on news aggregators, smaller outlets are exposed to negative, outgoing attention spillovers offsetting the positive, incoming attention spillovers, which ultimately draws attention and readership away from them. Similarly, if outlets are very focused in their content, they do not attract sufficient attention from their competitors' readers while losing some of their own readers to competitors.

Implications for Research. We contribute to the debate on the substitution between on- and off-platform sales channels as key driver of complementors' success (Kretschmer and Peukert 2020, Athey et al. 2021, Calzada and Gil 2020, Chiou and Tucker 2017) and to work arguing that the shift in the type of competition on platforms compared to traditional markets affects complementors' ability to capture value (Adner and Lieberman 2021, Cennamo 2021, Zhu and Liu 2018). Our findings stress the possibly asymmetric competitive effects complementors face on platforms when competing for consumer attention, and identify the relative importance of two main factors that drive these effects: scale and focus. We go beyond asking whether news excerpts on news aggregators substitute for reading the full article and study instead how the on-platform competitive relationships between complementors play out. Thus, we shift the focus from competition between aggregator and outlets to competition among news outlets on the same platform. This contributes to a broader research agenda on the largest group of actors of the platform economy, the complementors, and how their heterogeneity affects their benefits from platform membership. Thus, we add granularity to the analysis of platform markets by focusing on the effects of complementor heterogeneity (compared to market-level factors).

We also add to work on the relationship between platform dynamics and the range of firm (complementor) activities. Studies on firm scope in platform markets have mainly focused on the platform's decisions (Cennamo 2021, Gawer 2021, Giustiziero, Kretschmer, Somaya and Wu 2022), while complementors are often (implicitly) considered small and atomistic. With the exception of Tavalaei and Cennamo (2021), who study complementors' specialization strategies in the context of innovation platform ecosystems, studies on complementor scale and scope are rare. Our study suggests that complementors, especially those that join a platform after a period of independent operation, differ in dimensions that matter for their performance on the platform. Specifically, firms can obtain greater returns from scale as well as breadth. This may call the efficacy of "focus strategies" a la Porter (1980) into question and may imply another source of economies of scale through the ability to attract eyeballs on a crowded platform, especially if products are considered ex-ante homogenous by consumers. How, then, can firms create enough exposure to become the default option on such a platform? Which strategies are especially useful in the "war for attention" and how will competition play out in these arenas?

Our findings also speak to the debate on the role of algorithms in competition on platform markets. Note that we do not observe the underlying algorithms at play in our setting, but anecdotal evidence suggests that the algorithmic "bias" of platforms, i.e., the platform favoring popular content from large content providers, serves to reinforce existing asymmetries. Such a bias would then drive the selection of news articles listed in the clustered topics shown on news aggregators. Other, less-known options may not be selected by the algorithm for the given topic. While this bias does not dictate the readers' ultimate choices from the menu, it may steer readers to content by larger and broader outlets, which would serve to reinforce an already existing tendency of "winner-take-most" dynamics. This raises interesting questions about market efficiency. Does the market become more efficient and "effectively" reward efficient players (those that are large and broadly diversified) while penalizing players below a minimum viable scale? How does this drive incentives to provide quality content?

Relatedly, our work underlines the role of complementor heterogeneity. Rating systems and other tools introduced by platforms to reduce transaction costs and coordinate cross-side market interactions reduce information asymmetries and allow for signaling of high quality to consumers, which in turn affects the selection decisions of consumers (Chevalier and Mayzlin 2006, Sun 2012), arguably even more so for experience goods (Nelson 1970, Kumar et al. 2014). However, if attention has a large impact on consumer choice, such mechanisms may reward prior attention with more attention, which can raise entry barriers for better or more targeted alternatives. The link between competition for consumer attention, ratings, and quality of complements is central to understanding overall platform efficiency.

Another area where competition for attention can change our outlook is in the interplay of platform first-party complements and third-party complementors in a market niche. Zhu and Liu (2018), analyzing *Amazon's* entry decisions into its third-party sellers' product market space, show that *Amazon* is more likely to enter more popular product categories to appropriate value from successful complementors. However, they also find that demand for all products in the focal category increases after *Amazon*'s entry. From an attention-based perspective, *Amazon*'s entry appears to produce a dual effect of attention spillovers similar to the one we identified: *Amazon* will redirect some attention from third-party products to its own, but it will also attract more consumer attention to this product category, which in turn may spill over to third-party sellers in the same category. It would be interesting to see how our scale/focus dimensions would play out in this context.

Implications for Practice and Policy. Our study also raises important questions for practice and policy, particularly in light of the current policy debate around the impact of information aggregators and platforms on market efficiency, and on society more broadly. In the specific case of news outlets, what is the impact that these competitive dynamics might have at the societal level? Do they pose a risk to democracy, as some have vehemently advocated (Greenslade 2016), by reducing the plurality of news sources, leaving only big players in the competitive landscape? Relatedly, do these dynamics condemn us citizens to consume (attention-grabbing) content of lower quality? While our results would suggest that in a world where news would be consumed exclusively via platform aggregators, we would likely observe greater industry concentration with few large players contending the attention space, it is an open question if this is bad or good for content quality on the "production" side and whether this leads to more and better-informed readers on the "consumption" side. Given the spillovers across news outlets, readers may "multihome" easily across news outlets on the same platform and source similar (or different) information from different news outlets, putting constant pressure on the quality of news provided by different outlets (Peitz and Reisinger 2014). This might increase overall quality. Conversely, in chasing a larger audience, news outlets may promote attention-grabbing content at the cost of de-emphasizing other content that is high quality and socially beneficial, but of more limited appeal. If such niche content is indeed of higher average quality than popular content, we could see a drop in the average quality of content offerings. Then, policy interventions requiring news aggregators to include other dimensions of "quality" and societal relevance in their algorithms to preserve greater plurality of news sources may be needed to correct for a "race to bottom" of attention-generating content.

Avenues for Future Research. Our study opens up several avenues for future research. First, while our empirical context of online news content comes with a number of industry specificities (e.g. frequently free consumption, little repeat consumption of articles), it lets us isolate the mechanisms around competition for attention, and it creates ex-ante heterogeneity among complementors from the "outside world" (off-platform). We expect these mechanisms to apply also, for instance, to complementors selling their products on transaction platforms like *Amazon Marketplace*, or small restaurants reaching out to consumers via *UberEats*. While featuring on these platforms gives firms access to many potential consumers, it also puts them in direct competition with larger competitors who may attract more attention and ultimately capture more demand when products are displayed side-by-side. While the balance between (i) improving matches between consumer preferences and products and (ii) competition for user attention on the other hand may be more even on these transaction platforms, the mechanisms on the attention side will likely apply too, even if they are partly offset by match quality and thus not the only driver of competition.[22] The question of whether and to what extent smaller complementors can command attention when competing with larger complementors on a platform will matter for complementors irrespective of the product or service they offer. Second, we do not directly observe the behavior and decision-making process of readers. While our findings are in line with our theoretical predictions on the net effect of the incoming and outgoing attention spillovers, taking our study to the level of specific products consumed by individual users would be a promising avenue to further isolate the mechanisms at play.

[22] See e.g., the discussion on the relationship between competition for user attention and product rating systems.

References

Adner R, Lieberman M (2021) Disruption through complements. *Strategy Science*, 6(1):91-109.

Alaoui L, Germano F (2020) Time scarcity and the market for news. *Journal of Economic Behavior & Organization*, 174:173-195.

Alba JW, Hutchinson JW (1987) Dimensions of consumer expertise. *Journal of Consumer Research*, 13(4):411-454.

Anderson C (2004) The long tail. *Wired Magazine*, 12(10):170–177.

Athey S, Mobius M, Pal J (2021) The impact of aggregators on internet news consumption (No. w28746). *National Bureau of Economic Research*.

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

Boik A, Greenstein S, Prince J (2016) The empirical economics of online attention (No. w22427). *National Bureau of Economic Research*.

Bordalo P, Gennaioli N, Shleifer A (2013) *Competition for attention*. (No. w19076). *National Bureau of Economic Research*.

Brynjolfsson E, Hu Y, Smith MD (2003) Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science*, 49(11):1580-1596.

Brynjolfsson E, Hu Y, Simester D (2011) Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science*, 57(8):1373-1386.

Bundesanzeiger (2013) Achtes Gesetz zur Änderung des Urheberrechtsgesetzes. Retrieved June 12, 2022, https://www.bgbl.de/xaver/bgbl/start.xav#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl113s1161.pdf%27%5D__1541854256266.

Bundesrat (2013) Länder billigen Leistungsschutzrecht. Retrieved June 12, 2022, https://www.bundesrat.de/SharedDocs/pm/2013/080-2013.html;jsessionid=9023BDDC9225190A687D8D65512A7B93.1_cid382.

Calvano E, Polo M (2021) Market power, competition and innovation in digital markets: A survey. *Information Economics and Policy*, 54, 100853.

Calzada J, Gil R (2020) What Do News Aggregators Do? Evidence from Google News in Spain and Germany. *Marketing Science*, 39(1):134-167.

Cennamo C (2021) Competing in digital markets: A platform-based perspective. *Academy of Management Perspectives*, 35(2):265-291.

Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3):345-354.

Chiou L, Tucker C (2017) Content aggregation by platforms: The case of the news media. *Journal of Economics & Management Strategy*, 26(4):782-805.

Dellarocas C, Katona Z, Rand W (2013) Media, aggregators, and the link economy: Strategic hyperlink formation in content networks. *Management Science*, 59(10):2360-2379.

Dellarocas C, Sutanto J, Calin M, Palme E (2016) Attention allocation in information-rich environments: the case of news aggregators. *Management Science*, 62(9):2543-2562.

Elberse A (2008) The Long Tail Debate: A Response to Chris Anderson. *Harvard Business Review*. (July 2), https://hbr.org/2008/07/the-long-tail-debate-a-respons.

Elberse A, Oberholzer-Gee F (2006) Superstars and underdogs: An examination of the long tail phenomenon in video sales (p. 43). *Division of Research, Harvard Business School*.

Evans DS (2013) Attention rivalry among online platforms. *Journal of Competition Law & Economics*, 9(2):313-357.

Evans DS (2019) Attention platforms, the value of content, and public policy. *Review of Industrial Organization*, 54(4):775-792.

Gawer A (2021) Digital platforms' boundaries: The interplay of firm scope, platform sides, and digital interfaces. *Long Range Planning*, 54(5):102045.

George LM, Hogendorn C (2020) Local News Online: Aggregators, Geo-Targeting and the Market for Local News. *The Journal of Industrial Economics*, 68(4):780-818.

Giustiziero G, Kretschmer T, Somaya D, Wu B (2022) Hyperspecialization and Hyperscaling: A Resource-based Theory of the Digital Firm. *Strategic Management Journal*. Forthcoming.

Greenslade R (2016) Publishers call on government to help over Google and Facebook. *The Guardian*. (September 22), https://www.theguardian.com/media/greenslade/2016/sep/22/publishers-government-google-facebook-newspaper.

Hansen MT, Haas MR (2001) Competing for attention in knowledge markets: Electronic document dissemination in a management consulting company. *Administrative Science Quarterly*, 46(1):1-28.

Hirche T (2013) Weiter Zu- und Absagen von Verlagen an VG Media. *IGEL Initiative gegen ein Leistungsschutzrecht*. Retrieved June 12, 2022, https://leistungsschutzrecht.info/news/2013-11-25/weitere-zu-und-absagen-von-verlagen-an-vg-media.

Hong S (2011) A simple model of news aggregators, information cascades, and online traffic. Working paper, College of Social Sciences, Yonsei University, Korea, Harvard Institute for Quantitative Social Science, Boston.

Jacobides MG, Cennamo C, Gawer A (2018) Towards a theory of ecosystems. *Strategic management journal*, 39(8):2255-2276.

Jacquemin AP, Berry CH (1979) Entropy measure of diversification and corporate growth. *The Journal of Industrial Economics*, 359-369.

Jeon DS, Nasr N (2016) News aggregators and competition among newspapers on the internet. *American Economic Journal: Microeconomics*, 8(4):91-114.

Kahneman D (1973) *Attention and effort* (Vol. 1063, pp. 218-226). Englewood Cliffs, NJ: Prentice-Hall.

Klaiber S (2013) Neues Leistungsschutzrecht: ein Gesetz, viele Fragen. Retrieved June 12, 2022, https://www.focus.de/kultur/medien/tid-29839/google-verleger-und-der-bundestag-neues-leistungsschutzrecht-ein-gesetz-viele-fragen_aid_930403.html.

Kretschmer T, Leiponen A, Schilling M, Vasudeva G (2022) Platform ecosystems as meta-organizations: Implications for platform strategies. *Strategic Management Journal*, 43(3):405-424.

Kretschmer T, Peukert C (2020) Video killed the radio star? Online music videos and recorded music sales. *Information Systems Research*, 31(3):776-800.

Kruse J (2014) Leistungsschutzrecht und VG media - und raus bist du. Retrieved June 12, 2021, http://www.taz.de/!5033165/.

Kumar A, Smith MD, Telang R (2014) Information discovery and the long tail of motion picture content. *MIS Quarterly*, 38(4):1057-1078.

Kuri J (2014) Leistungsschutzrecht: VG Media nimmt Microsoft, Yahoo und Telekom ins Visier. Retrieved June 12, 2022, https://www.heise.de/newsticker/meldung/Leistungsschutzrecht-VG-Media-nimmt-Microsoft-Yahoo-und-Telekom-ins-Visier-2235726.html.

Lanham RA (2006) *The economics of attention: Style and substance in the age of information.* University of Chicago Press.

Loh J, Kretschmer T (2022) Online Communities on Competing Platforms: Evidence from Game Wikis. *Strategic Management Journal*. Forthcoming

Macdonald EK, & Sharp BM (2000) Brand awareness effects on consumer decision making for a common, repeat purchase product: A replication. *Journal of Business Research*, 48(1):5-15.

McIntyre DP, Srinivasan A (2017) Networks, platforms, and strategy: Emerging views and next steps. *Strategic Management Journal*, 38(1):141-160.

Media Landscapes (2021) Germany. Retrieved June 12, 2022, https://medialandscapes.org/country/germany/media/print.

Nelson P (1970) Information and consumer behavior. *Journal of Political Economy*, 78(2):311-329.

Newman N (2020) Digital News Report - Executive Summary and Key Findings of the 2020 Report. Retrieved June 12, 2021, https://www.digitalnewsreport.org/survey/2020/overview-key-findings-2020/.

Palepu K (1985) Diversification strategy, profit performance and the entropy measure. *Strategic Management Journal*, 6(3):239-255.

Panico C, Cennamo C (2022) User preferences and strategic interactions in platform ecosystems. *Strategic Management Journal*, 43(3):507–529.

Parker GG, Van Alstyne MW (2005) Two-sided network effects: A theory of information product design. *Management Science*, 51(10):1494-1504.

Peitz M, Reisinger M (2014) *The economics of internet media*. Working paper series, No 14-23, University of Mannheim. http://hdl.handle.net/10419/129576.

Piezunka H, Dahlander L (2015) Distant search, narrow attention: How crowding alters organizations' filtering of suggestions in crowdsourcing. *Academy of Management Journal*, 58(3):856-880.

Porter ME (1980) Competitive strategy: Techniques for analyzing industries and competitors. New York: Free Press.

Prat A, Valletti TM (2021) Attention oligopoly. *American Economic Journal: Microeconomics*, Forthcoming.

Rietveld J, Eggers JP (2018) Demand heterogeneity in platform markets: Implications for complementors. *Organization Science*, 29(2):304-322.

Rossi F, Rubera G (2021) Measuring Competition for Attention in Social Media: National Women's Soccer League Players on Twitter. *Marketing Science*, 40(6):1147-1168.

Rui H, Whinston A (2012) Information or attention? An empirical study of user contribution on Twitter. *Information Systems and e-Business management*, 10(3):309-324.

Sun M (2012) How does the variance of product ratings matter?. *Management Science*, 58(4):696-707.

Sun L, Rajiv S, Chu J (2016) Beyond the more the merrier: The variety effect and consumer heterogeneity in system markets. *International Journal of Research in Marketing*, 33(2):261-275.

Tajedin H, Madhok A, Keyhani M (2019) A theory of digital firm-designed markets: Defying knowledge constraints with crowds and marketplaces. *Strategy Science*, 4(4):323-342.

Tan TF, Netessine S, Hitt L (2017) Is tom cruise threatened? an empirical study of the impact of product variety on demand concentration. *Information Systems Research*, 28(3):643-660.

Tavalaei MM, Cennamo C (2021) In search of complementarities within and across platform ecosystems: Complementors' relative standing and performance in mobile apps ecosystems. *Long Range Planning*, 54(5):101994.

The Economist (2017) The battle for consumers' attention – Forget the long tail. *The Economist* (February 9), https://www.economist.com/special-report/2017/02/09/the-battle-for-consumers-attention.

Tucker C, Zhang J (2011) How does popularity information affect choices? A field experiment. *Management Science*, 57(5):828-842.

Zahavi T, Lavie D (2013) Intra-industry diversification and firm performance. *Strategic Management Journal*, 34(8):978-998.

Zentner A, Smith M, Kaya C (2013) How video rental patterns change as consumers move online. *Management Science*, 59(11):2622-2634.

Zhu F, Liu Q (2018) Competing with complementors: An empirical look at Amazon. com. *Strategic Management Journal*, 39(10):2618-2642.

Figures and Tables

Figure 1 Readership Distribution in the On-Platform Case

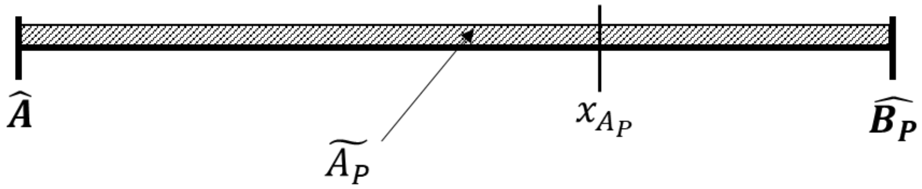


Figure 2 Baseline Setting in the On-Platform Case

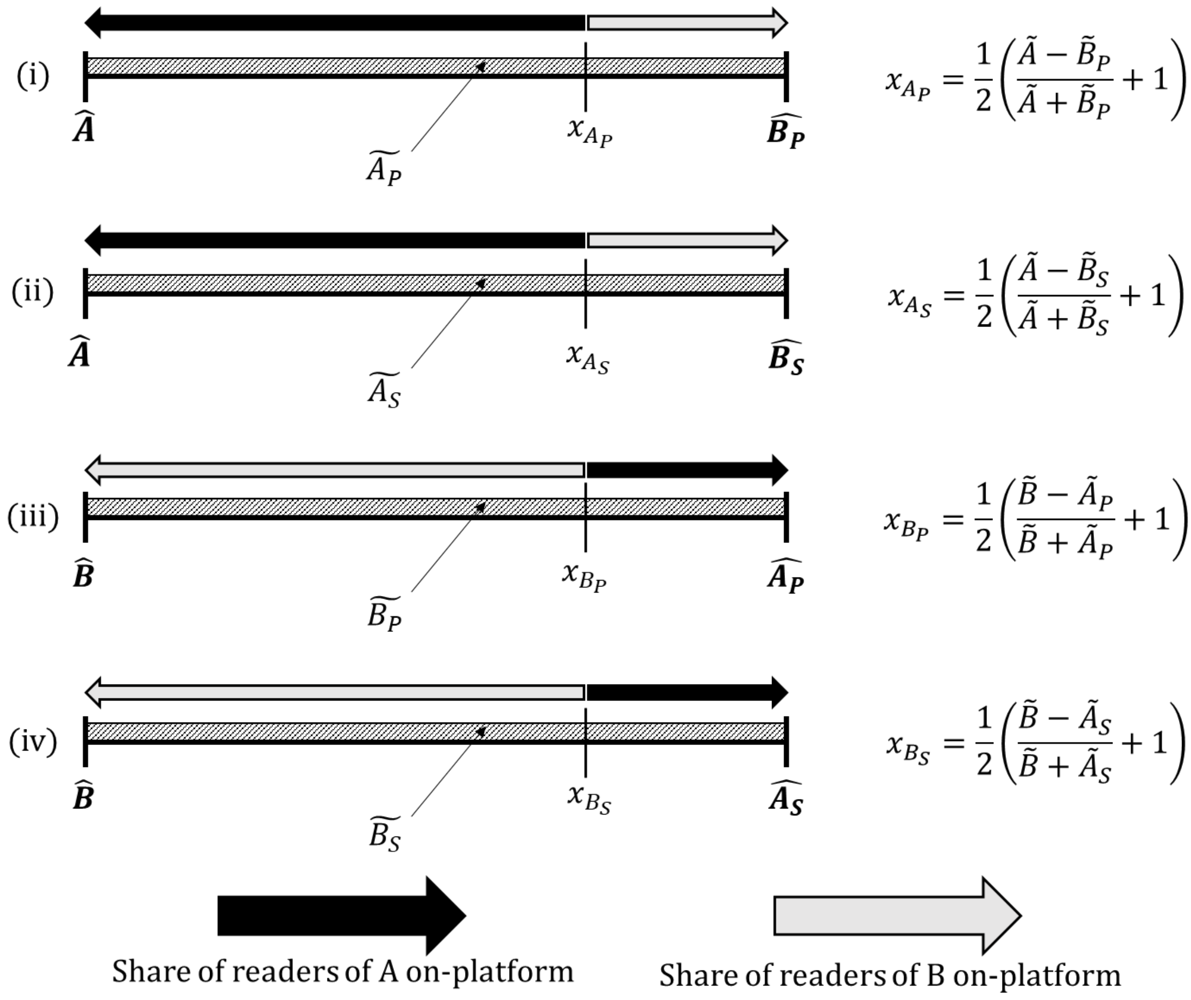


Table 1 Readership across Outlets and Categories (Off-Platform)

	Category P	*Category S*	*Total*
News Outlet A	$\widetilde{A_P}$	$\widetilde{A_S}$	$\tilde{A}$
News Outlet B	$\widetilde{B_P}$	$\widetilde{B_S}$	$\tilde{B}$
Total	$\tilde{P}$	$\tilde{S}$	

Table 2 Overview of VG Media Members and Non-Members in our Main Analysis

VG Media members	Non-members
Aachener Zeitung	Abendzeitung
Hamburger Abendblatt	Allgemeine Zeitung
Aichacher Zeitung	Badische Zeitung
Allgäuer Zeitung	Kreiszeitung Böblinger Bote
Westdeutsche Allgemeine Zeitung	Bocholter-Borkener Volksblatt
Augsburger Allgemeine	Bietigheimer Zeitung
Berliner Morgenpost	Backnanger Kreiszeitung
Berliner Kurier	Borkener Zeitung
Berliner Zeitung	Bürstädter Zeitung
Braunschweiger Zeitung	Cuxhavener Nachrichten
BZ Berlin	Der Patriot
Die Glocke	Deister- und Weserzeitung
Esslinger Zeitung	Delmenhorster Kreisblatt
Express	Mittelbayerische Zeitung
General Anzeiger	Donaukurier
Göttinger Tageblatt	Echo Online
Anzeiger für Harlingerland	Elbe-Jeetzel Zeitung
Hannoversche Allgemeine Zeitung	Dülmener Zeitung
Herfelder Zeitung	Frankfurter Neue Presse
Hessische Niedersächsische Allgemeine	Fränkische Nachrichten
Ibbenbürener Volkszeitung	Frankenpost
Jeversches Wochenblatt	Freie Presse
Kieler Nachrichten	Fuldaer Zeitung
Kreiszeitung	Gäubote
Schaumburger Zeitung	Gelnhäuser Tageblatt
Landeszeitung	General-Anzeiger Bonn
Lübecker Nachrichten	Giessener Allgemeine
Lausitzer Rundschau	Gmünder Tagespost
Leipziger Volkszeitung	Goslarsche Zeitung
Märkische Allgemeine Zeitung	Haller Tagblatt
Mainpost	Hellweger Anzeiger
Hamburger Morgenpost	Hildesheimer Allgemeine
Mitteldeutsche Zeitung	Idowa
Naumburger Tageblatt	Kreis-Anzeiger
Nordbayerischer Kurier	Lauterbacher Anzeiger
Nordkurier	Ludwigsburger Kreiszeitung
Ostfriesische Nachrichten	Main-Echo
Oberhessische Presse	Main-Spitze
Offenbach Post	Münchner Merkur
Ostfriesenzeitung	Mittelbayerische Zeitung
Ostsee Zeitung	Mittelhessen
Peiner Allgemeine Zeitung	Mannheimer Morgen
Rheinische Post	Mühlacker Tagblatt
Kölnische Rundschau	Das Nürnberger Land
Schwäbische Zeitung	Neue Deister-Zeitung
Schaumburger Nachrichten	Neue Osnabrücker Zeitung
Schleswig-Holsteinischer Zeitungsverlag	Niederelbe Zeitung
Südkurier	Nordbayern
Schweriner Volkszeitung	Nordsee-Zeitung
tz	Neue Presse

Volksfreund	Nürtinger Zeitung
Westfälischer Anzeiger	Neue Westfälische
Westfälische Nachrichten	Oberhessische Zeitung
Westfalenblatt	Oberpfalznetz
Waldeckische Landeszeitung	Oldenburgische Volkszeitung
Westdeutsche Zeitung	Oberbayerisches Volksblatt
	Passauer Neue Presse
	Pforzheimer Zeitung
	Remscheider General-Anzeiger
	Rhein-Zeitung
	Rhein-Neckar-Zeitung
	Schaumburger Zeitung
	Schwäbische Post
	Schwarzwälder Bote
	Siegener Zeitung
	Solinger Tageblatt
	Heilbronner Stimme
	Stuttgarter Zeitung
	Traunsteiner Tagblatt
	Südwest Presse
	Sächsische Zeitung
	Westfälische Nachrichten
	Der Teckbote
	Torgauer Zeitung
	Usinger Anzeiger
	Vaihinger Kreiszeitung
	Volksstimme
	Wilhelmshavener Zeitung
	Wetterauer Zeitung
	Wiesbadener Kurier
	Wiesbadener Tagblatt
	Weinheimer Nachrichten
	Wormser Zeitung

Table 3 Classification of Category Page Impressions

Category Classification before May 2014	Category Classification after May 2014	Aggregate Category Classification	Main Newspaper Category
News, Homepage	News	News	Yes
Economics & Finance	Economics & Finance, Job & Career	Economics & Finance	Yes
Sports	Sports	Sports	Yes
Entertainment & Lifestyle	Entertainment, Tabloid, Stars, Film, Music, Fashion & Beauty, Love & Relationships, Living, Real Estate, Garden, Domestic	Tabloid	Yes
Travel	Travel & Tourism	Travel	Yes
Family, Leisure, Health	Family, Kids, Self-Help Health, Food & Drink	Health & Family	Yes
Computer, Telecommunication, Consumer Electronics, Business Communication	Computer, Consumer Electronics. Telecommunication & Broadband	Computer & Electronics	Yes
Science, Technology, Education	Science, Education, Nature, Environment Art, Culture, Literature	Science & Literature	Yes
Erotics	Erotics	Erotics	No
Newsletters	Newsletters	Newsletters	No
Miscellaneous	Miscellaneous	Miscellaneous	No
E-commerce (Aggregate)	Onlineshops, Shopping Mall, Auctions, B2B Marketplaces, Real Estate, Classified Ads, Jobs Classified Ads, Vehicle Classified Ads, Other Classified Ads	E-commerce	No
Search Engines (Aggregate)	Search Engines, Indices & Information, Services	Search Engines	No
Social Networking	Social Networking (Private), Social Networking (Business), Dating, E-Mail, SMS, E- Cards, Messenger & Chat Other Networking & Communication	Communication	No
Games (Aggregate)	Games, General Gaming Site, Casual Games, Core Games, Other Games	Games	No

IVW changed its category classification in May 2014. *Table 3* illustrates how we matched categories before and after the change, and how we aggregated page impressions into 8 main news outlet content categories. We use page impressions by news outlet content category (as defined in the fourth column) to determine news outlet focus and scope.

Table 4 Summary Statistics

Variable	Obs	Mean	Median	SD	Min	Max
VGM	4570	0.4	0.0	0.5	0.0	1.0
Post	4570	0.5	0.0	0.5	0.0	1.0
log(Visits)	4560	13.1	13.1	1.4	4.7	16.5
log(avgVisits)	4570	13.0	13.0	1.4	10.3	16.3
Herfindahl	4570	0.5	0.5	0.1	0.1	0.8
log(Count)	4570	2.0	2.1	0.2	1.4	2.1

Table 5 Main Results

	(1) log(Visits)	(2) log(Visits)	(3) log(Visits)	(4) log(Visits)	(5) log(Visits)
Post * VGM	-0.00112	1.199***	0.490***	1.241**	1.137**
	(0.982)	(0.009)	(0.009)	(0.016)	(0.020)
Post * log(avgVisits)		-0.00516			-0.00398
		(0.856)			(0.886)
Post * VGM * log(avgVisits)		-0.0877**			-0.0557*
		(0.011)			(0.090)
Post * Herfindahl			0.273		0.271
			(0.415)		(0.422)
Post * VGM * Herfindahl			-1.100***		-0.835**
			(0.004)		(0.028)
Post * log(Count)				0.334	
				(0.130)	
Post * VGM * log(Count)				-0.630**	
				(0.016)	
News outlet FE	X	X	X	X	X
Time FE	X	X	X	X	X
Constant	13.10***	13.13***	13.04***	12.78***	13.06***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	4560	4560	4560	4560	4560
R^2	0.974	0.975	0.975	0.975	0.976

p-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the news outlet level (140 clusters). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix

A.1 Validity of the Parallel Trends Assumption

Our empirical strategy hinges on the validity of the parallel trends assumption. Put differently, we assume that the web traffic of *VGM* members would have developed parallel to non-members if they had remained on the news aggregator platforms. This section presents four arguments that support the assumption of parallel trends.

First, we can rule out that the legal dispute and the subsequent removal of news outlets from news aggregation platforms were anticipated and thereby affected their pre-trends in web traffic. The copyright bill underwent major changes during the week before it was passed; in particular, the controversial exemption of "short excerpts of text" was enacted last-minute. Thus, anticipation effects before March 2013 are extremely unlikely. To rule out anticipation effects between March 2013 and August 2014, we discard this time period from the analysis.

Second, being removed from news aggregators was not an active choice by *VGM* members. In fact, the objective of *VGM* was to collect royalty fees from news aggregators, which is impossible by definition if the content of their members is removed from news aggregators. While news outlets commissioned *VGM* with the enforcement of their claims around November 2013 (Hirche 2013), this was long before their removal from the news aggregators took place.

Third, we conduct a series of placebo regressions to support the plausibility of the parallel trends assumption. To this end, we augment *Equation (6)* to

$$\log(Visits)_{it} = \alpha_1(VGM_i * FakePost_t) + \alpha_2(S_i * FakePost_t) + \alpha_3(VGM_i * S_i * FakePost_t) + \varphi_i + \lambda_t + \varepsilon_{it} \mid Post_t = 0 \tag{7}$$

where S_i corresponds to complementor scale, focus, or scope, $FakePost_t$ is equal to one if $t \geq 3$ in the first placebo regression, $t \geq 4$ in the second placebo regression, and so on – in sum, we conduct 15 placebo regressions for each of the three specifications. If the web traffic of *VGM* members and nonmembers was on parallel trends before March 2013, the estimates for α_1 and α_3 should be zero and statistically insignificant. In three times fifteen placebo regressions, we find that α_1 and α_3 are weakly statistically significant at the 10%-level in just one instance, hence supporting the validity of the parallel trends assumption.

Fourth, we support the parallel trends assumption with a series of event studies. To this end, we augment *Equation (6)* to

$$\log(Visits)_{it} = \sum_{t=1}^{17} \gamma_{1,t} \left(VGM_i * Pre_t\right) + \sum_{t=19}^{36} \gamma_{1,t} \left(VGM_i * Post_t\right) + \gamma_2 (S_i * Pre_t) + \gamma_3 (S_i * Post_t) + \sum_{t=1}^{17} \gamma_{4,t} (VGM_i * S_i * Pre_t) + \sum_{t=19}^{36} \gamma_{4,t} (VGM_i * S_i * Post_t) + \varphi_i + \lambda_t + \varepsilon_{it}. \quad (8)$$

The idea is as follows. In *Equation (8)*, we have replaced the first and third occurrence of $Post_t$ in *Equation (6)* with a series of monthly dummies, using the month just before the copyright bill was passed (February 2013, *t=18*) as a baseline. This specification allows us to interpret the coefficients of the interaction terms as the effect of VGM_i and ($VGM_i * S_i$) on the web traffic of complementor *i* relative to a baseline month just before the copyright bill was passed. If the web traffic of *VGM* members was on parallel trends before March 2013 and only diverged afterwards, all estimates $\gamma_{\cdot,t}, t \leq 17$ must be close to zero and not statistically significant, while the estimates $\gamma_{\cdot,t}, t \geq 19$ should be negative and statistically significant.[23]

Figures A.1 (a) to *(c)* show our results for $\gamma_{1,t}$. The black dots connected by a solid line depict the estimates $\hat{\gamma}_{1,t}$, the grey dots connected by dashed lines depict a 95% confidence interval. In each case, the estimates $\hat{\gamma}_{1,t}, t \leq 17$ are close to zero and not statistically significant. In contrast to that, the estimates $\hat{\gamma}_{1,t}, t \geq 19$ are positive and most of them significant at the 5%-level.

Analogously, *Figures A.1 (d)* to *(f)* show our results for $\gamma_{4,t}$. Again, all estimates $\hat{\gamma}_{4,t}, t \leq 17$ are close to zero and not statistically significant. The estimates $\hat{\gamma}_{4,t}, t \geq 19$, in contrast, are negative and most of them are statistically significant at the 5%-level, hence supporting the plausibility of the parallel trends assumption.[24]

---

[23] Note: As the interaction term ($S_i * Post_t$) in *Equation (6)* is of minor importance when examining the validity of the parallel trends assumption, and because our number of independent complementor clusters is limited, we interact S_i in *Equation (7)* with just a single indicator for all $t \leq 17, Pre_t$, and one indicator for all $t \geq 19$, $Post_t$.
[24] One exception is *Figure A.1(f)*, where we consider complementor scope in terms of count. While the estimates $\hat{\gamma}_{4,t}, t \leq 17$ are indeed close to zero and statistically insignificant in most instances, the estimates $\hat{\gamma}_{4,t}, t \geq 19$ – though mostly negative as expected -- are not statistically significant at the 5%-level.

Insert *Figure A.1* about here

A.2 Robustness Checks

A.2.1. Alternative Dependent Variable

We use page impressions (as opposed to visits) as an alternative dependent variable. If a user accesses three articles of one news outlet, *IVW* would count this as three page impressions but only one visit. If a user accesses two articles, leaves the news outlet's domain for at least thirty minutes, and returns to access another article, *IVW* would count three page impressions and two visits. Again, we employ the logarithm of page impressions, as their distribution is heavily skewed. It can be seen from *Table A.1* that our results remain unchanged compared to *Table 5* if we use this alternative dependent variable. All estimates of interest carry the expected sign and are statistically significant at the 5%- or 1%-level. The only exception is column (5) where the p-value of our estimate for β_3 is slightly above the 0.1 threshold.

Insert *Table A.1* about here

A.2.2. Alternative Independent Variables

Scale

To assess the robustness of our findings, we use four alternative measures of news outlet scale.

The first one is the number of counties ($Counties_i$) that make up a news outlet's (offline) distribution area. The second one is the average circulation ($Circulation_i$) per month before March 2013, i.e., the average number of physical copies of a given news outlet that were sold per month before March 2013. Since both of these measures for news outlet scale are heavily skewed, we employ their logarithms in the empirical analysis.

Furthermore, the moderating effect of news outlet scale on the effect of being removed from news aggregators may depend on its scale *relative* to the scale of its competitors. To take a news outlet's relative scale into account, we define a competitor *k* of news outlet *i* as a news outlet whose distribution area overlaps with the distribution area of *i* by at least one county. Then, we compute the scale of the focal news outlet *i*'s competitors as $\sum_k scale_k$ and set it into relation to the scale of *i*.[25] This gives us our third alternative measure of scale:

$$RelVisits_i = \frac{scale_i}{\sum_k scale_k}. \tag{9}$$

If large-scale news outlets benefit from news aggregators because they are more well-known than their competitors, the moderating effect of *relative* news outlet scale on the effect of platform removal must be negative, too.

A news outlet's competitors may be close or distant. For instance, a news outlet that predominantly covers political news will consider a competing news outlet that is also specialized on political news as a much closer competitor than a news outlet that predominantly covers sports. To take this into account, we weight the scale of each competitor *k* of news outlet *i* by the extent to which their category compositions overlap and then compute our fourth alternative measure of news outlet scale:

$$RelWghtVisits_i = \frac{scale_i}{\sum_k overlap_k * scale_k}. \tag{10}$$

Insert *Table A.2* about here

The results we obtain with our alternative measures of scale can be seen in Columns (1) to (4) of *Table A.2*. As in the previous case, we find that all estimates for β_3 are negative and statistically significant, thus supporting Hypothesis 1. Again, we find that the total effect of being removed grows (i.e., becomes more negative) if news outlet scale increases, while it approaches zero and eventually flips its sign if news outlet scale decreases.

Focus

We use the entropy measure (Jacquemin and Berry 1979, Palepu 1985) as an alternative for the Herfindahl diversity measure that we use to measure focus in our main specification. Again, we compute the relative number of page impressions by news outlet that each of the eight main news outlet content

[25] The results remain similar when we consider the average competitor or the largest competitor instead of the sum of competitors of news outlet *i*.

categories in our data attracts and denote these fractions as $c_{i,k}$. The entropy measure can then be calculated as follows:

$$Entropy_i = \sum_{k=1}^{8} c_{i,k} \ln({}^{1}/_{c_{i,k}}) \tag{11}$$

It can be seen from Column (5) in *Table A.2* that our results are similar to the ones we obtained in Column (3) of *Table 5*, i.e. our main specification.

A.2.3. Time Trends

A potential concern might be that news outlets follow specific time trends that could drive our results. To rule out this possibility, we extended the difference-in-differences specification described in *Equations (5)* and *(6)* by adding news-outlet-specific time trends. It can be seen from *Table A.3* that our findings remain unchanged compared to the main results displayed in *Table 5*.

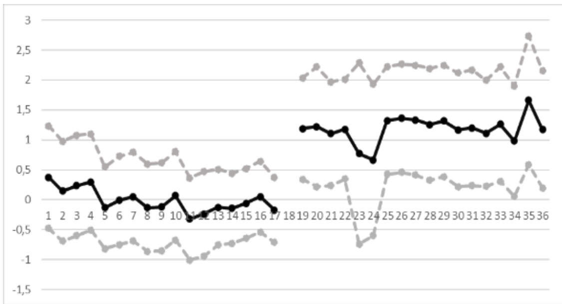
Insert *Table A.3* about here

A.2.4. Alternative Time Windows

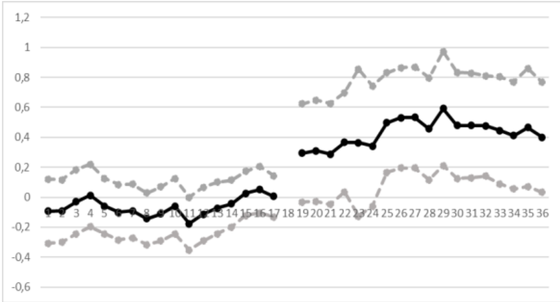
To confirm that the results in *Table 5* do not depend on a specific time window, we conduct three further robustness checks. First, we narrow the observation periods to just twelve months before March 2013 and after August 2014. Second, we extend it to 24 months before March 2013 and after August 2014. Third, we do not drop the seventeen months between March 2013 and August 2014 and consider them as post-treatment periods instead. *Tables A.4* to *A.6* demonstrate that our estimates remain qualitatively unchanged.

Insert *Tables A.4*, *A.5* and *A.6* about here

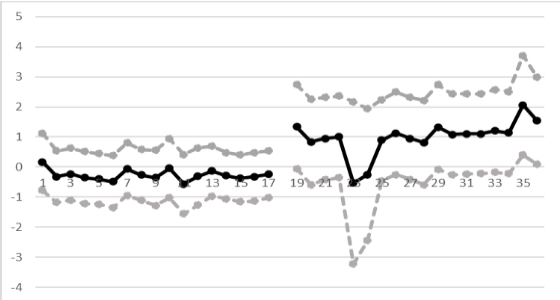
Figure A.1 Event Studies



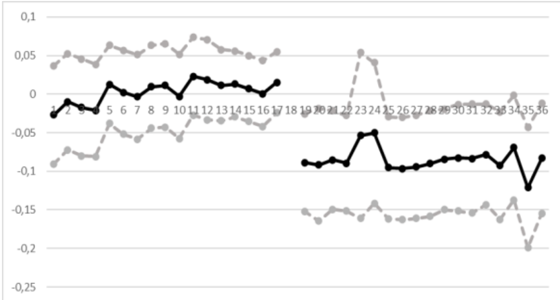
(a) Event study for $\hat{\gamma}_{1,t}$ and *Scale*



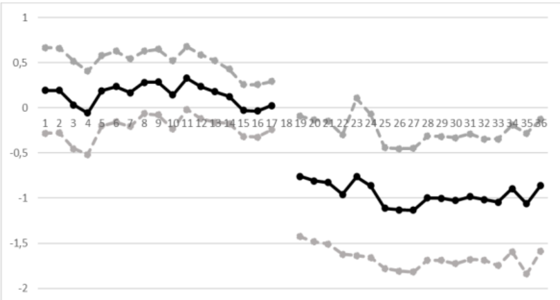
(b) Event study for $\hat{\gamma}_{1,t}$ and *Focus*



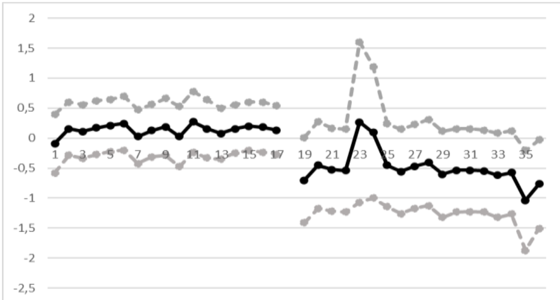
(c) Event study for $\hat{\gamma}_{1,t}$ and *Scope*



(d) Event study for $\hat{\gamma}_{4,t}$ and *Scale*



(e) Event study for $\hat{\gamma}_{4,t}$ and *Focus*



(f) Event study for $\hat{\gamma}_{4,t}$ and *Scope*

Table A.1 Alternative Dependent Variable – Page Impressions

	(1) log(PIs)	(2) log(PIs)	(3) log(PIs)	(4) log(PIs)	(5) log(PIs)
Post * VGM	0.0455	2.169***	0.516**	1.583***	1.936***
	(0.444)	(0.000)	(0.018)	(0.007)	(0.001)
Post * log(avgVisits)		0.0296			0.0297
		(0.349)			(0.351)
Post * VGM * log(avgVisits)		-0.158***			-0.118***
		(0.000)			(0.002)
Post * Herfindahl			0.0115		0.0227
			(0.976)		(0.953)
Post * VGM * Herfindahl			-1.074**		-0.696
			(0.017)		(0.119)
Post * log(Count)				0.358	
				(0.135)	
Post * VGM * log(Count)				-0.778***	
				(0.008)	
News outlet FE	X	X	X	X	X
Time FE	X	X	X	X	X
Constant	14.69***	14.50***	14.69***	14.35***	14.50***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	4560	4560	4560	4560	4560
R^2	0.975	0.976	0.976	0.975	0.977

p-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the news outlet level (140 clusters). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.2 Alternative Independent Variables

	(1) log(Visits)	(2) log(Visits)	(3) log(Visits)	(4) log(Visits)	(5) log(Visits)
Post * VGM	0.206**	1.266***	0.0506	0.0507	0.482**
	(0.028)	(0.008)	(0.338)	(0.337)	(0.015)
Post * log(Counties)	0.0169				
	(0.696)				
Post * VGM * log(Counties)	-0.126**				
	(0.015)				
Post * log(Circulation)		0.0469			
		(0.159)			
Post * VGM * log(Circulation)		-0.114***			
		(0.008)			
Post * RelVisits			0.101***		
			(0.003)		
Post * VGM * RelVisits			-0.212***		
			(0.001)		
Post * RelWghtVisits				0.101***	
				(0.003)	
Post * VGM * RelWghtVisits				-0.212***	
				(0.001)	
Post * Entropy					0.146
					(0.433)
Post * VGM * Entropy					-0.550***
					(0.010)
News outlet FE	X	X	X	X	X
Time FE	X	X	X	X	X
Constant	13.09***	12.86***	13.09***	13.09***	13.04***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	4472	4560	4436	4436	4560
R^2	0.975	0.975	0.975	0.975	0.975

p-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the news outlet level (140 clusters). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3 Main Results – With News Outlet Time Trends

	(1) log(Visits)	(2) log(Visits)	(3) log(Visits)	(4) log(Visits)	(5) log(Visits)
Post * VGM	-0.00120	1.192***	0.488***	1.223**	1.128**
	(0.981)	(0.009)	(0.009)	(0.017)	(0.021)
Post * log(avgVisits)		-0.00673			-0.00536
		(0.813)			(0.848)
Post * VGM * log(avgVisits)		-0.0872**			-0.0553*
		(0.012)			(0.092)
Post * Herfindahl			0.266		0.262
			(0.427)		(0.438)
Post * VGM * Herfindahl			-1.094***		-0.824**
			(0.004)		(0.030)
Post * log(Count)				0.331	
				(0.133)	
Post * VGM * log(Count)				-0.621**	
				(0.017)	
News outlet FE	X	X	X	X	X
Time FE	X	X	X	X	X
News Outlet Time Trend	X	X	X	X	X
Constant	13.25***	13.36***	13.19***	12.87***	13.26***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	4560	4560	4560	4560	4560
R^2	0.974	0.975	0.975	0.975	0.976

p-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the news outlet level (140 clusters). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4 Narrower Time Window

	(1) log(Visits)	(2) log(Visits)	(3) log(Visits)	(4) log(Visits)	(5) log(Visits)
Post * VGM	-0.000996	1.188***	0.469***	1.004*	1.128**
	(0.984)	(0.008)	(0.008)	(0.055)	(0.016)
Post * log(avgVisits)		-0.0121			-0.0111
		(0.655)			(0.677)
Post * VGM * log(avgVisits)		-0.0866***			-0.0574*
		(0.009)			(0.072)
Post * Herfindahl			0.220		0.216
			(0.486)		(0.500)
Post * VGM * Herfindahl			-1.054***		-0.754**
			(0.004)		(0.036)
Post * log(Count)				0.309	
				(0.123)	
Post * VGM * log(Count)				-0.511*	
				(0.053)	
News outlet FE	X	X	X	X	X
Time FE	X	X	X	X	X
Constant	13.10***	13.17***	13.05***	12.81***	13.12***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	3051	3051	3051	3051	3051
R^2	0.972	0.973	0.973	0.972	0.973

p-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the news outlet level (140 clusters). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5 Broader Time Window

	(1) log(Visits)	(2) log(Visits)	(3) log(Visits)	(4) log(Visits)	(5) log(Visits)
Post * VGM	0.137	0.620	0.587*	2.148**	1.854**
	(0.113)	(0.321)	(0.087)	(0.010)	(0.047)
Post * log(avgVisits)		-0.0580			0.0119
		(0.121)			(0.845)
Post * VGM * log(avgVisits)		-0.0349			-0.113
		(0.474)			(0.124)
Post * Herfindahl			0.425		0.435
			(0.459)		(0.444)
Post * VGM * Herfindahl			-0.975		-0.546
			(0.205)		(0.475)
Post * log(Count)				0.170	
				(0.569)	
Post * VGM * log(Count)				-1.003**	
				(0.018)	
News outlet FE	X	X	X	X	X
Time FE	X	X	X	X	X
Constant	12.79***	12.99***	12.60***	12.57***	12.56***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	20598	20333	16181	16181	16181
R^2	0.942	0.944	0.922	0.922	0.922

p-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the news outlet level (140 clusters). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6 Without Dropping 17 Months in Between

	(1) log(Visits)	(2) log(Visits)	(3) log(Visits)	(4) log(Visits)	(5) log(Visits)
Post * VGM	-0.00532	0.728**	0.202	0.970***	0.726**
	(0.874)	(0.026)	(0.145)	(0.004)	(0.045)
Post * log(avgVisits)		0.00222			0.00295
		(0.900)			(0.867)
Post * VGM * log(avgVisits)		-0.0545**			-0.0445*
		(0.028)			(0.080)
Post * Herfindahl			0.103		0.105
			(0.660)		(0.657)
Post * VGM * Herfindahl			-0.469*		-0.295
			(0.095)		(0.291)
Post * log(Count)				0.263*	
				(0.070)	
Post * VGM * log(Count)				-0.495***	
				(0.004)	
News outlet FE	13.04***	13.03***	13.02***	12.80***	13.00***
Time FE	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	13.10***	13.13***	13.04***	12.78***	13.06***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	4575	4575	4575	4575	4575
R^2	0.983	0.984	0.984	0.984	0.984

p-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the news outlet level (140 clusters). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$