Received: 9 November 2022 | Revised: 22 February 2024 | Accepted: 27 February 2024

DOI: 10.1002/smj.3600

RESEARCH ARTICLE

SMS | Strategic Management Journal

WILEY

Competing for attention on digital platforms: The case of news outlets

Tim Meyer[1] | Anna Kerkhof[2] | Carmelo Cennamo[3] | Tobias Kretschmer[4]

[1]Institute of Management & Strategy and Center for Financial Services Innovation, University of St. Gallen, St. Gallen, Switzerland

[2]Center for Economic Studies, ifo Institute for Economic Research and LMU Munich, Munich, Germany

[3]Department of Strategy and Innovation, Copenhagen Business School, Frederiksberg, Denmark

[4]Institute for Strategy, Technology and Organization, LMU Munich School of Management and CEPR, Munich, Germany

Correspondence
Tobias Kretschmer, Institute for Strategy, Technology and Organization, LMU Munich School of Management and CEPR, Munich, Germany.
Email: t.kretschmer@lmu.de

Abstract

Research Summary: Platforms are often assumed to benefit firms, especially smaller ones, by facilitating access to a broader consumer base and increasing visibility. However, this logic relies on platforms" ability to match consumer preferences to complement characteristics. In addition to this matching mechanism, we posit that platforms also broker consumer attention towards complements, which then compete for this attention. We propose that this attention mechanism is particularly prominent in settings where complement characteristics cannot be observed ex-ante, and argue that complementors' with larger scale and broader scope are better positioned to capture attention than smaller and less broad ones. We formalize and test this intuition in the context of news aggregators, highlighting the significance of complementors' ability to draw attention in evaluating their benefits from platform participation.

Managerial Summary: Small firms are often assumed to benefit most from joining a platform to expand their market reach and visibility. However, this will only be the case if the main function of platforms is to match consumer preferences to product characteristics. We argue that platforms also direct attention towards some products at the expense of others on the platform. This "attention mechanism" is particularly important whenever product characteristics cannot be observed prior to consumption, and we propose that in such settings, larger scale and broader scope of products drive attention towards specific firms on the platform. We test these predictions in the context of online news aggregators, which feature news content by newspapers of different sizes and with different range of articles. We find that indeed large and generalist newspapers benefit most from being on a news aggregators, while small and focused newspapers perform better when they are not featured on the news aggregator at all.

KEYWORDS

competition for attention, complementor heterogeneity, consumer attention, digital platforms, news aggregators, news content

1 | INTRODUCTION

Digital platforms connect multiple independent parties and have been described as novel organizational forms distinct from markets and hierarchical organizations (Gulati et al., 2012; Jacobides et al., 2024; Kretschmer et al., 2022). Participating in platforms can offer notable advantages to firms, particularly smaller ones, by providing them with access to a wider set of potential consumers than they could reach on their own (Brynjolfsson et al., 2003, 2011; Kumar et al., 2014). This is due to platforms facilitating efficient product search and discovery, and helping consumers find their best match based on the products' characteristics (Bakos, 1997; Cennamo, 2021; Tajedin et al., 2019). Digital platforms thus provide a *matching mechanism* that leads to better matches between consumer preferences and product characteristics, and ultimately more transactions.

However, many products offered on digital platforms (i.e., complements[1]) are experience goods whose characteristics cannot be assessed prior to consumption. Thus, matching consumer preferences to these concealed complement characteristics is challenging. In these contexts, digital platforms coordinate interactions in a different way. They serve as *attention brokers* (Boik et al., 2016; Evans, 2019; Prat & Valletti, 2022) that forge connections between consumers and complements by directing consumer attention towards possible consumption options. By aggregating a diverse set of complements, platforms draw more consumer attention than individual firms could do independently, creating potential value for complementors (i.e., firms providing complements on-platform) if they can capture a share of this aggregated consumer attention. While the characteristics of complements (and their potential match to consumers' preferences)

[1]Products listed on platforms are typically referred to as "complements," while the firm providing such complements is typically referred to as a "complementor." In this article, we use the respective terms interchangeably, as our focus is on firms switching between on-platform and off-platform settings.

can be hard, or costly, to assess ex-ante, complements still vary in their ability to draw consumer attention. The attention of consumers is often captured by other cues (such as whether the complementor is well-known), which consumers use to infer the prospective value of a focal complement. We call this the *attention mechanism*. While both *matching* and *attention* mechanism coexist on most platforms, their relative importance depends on platform and complement characteristics. For instance, e-commerce platforms such as *Amazon Marketplace* organize mainly around matching, while social media platforms or news aggregators (our empirical setting) are primarily driven by the attention mechanism.

Most prior work has focused on the matching mechanism, showing how platforms expand the potential market for all complementors and particularly benefit complements in the "long tail" of smaller and more specialized products (Brynjolfsson et al., 2011; Kumar et al., 2014). However, for experience goods, this matching mechanism based on complement characteristics is muted, and who will draw most consumer attention matters more. In other words, while complements may be perceived as homogenous in their *characteristics*, they are not homogenous in the *attention* they draw. The attention received can depend on *complement*-level factors (Elberse & Oberholzer-Gee, 2006; Tan et al., 2017)[2] but also on *complementor*-level factors, particularly if the latter are easier to observe than the former. Specifically, heterogeneity across complementors in size (*scale*) and whether their complements are specialized or span multiple categories (*scope*) can drive the attention a complement draws and create heterogeneity among complements whose characteristics are otherwise hard to discern. Yet, how complementor scale and scope affect their ability to draw attention and ultimately transactions remains a largely overlooked question.

We focus on *complementor*-level drivers of the attention mechanism and study the success of complementors differing in scale and scope on a platform for experience goods. In this context, characteristics and match quality are not discernable ex-ante, which largely mutes the matching mechanism. Thus, our key question is: "*How and to what extent are the benefits of platform membership moderated by complementor characteristics, specifically their scale and scope?*"

We argue that a platform's attention mechanism has a dual impact, characterized by both "incoming attention spillover" and "outgoing attention spillover" effects for firms on the platform: A share of a focal firm's (i.e., complementor's) consumer base will be attracted to other complementors on the platform, while simultaneously, the focal complementor will attract consumers of other complementors. The net effect depends on the complementor's relative capacity to attract consumer attention, which, we argue, is affected by its scale and scope. We think of scale and scope as broad concepts that reflect several aspects that might draw consumer attention, such as brand awareness or reputation. Complementors with larger scale benefit more from the positive "incoming attention spillover" effect as they are more likely to attract attention. Conversely, smaller complementors may face an "outgoing attention spillover" effect that outweighs the potential benefits of the incoming effect. Further, complementors with broader scope enjoy attention spillovers across multiple product categories and do not disproportionally lose attention in any specific category, which increases the overall number of consumers they attract.

We formalize and empirically test our reasoning in the context of local news outlets (complementors) listed on online news aggregators (platforms). News aggregators such as *Google News*, *Apple News*, or *Yahoo! News* display headlines and small excerpts of news articles

[2]For instance, a movie that features a famous actor may draw more attention than its competitors.

(complements) produced by online news outlets. While local news outlets operate in a limited geographic region, their content is also of interest outside this region. Therefore, being listed on news aggregators could be beneficial for them. We exploit a legal dispute following a policy change in Germany that led to a group of German news outlets being removed from several news aggregators. We compare local news outlets that were removed to those that were not before and after the legal dispute. Using web traffic data on 140 outlets, we find that news outlets with larger scale and broader scope suffer more from being delisted from news aggregators than smaller and less broad ones. In other words, news outlets with larger scale and broader scope benefit more from the attention mechanism on platforms and thus lose more when their content is moved off platform. Specifically, a large news outlet (75th size percentile) loses about 56.300 (or 4.08%) monthly visits post-removal, while a small news outlet (25th percentile) attracts around 21.240 (or 16.96%) more monthly visits. Similarly, a broad news outlet (75th scope percentile) loses about 250.900 (or 12.4%) monthly visits post-removal, while a specialized news outlet (25th percentile) gains around 108.000 (or 9.9%) monthly visits.

Conceptually, we contribute to platform research by distinguishing between the matching and attention mechanisms, and by highlighting the role of attention in the success of heterogenous complementors. Our formalization provides a workhorse for future work on platforms in which both mechanisms are present. Empirically, we show how competing for attention can have ambiguous effects for complementors *on* the platform and expose complementors with smaller scale and narrower scope to its negative effects. The starting point of previous work on news aggregators has often been the potential substitution effect of news aggregators on news outlets, weighted against the potential market expansion that news aggregators can bring about, and thus competition *for* the market for attention between aggregator and outlets (Athey et al., 2021; Calzada & Gil, 2020; Dellarocas et al., 2016; Peitz & Reisinger, 2014). Conversely, our empirical setting holds other effects of news aggregators like market expansion largely constant and lets us focus on competition between complementors *in* the platform market, that is, outlets competing for attention on the same platform (Evans, 2013). Moreover, by focusing on *local* outlets, we provide insights on a group of outlets in particular danger of disappearing in the process of digitization. Somewhat contrary to findings on the "long tail" effect of platform matching mechanisms, our study reveals that smaller complementors face challenges from the *attention mechanism*, as larger competitors attract more consumer attention than smaller ones, and thus attract consumers away from those complementors. This effect is likely to extend to other platforms, albeit mitigated by the *matching mechanism* which lets long tail complementors expand their market reach through enhanced match quality.[3]

2 | RELATED LITERATURE

2.1 | Costs and benefits of platform membership

Platforms "mediate transactions between two or more sides, such as […] complementors and users" (McIntyre & Srinivasan, 2017, p. 143), creating indirect network effects: The more

[3]Most platforms have both a matching element and an attention element to them. In our setting, the experience good nature of news content (I do not know if I like an article before I read it) means that prospective readers do not observe the article's characteristics, and the defining features of heterogeneity relate to newspapers' ability to attract attention on a news aggregator platform. Hence, our setting is close to the "attention-based platform" end of a continuum.

complementors are on the platform, the more attractive the platform becomes to users[4] and vice versa (Parker & van Alstyne, 2005; Rochet & Tirole, 2006). Platforms share some characteristics with markets, such as the legal independence of economic actors, and with hierarchical organizations, since some organizational mechanisms are actively designed by platform owners (Gulati et al., 2012; Kretschmer et al., 2022; Spulber, 2019). The coordination and aggregation of independent actors by the platform allow for better value creation and capture than complementors can achieve independently (Jacobides et al., 2018, 2024), which is a key incentive for complementors to join the platform (Kretschmer et al., 2022), particularly for smaller complementors which have access to a limited market and face challenges in attracting potential consumers.

There are two mechanisms that attract complementors to platforms: First, if product characteristics can be assessed prior to consumption, platforms expand the market for potential consumers by helping consumers search and find complements that match their specific preferences–the *matching mechanism*. Second, by aggregating a variety of complements, platforms can draw more consumer attention than individual complementors could do independently, and broker this attention toward possible consumption options. Complementors, therefore, join a platform to benefit from the attention bundled on the platform—the *attention mechanism*. While these mechanisms are not mutually exclusive, they work differently.

The matching role of platforms is particularly important if complements and consumer preferences are heterogeneous (Panico & Cennamo, 2022; Rietveld & Eggers, 2018; Sun et al., 2016). Here, platforms facilitate matches between consumers and complements (Cennamo, 2021; Tajedin et al., 2019) by reducing search costs compared with conventional markets (Bakos, 1997). Potential consumers can choose among a wider set of complements than they could off the platform, leading to better matches. This logic implies that digital platforms especially benefit complements at the lower end of the sales distribution (the so-called long tail) (Anderson, 2004) since platforms make discovering these complements easier and, thus, help increase demand for these complementors. Indeed, consumers are more likely to buy long-tail products when moving from physical to online channels (Zentner et al., 2013). Recommender systems facilitate discovery of long-tail products in these settings (Brynjolfsson et al., 2011) and the sales distribution is less skewed as consumers receive more information about products, improving match quality (Kumar et al., 2014; Tucker & Zhang, 2011).

However, this logic only holds if consumers can observe complements' characteristics sufficiently well before buying or consuming them (i.e., before the match between consumers and complements happens), so that they can assess the potential match with their personal preferences. This assumption is appropriate for platforms such as *Amazon Marketplace*, in which potential consumers can easily access ample information on complement characteristics (from product descriptions or reviews) prior to a transaction and often devote significant effort to assessing whether these characteristics match their preferences. Moreover, consumers actively search for complements that match their specific needs.

Regarding the attention mechanism, on platforms such as social networks, search engines, and news aggregators, consumers do not necessarily search for a particular match. Instead, they choose from a wide set of informational content (complements[5]) whose exact characteristics and match to personal preferences are hard to assess prior to consumption, either because these complements are experience goods (Nelson, 1970) or because information on complement

[4]We use "users" and "consumers" interchangeably in this article.

[5]On platforms such as social networks, search engines and news aggregators, complements are referred to as "content."

characteristics are difficult or costly to gather, for example, if there are too many options (and thus too much information on characteristics) to evaluate and/or consumers do not have sufficient resources (e.g., time) to gather and process such information (i.e., information overload).

Content on these platforms can often be consumed for free, so that decision-making may be relatively fast paced, as gathering information on the content is difficult (i.e., costly) ex-ante, while the cost of consuming content is low. Consumers may be attracted to novel content that others pay attention to, not necessarily limiting their search and selection criteria to content they are interested in ex-ante. Consumer choice here is mainly driven by factors that attract consumers' (limited) attention (Boik et al., 2016; Evans, 2019) and not by the (likely heterogeneous, yet difficult to observe) characteristics of the content. Cues that attract consumer attention can be present at the complement level (e.g., a movie featuring a famous actor may attract more attention) or at the complementor level (e.g., complementor reputation that spills over to individual complements). Yet, in these settings, complements will vary in the attention they draw mainly because of complementor characteristics, which are often easier to observe by consumers than individual complement characteristics. We thus focus on the complementor-level factors and study how *complementors* compete for the attention of consumers, with platforms acting as "attention brokers" between the two (Prat & Valletti, 2022).

We posit two opposite—incoming and outgoing—spillover effects on such platforms. First, by attracting and aggregating attention around focal content, platforms can help consumers discover content they would miss otherwise. Complementors competing for attention with others may thus benefit from "*incoming attention spillover*" effects. Second, information processing requires cognitive resources (Kahneman, 1973), making attention a scarce and rivalrous good (Calvano & Polo, 2021; Evans, 2013; Lanham, 2006). Thus, complementors competing for consumer attention on a platform may suffer from "*outgoing attention spillover*" effects. How competition for attention affects a specific complementor depends on how these opposing effects ultimately play out.

The ability of complementors to attract consumers thus depends on how much attention they can draw. Complementors that are larger in scale (i.e., offer more content) and broader in scope (i.e., offer less specialized content, which spans multiple categories) are more likely to draw attention than smaller and more narrow complementors. Scale and scope reflect several factors observable to consumers that draw attention. For instance, complementors larger in scale and broader in scope are typically better known to potential consumers due to previous exposure through consumption or advertising (Alba & Hutchinson, 1987), either within a given (content) category (amplified by *scale*) or across multiple (content) categories (amplified by *scope*). These complementors have likely built their scale and scope by "delivering quality over time" (George et al., 2016, p. 1) within and across categories, which helped them become more reputable. Potential consumers might use this reputation as a cue when allocating their attention. Especially in settings where information on specific complement characteristics is lacking, factors like reputation can be important drivers of consumer choice (Cho & Zhou, 2021; Shapiro, 1983; Washington & Zajac, 2005; Weigelt & Camerer, 1988). Content by complementors that are larger in scale and broader in scope hence becomes the "default option" (Macdonald & Sharp, 2000), while other options require additional cognitive effort. This effect might be even stronger if more complements are available, as consumers turn to well-known information sources on crowded platforms (Piezunka & Dahlander, 2015).

Hence, while novel content becomes easier to discover on platforms (compared with an off-platform setting), complementors with larger scale and broader scope capture the bulk of consumer attention, resembling "hit" or "superstar" products (Elberse, 2008; Kumar et al., 2014).

Demand remains highly concentrated even if content variety increases (Tan et al., 2017) and consumers transact more through platforms (Elberse & Oberholzer-Gee, 2006). Niche content ("long tail" products) is likely to be consumed mostly by consumers who seek variety after extensively consuming mainstream content or by a small number of connoisseurs with specific tastes (Elberse, 2008). Thus, it is not obvious if digital platforms featuring experience goods do indeed promote smaller complementors or if the dynamics lead to winner-take-all outcomes favoring large, well-known complementors.

The attention mechanism may be further reinforced by platform algorithms used to select and display complements on the platform. Prior work has highlighted that factors which draw consumer attention, such as brand awareness or reputation, can affect consumer choice on platforms in two ways (i.e., directly and indirectly) by increasing the probability that consumers choose a given complement and by leading to better placement in the list of search results (Baye et al., 2016). Specifically, brand awareness or reputation likely affect the so-called "static ranking" of complementors, which reflects a complementor's "experience, expertise, authoritativeness and trustworthiness" (Calzada et al., 2023, p. 9) and, in turn, determines its placement in the list of search results. However, algorithms merely reinforce the attention mechanism but do not generate it in the first place, as they typically generate rankings based on previous performance (Baye et al., 2016). Thus, outlets need to attract attention initially, for instance due to larger scale and broader scope, to be picked up by the algorithm and placed higher in the search rankings.

Prior work on competing for attention has focused mostly on competition *between* different platforms (Boik et al., 2016; Evans, 2013; Peitz & Reisinger, 2014) or on the role of advertising in the business model of these platforms (Evans, 2019; Prat & Valletti, 2022). Work on competition for attention *within* platforms has focused on incentives for complementors to contribute (Loh & Kretschmer, 2023; Rui & Whinston, 2012) and how contribution behavior relates to competition among complementors, for instance on social media (Rossi & Rubera, 2021) or intra-organizational knowledge platforms (Hansen & Haas, 2001). We take a demand-side perspective to study how differences in scale and scope affect the demand for complementors. Previous research on "long-tail" and "superstar" effects has often compared digital to non-digital settings (Brynjolfsson et al., 2011; Elberse & Oberholzer-Gee, 2006) and focused on the product level (Brynjolfsson et al., 2011; Elberse & Oberholzer-Gee, 2006) to study market category expansion. However, this research stream does not consider the role of complementor-level characteristics in content success, and how such characteristics drive competitive dynamics at the platform level. Whenever attention plays a major role in consumer choice, the capacity of complements to attract consumers may largely be driven by the extent to which its *provider* (the complementor) draws attention, which in turn depends on its role in the broader market space, and thus its scale and scope. This is the focus of our study.

2.2 | News aggregators

News aggregators such as *Google News*, which aggregate news articles by online news outlets and make short excerpts (so-called snippets) of these news articles available to potential readers, are a textbook example of platforms with a pronounced attention mechanism. Prior work has often focused on potential substitution between news aggregator and online news outlets, that is, whether the headlines and news articles' excerpts on news aggregators give sufficient information for consumers and ultimately substitute for reading the full article on the news outlet's website (Athey et al., 2021; Dellarocas et al., 2016). We refer to this as competition *for* the market for attention between the news aggregator and news outlets. Indeed, Dellarocas et al. (2016) show that snippets can substitute for full articles in some cases. However, news aggregators can also expand the market and *increase* the readership of online outlets, as aggregators help consumers discover new content they may otherwise not be aware of. Previous studies have shown that the net effect of these two countervailing effects is positive, suggesting that aggregators have a positive aggregate effect on outlet readership (Athey et al., 2021; Calzada & Gil, 2020).

Moreover, news aggregators can redistribute visits from one type of outlet to another, benefiting some outlets more than others. We refer to this as competition *in* the market, that is, competition among news outlets for the same readers on news aggregators. Clearly, this effect often occurs in parallel with the previous one, which is why some studies also study competition in the market even if their starting point may be competition for the market (Athey et al., 2021; Calzada & Gil, 2020). These studies show that some types of outlets do indeed benefit more from news aggregators, for instance horizontally or vertically differentiated ones (Chiou & Tucker, 2017), publishers whose content is hard to find (Athey et al., 2021) or lower-performing websites and local news outlets (Calzada & Gil, 2020), in line with the idea that news aggregators facilitate discovery of unknown content. Conversely, research also found that larger outlets such as *Axel Springer* members gained significantly from being on news aggregators (Calzada & Gil, 2020), and aggregators may not always steer consumers towards new content (George & Hogendorn, 2020).

We complement prior work through a nuanced empirical analysis of competition *in* the market for attention. Our empirical setting lets us isolate competition *in* the market, as news aggregators are not completely shut down and consumers still have a sizable amount of content to discover. By using data on a set of (local) news outlets that differ in their scale and scope but are otherwise largely homogenous, and a setting where total readership remains fairly constant before and after the shock we observe, we can study the moderating role of outlet scale and scope and their effect on readership redistribution across outlets.

3 | THEORY AND HYPOTHESES

3.1 | Intuition of the formal model

We develop a simple formal model describing the behavior of heterogenous consumers and derive hypotheses on the effect of firms (=complementors) being on or off the platform. Before delving into the details of the model, we outline the intuition of the effects it captures. We want to compare the number of consumers off- and on-platform for firms of different scale and scope. Off-platform, consumers choose products from their existing suppliers without considering other sources. However, once firms are listed on the platform, consumers of the focal firm become aware of products by other firms and include them in their choice set. Moreover, the products of the focal firm now enter the choice set of consumers of other firms' products. Firms listed on the platform thus face two opposing forces: First, their existing consumers are "at risk" of choosing an alternative product now that it is readily available to them (the "outgoing spillover" effect). Second, the focal firm can attract consumers that previously chose a different product (the "incoming spillover" effect). The net effect of these two forces determines whether a firm stands to benefit from being listed on the platform or not. We posit that this outcome depends on the degree of attention the focal firm attracts relative to its competitors on the platform. Attention, in this context, is relative to the relevant alternatives that consumers consider, with our assumption being that a larger existing consumer base commands more attention. This leads to two testable hypotheses on the net effect of being on-platform.

Our first hypothesis posits that firms with larger scale (i.e., a larger consumer base off-platform), benefit more from being on-platform than smaller ones. This is because their existing consumers are less likely to change their choices (resulting in a smaller outgoing spillover). At the same time, these firms can attract consumers from other firms due to their larger consumer base (leading to a larger incoming spillover). The second hypothesis states that firms with broader scope (i.e., less specialization and a broader coverage of multiple categories) benefit more from being on-platform than firms that are narrow in scope. The intuition is that firms with broader scope lose comparatively little in categories with a more focused competitor (thus incurring a smaller outgoing spillover). Conversely, they can take advantage of that competitor's weakness in other categories and draw consumers towards their own offerings (resulting in a larger incoming spillover). We develop these two hypotheses formally in an analytical model and test them empirically. Specifically, we assess the performance effect from the unexpected removal of some newspapers (which are heterogeneous mostly in scale and scope) from German news aggregators.

3.2 | General theoretical framework

We propose a simple theoretical framework that is based on, and extends, the classic Hotelling (1929) model, which describes consumer choice between two options $i = A,B$ that represent, for instance, physical products, stores or, as in our setting, digital products (i.e., online news content). Consumers are uniformly distributed along a line of length one. The options offered to those consumers are located at the extremes of the line,[6] and each consumer consumes exactly one of these options (due to their limited time budget). A consumer's location $x \in [0,1]$ on the line represents their *individual* inclination to choose option A or B.

Which of the two options consumers will choose depends on the net attraction U_i that each option $i \in \{A,B\}$ exerts on consumer x, where U_i is determined by the baseline attraction of option i (exerted equally on all consumers on the line) and the individual transportation cost of consumer x to option i. In the classic Hotelling model, the baseline attraction of option i is determined by observable characteristics such as quality or overall performance. We refer to this as characteristics-based attraction v_i. Individual transportation cost is denoted by $t \cdot x$ and stems from overcoming consumer x's inclination (or lack thereof) to choose one option over the other. It is incurred by consumers who choose an option whose characteristics do not closely meet their preferences. The net attraction U_i of option i on consumer x is thus:

$$U_i = v_i - t \cdot x.$$

[6]Hotelling (1929) also discusses the optimal location for sellers. We take the location of sellers as given at the extremes of the line, as we focus on consumer choice rather than seller location. This reflects our empirical setting (and many others), where complementors are often monopolists in a given submarket. Consumers are distributed along this line, with consumers that have the weakest inclination to choose that complementor being located the furthest away.

This classic reasoning is based on the (implicit) assumption that the characteristics of an option can be observed by consumers prior to consumption, which will then result in characteristics-based attraction v_i as the key driver of consumer choice. However, in many settings, a product's baseline attraction can also be determined by the *attention* it draws. We account for this *attention mechanism* by augmenting the classic Hotelling model by an additional type of attraction, a_i, the *attention-based attraction* of option i. The attention-based attraction of option i can be more or less pronounced (compared with characteristics-based attraction). The parameter $\beta \in [0,1]$ measures the extent to which the baseline attraction of option i depends on characteristics-based or attention-based attraction. Net attraction U_i in our model thus becomes:

$$U_i = (1-\beta) \cdot v_i + \beta \cdot a_i - t \cdot x.$$

Note that while options A and B exert the same characteristics- and attention-based attraction on all consumers, the inclination to choose A or B (i.e., the location on the line) will still differ by consumer. This could be, for instance, because some consumers prefer the blue color of product B over the green color of product A, or because they are more susceptible to devoting their attention to content of outlet A than outlet B because its reporting better reflects their political orientation. This means that the choice of consumer x is driven by both *individual*-level factors (i.e., consumer x's inclination to choose A or B), as well as the options' characteristics-based and attention-based attraction, which are equal for *all consumers*.

Characteristics-based and attention-based attraction will not be equally important in driving consumer choice across all platforms. On platforms such as *Amazon Marketplace*, consumers can gather information on complement characteristics ex-ante at low cost due to advanced search functions and filters, and make their choice accordingly. For instance, faced with two options A and B, consumers can evaluate the characteristics of the options (e.g., by looking at their technical specifications or customer reviews), which might result in a higher characteristics-based attraction of, say, option A. Characteristics-based attraction therefore constitutes a large share of the baseline attraction on platforms like *Amazon Marketplace* (i.e., β will be relatively small). Conversely, many complements such as news articles, our empirical context, are experience goods whose characteristics are unknown prior to consumption. Consumers perceive the characteristics of options A and B as ex-ante similar and cannot base their consumption choice on them. Hence, the attraction of complements on platforms such as news aggregators will be driven mostly by attention-based attraction a_i of each option on consumers (i.e., β will be relatively large, see also Prat & Valletti, 2022). Note that consumer choice on the platforms mentioned above need not be driven exclusively by *either* v_i (i.e., characteristics-based attraction) *or* a_i (i.e., attention-based attraction). Rather, these are the *main* drivers of consumer choice on the respective platforms. On some platforms, both factors may be important (e.g., $\beta = 0.5$), which makes our distinction a matter of degree rather than a binary categorization.

This augmented Hotelling model lets us study differences between on-platform and off-platform settings. We think of the off-platform setting as a setting where consumers are only aware of one option A (i.e., their choice set only contains this option) and can decide whether or not to choose this one. Consumers with the highest inclination to choose A (e.g., because it closely matches their preferences) are located close to A on the Hotelling line, while consumers with the lowest inclination to choose A are located at the other end of the Hotelling line. A thus exerts the lowest net attraction U on consumers with a low inclination to choose it, as their transportation cost $t{\cdot}x$ is highest.[7] Platforms make consumers aware of additional options. On-platform, consumers can choose between the existing option A and the next-best option B, that is, the additional option with the strongest attraction on them. Among consumers on the Hotelling line, the next-best option B will exert a strong net attraction U_i on consumers with a weak inclination to choose the original option A. These consumers will feel relatively more inclined to choose option B as their inclination to choose the original option A is particularly weak. On platforms like *Amazon Marketplace*, where characteristics-based attraction plays a particularly important role (small β), this will help consumers become aware of (and potentially choose) options whose characteristics better match their individual preferences (i.e., options that they are more inclined to choose[8]). This reflects the logic in prior research, which argues that platforms are particularly beneficial for consumers whose preferences have not been closely matched by off-platform options. We refer to this characteristics-based mechanism as *matching mechanism*.

Conversely, on platforms like news aggregators, characteristics-based attraction is muted (large β) because the characteristics of different options are difficult to observe ex-ante. These platforms can make consumers aware of options they are more susceptible to pay attention to. For instance, consumers who, off-platform, have only been exposed to content which does not match their political stance will be susceptible to devoting their attention to content from a (newly added) news outlet with a different political orientation. We refer to this mechanism, closely related to attention-based attraction a_i, as *attention mechanism*.

3.3 | Theoretical framework in the context of online news outlets

Based on our general framework, we describe a stylized scenario where two online (news) outlets offer two types of (news) categories, respectively. The outlets compete for the attention of a given set of readers to maximize their page views (in an effort to maximize advertising revenues; see Anderson & Jullien, 2015).[9]

We introduce the off-platform setting as benchmark and contrast it with the on-platform case. Off-platform, outlets operate as local monopolists in an online market of given size. On-platform, outlets gain access to a broader set of potential readers but must compete with the other outlet. Since news articles are experience goods, we assume that the attention-based mechanism is the major force that determines readers' demand for one outlet or the other. We keep the overall number of readers constant in both scenarios to abstract from market expansion effects and to isolate spillover effects between outlets competing for reader attention.[10]

[7]Consumers with a net attraction below zero are located outside the Hotelling line. The relevant market we consider is thus represented by the consumers located on the Hotelling line.

[8]For instance, consumers may initially only be aware of blue T-shirts, which some consumers may not particularly like. On platform, consumers may become aware of green T-shirts, which those consumers unhappy with their initial blue T-shirts will be relatively more keen to purchase now they became available.

[9]While we use our model to describe a specific setting (i.e., news aggregators), it can be applied more generally. In the Appendix A, we provide an overview of such settings along with the corresponding application of our model.

[10]Consistent with this reasoning, we do not find any empirical evidence for a market expansion effect in our context.

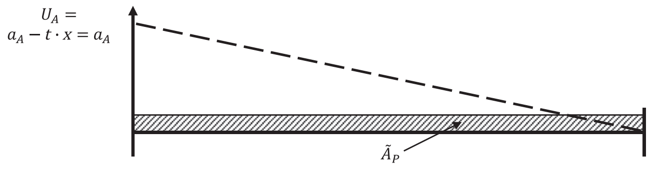


FIGURE 1 Hotelling line for submarket $\tilde{A}_P$ (off-platform). The figure shows the uniform distribution of off-platform readers of outlet A in category P (i.e., $\tilde{A}_P$, represented by the shaded area) along the Hotelling line. *Off-platform*, these readers only have one option to choose from, that is, outlet A, which exerts a certain attraction on readers depending on their location. For readers located at the extreme left of the line (where the inclination to read outlet A is highest), the net attraction can be described as $U_A = a_A - t \cdot x = a_A$ where a_A is the attention-based attraction exerted by outlet A and tx are the transportation costs, which are equal to 0 for these readers. For readers located further on the right of the line, the net attraction U_A exerted on them decreases due to transportation cost (the net attraction is represented by the dashed line).

3.3.1 | Off-platform

Consider two outlets $i = A,B$.[11] Each outlet i is active in two categories $j \in \{P,S\}$[12] and faces an off-platform readership $\tilde{i}_j \geq 0$ in each of the two categories, respectively. Each of the four outlet-category pairs $\tilde{i}_j$ can be thought of as a separate submarket. Readers face an attention budget constraint such that each reader consumes at most one article (unit demand).[13] The total mass of readers in all four submarkets is equal to one, that is, $\tilde{A}_P + \tilde{A}_S + \tilde{B}_P + \tilde{B}_S = 1$. Prices and (fixed and marginal) costs are normalized to zero.

Readers' off-platform choice sets include just one outlet i, for example, because search costs for another outlet are prohibitively high or because outlets have limited geographic reach.[14] We thus take the off-platform readership sizes as exogenously given and assume that outlets are monopolists in the respective submarkets $\tilde{i}_j$. More broadly, this means that readers take the presence and position on one specific outlet's Hotelling line as given and decide whether to consume content from that outlet. In our empirical context, this reflects the typical consumption pattern where readers first access an outlet's landing page and then choose whether to read the available content.

Each off-platform readership $\tilde{i}_j$ is uniformly distributed on a Hotelling line of length $\tilde{i}_j$ where a reader's position x on the line corresponds to their inclination to consume content by outlet i in category j. Following Section 3.2, we assume that the attention-based rather than the characteristics-based mechanism dominates in the market for news. For simplicity, we assume

[11]Note that, even in the off-platform setting, these are *online* news outlets. We do not study *offline* settings with physical newspapers.

[12]For example, politics and sports.

[13]In practice, consumers may read different outlets in sequential sessions or engage in multihoming (i.e., consume articles from different outlets in parallel). We abstract from this possibility as it is beyond the scope of our article.

[14]This assumption provides a strong contrast to the on-platform case and lets us focus on our main mechanisms. Note that, without search costs, readers might consume articles from both outlets even in the off-platform case. For local news outlets, however, it is reasonable to assume that their readership is primarily determined by geographic factors. The local outlets in our sample typically target a specific geographic region and have often done so for many years. Their median founding year is 1945, with the oldest outlet being founded in 1705 (*Hildesheimer Allgemeine Zeitung*), suggesting that the outlets have a relatively stable readership in their respective local markets.

$\beta=1$, that is, all weight lies on the attention-based mechanism. The net attraction of content j by outlet i is thus given by

$$U_{ij}=a_{ij}-t\cdot x,$$

where we assume for simplicity that $U_{ij}>0$ for all readers in submarket $\tilde{i}_j$, that is, all readers prefer consuming category j by outlet i over consuming no content at all.

Figure 1 illustrates the distribution of one of the four readerships (i.e., $\tilde{A}_P$) and the net attraction that outlet A exerts on the respective readers. The readership $\tilde{A}_P$ is uniformly distributed on a Hotelling line of length 1 and the mass of $\tilde{A}_P$[15] is represented by the shaded area. The dashed line indicates the net attraction exerted on readers located on the line: net attraction is highest for readers at the far left of the line ($U_A=a_A$) and decreases (due to transportation cost) as readers are located further right.

3.3.2 | On-platform

On-platform, outlets A and B are listed on a (news) aggregator.[16] This means that articles from the other outlet now also enter the choice set of readers and exert (attention-based) attraction on them. Hence, compared with the off-platform setting, both outlets gain access to a broader set of potential readers but must compete with the other outlet for readers' attention.[17] Moreover, readers now have immediate access to all available articles, without being restricted to content from a given outlet. When both outlets A and B are listed on the aggregator, readers need to choose among the (newly) available options—$\widehat{A}$ and $\widehat{B}$—and their choice will be determined by the net attraction U_i of the competing options, which is given by:

$$U_A=a_A-t\cdot x$$

and

$$U_B=a_B-t\cdot(1-x)$$

and reader x remains with outlet A if $U_A>U_B$. Again, a reader's position x captures their inclination to consume each of the options and we set $\beta=1$, that is, only the attention-based mechanism matters.

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

[16]On-platform, news outlets thus become "complementors" on the platform (i.e., the news aggregator) and their news articles (or, more broadly, their content) become the respective "complements."

[17]Off-platform, we assume that readers directly access an outlet's website to read the respective content. We denote visits generated in this process as "off-platform" visits (or, more broadly, readership). Conversely, on-platform, readers first access the news aggregator (where they choose the content they want to consume) and are then redirected to the respective outlet's website. Visits generated in this process are "on-platform" visits. Both scenarios focus on readers who ultimately access an outlet's *own* web page. While some readers may only consume the news excerpts (snippets) on the aggregator in the on-platform case, we consider the *net* effect on each outlet's readership. This lets us focus on competition between outlets rather than competition between outlets and news aggregators, and it accounts for the fact that outlets only care about visits on their own web page as only these visits yield advertising revenues.

We assume that a reader from off-platform submarket $\tilde{i}_j$ can now decide between a broader set of relevant alternatives (i.e., alternatives that can potentially attract at least some attention on-platform). Specifically, we assume that she will be aware of all content (i.e., in both categories P and S) of the outlet she would consume off-platform (e.g., outlet A), and that this content will draw some attention (i.e., be a relevant alternative). This may be, for instance, because she is accustomed to this outlet from her off-platform consumption and thus aware of its content. Additionally, on-platform, she will become aware of content of the other outlet (i.e., outlet B) in a given category (say, P) once she searches for (or simply sees) content in that particular category on the news aggregator. This content is thus also a relevant alternative.

Hence, the content of outlet i in category j is most likely to compete for reader attention with alternatives in the *same category* or with alternatives in the *same outlet*, as these are the closest alternatives, but it is unlikely to compete with alternatives in other categories *and* other outlets. We do not distinguish between the two categories P and S of the focal news outlet A to keep the analysis at the outlet level and to abstract from within-outlet readership flows. In our model, this means that readers from, say, off-platform submarket $\tilde{A}_P$, consider outlet A as on-platform consumption option on one extreme of the Hotelling line, and category P from the competing outlet B at the other extreme. The available options for consumer x are thus either remaining with outlet A (and read either P or S) or churning to outlet B (and read P).[18]

The attention-based attraction a_i of a specific on-platform option depends on the attention it can draw relative to all relevant alternatives. Content with a large *off*-platform readership draws more attention *on*-platform. Hence, on-platform attention-based attraction a_i depends on the size of outlet i's relevant off-platform readership (given by the numerator in Equation (1)) relative to the off-platform readership of all relevant alternatives on the platform (given by the denominator in Equation (1)). Consider again the off-platform readership $\tilde{A}_P$. The relevant on-platform attention-based attractions a_A and a_B are:

$$\begin{aligned} a_A &= \frac{\tilde{A}_P + \tilde{A}_S}{\tilde{A}_P + \tilde{A}_S + \tilde{B}_P} = \frac{\tilde{A}}{\tilde{A} + \tilde{B}_P} \\ a_B &= \frac{\tilde{B}_P}{\tilde{A}_P + \tilde{A}_S + \tilde{B}_P} = \frac{\tilde{B}_P}{\tilde{A} + \tilde{B}_P}. \end{aligned} \tag{1}$$

See Figure 2 for further illustration.

For each submarket $\tilde{i}_j$, deriving the indifferent reader $\bar{x}_{ij}$ for whom the relevant on-platform options are equally attractive yields the respective shares of remaining and churning readers. Sticking to the above example, equating U_A and U_B and setting $t=1$ yields

$$\bar{x}_{A_P} = \frac{1}{2}\left(\frac{\tilde{A} - \tilde{B}_P}{\tilde{A} + \tilde{B}_P} + 1\right),$$

where the share $\bar{x}_{A_P}$ of readers from submarket $\tilde{A}_P$ remains with outlet A, and the share $(1 - \bar{x}_{A_P})$ of readers from submarket $\tilde{A}_P$ churns to outlet B. Repeating this procedure for all off-

[18]The fact that reader attention is drawn by content of A in both categories but only by content of B in the focal category can be seen as a form of "stickiness" that ties readers to the outlet they would read off-platform. This, in turn, leads to asymmetry in the extent to which readers' off-platform choice and the newly available alternatives can draw attention.

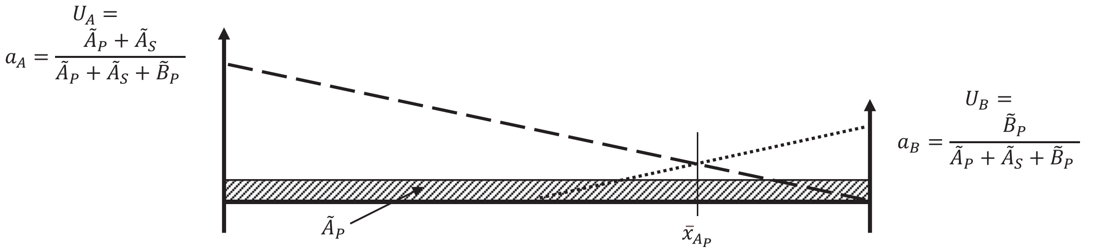


FIGURE 2 Hotelling line for submarket $\tilde{A}_P$ (on-platform). The figure shows the uniform distribution of off-platform readers of outlet A in category P (i.e., $\tilde{A}_P$, represented by the shaded area) along the Hotelling line. *On*-platform, these readers can choose between two options, that is, outlet A or B, which each exert a certain net attraction U_i on readers depending on their location. For readers located at the extreme left of the line (where the inclination to read outlet A is highest), this net attraction can be described as $U_A = a_A - t \cdot x = a_A$, where a_A is the attention-based attraction exerted by outlet A in category P and tx are the transportation costs, which are equal to 0 for these readers. Similarly, for readers located at the extreme right of the line (where the inclination to read outlet A is lowest), this attraction can be described as $U_B = a_B - t \cdot (1-x) = a_B$. On-platform, a portion of the total reader attention is drawn by each outlet A and B (i.e., $a_A \geq 0$ and $a_B \geq 0$, with $a_A + a_B = 1$). For readers that are located further on the right (left) of the line, the attraction exerted on them by outlet A (B) decreases due to transportation cost (the net attraction is represented by the dashed (dotted) line). The indifferent reader between A and B is represented by $\bar{x}_{A_P}$.

platform readerships $\tilde{i}_J$ gives the overall number of remaining and churning readers for outlets A and B in the on-platform case.

While our simple model omits many complexities underlying media choice, it is a useful workhorse to generate testable hypotheses. Specifically, we hypothesize that on-platform market shares are affected by two drivers of reader attention: *scale* and *scope*. While these dimensions may be correlated (e.g., an outlet relatively large in scale may also be broader in scope), we study their mechanism and impact separately.

3.4 | Hypotheses

We are interested in understanding the effect of being delisted from ("struck off") a news aggregator platform, specifically, how the web traffic of *delisted* outlets is affected compared with the traffic of outlets whose content remained on the aggregators. This scenario reflects our empirical context, local outlets in the German market, in which several news aggregators, following a copyright bill passed in August 2013, decided to delist content from a subset of outlets (i.e., members of the *VGM* association[19]) to avoid a possible dispute with them (see more details in Section 4.1). Accordingly, we use our model to develop two hypotheses on the effect of being *delisted* from an aggregator. However, the analogous logic (with inverse sign) applies (by definition) for outlets (complementors) joining a news aggregator (platform). Consider the baseline setting in Figure 3, where each of the four off-platform readerships $\tilde{i}_J$ is uniformly distributed on a Hotelling line of length 1, with the relevant on-platform options located at the

[19]*VGM* (*VG Media*, *Gesellschaft zur Verwertung der Urheber- und Leistungsschutzrechte von Sendeunternehmen und Presseverlegern mbH*), which renamed itself into *Corint Media* in 2021, is a German copyright collecting society of privately owned broadcasters and press publishers.

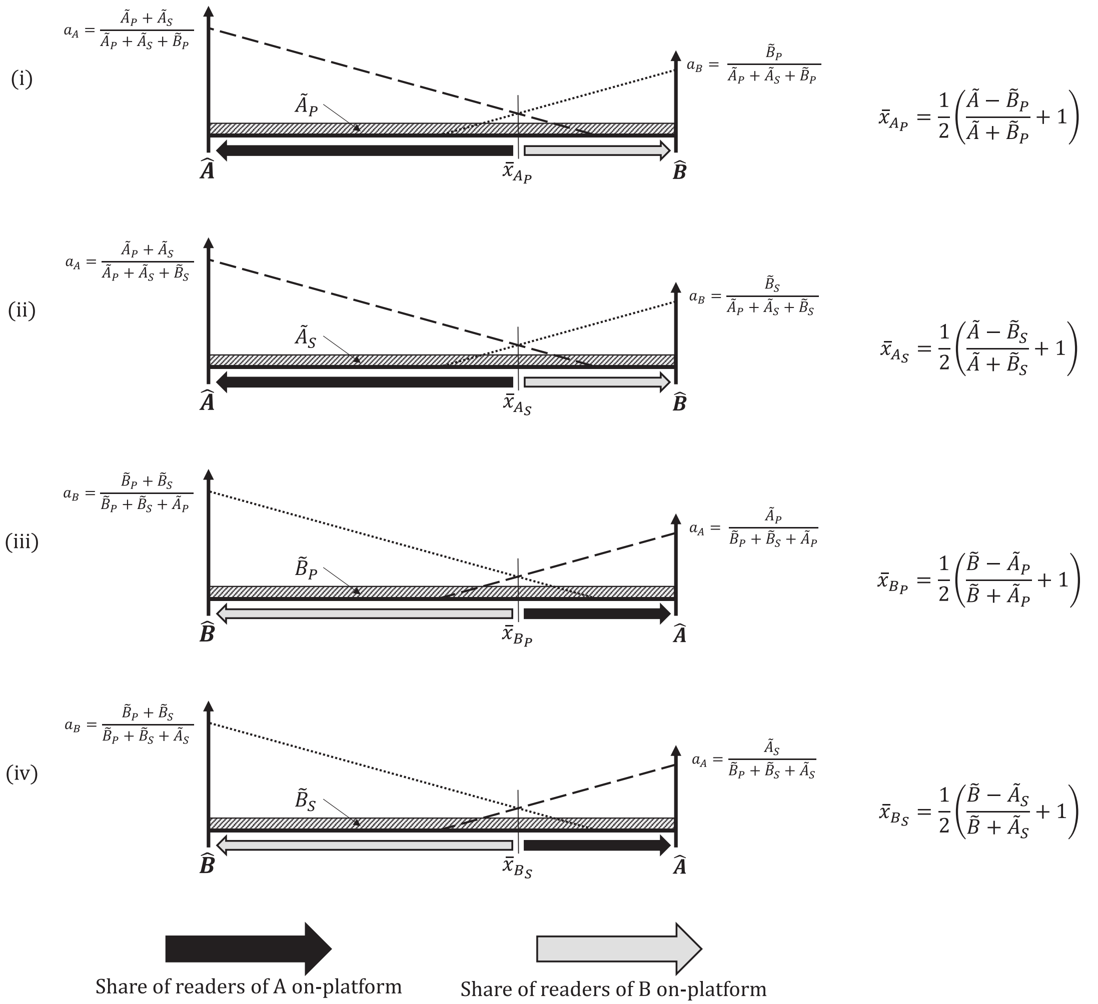


FIGURE 3 Baseline setting in the on-platform case. The figure shows the uniform distribution of off-platform readers of outlet i in category j (i.e., $\tilde{i}_j$, represented by the shaded areas) along the respective Hotelling lines. *On*-platform, these readers can choose between two options, that is, outlet A or B, which each exert a certain net attraction on readers depending on their location. For readers located at the extremes of the lines, the net attraction of the respective outlet is highest (i.e., it only consists of the attention-based attraction a_i, while transportation costs are 0) and for readers located further towards the opposite extreme, attraction decreases due to transportation cost. The indifferent reader between the two options is represented by $\bar{x}_{i_j}$. Depending on their location relative to $\bar{x}_{i_j}$, readers will either choose outlet A (dark arrow) or B (light arrow).

extremes of the line. The shaded area gives the mass of $\tilde{i}_J$, and reader $\bar{x}_{i_j}$ is indifferent between 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, that is, $\tilde{A}_P = \tilde{A}_S = \tilde{B}_P = \tilde{B}_S = \frac{1}{4}$. The number of “churning” readers then equals the number of “incoming” readers for each outlet[20] and total on-platform readership equals total off-platform readership.

[20]“Incoming” readers are those who churn away from the other outlet.

3.4.1 | Outlet scale

Consider now the case where we deviate from the baseline setting and 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 remain 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 3 helps us illustrate how this deviation from the baseline setting, and the related changes in off-platform readerships $\tilde{i}_J$, affect $\widehat{A}$ and $\widehat{B}$. First, the on-platform options by outlet A receive relatively more attention than the options by B. When $\tilde{A}_P$ and $\tilde{A}_S$ increase while $\tilde{B}_P$ and $\tilde{B}_S$ decrease, the indifferent consumers $\overline{x}_{A_P}$ (in Figure 3, (i)) and $\overline{x}_{A_S}$ (in Figure 3, (ii)) move to the right (increasing A's share of remaining readers), and x_{B_P} (in Figure 3, (iii)) and x_{B_S} (in Figure 3, (iv)) move to the left (increasing A's share of incoming respectively 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 Figure 3, (i,ii), the shaded areas above the Hotelling lines grow and in Figure 3, (iii,iv), the shaded areas above the Hotelling lines shrink). However, since A increases its on-platform share of readers in every market, the first effect (i.e., options by outlet A receiving relatively more attention than options by B) dominates the second effect (i.e., the overall effect of outlet scale).

Intuitively, the increased attention resulting from a larger off-platform readership helps outlets stand out from other options in the choice set of readers and thus increases the portion of readers (compared with the baseline case described above) they can attract in a given submarket. This is not only true for submarkets in which the larger outlet would be the only off-platform choice (in this case $\tilde{A}_P$ and $\tilde{A}_S$) but also for submarkets where the larger outlet enters the choice set of readers on-platform (in this case $\tilde{B}_P$ and $\tilde{B}_S$).

In sum, larger outlets capture more demand and attract readers from smaller competitors in the on-platform case, which implies that they suffer more from being removed from aggregators.

> **Hypothesis 1.** The larger in scale a news outlet is, the more negatively it will be affected by being removed from news aggregators.

3.4.2 | Outlet scope

Consider now the case where outlets A and B are equally large in scale ($\tilde{A}=\tilde{B}$), but outlet A is more narrow in scope ($\tilde{A}_P>\tilde{A}_S$ and $\tilde{B}_P=\tilde{B}_S$), that is, 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 off-platform readership affect on-platform readership in two ways. First, outlet A draws relatively more attention in category P and relatively less in S. Hence, the indifferent consumer x_{B_P} (Figure 3, (iii)) moves to the left (increasing A's share of incoming readers), while x_{B_S} (Figure 3, (iv)) 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} less 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. 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} (Figure 3, (i,ii)) are unaffected. Further, the relative importance of 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 (Figure 3, (i,ii)).

The intuition can be described as follows: An outlet with a narrow scope has a larger readership in one of the categories (i.e., the "strong" category it focuses on). Consequently, it draws more attention in that specific category. However, this advantage comes at the expense of a smaller readership and thus less attention in the other category (i.e., the "weak" category it *does not* focus on), keeping overall readership constant. These two effects are not symmetric in their magnitude. Compared with an outlet with broader scope, a narrower outlet loses more readers in its weak category than it can gain in its strong category. This is because more readership in the strong category does not only lead to more attention for the outlet's content in that category, but also results in an increase in the total attention generated by all relevant alternatives (including the focal one) the focal content is compared with.[21] Conversely, less readership in the weak category also results in a decrease in the total attention generated by all relevant alternatives (including the focal one) that the focal content is compared with. That is, the increase in readership (compared with the baseline setting with equal distribution across categories) in the strong category is compared with a *larger* comparison group, while the same decrease in readership (in absolute terms) in the weak category is compared with a *smaller* comparison group. Thus, the potential upside of drawing more attention in the strong category is smaller in absolute terms than the associated downside of losing attention in the weak category. An outlet with a broader scope is strategically insulated from such competitive forces as it avoids having any "weak" categories. This implies that such outlets derive greater benefits from their presence on the platform (and suffer more from being removed from the platform). Hence, we expect:

Hypothesis 2. The broader in scope a news outlet is, the more negatively it will be affected by being removed from news aggregators.

4 | DATA AND METHODS

4.1 | Empirical setting

Our empirical setting is the German newspaper industry. As in many other countries, the German newspaper industry has undergone drastic digitization in the past years. Most print media outlets do not just produce physical newspapers but make (some) of their content available online as well. This content is often collected by news aggregators, which collect articles from different outlets and provide links to the original content producers. Since outlets could initially not opt out of this procedure, they viewed it as a potential threat. For this reason, the German government introduced the so-called "ancillary copyright for press publishers" (*Leistungsschutzrecht für Presseverleger*), which lets print media companies 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 surprisingly exempted "short excerpts of text" from the

[21]Recall that the relevant alternatives the focal content is compared to are given by the denominator in Equation (1).

regulation the week before the bill was passed, arguing that it would constrain the public's basic 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 royalty fees, outlets insisted. In particular, *VGM* (*VG Media*, *Gesellschaft zur Verwertung der Urheber- und Leistungsschutzrechte von Sendeunternehmen und Presseverlegern mbH*), 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. *VGM* proposed a pricing schedule that would let licensees reuse its members' original content and threatened to file lawsuits against news aggregators that did not comply (Kuri, 2014). At the time of these events, *VGM* members included most private German television and radio broadcasters, and several press publishers with their online outlets. Prominent examples include *Axel Springer*, *Funke Mediengruppe*, and *ProSiebenSat.1*. Table A.13 gives an overview of all *VGM* members in our 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 use this unexpected removal of *VGM* news articles from several news aggregators as an exogenous shock to study how the web traffic of *VGM* outlets is affected compared with the traffic of outlets whose articles remained on the aggregators.[22] Compared with similar events in Spain, which have been used as the empirical setting in previous studies (Athey et al., 2021; Calzada & Gil, 2020), a key characteristic of our setting is that news aggregators are not completely shut down and some outlets remain on the aggregators. Hence, the market expansion effect of aggregators is muted, which lets us focus on competition in the market for attention among outlets.

We focus on *local* outlets for several reasons. First, these outlets supposedly have a particularly strong potential to increase their access to consumers when joining a news aggregator, as they find attracting readers outside their limited market difficult in the absence of news aggregators. Local outlets traditionally play a much bigger role in the German news landscape than in many other countries, both in terms of the sheer number of outlets and the extent to which readers appreciate their content (Media Landscapes, 2021; Newman, 2020). This also means that most readers already consume content from *some* local outlet. If local outlets are listed on platforms, this will likely lead mostly to a redistribution across different *local* outlets rather than to a discovery of local outlets by readers who were previously only aware of national outlets.

Most outlets in our sample have existed for several decades, with the oldest founded in 1705 (median founding year = 1945). However, the number of subscribers to local outlets is declining, raising the question if the digitization of the news industry and the increasing availability of free content contribute to this trend (Media Landscapes, 2021). All this makes the German newspaper industry a good setting to explore the effects from being on or off a platform.

[22]A similar setting has been used by Calzada and Gil (2020) to study the effect of news aggregators on outlets. Their setting differs in two main points. First, they study a 2-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 more limited sample of domains that include local, national, business and sports outlets, we cover a much larger sample of local outlets.

Second, focusing on local outlets lets us draw on a set of outlets that mainly differ in their scale and scope but are otherwise considered fairly homogenous. We can, therefore, abstract from other differences between outlets like differences in characteristics, capabilities or business models between *national* and *local* outlets, which could potentially drive our results.

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 them 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 high due to 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 include information on most German media outlets and can be accessed through the *IVW* website. Compared with other web traffic data from sources like *SimilarWeb* or *Alexa*, the *IVW* data represent the actual traffic on a given website. We complement these data with additional information gathered from *Onlineatlas der Zeitungen* (*Online Atlas of Newspapers*), a website providing information on an outlet's (offline) distribution area.

4.2.1 | Dependent variable

We use an outlet's number of visits per month ($Visits_{it}$) as our dependent variable, web traffic, where a visit is defined as an entire user session.[23] If, for instance, a user accesses three articles of one outlet, *IVW* counts one visit. If a user accesses two articles, leaves the outlet's domain for at least 30 min, and returns to access another article, *IVW* counts two visits.[24] As few outlets attract the lion's share of user attention, we use the logarithm of visits as dependent variable.

4.2.2 | Independent variables

Treated variable

We retrieve a list of all *VGM* members from the association's website *vg-media.de* to generate our treated variable (VGM_i). Our analysis includes 57 members and 83 nonmembers (see Table A.13 and the following section for details on the selection of observations).

[23]These numbers refer to visits of the outlet's *own* web page, not the outlet's content on the news aggregator. Our measure thus captures the *net* effect on the web traffic outlets attract to their website. The *IVW* data do not discriminate between direct visits and search visits, that is, we cannot observe how a user was directed to an outlet's website.

[24]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 outlet, *IVW* would count this as three page impressions.

Scale

We use the average number of visits per month *before* March 2013 ($avgVisits_i$) as our measure for outlet scale.[25] Like the dependent variable, our measure for outlet scale is highly skewed, so we use its logarithm in our analysis. Our results are robust to using alternative measures for outlet scale, such as average (offline) circulation per month before March 2013, the number of counties that make up an outlet's (offline) distribution area or the relative scale compared with its local competitors (see Appendix A.2.2).

Scope

To measure outlet scope, that is, the extent to which outlets are present in different content categories (e.g., news or sports), we use their mean monthly number of page impressions before March 2013 by category (see Table A.14). Although these category page impressions do not directly measure an outlet's content category composition, they are a valid proxy as a relatively large number of page impressions in a specific content category indicates that the outlet concentrates more on that category.

For each outlet i, we compute the relative number of page impressions for each of the eight main categories listed in Table A.14. Denote these fractions as $c_{i,j}$, with $j=1, ..., 8$ and $\sum_j c_{i,j}=1$. Using the fractions $c_{i,j}$, we then compute an Inverse Herfindahl diversity measure, based on a well-known measure of within-firm diversity (Montgomery, 1982; Zahavi & Lavie, 2013):[26]

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

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 in August 2014 (i.e., we discard the 17 months in between).[27] By considering only local outlets, we reduce (potentially unobserved) heterogeneity beyond scale and scope (e.g., in terms of quality or news content). Table 1 summarizes all our variables.

4.3 | Empirical strategy

4.3.1 | Baseline specification

To isolate the causal effect of platform removal on local outlets' web traffic, we use the unexpected removal of *VGM* members from the aggregators' platforms following the passage of the copyright bill in March 2013 and the subsequent legal dispute between *VGM* and news

[25]Since all outlets in our sample were listed by news aggregators before March 2013, our measure of scale incorporates not only direct visits, but also visits that were generated through the news aggregators.

[26]We use the entropy measure (Jacquemin & Berry, 1979; Palepu, 1985) as well as a simple count of categories covered (Hashai, 2015) as robustness checks in the Appendix A.

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

TABLE 1 Summary statistics.

Sample	Variable	Obs	Mean	Median	SD	Min	Max
All outlets	VGM	4560	0.42	0.00	0.49	0.00	1.00
	Post	4560	0.48	0.00	0.50	0.00	1.00
	Visits	4560	1,190,834.54	493,470.00	1,803,334.84	111.00	14,348,816.00
	log(Visits)	4560	13.10	13.11	1.42	4.71	16.48
	avgVisits	4560	1,154,431.58	446,753.94	1,881,663.93	30,198.75	12,500,797.00
	log(avgVisits)	4560	13.03	13.01	1.42	10.32	16.34
	Inv_Herfindahl	4560	0.46	0.49	0.14	0.12	0.76
VGM members	Post	1907	0.51	1.00	0.50	0.00	1.00
	Visits	1907	1,865,636.20	1,046,265.00	2,421,505.28	11,079.00	14,348,816.00
	log(Visits)	1907	13.68	13.86	1.35	9.31	16.48
	avgVisits	1907	1876 t,420.16	917,313.13	2,596,555.29	37,546.58	12,500,797.00
	log(avgVisits)	1907	13.61	13.73	1.41	10.53	16.34
	Inv_Herfindahl	1907	0.44	0.48	0.16	0.12	0.73
VGM non-members	Post	2653	0.46	0.00	0.50	0.00	1.00
	Visits	2653	705,781.11	273,147.00	901,899.85	111.00	6,440,609.00
	log(Visits)	2653	12.68	12.52	1.31	4.71	15.68
	avgVisits	2653	635,459.77	273,643.19	772,556.60	30,198.75	3,441,650.75
	log(avgVisits)	2653	12.61	12.52	1.27	10.32	15.05
	Inv_Herfindahl	2653	0.48	0.50	0.13	0.17	0.76

aggregators as an exogenous shock. Specifically, 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 model, we omit the period between these dates to prevent that unobserved anticipation effects on behalf of *VGM* members or news aggregators, or the short-term removal of *VGM* members from *Google News* (Calzada & Gil, 2020) confound our results. We provide supporting evidence for this decision in Appendix A.2.5, where we also demonstrate that our results are robust to considering this time period. The baseline regression is:

$$\log(Visits)_{it} = \beta(VGM_i * Post_t) + \varphi_i + \lambda_t + \varepsilon_{it}, \tag{3}$$

where the dependent variable is the web traffic (in visits) of outlet i in month t, the treated variable VGM_i is a dummy equal to one for all *VGM* members, the treatment variable $Post_t$ is a dummy equal to one for all time periods after the passage of the copyright bill, and φ_i and λ_t are outlet and monthly fixed effects, respectively. The parameter of interest in Equation (3), β, gives the average change in web traffic of *VGM* members after August 2014 relative to the change in web traffic of non-members.

We include monthly and outlet fixed effects in our analysis. Monthly fixed effects control for general changes in all outlets' web traffic, including seasonality and a growing online audience. Outlet fixed effects capture (time-invariant) unobserved heterogeneity between our observations, including differences in quality, style, political orientation, specific features of the local market they operate in, and potential differences in exposure to the news aggregators.[28] For example, high-quality, well-organized, or visually attractive outlets could tend to join *VGM*. As the web traffic of such outlets is likely to be larger, too, omitting outlet fixed effects could lead to overestimating the impact of platform removal. Including outlet fixed effects mitigates omitted variable bias from unobserved time-invariant heterogeneity. Moreover, outlet fixed effects bring our empirics closer to the theoretical framework in Section 3. Specifically, our formal model considers the ceteris paribus effects of scale and scope: we isolate the impact of both moderators holding everything else fixed. Thus, using outlet fixed effects to capture time-invariant heterogeneity among outlets and monthly fixed effects to control for general changes in all outlets' web traffic brings our empirical analysis as close to this ideal as possible.

4.3.2 | Scale and scope as moderators

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

$$\log(Visits)_{it} = \beta_1(VGM_i * Post_t) + \beta_2(S_i * Post_t) + \beta_3(VGM_i * S_i * Post_t) + \varphi_i + \lambda_t + \varepsilon_{it}, \tag{4}$$

[28] Quality, style, and features of a local media market may gradually change over time, but our observation period is sufficiently short to consider them time-invariant. In Appendix A.2.3, we show that our results are robust to including outlet-specific linear time trends, as well as including outlet times quarter of the year fixed effects to account for different seasonality at the outlet level.

where S_i refers to outlet scale and scope as defined in Section 4.2. The parameter of interest in (4) is β_3, the moderating effect of outlet scale and scope on the effect of being removed from aggregators in August 2014.

5 | RESULTS

5.1 | Baseline specification

Column 1 of Table 2 shows OLS estimates for regression (3). All specifications in Table 2 include outlet and monthly 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.[29]

5.2 | Scale as moderator

Column 2 of Table 2 displays the OLS estimate of the triple difference-in-differences (4) capturing the moderating effect of outlet scale. Consistent with H1, we find that the estimate for β_3 is negative and statistically significant (*coeff.* = −0.088, $p = .011$).

The point estimate can be interpreted as follows. According to $\widehat{\beta}_3$, 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 (*coeff.* $= 1.199$, $p = .009$). To interpret our estimates, note that β_1 captures the baseline effect of being removed from an aggregator for a (fictional) outlet of zero scale. Clearly, examining $\widehat{\beta}_1$ in isolation is not meaningful. Thus, we jointly evaluate $\widehat{\beta}_1$ and $\widehat{\beta}_3$ at meaningful margins of outlet scale. For instance, consider the median *VGM* member (in terms of scale) in our sample with $\log(avgVisits)_i \approx 13.73$. For this outlet, the overall treatment effect is negative and equal to -0.009 (i.e., $1.199 - 0.088 \times 13.73$). In other words, the median *VGM* member's web traffic decreases by about 0.9% after August 2014. While this effect grows (i.e., becomes more negative) as outlet scale increases, it approaches zero and eventually switches sign as outlet scale shrinks. Specifically, a small outlet (25th size percentile) attracts around 21.240 (or 16.96%) more monthly visits post-removal, while a large outlet (75th size percentile) loses about 56.300 (or 4.08%) of its monthly visits. The reversal of the treatment effect occurs around the 45th 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.

Thus, in addition to supporting Hypothesis 1, our analysis on outlet scale shows that the sign of the overall treatment effect differs between small- and large-scale outlets, which explains the zero average effect we find in Column 1 of Table 2. Moreover, the effect heterogeneity supports our theory from Section 3.4.1, where we argue that larger-scale outlets are more likely to benefit from aggregators as they are better able to attract additional demand

[29]Note that our panel is not strongly balanced, because some outlets did not report their web traffic to *IVM* in each month (e.g., for technical reasons) or joined *IVM* later. When we restrict the analysis to a fully balanced panel (103 outlets), we obtain very similar results with even larger coefficients of interest.

TABLE 2 Main results.

	(1)	(2)	(3)	(4)
	log(Visits)	**log(Visits)**	**log(Visits)**	**log(Visits)**
Post * VGM	−0.001	1.199	0.490	1.137
	(0.982)	(0.009)	(0.009)	(0.020)
Post * log(avgVisits)		−0.005		−0.004
		(0.856)		(0.886)
Post * VGM * log(avgVisits)		−0.088		−0.056
		(0.011)		(0.090)
Post * Inv_Herfindahl			0.273	0.271
			(0.415)	(0.422)
Post * VGM * Inv_Herfindahl			−1.100	−0.835
			(0.004)	(0.028)
Outlet FE	X	X	X	X
Time FE	X	X	X	X
Constant	13.10	13.13	13.04	13.06
	(0.000)	(0.000)	(0.000)	(0.000)
N	4560	4560	4560	4560
R^2	.974	.975	.975	.976

Note: The *p*-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the outlet level (140 clusters).

from the platforms than smaller-scale outlets, which may even suffer from more competition for readers on platforms. Hence, smaller-scale outlets may benefit from being off the platform.

5.3 | Scope as moderator

Column 3 of Table 2 shows OLS estimates of the triple difference-in-differences (4) for the moderating effects of outlet scope (Column 3). Analogous to Column 2, we find that the estimate for β_3 is negative and statistically significant (*coeff.* $= -1.100$, $p = .004$), supporting Hypothesis 2.

Again, we interpret the point estimates using the estimates from Column 3. According to $\widehat{\beta}_3$, a one standard deviation increase in outlet scope (SD = 0.16) corresponds to a 17.6% decrease in *VGM* members' web traffic after August 2014. The median *VGM* member in terms of scale attracted an average of around 917,000 visits per month before March 2013. If the scope of this median outlet increases by one standard deviation, its web traffic would decrease by about 161,000 visits per month. An outlet with narrow scope (25th scope percentile) gains around 108.000 (or 9.9%) monthly views post-removal, while an outlet with broad scope (75th scope percentile) loses about 250.900 (or 12.4%) monthly views.

Our estimate for β_1 is positive in Column 3 and statistically significant (*coeff.* $= 0.490$, $p = .009$). Again, we jointly evaluate $\widehat{\beta}_1$ and $\widehat{\beta}_3$ at meaningful margins of outlet scope. Consider the median *VGM* member in our sample in terms of scope ($Inv\_Herfindahl_i = 0.48$). For this outlet, the full treatment effect equals -0.038 (i.e., $0.490 - 1.100 \times 0.48$). While this effect grows (i.

e., becomes more negative) if outlet scope increases, it approaches zero and eventually turns positive if outlet scope decreases. The reversal of the effect again occurs around the 45th percentile of outlet scope distribution. Hence, the overall effect of being removed from aggregators is negative for the majority of *VGM* members in our sample.

5.4 | Scale and scope as moderators

Of course, large-scale outlets may also have a broad scope. In other words, it could be that outlet scale and scope do not operate as independent moderators, but simultaneously characterize the same outlets. The unconditional correlation between scale and scope is .1687 (significant at $p < .000$), indicating moderate collinearity. To evaluate the relationship between our two main moderators, we include both scale and scope in our regression in Column 4 of Table 2. Our estimates for the moderating effects of scale and scope remain negative, albeit at reduced statistical significance ($p = .011 \rightarrow p = .090$ and $p = .004 \rightarrow p = .028$ for scale and scope, respectively), and their magnitudes drop by 36% and 24%, respectively. Thus, despite some collinearity, outlet scale and scope capture different dimensions of heterogeneity.[30]

5.5 | Validity and robustness checks

We perform several validity and robustness checks, summarized here and explained further in Appendix A.

5.5.1 | Validity checks

First, we demonstrate the validity of the parallel trends assumption. Specifically, our empirical approach compares the development of web traffic of *VGM* members and nonmembers before and after the passage of the copyright bill in a (triple) difference-in-differences framework. We thereby assume that the web traffic of *VGM* members would have developed in parallel to the web traffic of nonmembers if their content had not been removed from the news aggregators. Moreover, given that we study outlet scale and scope as moderators, we also assume that the web traffic of *VGM* members that are small or large in scale (narrow or broad in scope) would have developed parallel to traffic of nonmembers comparable in scale or scope.

We present several arguments that speak against selection or anticipation effects on behalf of the *VGM* members (Appendix A.1.1). Moreover, we conduct two types of placebo regressions that support the parallel trends assumption (Appendix A.1.2). Specifically, we first focus on the time period *before* the copyright bill was passed and examine the impact of a series of fake treatment dates. The idea is that while the web traffic of *VGM* members and nonmembers has developed differently *after* the content of *VGM* members was removed from the news aggregators, these differences should not have occurred *before* the bill was passed. Hence, all estimates for

[30]Figure A.2 in the Appendix plots all time fixed effects from the four regressions in Table 2, that is, the general time trend in news outlets' web traffic. All curves are upwards trending, showing that the online audience has grown over time. Moreover, all curves follow the same parallel trend. The differences in levels can be explained through different amounts of variation in web traffic that can be explained by our moderators, scale, and scope.

the fake treatments *before* March 2013 should be close to zero and not statistically significant, which is indeed the case. Second, we replace the dependent variable $\log(Visits)_{it}$ in Equation (4) with the log number of print copies as readers of an outlet's print version typically do not consume its online version. Thus, the number of print copies should be unaffected if an outlet's online content is removed from news aggregators. Conversely, any divergence in the development of printed copies between *VGM* members and nonmembers after March 2013 could indicate that the parallel trends assumption is violated. Reassuringly, our regression estimates with the "fake" dependent variable are close to zero and not statistically significant, further corroborating our empirical strategy.

5.5.2 | Robustness checks

We also document the robustness of our results along multiple dimensions. We first demonstrate that our results do not hinge on our variable specifications by showing that our results are robust to using alternative measures for our dependent variable web traffic (Appendix A.2.1) and our moderators scale and scope (Appendix A.2.2). Next, we show that our results are robust to including both linear outlet specific time trends as well as outlet times seasonality fixed effects into our analyses (Appendix A.2.3). Hence, we can rule out that our results are driven by unobserved outlet characteristics that vary over time.

Importantly, our results are also robust to considering different time periods. First, they hold for using both shorter and longer pre- and post-treatment observation periods (Appendix A.2.4).

TABLE 3 Without dropping 17 months in between.

	(1)	(2)	(3)	(4)
	log(Visits)	**log(Visits)**	**log(Visits)**	**log(Visits)**
Post * VGM	0.00451	0.942	0.343	0.940
	(0.902)	(0.006)	(0.022)	(0.013)
Post * log(avgVisits)		0.00166		0.00319
		(0.932)		(0.868)
Post * VGM * log(avgVisits)		−0.0695		−0.0509
		(0.007)		(0.038)
Post * Inv_Herfindahl			0.208	0.211
			(0.438)	(0.438)
Post * VGM * Inv_Herfindahl			−0.761	−0.555
			(0.013)	(0.064)
Outlet FE	X	X	X	X
Time FE	X	X	X	X
Constant	13.09	13.08	13.03	13.00
	(0.000)	(0.000)	(0.000)	(0.000)
N	6053	6053	6053	6053
R^2	.973	.974	.974	.974

Note: The *p*-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the outlet level (140 clusters).

TABLE 4 Two treatment periods.

	(1) log(Visits)	(2) log(Visits)	(3) log(Visits)	(4) log(Visits)
Post_1 * VGM	−0.0104	0.666	0.194	0.662
	(0.753)	(0.041)	(0.152)	(0.063)
Post_2 * VGM	0.0238	1.302	0.529	1.283
	(0.637)	(0.005)	(0.006)	(0.010)
Post_1 * log(avgVisits)		0.00467		0.00595
		(0.789)		(0.734)
Post_1 * VGM * log(avgVisits)		−0.0502		−0.0396
		(0.042)		(0.116)
Post_2 * log(avgVisits)		−0.00142		0.000237
		(0.958)		(0.993)
Post_2 * VGM * log(avgVisits)		−0.0946		−0.0650
		(0.006)		(0.050)
Post_1 * Inv_Herfindahl			0.122	0.125
			(0.585)	(0.580)
Post_1 * VGM * Inv_Herfindahl			−0.457	−0.312
			(0.095)	(0.256)
Post_2 * Inv_Herfindahl			0.314	0.316
			(0.361)	(0.363)
Post_2 * VGM * Inv_Herfindahl			−1.133	−0.846
			(0.004)	(0.031)
News outlet FE	X	X	X	X
Time FE	X	X	X	X
Constant	13.09	13.07	13.03	13.00
	(0.000)	(0.000)	(0.000)	(0.000)
N	6053	6053	6053	6053
R^2	.973	.974	.974	.974

Note: The *p*-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the news outlet level (140 clusters).

Second, they are robust to including the 17 months between the passage of the copyright bill and the ultimate removal of *VGM* members' content from news aggregators as a distinct treatment period. We run two robustness checks along these lines: First, we add the time period from March 2013 to August 2014 to our original post-treatment observations and estimate the average effect of being delisted from the news aggregators for this extended time period (see Table 3). Second, we consider two distinct post-treatment periods, where the first corresponds to March 2013 to August 2014, and the second corresponds to the 18 months after that, i.e., to our original post-treatment observation period (see Table 4). Tables 3 and 4 demonstrate that the resulting estimates are very similar to our main results.

TABLE 5 Scale and scope by quintiles.

	(1) log(Visits)	(2) log(Visits)	(3) log(Visits)
Post * VGM	0.212	0.206	0.344
	(0.006)	(0.054)	(0.013)
Post * log(avgVisits)	−0.0178		−0.0114
	(0.507)		(0.669)
Post * VGM * Q2_log(avgVisits)	−0.183		−0.177
	(0.034)		(0.076)
Post * VGM * Q3_log(avgVisits)	−0.127		−0.166
	(0.234)		(0.155)
Post * VGM * Q4_log(avgVisits)	−0.147		−0.103
	(0.182)		(0.391)
Post * VGM * Q5_log(avgVisits)	−0.348		−0.268
	(0.007)		(0.053)
Post * Inv_Herfindahl		0.149	0.177
		(0.642)	(0.583)
Post * VGM * Q2_Inv_Herfindahl		−0.237	−0.154
		(0.016)	(0.151)
Post * VGM * Q3_Inv_Herfindahl		−0.154	−0.132
		(0.172)	(0.234)
Post * VGM * Q4_Inv_Herfindahl		−0.334	−0.314
		(0.018)	(0.020)
Post * VGM * Q5_Inv_Herfindahl		−0.426	−0.319
		(0.013)	(0.055)
News outlet FE	X	X	X
Time FE	X	X	X
N	4560	4560	4560
R^2	.975	.975	.976

Note: The *p*-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the news outlet level (140 clusters).

Next, we generate dummy variables for quintiles of the distribution of scale and scope (Appendix A.2.6). Using these dummy variables instead of the continuous measures for scale and scope in Equation (4) gives us the average impact of belonging to a certain part of the scale or scope distribution on the change in web traffic (the baseline is the lowest quintile) to capture potential non-linearities in the impact of scale and scope. Table 5 shows that, in line with the results from our main specification, outlets that are large (small) in scale and broad (narrow) in scope lose (gain) web traffic when their content is being removed from the news aggregators. Table 5 also shows that the effect is mainly driven by the larger outlets. Note, however, that we cannot directly compare the magnitude of the coefficients in Table 5 to our main results in Table 2. In particular, while the coefficients for the triple interaction terms in Table 2 measure the average change in web traffic if scale and scope increase by 1%, the coefficients for the triple interaction terms in Table 5 must be interpreted as the average percentage change in web traffic for outlets in a specific quintile of the distribution of scale and scope relative to outlets in the lowest quintile.

The political position of news outlets could also mitigate or reinforce the moderating effects of scale and scope (Appendix A.3.1). In particular, outlets positioned at the center of the political spectrum might better attract readers' attention when listed by news aggregators than outlets positioned at the political extremes. Consistent with that reasoning, we find that the negative impact of outlet scale and scope when being delisted from news aggregators is reinforced when we consider only outlets close to the center of the political spectrum, and that it is mitigated when using only outlets positioned more at the extremes.

Finally, we study the *Google* dispute that took place in October and November 2014 (Appendix A.3.2). In line with Calzada and Gil (2020), we find that its impact on web traffic was limited.

6 | DISCUSSION AND CONCLUSION

We explore on-platform effects of competition for consumer attention among complementors. While platforms can grant smaller complementors access to larger potential demand, they also expose them to more intense competition from other complementors, all battling to capture the same limited consumer attention. Studying local news outlets, which traditionally only had access to a limited market, we ask how the performance of a subset of these outlets that were removed from several news aggregators in Germany after a legal dispute evolved compared with local outlets that remained.

We argue (and show empirically) that outlets larger in scale and/or broader in scope benefit more from being on aggregators, as they can better capture consumer's attention on platforms. When displayed side-by-side, articles by larger outlets will attract a larger share of consumer attention compared with those of smaller outlets. We posit that content by larger-scale outlets is the default option for potential readers whereas turning to smaller-scale outlets would require additional cognitive effort from readers to move past their default. Outlet scale directs consumer attention, which helps larger outlets capture more of the demand on news aggregators. Outlets that broadly cover different types of content also capture a larger share of on-platform demand because they do not disproportionally lose attention by readers in specific categories.

Indeed, we find that outlets with small scale and/or narrow scope are better off not being on aggregators at all. By featuring their content on aggregators, smaller outlets are exposed to negative, "outgoing attention spillovers." These adverse effects offset the positive impact of "incoming attention spillovers," ultimately drawing attention and readership away from them. Similarly, outlets with narrow scope do not attract sufficient attention from their competitors' readers while losing a portion of their own readership to competitors in categories they are comparably weak in.

6.1 | Implications for research

We contribute to the debate on the substitution between on- and off-platform sales channels as a driver of complementors' success (Athey et al., 2021; Calzada & Gil, 2020; Chiou & Tucker, 2017; Kretschmer & Peukert, 2020) and to work arguing that the shift in the type of competition on platforms compared with traditional markets affects complementors' ability to capture value (Adner & Lieberman, 2021; Cennamo, 2021; Zhu & Liu, 2018). Our findings stress the possibly asymmetric competitive forces complementors face when competing for consumer attention on platforms and identify the relative importance of two factors that drive these forces: scale and scope. This complements prior work on news aggregators and their effect on outlets (Athey et al., 2021; Calzada & Gil, 2020; Chiou & Tucker, 2017) by focusing specifically on how the on-platform competitive relationships between outlets play out. By drawing on a setting with a large set of fairly homogenous (except for their scale and scope) local outlets where the market expansion effect is held somewhat constant and in which most readers already consume local news, we can focus on the effect of competition among outlets on the same platform and the resulting redistribution of readers across them. We thus complement research that has compared the on-platform performance of more diverse groups of outlets (such as local and non-local outlets) (Athey et al., 2021; Calzada & Gil, 2020; Chiou & Tucker, 2017) by highlighting the moderating effect of scale and scope *within* a group of outlets that is less diverse in terms of its idiosyncratic characteristics. This feeds into a broader research agenda on the largest group of actors of the platform economy, complementors, and how their heterogeneity affects the benefits they can attain from platform membership. Specifically, while market expansion effects and competition among complementors are often hard to disentangle, we can isolate the latter. Particularly for more mature platforms where user growth declines, such competition among complementors becomes the main mechanism of competitive dynamics. While smaller complementors may benefit from market expansion effects in eras of growth, they may be hurt by competition in the market when user growth slows down, and competition intensifies. Thus, we add nuance to the analysis of platform markets by focusing on complementor heterogeneity (compared with market-level factors).

We also shed light on the relationship between platform dynamics and complementor scope. Studies on firm scope in platform markets often focus on platform decisions (Cennamo, 2021; Gawer, 2021; Giustiziero et al., 2023), while complementors are often considered small and atomistic. However, some recent work has studied complementor characteristics such as scale and scope. Prior studies mostly study complementor scope *across* platforms in the form of multi-homing (Cennamo et al., 2018; Chung, Zhou, & Ethiraj, 2023; Li & Zhu, 2021; Tavalaei & Cennamo, 2021), or even the scope of complementors within (diversifying) *and* across platforms (Chung, Zhou, & Choi, 2023). We extend this stream of research by focusing on complementor scope *within* platforms.

Specifically, we suggest that complementors, especially those joining a platform after a period of independent operation, differ in dimensions that affect their on-platform performance and can obtain greater returns from scale and scope. This may call the efficacy of “focus strategies” in the manner of Porter (1980) into question and hints at another source of economies of scale through the ability to attract eyeballs on a crowded platform. Our study on *firm*-level drivers of attention thus complements previous *product*-level studies on “hit” or “superstar” products (Elberse, 2008; Elberse & Oberholzer-Gee, 2006; Kumar et al., 2014; Tan et al., 2017), which find evidence of demand concentration around better-known products. We expect such patterns to be particularly pronounced if the exact characteristics of complements (and their potential match with consumer preferences) are hard to assess before consumption. Conversely, a different pattern may emerge if consumers can use information on complement heterogeneity in their decision making. In which settings, then, will each of these patterns be dominant and how will competition play out? What strategies work best for complementors on platforms where complement characteristics are hard to assess?

Our findings also speak to the debate on the role of algorithms in competition on platform markets. While we do not observe the underlying algorithms at play in our setting, evidence suggests that algorithms and recommender systems mostly reinforce existing selection patterns by placing those complements higher up in the ranking that have been popular in the past (Fleder & Hosanagar, 2009) or, more generally, those that have a higher probability of being selected (Ursu, 2018). Such a bias would then drive the selection of news articles in the topic clusters shown on news aggregators. Less-known options may not be selected by the algorithm and the platform may favor popular content from large complementors. While this bias does not force the readers' ultimate choices from the menu, it steers readers to content by larger and broader outlets, which would reinforce an already existing tendency of "winner-takes-most" dynamics. However, note that, while algorithms might reinforce such tendency, they are unlikely to generate the underlying effect in the first place. Algorithms mostly base their ranking on previous performance (Baye et al., 2016), which by definition requires outlets larger in scale and scope to attract more attention (and ultimately clicks) initially before they can be favored by the algorithm at a later point in time. This raises interesting questions about market efficiency and the power of news aggregators. Does the market reward efficient players (e.g., large and broadly diversified ones) while penalizing players below a minimum viable scale? Does this affect incentives to provide quality? Relatedly, search engines have significant market power and can potentially introduce bias (De Corniere & Taylor, 2014) or affect the concentration of web traffic (Calzada et al., 2023).

Our work also emphasizes the role of complementor heterogeneity. Platforms use rating systems and other tools to reduce transaction costs and coordinate cross-side market interactions, which reduces information asymmetries and enables effective signaling of high quality to consumers (Chevalier & Mayzlin, 2006; Sun, 2012), especially for experience goods (Kumar et al., 2014; Nelson, 1970). However, if attention strongly affects consumer choice, such mechanisms reward prior attention with more attention and may raise entry barriers for better or more targeted alternatives. The link between competition for consumer attention, ratings, and complement quality is central to understanding overall platform efficiency.

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

6.2 | Implications for practice and policy

Our study also raises important questions for practice and policy, particularly regarding the debate around the impact of information aggregators and platforms on market efficiency, and on society at large. Regarding news outlets, what societal impact could these competitive dynamics have? Do they pose a risk to democracy, as suggested by some observers (Greenslade, 2016), by reducing the plurality of news sources, leaving only big players to deliver news? Our results suggest that in our sample of local outlets that are relatively small to begin with, smaller and narrower outlets are indeed the ones who are harmed most by aggregators. This is consistent with the general perception that small outlets may suffer most by ongoing digitization in the media industry. Additional analyses (reported in Appendix A.3.1) suggest that the effect of scale and scope is more pronounced for less polarized outlets than for more polarized ones, suggesting that the attention mechanism is more pronounced in the case of less polarized outlets. Hence, more generalist outlets can better attract readers' attention when listed by news aggregators and may benefit more from them. Conversely, if the media landscape is more polarized and outlets are known to be polarized ex-ante, the matching mechanism may become more pronounced. Hence, whether a setting is characterized by "rich get richer" dynamics that make big players strive depends on complementor heterogeneity and, more importantly, the ability of potential consumers to perceive it.

Relatedly, do the dynamics we document condemn citizens to consume low-quality, attention-grabbing content? While our results suggest that if platform aggregators play a key role in news consumption we would observe a few large players competing for attention, it is an open question if this is bad or good for content quality on the "production" side and if this leads to more and better-informed readers on the "consumption" side. Given the spillovers across outlets, readers may "multihome" easily across outlets on the same platform and source similar (or different) information from different outlets, putting constant pressure on the quality of news provided by different outlets (Peitz & Reisinger, 2014). This competitive pressure for consumer attention might increase overall quality. Conversely, in chasing a larger audience, outlets may promote attention-grabbing content at the cost of de-emphasizing other content that is of high quality and socially beneficial, but of more limited appeal. If such niche content is of higher average quality than popular content, average content quality could drop. Then, policy interventions mandating 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 stop a "race to bottom" of attention-generating content.

6.3 | Avenues for future research

Our study opens rich avenues for future work. First, while our empirical context of online news content comes with several specificities (e.g., often free consumption, little repeat consumption of articles), it lets us isolate the mechanisms around competition for attention and creates ex-ante heterogeneity among complementors from the "outside world" (off-platform). We expect these mechanisms to apply also, for instance, to complementors on transaction platforms like *Amazon Marketplace*, or small restaurants on *UberEats*. While featuring on these platforms gives firms access to a larger market, it also puts them in direct competition with larger competitors who may attract more attention and capture more demand when complements are displayed side-by-side. How these two mechanisms interact, and which one will dominate in which situations remains unclear. Whether and to what extent smaller complementors can command attention when competing with larger ones on a platform will matter for complementors on any type of platform. Further, getting a better understanding of how long the attention-based advantage of larger outlets lasts after an increase in search costs is another interesting aspect to study. We find that readers consume more content from outlets that are smaller in size and less broad in scope after outlets get removed from aggregators, which suggests that the advantage of larger and broader outlets dissipates quickly. However, we do not observe how this process unfolds in detail. Finally, we do not directly observe individual consumer behavior and decision-making processes. While our findings are in line with our predictions on the net effect of incoming and outgoing attention spillovers, research at the level of specific products and individuals would be a promising path to isolate the mechanisms at play.

ACKNOWLEDGEMENT

Open Access funding enabled and organized by Projekt DEAL.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at https://osf.io/35f48.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request. A replication package of the data and the code is available at https://osf.io/35f48/.

ORCID

Tim Meyer https://orcid.org/0009-0003-3562-9327
Carmelo Cennamo https://orcid.org/0000-0003-1050-7713
Tobias Kretschmer https://orcid.org/0000-0002-7830-7825

REFERENCES

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

Alba, J. W., & Hutchinson, J. W. (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.

Anderson, S. P., & Jullien, B. (2015). The advertising-financed business model in two-sided media markets. In S. P. Anderson, J. Waldfogel, & D. Strömberg (Eds.), *Handbook of media economics* (Vol. 1, pp. 41–90). North-Holland.

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

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

Baye, M. R., De los Santos, B., & Wildenbeest, M. R. (2016). Search engine optimization: What drives organic traffic to retail sites? *Journal of Economics & Management Strategy*, *25*(1), 6–31.

BDZV. (2022). Werbewirkung E-Paper. Bundesverband Digitalpublisher und Zeitungsverleger e.V. https://www.bdzv.de/alle-themen/marktdaten/werbewirkung-e-paper

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

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.

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

Bundesanzeiger. (2013). Achtes Gesetz zur Änderung des Urheberrechtsgesetzes. 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. 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., Duch-Brown, N., & Gil, R. (2023). Do search engines increase concentration in media markets? Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4592958

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.

Cennamo, C., Özalp, H., & Kretschmer, T. (2018). Platform architecture and quality trade-offs of multihoming complements. *Information Systems Research*, *29*(2), 461–478.

Chevalier, J. A., & 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.

Cho, N. E., & Zhou, Y. M. (2021). Profits, reputation, and the risk of medical malpractice liability. *Strategy Science*, *6*(3), 228–243.

Chung, H. D., Zhou, Y. M., & Choi, C. (2023). When Uber eats its own business, and its competitors' too: Platform diversification and cross-platform cannibalization. Working Paper.

Chung, H., Zhou, Y., & Ethiraj, S. (2024). Platform governance in the presence of within-complementor interdependencies: Evidence from the rideshare industry. *Management Science*, *70*(2), 799–814.

De Corniere, A., & Taylor, G. (2014). Integration and search engine bias. *The Rand Journal of Economics*, *45*(3), 576–597.

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*. 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, D. S. (2013). Attention rivalry among online platforms. *Journal of Competition Law & Economics*, *9*(2), 313–357.

Evans, D. S. (2019). Attention platforms, the value of content, and public policy. *Review of Industrial Organization*, *54*(4), 775–792.

Fischer, G. (2018). Online & Print: Wie organisieren sich die Redaktionen? Einfacherdienst. https://einfacherdienst.de/online-print-wie-organisieren-sich-die-redaktionen

Fleder, D., & Hosanagar, K. (2009). Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management Science*, *55*(5), 697–712.

Fricker, T. (2022). Wie sich die Print- und Onlineredaktion der BZ langsam annäherten. Badische Zeitung. https://www.badische-zeitung.de/wie-sich-die-print-und-onlineredaktion-der-bz-langsam-annaeherten?mode=in

Garz, M., Sörensen, J., & Stone, D. F. (2020). Partisan selective engagement: Evidence from Facebook. *Journal of Economic Behavior & Organization*, *177*, 91–108.

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

George, G., Dahlander, L., Graffin, S. D., & Sim, S. (2016). Reputation and status: Expanding the role of social evaluations in management research. *Academy of Management Journal*, *59*(1), 1–13.

George, L. M., & 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. (2023). Hyperspecialization and hyperscaling: A resource-based theory of the digital firm. *Strategic Management Journal*, *44*(6), 1391–1424.

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

Gulati, R., Puranam, P., & Tushman, M. (2012). Meta-organization design: Rethinking design in inter-organizational and community contexts. *Strategic Management Journal*, *33*(6), 571–586.
Hansen, M. T., & Haas, M. R. (2001). Competing for attention in knowledge markets: Electronic document dissemination in a management consulting company. *Administrative Science Quarterly*, *46*(1), 1–28.
Hashai, N. (2015). Within-industry diversification and firm performance—An S-shaped hypothesis. *Strategic Management Journal*, *36*(9), 1378–1400.
Hirche, T. (2013). Weiter Zu- und Absagen von Verlagen an VG Media. IGEL Initiative gegen ein Leistungsschutzrecht. https://leistungsschutzrecht.info/news/2013-11-25/weitere-zu-und-absagen-von-verlagen-an-vg-media
Hotelling, H. (1929). Stability in competition. *The Economic Journal*, *39*(153), 41–57.
Jacobides, M. G., Cennamo, C., & Gawer, A. (2018). Towards a theory of ecosystems. *Strategic Management Journal*, *39*(8), 2255–2276.
Jacobides, M. G., Cennamo, C., & Gawer, A. (2024). Externalities and complementarities in platforms and ecosystems: From structural solutions to endogenous failures. *Research Policy*, *53*(1), 104906.
Jacquemin, A. P., & Berry, C. H. (1979). Entropy measure of diversification and corporate growth. *The Journal of Industrial Economics*, *27*, 359–369.
Kahneman, D. (1973). *Attention and effort* (Vol. 1063, pp. 218–226). Prentice-Hall.
Klaiber, S. (2013). Neues Leistungsschutzrecht: ein Gesetz, viele Fragen. 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. http://www.taz.de/!5033165/
Kumar, A., Smith, M. D., & 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. https://www.heise.de/newsticker/meldung/Leistungsschutzrecht-VG-Media-nimmt-Microsoft-Yahoo-und-Telekom-ins-Visier-2235726.html
Lanham, R. A. (2006). *The economics of attention: Style and substance in the age of information*. University of Chicago Press.
Li, H., & Zhu, F. (2021). Information transparency, multihoming, and platform competition: A natural experiment in the daily deals market. *Management Science*, *67*(7), 4384–4407.
Loh, J., & Kretschmer, T. (2023). Online communities on competing platforms: Evidence from game wikis. *Strategic Management Journal*, *44*(2), 441–476.
Macdonald, E. K., & Sharp, B. M. (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, D. P., & Srinivasan, A. (2017). Networks, platforms, and strategy: Emerging views and next steps. *Strategic Management Journal*, *38*(1), 141–160.
Media Landscapes. (2021). Germany. https://medialandscapes.org/country/germany/media/print
Montgomery, C. A. (1982). The measurement of firm diversification: Some new empirical evidence. *Academy of Management Journal*, *25*(2), 299–307.
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. 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, G. G., & van Alstyne, M. W. (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, M. E. (1980). *Competitive strategy: Techniques for analyzing industries and competitors*. Free Press.

Prat, A., & Valletti, T. M. (2022). Attention oligopoly. *American Economic Journal: Microeconomics*, *14*(3), 530–557.

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

Rochet, J. C., & Tirole, J. (2006). Two-sided markets: A progress report. *The Rand Journal of Economics*, *37*(3), 645–667.

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.

Shapiro, C. (1983). Premiums for high quality products as returns to reputations. *The Quarterly Journal of Economics*, *98*(4), 659–679.

Skogerbø, E., & Winsvold, M. (2011). Audiences on the move? Use and assessment of local print and online newspapers. *European Journal of Communication*, *26*(3), 214–229.

Spulber, D. F. (2019). The economics of markets and platforms. *Journal of Economics & Management Strategy*, *28*(1), 159–172.

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.

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

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, T. F., 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, M. M., & 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.

Thurman, N. (2018). *Time spent Reading news-brands: Print–digital performance, online-only effects, and long-term trends*. World Association of Newspapers and News Publishers (WAN-IFRA).

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

Ursu, R. M. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science*, *37*(4), 530–552.

Washington, M., & Zajac, E. J. (2005). Status evolution and competition: Theory and evidence. *Academy of Management Journal*, *48*(2), 282–296.

Weigelt, K., & Camerer, C. (1988). Reputation and corporate strategy: A review of recent theory and applications. *Strategic Management Journal*, *9*(5), 443–454.

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.

How to cite this article: Meyer, T., Kerkhof, A., Cennamo, C., & Kretschmer, T. (2024). Competing for attention on digital platforms: The case of news outlets. *Strategic Management Journal*, 1–60. https://doi.org/10.1002/smj.3600

APPENDIX A

A.1 | VALIDITY OF THE PARALLEL TRENDS ASSUMPTION

Our empirical analysis hinges on the validity of the parallel trends assumption. In particular, we compare the development of web traffic of *VGM* members and nonmembers before the passage of the Copyright Bill in March 2013 and after the subsequent removal from a subset of aggregators in a (triple) difference-in-differences framework. We thereby assume that the web traffic of *VGM* members would have developed parallel to the web traffic of nonmembers if their content had not been removed from the news aggregators. Moreover, given that we study outlet scale and scope as moderators of the effect, we also assume that the web traffic of *VGM* members that are small or large in scale (narrow or broad in scope) would have developed parallel to the web traffic of nonmembers that are similarly small or large in scale (narrow or broad in scope). To support the plausibility of these assumptions, we present multiple arguments that speak against selection or anticipation effects on behalf of the *VGM* members in Appendix A.1.1, and we conduct a series of placebo regressions and event studies in Appendix A.1.2.

A.1.1. | No selection or anticipation effects

Our empirical strategy assumes that the web traffic of nonmembers constitutes a valid benchmark for the development of web traffic of VGM members. This assumption could be violated if certain outlets selected into VGM after the passage of the copyright bill and if the web traffic of those who decided to join the association would have developed differently from the web traffic of those who did not. Several arguments speak against such concerns, though.

First, selection into the treatment is implausible. *VGM* was already founded in 1997 for unrelated reasons, and many members joined before the copyright bill came into action. While its members commissioned *VGM* with the enforcement of their claims against news aggregators around November 2013, that is, following the passage of the bill (Hirche, 2013), this was long before their removal from the news aggregators took place. More importantly, the removal of *VGM* members' content from the news aggregators was not anticipated and by no means an active choice of *VGM* (i.e., they did not choose to receive the treatment of being removed), ruling out concerns about potential selection effects. In fact, if the content of *VGM* members is removed from news aggregators, this makes it impossible (by definition) to collect royalty fees, and thus works against the initial objective of *VGM* members. In other words, the unexpected removal of content from the news aggregators can be considered as an exogenous shock to *VGM* members' web traffic that lets us identify the causal effect of being removed from the platform.

Second, anticipation effects on behalf of the outlets—that is, changes in the development of web traffic even before March 2013, in anticipation of the passage of the copyright bill—are unlikely. As argued in Section 4.1, 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. These unexpected changes prevent that the behavior of outlets and news aggregators was affected even before the Bill was passed. Thus, the observations from before March 2013 do not exhibit any anticipation effects and thereby constitute a valid benchmark for the development of web traffic after March 2013 in our (triple) difference-in-differences estimation. Relatedly, we can use the time period from before March 2013 to document that the web traffic of *VGM* members and nonmembers was indeed on parallel trends before the Copyright Bill was passed (see Appendices A.1.2 and A.1.3 below). To prevent potential anticipation effects that may have occurred between March 2013 and the ultimate removal of *VGM* members from the news aggregators in August 2014, we discard the 17 months in between (see Appendix A.2.5 for further discussion and for robustness checks that include this time period).

A.1.2. | Placebo regressions

Next, we conduct two types of placebo regressions to validate our triple difference-in-differences estimation. First, we conduct a series of placebo regressions to document that the web traffic of *VGM* members and nonmembers was indeed on parallel trends before the Copyright Bill was passed, and that this development was parallel for *VGM* members and nonmembers of similar scale and scope. To this end, we augment Equation (4) to

$$\log(Visits)_{it} = \alpha_1(VGM_i * FakePost_t) + \alpha_2(S_i * FakePost_t) + \alpha_3(VGM_i * S_i * FakePost_t) + \varphi_i + \lambda_t + \varepsilon_{it} \mid Post_t = 0, \tag{A.1}$$

where in the first placebo regression $FakePost_t$ is a dummy variable equal to one for all months after December 2011, in the second placebo regression $FakePost_t$ is a dummy variable equal to one for all months after January 2012, and so on—in sum, we conduct 15 placebo regressions using scale and scope as moderator, respectively. The idea is that while the web traffic of *VGM* members and nonmembers has developed differently *after* the content of *VGM* members was removed from the news aggregators, these differences should not have occurred *before* the Copyright Bill was passed. Hence, the coefficient for α_1—that is, the coefficient for the interaction term of our *VGM* indicator and the $FakePost_t$ dummies, which measures differences in the development of web traffic between *VGM* members and nonmembers after the respective date—should be close to zero and statistically insignificant in all placebo regressions. Similarly, *VGM* members and nonmembers of specific scale and scope should not exhibit diverging trends in web traffic before March 2013, so all coefficients for α_3 should also be close to zero and statistically insignificant in all placebo regressions. Note that it is crucial to consider only observations from *before* March 2013 here. Otherwise, the placebo regressions might pick up part of the true treatment effect, which would prevent a clean check of parallel trends in web traffic. Note also that our (triple) difference-in-differences framework requires parallel *trends* only. In other words, the web traffic of *VGM* members and nonmembers may differ in terms of levels as long as their *development* is parallel, which is exactly what α_1 and α_3 are capturing.

Table A.1 shows the coefficients and standard errors for α_1 and α_3 in our two times 15 placebo regressions. The coefficients are several times smaller than their counterparts in Table 2, and none of them is statistically significant, which confirms that the web traffic of *VGM* members and nonmembers was indeed on parallel trends before the content of *VGM* members was removed from the news aggregators.

As a second type of placebo regression, we estimate the original triple difference-in-differences Equation (4), but replace the dependent variable $\log(Visits)_{it}$ with $\log(Copies)_{it}$. The idea is that news outlets' online and offline audiences are very distinct, that is, readers of an outlet's print version typically do not consume its online version (BDZV, 2022; Skogerbø &

Winsvold, 2011; Thurman, 2018); moreover, many outlets had separate online and print editorial offices during our observation period (Fischer, 2018; Fricker, 2022). Consequently, an outlet's number of *printed copies* should be unaffected if its online content is removed from the news aggregators' platforms. Any divergence in the development of printed copies between *VGM* members and nonmembers of different scale and scope after the passage of the copyright bill in March 2013 could indicate that there are unobserved differences between *VGM* members and nonmembers though, in which case nonmembers could not serve as a valid control group. However, Table A.2 confirms that all coefficients for our triple interaction terms are close to zero and not statistically significant when we use $\log(Copies)_{it}$ as dependent variable in Equation (4), which further supports the validity of our empirical strategy.

A.1.3. | Event studies

Finally, we support the validity of our triple difference-in-differences estimation with a series of event studies. Similar to the first type of placebo regressions (i.e., with the original dependent variable $\log(Visits)_{it}$), the idea is to document that there were no differences in the development of web traffic between *VGM* members and nonmembers before the Copyright Bill was passed, but that the development diverged after the content of *VGM* members was removed from the news aggregators. To this end, we augment Equation (4) to

$$\log(Visits)_{it} = \sum\nolimits_{t=1}^{17} \gamma_{1,t}(VGM_i * Pre_t) + \sum\nolimits_{t=19}^{36} \gamma_{1,t}(VGM_i * Post_t) + \gamma_2(S_i * Pre_t) + \gamma_3(S_i * Post_t) + \sum\nolimits_{t=1}^{17} \gamma_{4,t}(VGM_i * S_i * Pre_t) + \sum\nolimits_{t=19}^{36} \gamma_{4,t}(VGM_i * S_i * Post_t) + \varphi_i + \lambda_t + \varepsilon_{it}. \tag{A.2}$$

The idea is as follows. In Equation (A.2), we replace the indicator $Post_t$ from Equation (4) with a series of monthly dummies, using the month just before the Copyright Bill was passed (February 2013, $t = 18$) as baseline. In other words, we interact the indicator for *VGM* membership, VGM_i, as well as the moderating term $(VGM_i * S_i)$ with a monthly dummy for each month before the Copyright Bill was passed (17 monthly dummies in total), and with a monthly dummy for each month after the content of *VGM* members was removed from the news aggregators (18 monthly dummies in total). We can thereby interpret each coefficient $\hat{\gamma}$ for each of these novel interaction terms as the impact of *VGM* membership, moderated by scale or scope, on the development of web traffic relative to the baseline month of February 2013. If there was no difference in the development of web traffic of *VGM* members and nonmembers (of similar scale and scope) before the Copyright Bill was passed, all coefficients for $\gamma_{\cdot,t}, t \leq 17$ must be close to zero and statistically insignificant. In contrast to that, we expect all coefficients for $\gamma_{\cdot,t}, t \geq 19$ to be unequal to zero and statistically significant, since our main results show that the removal of *VGM* members' content from the news aggregators did affect their web traffic, and this effect was moderated by their scale and scope.

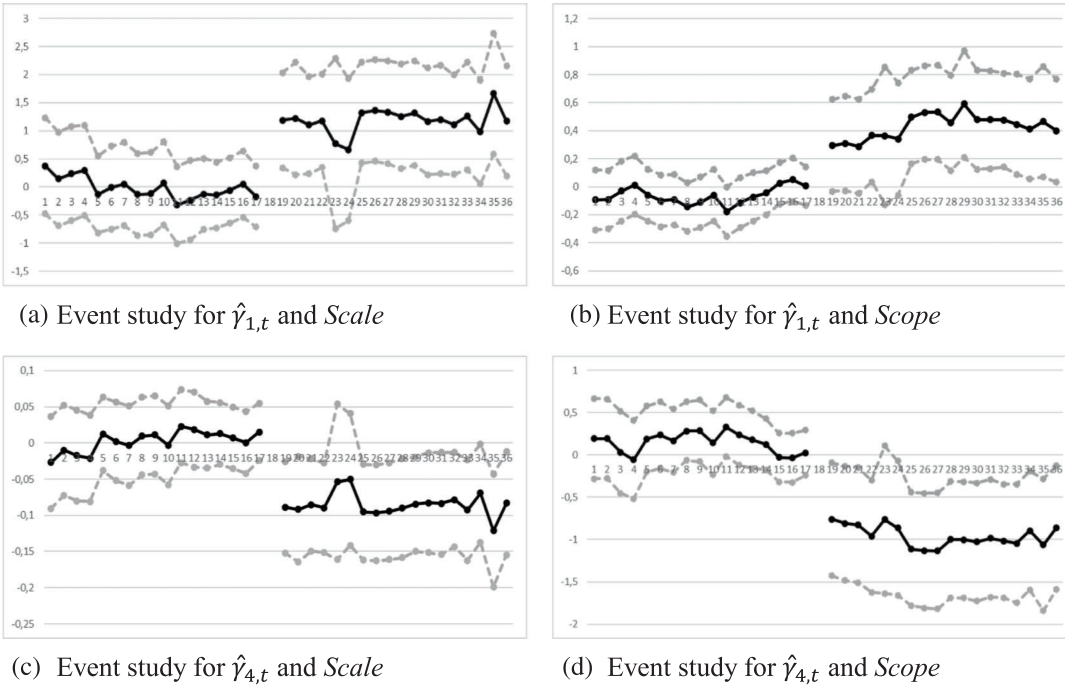
Figure A.1a,b display the coefficients for $\gamma_{1,t}$, that is, for the baseline interaction term. Analogously, Figure A.1c,d displays the coefficients for $\gamma_{4,t}$, that is, the triple interaction term that captures the moderating impact of scale and scope. In each graph, the black dots connected by a solid line depict the point estimates, and the gray dots connected by dashed lines depict a 95%

TABLE A.1 Placebo regressions—fake treatment.

	α_1^{Scale}	α_3^{Scale}	α_1^{Scope}	α_3^{Scope}
Fakepost(t≥3)	−0.037	0.002	−0.026	0.102
	(0.238)	(0.018)	(0.072)	(0.175)
Fakepost(t≥4)	−0.159	0.012	−0.049	0.156
	(0.245)	(0.019)	(0.074)	(0.181)
Fakepost(t≥5)	−0.125	0.009	−0.029	0.091
	(0.206)	(0.016)	(0.068)	(0.161)
Fakepost(t≥6)	−0.108	0.008	−0.008	0.041
	(0.192)	(0.015)	(0.066)	(0.153)
Fakepost(t≥7)	−0.065	0.004	−0.009	0.012
	(0.188)	(0.014)	(0.064)	(0.146)
Fakepost(t≥8)	0.003	−0.000	0.011	−0.024
	(0.194)	(0.015)	(0.061)	(0.140)
Fakepost(t≥9)	0.037	−0.003	0.023	−0.028
	(0.202)	(0.015)	(0.060)	(0.136)
Fakepost(t≥10)	0.096	−0.008	0.022	−0.075
	(0.120)	(0.015)	(0.061)	(0.141)
Fakepost(t≥11)	0.139	−0.011	0.050	−0.155
	(0.120)	(0.015)	(0.062)	(0.144)
Fakepost(t≥12)	0.221	−0.017	0.089	−0.155
	(0.203)	(0.015)	(0.063)	(0.146)
Fakepost(t≥13)	0.240	−0.018	0.097	−0.177
	(0.203)	(0.015)	(0.063)	(0.147)
Fakepost(t≥14)	0.300	−0.023	0.104	−0.201
	(0.205)	(0.016)	(0.064)	(0.147)
Fakepost(t≥15)	0.300	−0.023	0.088	−0.164
	(0.216)	(0.017)	(0.064)	(0.155)
Fakepost(t≥16)	0.147	−0.012	0.058	−0.120
	(0.222)	(0.017)	(0.068)	(0.153)
Fakepost(t≥17)	0.189	−0.015	0.055	−0.117
	(0.223)	(0.017)	(0.068)	(0.150)

Note: Heteroskedasticity-robust standard errors in parentheses. All standard errors are clustered on the outlet level (140 clusters). Each line represents the coefficients from two placebo regressions using scale and scope as moderators, respectively. All placebo regressions are: based on observations before March 2013.

confidence interval. In each case, the coefficients for the interaction terms using monthly dummies from before the passage of the Copyright Bill are close to zero and not statistically significant. In contrast to that, all monthly interactions after removing the content of *VGM* members are unequal to zero, most of them are statistically significant at the 5%-level, and their sign is in line with our main results in Table 2. We can thus conclude that the event studies support our triple difference-in-differences estimation in that the web traffic of *VGM* members



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

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

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

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

FIGURE A.1 Event studies.

(of similar scale and scope) was indeed on parallel trends before the Copyright Bill was passed, and only diverged afterwards.

A.2 | ROBUSTNESS CHECKS

A.2.1. | Alternative dependent variable

We use monthly page impressions (as opposed to visits) as an alternative dependent variable (PIs_{it}). If a user accesses three articles of one outlet, *IVW* counts this as three page impressions but only one visit. If a user accesses two articles, leaves the outlet's domain for at least 30 min, and returns to access another article, *IVW* would count three page impressions and two visits. Again, we employ the logarithm of page impressions. Table A.3 shows that our results remain unchanged to Table 2 if we use this alternative dependent variable. All estimates of interest carry the expected sign (*coeff.* = −0.158 and *coeff.* = −1.074 in columns (2) and (3), respectively) and are statistically significant ($p = .000$ and $p = .017$ in columns (2) and (3), respectively). We also obtain similar results to Table 2 when including both independent variables of interest in the same model (*coeff.* = −0.118; $p = .002$ and *coeff.* = −0.696; $p = .119$ in column (4)).

A.2.2. | Alternative independent variables

A.2.2.1. | Scale

To probe the robustness of our main findings, we use four alternative measures of outlet scale. The first one is the number of counties ($Counties_i$) that represent an outlet's (offline)

TABLE A.2 Placebo regressions—printed copies.

	(1)	(2)	(3)
	log(Copies)	**log(Copies)**	**log(Copies)**
Post * VGM	0.0878	0.0169	0.0395
	(0.320)	(0.578)	(0.647)
Post * log(avgVisits)	−0.0085		−0.0086
	(0.161)		(0.157)
Post * VGM * log(avgVisits)	−0.0078		−0.0027
	(0.268)		(0.694)
Post * Inv_Herfindahl		−0.0331	−0.0344
		(0.519)	(0.512)
Post * VGM * Inv_Herfindahl		−0.103	−0.0515
		(0.102)	(0.415)
Outlet FE	X	X	X
Time FE	X	X	X
Constant	10.96	10.92	10.97
	(0.000)	(0.000)	(0.000)
N	4553	4553	4553
R^2	.999	.999	.999

Note: The *p*-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the outlet level (140 clusters).

distribution area. The second one is the monthly mean circulation ($Circulation_i$) before March 2013, that is, the average number of an outlet's physical copies sold per month before March 2013. To account for skewness, we employ their logarithms in the empirical analysis.

Furthermore, the moderating effect of 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 an outlet's relative scale into account, we define a competitor *k* of outlet *i* as an outlet whose distribution area overlaps with the distribution area of *i* by at least one county. Then, we compute the scale of the focal outlet *i*'s competitors as $\sum_k scale_k$ and set it into relation to the scale of *i*.[31] This gives us our third alternative measure of scale:

$$RelVisits_i = \frac{scale_i}{\sum_k scale_k}. \tag{A.3}$$

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

Finally, an outlet's competitors may be close or distant. For instance, an outlet that predominantly covers political news will consider a competing outlet also specialized on political news

[31]Results are similar if we take the average or the largest competitor instead of the sum of competitors of outlet *i*.

TABLE A.3 Alternative dependent variable—page impressions.

	(1)	(2)	(3)	(4)
	log(PIs)	log(PIs)	log(PIs)	log(PIs)
Post * VGM	0.046	2.169	0.516	1.936
	(0.444)	(0.000)	(0.018)	(0.001)
Post * log(avgVisits)		0.030		0.030
		(0.349)		(0.351)
Post * VGM * log(avgVisits)		−0.158		−0.118
		(0.000)		(0.002)
Post * Inv_Herfindahl			0.012	0.023
			(0.976)	(0.953)
Post * VGM * Inv_Herfindahl			−1.074	−0.696
			(0.017)	(0.119)
Outlet FE	X	X	X	X
Time FE	X	X	X	X
Constant	14.69	14.50	14.69	14.50
	(0.000)	(0.000)	(0.000)	(0.000)
N	4560	4560	4560	4560
R^2	.975	.976	.976	.977

Note: The p-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the outlet level (140 clusters).

as a closer competitor than an outlet predominantly covering sports. To reflect this, we weigh the scale of each competitor k of outlet i by the extent to which their category compositions overlap and then compute our fourth alternative measure of outlet scale:

$$RelWghtVisits_i = \frac{scale_i}{\sum_k overlap_k * scale_k}. \tag{A.4}$$

The results we obtain with our alternative measures of scale can be seen in Columns (1) to (4) of Table A.4. As in the previous case, we find that all estimates for β_3 are negative and statistically significant (p-values between .001 and .015), thus supporting Hypothesis 1. Again, we find that the total effect of being removed grows (i.e., becomes more negative) if outlet scale increases, while it approaches zero and eventually flips its sign if outlet scale decreases.

A.2.2.2. | Scope

We first use the entropy measure (Jacquemin & Berry, 1979; Palepu, 1985) as an alternative for the Inverse Herfindahl diversity measure we use to measure scope in our main specification. We compute the relative number of page impressions by outlet for each of the eight main outlet

TABLE A.4 alternative independent variables.

	(1) log (Visits)	(2) log (Visits)	(3) log (Visits)	(4) log (Visits)	(5) log (Visits)	(6) log (Visits)
Post * VGM	0.206	1.266	0.051	0.051	0.482	1.241
	(0.028)	(0.008)	(0.338)	(0.337)	(0.015)	(0.016)
Post * log(Counties)	0.017					
	(0.696)					
Post * VGM * log(Counties)	−0.126					
	(0.015)					
Post * log(Circulation)		0.047				
		(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)	
Post * log(Count)						0.334
						(0.130)
Post * VGM * log(Count)						−0.630
						(0.016)
Outlet FE	X	X	X	X	X	X
Time FE	X	X	X	X	X	X
Constant	13.09	12.86	13.09	13.09	13.04	12.78
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	4472	4560	4436	4436	4560	4560
R^2	.975	.975	.975	.975	.975	.975

Note: The *p*-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the outlet level (140 clusters).

TABLE A.5 Main results—with outlet time trends.

	(1)	(2)	(3)	(4)
	log(Visits)	log(Visits)	log(Visits)	log(Visits)
Post * VGM	−0.001	1.192	0.488	1.128
	(0.981)	(0.009)	(0.009)	(0.021)
Post * log(avgVisits)		−0.007		−0.005
		(0.813)		(0.848)
Post * VGM * log(avgVisits)		−0.087		−0.055
		(0.012)		(0.092)
Post * Inv_Herfindahl			0.266	0.262
			(0.427)	(0.438)
Post * VGM * Inv_Herfindahl			−1.094	−0.824
			(0.004)	(0.030)
Outlet FE	X	X	X	X
Time FE	X	X	X	X
Outlet-specific time trend	X	X	X	X
Constant	13.25	13.36	13.19	13.26
	(0.000)	(0.000)	(0.000)	(0.000)
N	4560	4560	4560	4560
R^2	.974	.975	.975	.976

Note: The *p*-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the outlet level (140 clusters).

content categories in our data and denote these fractions as $c_{i,k}$. Entropy is calculated as follows:

$$Entropy_i = \sum_{k=1}^{8} c_{i,k} \ln\left({}^{1}\!/\!_{c_{i,k}} \right). \tag{A.5}$$

Column (5) in Table A.4 shows that our results are similar to the ones we obtained in Column (3) of Table 2, that is, our main specification.

Second, as an alternative (coarser) measure of scope, we consider the number of categories an outlet covers (Hashai, 2015). An outlet covering many different topics ranging from, say, political news to celebrities is broader in scope than an outlet covering a single topic (e.g., sports). We use (the log of) a simple count of the number of an outlet's active categories:

$$Count_i = Count\left(c_{i,j} \neq 0\right). \tag{A.6}$$

In Column (6) in Table A.4 we see that our results are robust to this alternative specification.

TABLE A.6 Outlet × quarter of the year fixed effects.

	(1)	(2)	(3)	(4)
	log(Visits)	log(Visits)	log(Visits)	log(Visits)
Post * VGM	0.150	1.417	0.500	1.323
	(0.000)	(0.004)	(0.013)	(0.011)
Post * log(avgVisits)		0.00271		0.00319
		(0.927)		(0.913)
Post * VGM * log(avgVisits)		−0.104		−0.0700
		(0.004)		(0.042)
Post * Inv_Herfindahl			0.211	0.211
			(0.557)	(0.557)
Post * VGM * Inv_Herfindahl			−1.112	−0.825
			(0.007)	(0.041)
Outlet FE	X	X	X	X
Time FE	X	X	X	X
Outlet × quarter FE	X	X	X	X
Constant	13.003 (0.000)	12.878 (0.000)	12.886 (0.000)	12.882 (0.018)
N	4560	4560	4560	4560
R^2	.147	.273	.276	.287

Note: The *p*-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the outlet level (140 clusters).

A.2.3. | Time trends

A potential concern might be that the web traffic of outlets follows individual time trends, for example, because some unobserved characteristics such as quality, style, or political orientation vary over time. To rule out this possibility, we extended the difference-in-differences specification described in Equations (3) and (4) by adding outlet-specific time trends. It can be seen from Table A.5 that our findings remain unchanged compared with the main results displayed in Table 2.

Relatedly, the impact of seasonality might vary on the outlet level, in which case it would not be well captured by the monthly and outlet fixed effects. We thus generate outlet × quarter of the year fixed effects and add them to Equations (3) and (4). Table A.6 shows that the estimates are similar to our main results in Table 2, and the standard errors tend to be even smaller.

A.2.4. | Alternative time windows

To confirm that the results in Table 2 do not depend on a specific time window, we run two further robustness checks. First, we shorten the observation periods to 12 months before March 2013 and after August 2014. Second, we extend it to 24 months before March 2013 and after August 2014. Tables A.7 and A.8 show that our results are qualitatively unchanged.

TABLE A.7 Narrower time window.

	(1)	(2)	(3)	(4)
	log(Visits)	log(Visits)	log(Visits)	log(Visits)
Post * VGM	−0.001	1.188	0.469	1.128
	(0.984)	(0.008)	(0.008)	(0.016)
Post * log(avgVisits)		−0.012		−0.011
		(0.655)		(0.677)
Post * VGM * log(avgVisits)		−0.087		−0.057
		(0.009)		(0.072)
Post * Inv_Herfindahl			0.220	0.216
			(0.486)	(0.500)
Post * VGM * Inv_Herfindahl			−1.054	−0.754
			(0.004)	(0.036)
Outlet FE	X	X	X	X
Time FE	X	X	X	X
Constant	13.10	13.17	13.05	13.12
	(0.000)	(0.000)	(0.000)	(0.000)
N	3051	3051	3051	3051
R^2	.972	.973	.973	.973

Note: In *p*-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the outlet level (140 clusters).

A.2.5. | Alternative treatment periods

As argued above, we discard the 17 months in between the passage of the Copyright Bill in March 2013 and the ultimate removal of *VGM* members' content from the news aggregators in August 2014 from our main analysis to measure the causal effect on *VGM* members' web traffic as cleanly as possible. First, it is unclear whether these 17 months should be considered as pre- or as post-treatment. On the one hand, *VGM* members' content was not removed immediately after the passage of the Bill. On the other hand, it is possible that *VGM* members or news aggregators anticipated the upcoming legal dispute and adapted their behavior even before the ultimate content removal in August 2014. Second, it is not clear when exactly the news aggregators removed the content of *VGM* members from their platforms—the only available information is that all of them had definitely done that by August 2014 (Athey et al., 2021; Kruse, 2014). Excluding observations from March 2013 to August 2014 has thus the advantage that potential heterogeneity in terms of staggered removal of *VGM* members' content is of no concern.

To demonstrate that our main results do not depend on the decision to entirely discard the time period in between March 2013 and August 2014, we conduct two robustness checks where we include these observations into the empirical analysis. First, we add the time period from March 2013 to August 2014 to our original post-treatment observations and estimate the average effect of being delisted from the news aggregators for this extended time period. Second, we consider two distinct post-treatment periods, where the first corresponds to March 2013 to

TABLE A.8 Broader time window.

	(1)	(2)	(3)	(4)
	log(Visits)	log(Visits)	log(Visits)	log(Visits)
Post * VGM	−0.0230	0.455	0.547	1.332
	(0.706)	(0.395)	(0.010)	(0.019)
Post * log(avgVisits)		−0.0566		0.00148
		(0.109)		(0.965)
Post * VGM * log(avgVisits)		−0.0342		−0.0707
		(0.389)		(0.079)
Post * Inv_Herfindahl			0.359	0.360
			(0.352)	(0.354)
Post * VGM * Inv_Herfindahl			−1.216	−0.922
			(0.006)	(0.035)
Outlet FE	X	X	X	X
Time FE	X	X	X	X
Constant	13.29	13.65	13.00	12.99
	(0.000)	(0.000)	(0.000)	(0.000)
N	7092	6946	5927	5927
R^2	.969	.974	.970	.970

Note: The *p*-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the outlet level (140 clusters).

August 2014, and the second corresponds to the 18 months after that, that is, to our original post-treatment observation period. This specification allows us to document potential effects on web traffic that already occurred before August 2014 separate from the effect on web traffic due to the removal of *VGM* members' content from the news aggregators afterwards.

Tables 3 and 4 show the results. In Table 3, we have added the formerly discarded months to our original post-treatment period and estimate the average effect of being delisted from news aggregators based on that. Plausibly, we find that the estimates are qualitatively similar but smaller than our main results in Table 2: Given that major changes in web traffic are unlikely to occur between March 2013 and August 2014, considering these observations as post-treatment drags the average treatment effect closer towards zero.

Table 4 supports this line of thought. In particular, the estimates for the first post-treatment period (March 2013 to August 2014) are much smaller, but qualitatively in line with the effects from the second post-treatment period after *VGM* members' content was ultimately removed from the news aggregators. This confirms our concerns that the web traffic of *VGM* members was indeed affected even before August 2014 (but not before March 2013, see Appendix A.1) and supports our decision to discard observations from March 2013 to August 2014 from our main analysis.

A.2.6. | Scale and scope quintiles

In our main triple difference-in-differences equation, we use continuous measures for outlet scale and scope, respectively. This specification allows us to estimate the average marginal impact of outlet scale and scope on the effect of being delisted from news aggregators, that is, we can interpret the coefficients as the average change in web traffic if scale and scope increase by 1%, respectively. Alternatively, one could consider the distribution of outlet scale and scope, identify specific parts of it (e.g., quartiles or quintiles), and generate a series of dummy variables that indicate if an outlet belongs to a specific part of the distribution of scale and scope or not. Using these dummy variables instead of the continuous measures for scale and scope in Equation (4) would then give us the average impact of belonging to a certain part of the scale or scope distribution on the change in web traffic, relative to a baseline category (e.g., the downmost quartile or quintile). The advantage of this specification is that it can better capture potential non-linearities in the impact of scale and scope, as we do not estimate one single average effect for the entire sample of outlets but obtain distinct coefficients for outlets from different parts of the scale and scope distribution. On the other hand, using dummy variables prohibits a marginal interpretation of the coefficients for outlet scale and scope. Moreover, we would implicitly assume that the impact of scale and scope is equivalent for all outlets within a specific part of the distribution, which is especially problematic if the partition is coarse. While it is principally possible to mitigate that problem by generating a relatively large number of dummy variables, we are somewhat constrained by our small sample size. We therefore decided to generate dummy variables that indicate *quintiles* of the distribution of outlet scale and scope, respectively, and estimate Equation (4) using these indicators instead of the continuous measures.[32]

Table 5 shows the results. In line with the results from our main specification, we find that outlets that are large in scale and broad in scope lose web traffic when their content is being removed from the news aggregators. Specifically, we find that the coefficients for all triple interaction terms using our quintile indicators are negative, and most of them are statistically significant. In contrast to that, the coefficients for the baseline (downmost) quintiles are positive and statistically significant. Thus, analogous to our main results, we find that outlets that are relatively small in scale and narrow in scope benefit from being removed from the news aggregators, while outlets that are large in scale and broad in scope suffer. As before, the effect prevails when we consider scale and scope simultaneously (Column 4 in Table 5). Note, however, that we cannot directly compare the magnitude of the coefficients in Table 5 to our main results in Table 2. In particular, while the coefficients for the triple interaction terms in Table 2 measure the average change in web traffic if scale and scope increase by 1%, the coefficients for the triple interaction terms in Table 5 must be interpreted as the average percentage change in web traffic for outlets in a specific quintile of the distribution of scale and scope relative to outlets in the downmost quintile.

A.3 | FURTHER ANALYSES AND ADDITIONAL INFORMATION

A.3.1. | Political position

The political position of outlets could mitigate or reinforce the moderating effects of scale and scope. In particular, outlets that position themselves at the center of the political spectrum

[32]The results are qualitatively similar when we use terciles or quartiles.

could be better able to attract readers' attention when they are listed by news aggregators than outlets which position themselves more at the extremes and thus fail to match the political preferences of a wide audience. In other words, a more extreme political position could counteract the negative impact of scale and scope when the content of outlets is being removed from the news aggregators, while a position at the center of the political spectrum might reinforce it.

To study if this is indeed the case, we leverage information from Garz et al. (2020), who provide the most sophisticated dataset on the political position of German media outlets till date, including 37 outlets that we consider, too. Specifically, Garz et al. (2020) compute a score ranging from −1 to 1 for each outlet, where more negative values indicate a political position that is more to the left, and more positive values indicate a political position that is more to the right of the political spectrum. Since we are interested in whether an outlet positions itself at the center or at the extremes of the political spectrum (and not in the left- to right-dimension), we consider the *absolute* values of that score. Thus, an absolute score near zero indicates that an outlet is positioned close to the center, and growing absolute scores indicate positions more at the extremes of the political spectrum. Given that a four-way-difference-in-differences estimation on 37 outlets is not advisable from an econometric perspective, we split the subsample at the median value of the absolute score and conduct our main analyses on each of those two subsamples, respectively.

Tables A.9 and A.10 show the results. As expected, we find that the negative impact of outlet scale and scope when being delisted from news aggregators is reinforced when we consider only outlets close to the center of the political spectrum (Table A.9), and that it is mitigated when we consider only outlets that position themselves more at the extremes (Table A.10). In particular, the point estimates that measure the moderating effect of outlet scale and scope are larger than their counterparts in Table 2 and statistically significant at the 10%- or at the 5%-level despite the small sample size. In contrast to that, the respective point estimates in Table A.10 are smaller than their counterparts in Table 2 and not statistically significant. However, we must interpret these results with great care. First, we consider a small and selected subsample of outlets. In particular, the outlets on which information on their political orientation is available are on average 10 times larger (in terms of web traffic) than the outlets for whom such data do not exist. Second, we find that within the subsample of 37 outlets that we consider here, political orientation is negatively correlated with outlet size ($-0.23, p<.000$), that is, the larger outlets are plausibly also those that position themselves at the center of the political spectrum. In contrast to that, we find just a small correlation between outlet scope and political orientation ($-0.07, p<.013$).

A.3.2. | Google news dispute

In October and November 2014, *VGM* had a two-week dispute with *Google News*, which, as a result, temporarily delisted the content of *VGM* members (Calzada & Gil, 2020). To study if this event had an additional detrimental impact on the web traffic of *VGM* members relative to nonmembers, we generate a dummy variable equal to one for October and November 2014, interact it with our *VGM* indicator as well as with the interaction of *VGM* and outlet scale or scope, respectively, and add these interaction terms to our original triple difference-in-differences specification. Thus, the novel interaction terms can be interpreted as the additional effect of the *Google News* dispute on *VGM* members' web traffic.

Table A.11 shows that—in line with and supporting the results by Calzada and Gil (2020)—the short removal of *VGM* members' content from *Google News* did not systematically reduce their web traffic. In particular, we observe that VGM members' web traffic decreased on average during the time of the *Google* dispute (Column 1), but the effect shrinks and becomes statistically insignificant when we include outlet scale and scope as moderators, respectively. Crucially, controlling for the *Google News* dispute does not deteriorate our main results. Rather, our coefficients of interest tend to become even larger and more statistically significant than the main results in Table 2. In addition to that, Calzada and Gil (2020) report that the short removal from *Google News* only had a negative impact on outlets owned by the publisher *Axel Springer*. These are primarily national outlets (e.g., *WELT*, *BILD*) and therefore not part of our sample. However, national outlets tend to be large in scale and broad in scope, whereby the results by Calzada and Gil (2020) are in accordance to what we find in the context of being removed from smaller news aggregators in Germany.

A.3.3. | Application of the formal model across different settings

In this section, we present eight different settings in which our model could be used to describe consumer choice. We describe the exact specifications of our model in each of the settings in Table A.12.

A.3.3.1. | Setting 1: Two stands selling ice cream (off-platform)

In a simple (off-platform) scenario with two ice cream stands that each sell one type of ice cream, we can think of each ice cream stand as serving a separate submarket. Consumers in each submarket are only aware of the respective ice cream in that submarket and vary in their preference for that ice cream. While ice cream stands may differ in their ability to draw attention, the main driver of attraction will be characteristics-based attraction v_i (i.e., β will be low), as ice cream characteristics can be observed ex-ante. Scale and scope (which drive complementors' ability to draw attention in the context of news outlets) do not play a major role, as characteristics are the main driver of attraction. Off-platform, all consumers in a given submarket will consume ice cream from the corresponding ice cream stand, as this is the only option available to them.

A.3.3.2. | Setting 2: Two stands selling ice cream (on-platform)

If the previous setting is moved on-platform, consumers in each submarket will also become aware of the other ice cream (i.e., there are now two options per submarket). Their choice however will still be mainly driven by ice cream characteristics (i.e., v_i). Ice cream stand A might gain additional consumers, as consumers who were only aware of stand B are now also aware of A, or vice versa. Ice cream stands that generate more characteristics-based attraction v_i (e.g., through better quality or bigger portion sizes) will be better able to capture additional demand compared with the off-platform setting.

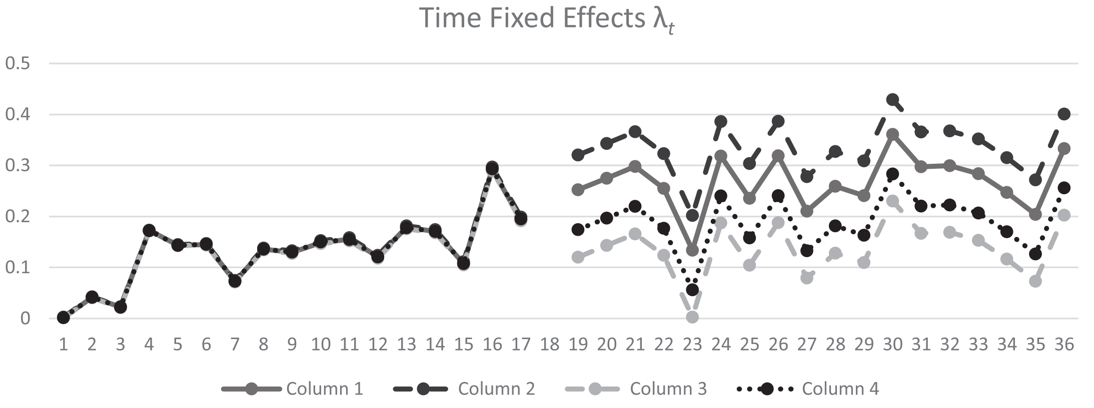


FIGURE A.2 Plotted time fixed effects from our main analyses in Table 2.

TABLE A.9 Political orientation—center outlets.

	(1) log(Visits)	(2) log(Visits)	(3) log(Visits)	(4) log(Visits)
Post * VGM	−0.106	2.495	0.842	3.695
	(0.470)	(0.074)	(0.088)	(0.057)
Post * log(avgVisits)		−0.123		−0.0751
		(0.135)		(0.198)
Post * VGM * log(avgVisits)		−0.167		−0.244
		(0.097)		(0.058)
Post * Inv_Herfindahl			0.650	0.356
			(0.370)	(0.672)
Post * VGM * Inv_Herfindahl			−1.908	−0.157
			(0.028)	(0.879)
Outlet FE	X	X	X	X
Time FE	X	X	X	X
Constant	14.87	15.76	14.72	15.33
	(0.000)	(0.000)	(0.000)	(0.000)
N	609	609	609	609
R^2	.964	.978	.973	.978

Note: The *p*-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the outlet level (17 clusters).

A.3.3.3. | Setting 3: Two stands, each selling ice cream and milkshake (off-platform)

If each stand sells not only ice cream but also milkshake, the number of submarkets will double to 4 (i.e., one for each combination of product and stand). While consumers who want to consume ice cream might also have some probability to be "distracted" and consume milkshake instead (and vice versa), this is unlikely to be a major driver of sales. Thus, all consumers in each submarket will only have one combination of product and stand to choose from, and consume this option.

TABLE A.10 Political orientation—fringe outlets.

	(1)	(2)	(3)	(4)
	log(Visits)	log(Visits)	log(Visits)	log(Visits)
Post * VGM	−0.0124	1.806	0.243	1.341
	(0.894)	(0.375)	(0.411)	(0.514)
Post * log(avgVisits)		−0.0170		0.00138
		(0.862)		(0.989)
Post * VGM * log(avgVisits)		−0.125		−0.0842
		(0.376)		(0.579)
Post * Inv_Herfindahl			−0.326	−0.329
			(0.270)	(0.308)
Post * VGM * Inv_Herfindahl			−0.455	−0.217
			(0.399)	(0.729)
Outlet FE	X	X	X	X
Time FE	X	X	X	X
Constant	14.38	14.51	14.46	14.45
	(0.000)	(0.000)	(0.000)	(0.000)
N	700	700	700	700
R^2	.950	.954	.954	.955

Note: The *p*-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the outlet level (20 clusters).

A.3.3.4. | Setting 4: Two stands, each selling ice cream and milkshake (on-platform)

If the previous setting is moved on-platform, consumers in each submarket will become aware of additional options and thus have two options to choose from: either the two ice creams offered by *A* and *B*, respectively, or the two milkshakes offered by *A* and *B*. The same process as in *Setting 2* will unfold in each of the four submarkets. Ice stands may gain or lose consumers (compared with the off-platform setting), depending on the characteristics-based attraction v_i their ice cream and milkshake can generate.

A.3.3.5. | Setting 5: Two bookstores, each selling one different novel and political nonfiction (off-platform)

Much like *Setting 3*, this setting will consist of 4 submarkets, one for each combination of book and store, where consumers only have one option to choose from (and choose that option) in each submarket. The main difference compared with *Setting 3* is that attraction will now be generated mainly by attention-based attraction a_i (i.e., β will be high), as consumers are not able to assess product characteristics before the purchase (i.e., books are experience goods). Moreover, since consumers now have much less information about the product (i.e., the book) to base their decision-making on, the role of bookstores in the decision-making process becomes relatively more important, as they may provide additional cues. For instance, books

TABLE A.11 Google news dispute.

	(1)	(2)	(3)	(4)
	log(Visits)	log(Visits)	log(Visits)	log(Visits)
Post * VGM	0.00837	1.248	0.507	1.185
	(0.868)	(0.008)	(0.007)	(0.017)
Google * VGM	−0.0834	−0.408	−0.154	−0.410
	(0.002)	(0.109)	(0.059)	(0.101)
Google * log(avgVisits)		−0.0284		−0.0282
		(0.104)		(0.105)
Google * VGM * log(avgVisits)		0.0258		0.0192
		(0.179)		(0.345)
log(avgVisits)		−0.00169		−0.000590
		(0.954)		(0.984)
Post * VGM * log(avgVisits)		−0.0909		−0.0580
		(0.011)		(0.087)
Google * Inv_Herfindahl			−0.0827	−0.0922
			(0.596)	(0.555)
Google * VGM * Inv_Herfindahl			0.155	0.203
			(0.383)	(0.280)
Post * Inv_Herfindahl			0.281	0.279
			(0.399)	(0.405)
Post * VGM * Inv_Herfindahl			−1.116	−0.857
			(0.004)	(0.024)
Outlet FE	X	X	X	X
Time FE	X	X	X	X
Constant	13.10	13.13	13.04	13.06
	(0.000)	(0.000)	(0.000)	(0.000)
N	4560	4560	4560	4560
R^2	.974	.975	.975	.976

Note: The *p*-values in parentheses. Heteroskedasticity-robust standard errors are clustered on the outlet level (140 clusters).

offered by bookstores that are larger in scale may be better able to generate attention-based attraction a_i. Thus, much like in the case of news aggregators, the bookstore will be determined first, and consumers then choose a specific book within that store.

A.3.3.6. | Setting 6: Two bookstores, each selling one different novel and political nonfiction (on-platform)

Moving on-platform again increases the number of options that consumers can choose from, as each of them will now be able to choose either between the two novels or between the two political nonfictions. Moreover, the importance of bookstores in driving consumer choice is

TABLE A.12 Application of the formal model across different settings.

Setting			Main driver of attraction	β	Relevance of complementor in generating attraction	No of submarkets (Hotelling lines)	No of options per submarket	Relevance of scale and scope
1	*2 stands selling ice cream*	*Off-platform*	Characteristics-based attraction v_i	Low	Low (ice cream stand)	2	1	Limited
2	*2 stands selling ice cream*	*On-platform*	Characteristics-based attraction v_i	Low	Low (ice cream stand)	2	2	Limited
3	*2 stands selling ice cream and milkshake*	*Off-platform*	Characteristics-based attraction v_i	Low	Low (ice cream/milkshake stand)	4	1	Limited
4	*2 stands selling ice cream and milkshake*	*On-platform*	Characteristics-based attraction v_i	Low	Low (ice cream/milkshake stand)	4	2	Limited
5	*2 bookstores selling novels and pol. nonfictions*	*Off-platform*	Attention-based attraction v_i	High	High (bookstore/ publisher)	4	1	High relevance of scale, no relevance of scope
6	*2 bookstores selling novels and pol. nonfictions*	*On-platform*	Attention-based attraction v_i	High	High (publisher)	4	2	High relevance of scale, no relevance of scope
7	*2 newsstands selling outlets with politics and sports*	*Off-platform*	Attention-based attraction v_i	High	High (news stand/news outlet)	4	1	High relevance of scale and scope
8	*2 newsstands selling outlets with politics and sports*	*On-platform*	Attention-based attraction v_i	High	High (news outlet)	4	2	High relevance of scale and scope

TABLE A.13 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

TABLE A.13 (Continued)

VG media members	Non-members
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 A.14 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

Note: The *IVW* changed its category classification in May 2014. Table A.14 illustrates how we matched categories before and after the change, and how we aggregated page impressions into eight main news outlet content categories. We use page impressions by outlet content category (as defined in the fourth column) to determine outlet scope.

reduced. While bookstores acted as a type of landing page that determines (and often limits) the choices available to consumers in the off-platform setting, consumers can now access all books directly through the platform. Instead, attention may now be generated by publishers, with larger publishers drawing more attention to their books. Readers' choices are however less likely to be affected by books offered by the same publisher in a different category (compared with the empirical setting in our paper), which means that the scope of publishers is less relevant.

A.3.3.7. | Setting 7: Two newsstands, each selling one different newspaper with two sections (one on politics and one on sports) (off-platform)

Much like in *Setting 5*, there are four submarkets in this setting, and consumer choice is driven by attention-based attraction. Again, the newsstand will be determined first and limit the subsequent choice of news content. While consumers may be interested in a specific content category, there is some probability that some of them are "distracted" and decide, for instance, to read sports content instead of politics content. In contrast to the bookstore-setting however, this will be less relevant for the publisher of the outlet as consumers will buy the same product (due to bundling) in either case.

A.3.3.8. | Setting 8: Two newsstands, each selling one different newspaper with two sections (one on politics and one on sports) (on-platform)

This setting closely resembles the setting in our paper. When moving on-platform, readers will not need to go through (the landing page of) the outlet anymore before accessing content. Content that was previously bundled by the outlet (i.e., politics and sports content) can now be accessed independently. Much like in our paper, we expect outlet scale and scope to affect how much attention (and thus readership) each outlet can attract (or lose).

A.3.4. | Additional information on the sample

Our sample includes 57 *VGM* members and 83 *VGM* nonmembers. An overview of the outlets in our sample can be found in Table A.13. To calculate our measure(s) of *Scope*, we look at the extent to which outlets are present in different content categories. An overview of how content is grouped into different categories can be found in Table A.14.