

August 2021

“Social Media and News: Content Bundling and
News Quality”

Alexandre de Cornière and Miklos Sarvary

Social Media and News: Content Bundling and News Quality

Alexandre de Cornière and Miklos Sarvary*

August 2021

Abstract

The growing influence of internet platforms acting as content aggregators is one of the most important challenges facing the media industry. We develop a simple model to understand the impact of third-party content bundling by a social platform that has a monopoly on showing user-generated content to consumers. In our model consumers can access news either directly through a newspaper's website, or indirectly through a platform, which also offers social content. We show that content bundling, when unilaterally implemented by the platform, tends to harm publishers and to increase the dispersion of quality across outlets, with initially high-quality outlets investing more and low-quality ones investing less. With many heterogeneous newspapers, the result is robust even if each newspaper can prevent the platform from using its content. When content bundling follows an agreement between the platform and publisher, its effects are reversed, as publishers' profits go up while quality dispersion goes down. In a setup with heterogeneous consumers, we also show that the platform's ability to personalize the mix of content it shows to users induces publishers to invest more in the quality of their content.

1 Introduction

With hundreds of millions of daily active users, a few large social networks have become the dominant online media outlets for most people. The largest among these, Facebook

*Alexandre de Cornière is Assistant Professor of Economics at the Toulouse School of Economics, Université Toulouse I Capitole, 1 esplanade de l'université, 31000 Toulouse, France. email: alexandre.de-corniere@tse-fr.eu, and Miklos Sarvary is the Carson Family Professor of Business at Columbia Business School, 3022 Broadway, New York, NY 10027-6902, USA. Tel.: +1 212 851-0165, email: miklos.sarvary@columbia.edu. We thank Charles Angelucci, Alessandro Bonatti, Guillermo Caruana, Anna D'Annunzio, Bruno Jullien, Juanjo Ganuza, Benjamin Ho, Gerard Llobet, Bang Nguyen, Markus Reisinger, Martin Peitz, Paul Seabright, Sandro Shelegia, Yossi Spiegel and participants at various seminars and conferences for helpful comments and discussions. We are grateful to the Net Institute (www.netinst.org) for financial support. De Cornière acknowledges funding from ANR under grant ANR-17-EUR-0010 (Investissements d'Avenir program), the TSE Digital Center, and the Artificial and Natural Intelligence Toulouse Institute (ANITI).

has reached over two billion active members across the globe who, on average, spend about an hour each day on the platform. In line with its significant consumer attention share, Facebook captured over \$80 billion worth of advertising revenues in 2020, corresponding to about one fifth of total worldwide digital advertising. Other successful social platforms include Tencent’s WeChat, Snap and ByteDance’s Tik Tok, among others. If, in their early days, social networks were mostly used as a way for users to share personal stories and pictures (which we refer to as user-generated content, or UGC, throughout the paper), their role has progressively evolved into one of content aggregation: an important share of the content displayed on their websites is produced by third-party publishers, who use the platforms as an alternative to their own website to reach consumers.

For many third-party content providers, large social networks represent an important distribution outlet. The news industry, in particular has been strongly affected by such ‘content bundling’ and given the industry’s public calling this impact has been relatively well documented.

In the U.S., 55% of adults got their news from social media “often” or “sometimes” in 2019, up from 44% in 2016 (Pew Research Center, 2019). Facebook alone accounts for a substantial part of traffic to publishers: a report by the U.K. Competition and Markets Authority puts the figure at 13%.¹ Social media is the second main gateway to news, behind direct traffic to publishers, with shares of 26% and 28% respectively (Reuters, 2020). For consumers aged 18-24, social media is the main gateway (38% against 16% for direct traffic, with search coming in second at 25%).

An important feature of the market is the curating role of platforms: the content seen by users is determined in large part by algorithms, whose design is a key strategic decision for the platform. In particular, a platform like Facebook can decide how much news to display on users’ newsfeed, and how much prominence to grant to each publisher.²

The central question that we ask in this paper is: how does content bundling by a platform affect publishers’ incentives to invest in the quality of their content? From an economic standpoint, the situation is a double-edged sword for publishers: while platforms allow newspapers to reach a large number of users, newspapers regularly complain that their business is hurt by social media, citing for instance the fact that Facebook’s market share (including Instagram) on the online display advertising market is around 40% while newspapers’ advertising revenues have plummeted. Another set of concerns is that platforms have “commoditized” news: in one survey, nearly half of U.S. respondents were

¹For mobile the figure is 17% with some heterogeneity among publishers, (see Competition and Markets Authority, 2020, Appendix S), but this number understates Facebook’s role, as a large number of consumers (47% in the U.K. in 2016) browse and read news on social media without clicking on links to access whole articles.

²Facebook reportedly uses a “news ecosystem quality score” to rank news publishers, and sometimes promotes certain articles to users who are particularly likely to share or discuss them (Roose and Dance, 2020). See Ershov and Morales (2021) for some empirical evidence on the effect of Facebook and Twitter’s algorithms on the sharing of news content.

unable to identify the source of a news story they had accessed through social media (Reuters, 2017), leading to fears of a phenomenon of brand dilution. Moreover, publishers also complain that platforms only give them access to aggregated data generated by consumers interacting with their content through the platforms, preventing them from offering better targeting services (Competition and Markets Authority, 2020, Appendix S).

For the above reasons, the general view is that content bundling has a negative effect on newspapers' incentive to invest in news quality. An often mentioned manifestation of lower news quality is the documented decline of newspapers' newsroom sizes and of the overall number of journalists (see discussions by Cagé, 2016; Fan, 2013). In fact, such concerns have led lawmakers and regulators to take actions aimed at ensuring that news publishers get a larger share of the industry revenue: In 2019 the European Parliament adopted a Directive on Copyright in the Digital Single Market whose article 15, referred to as the "link tax" article, grants publishers direct copyright over "online use of their press publications by information society service providers" (following similar - unsuccessful - attempts in Spain and Germany).³ Similarly, in 2021 the Australian government introduced the "News Media and Digital Platforms Mandatory Bargaining Code", according to which digital platforms and news organizations must negotiate in good faith over a financial remuneration for the use of news content, with the threat of arbitration if they fail to do so.

Another characteristic of social platforms is the large amount of personal data they collect, thanks to which they are able to offer a personalized mix of content to their users. This then leads consumers to spend more time on platforms, and to rely on them as "personalized curators" of content. It is not clear a priori whether the trend towards more personalization will lead to a softened or intensified competition between platforms and publishers (and thus to lower or higher news quality).

To study these issues, we develop a model in which a social platform and a newspaper (or publisher), both advertising-supported, compete for consumers' attention. The newspaper produces news stories and maintains a website, which only offers news content. The social platform relies on its users to produce UGC, such as personal stories or pictures. On its website, the platform can bundle UGC with content produced by the newspaper, in which case the platform and the newspaper share advertising revenues. The platform can choose how much news to display along its own content, even though it needs to show a minimal share of UGC.

We assume that the demand for news depends on content quality, which in turn is the result of an investment by the newspaper. Consumers can freely allocate their attention across the two websites, but, when on the platform, have to consume the mix of content that is offered to them, i.e. the platform plays a content curator role.

³Applying this directive, in 2020 the French Competition Authority enjoined Google to negotiate payments with publishers "in good faith".

To understand the basic competitive forces, we first analyze a baseline model with two simplifying characteristics, namely, (i) that consumers are homogeneous in their demand for news and, (ii) that the platform can unilaterally decide the implementation of content bundling (i.e. the publisher cannot prevent content bundling). In subsequent analyses, we relax these assumptions to reveal further insights on the impact of content bundling.

We start by comparing a benchmark where the platform cannot show news to a set-up where it can adopt a unilateral content bundling strategy. In this simple model content bundling leads consumers to spend more time on the platform without expanding news consumption, so that it makes the publisher worse-off. However, the fact that content bundling makes the newspaper less profitable does not lead to an unambiguous decline in content quality. In fact, we find that content bundling has a negative effect on quality if and only if the cost of quality is large. In other words, our model predicts that a high-quality newspaper (with a lower cost of producing quality) should invest more under content bundling, while a low-quality one should invest less. Intuitively, when the cost of quality is high, and thus the quality of news is low, consumers have a low demand for the publisher's content, and this demand can be satisfied by a few links on the platform. In this case, consumers only visit the platform's website (they single-home). The returns to investment in quality are relatively small, because a higher quality only increases the share of news content shown on the platform rather than attracting consumers to the publisher's website. When the cost of quality is low and quality is therefore high, consumers' demand for the publisher's content cannot be entirely satisfied by the platform (who cannot merely replicate the whole newspaper's content). Consumers then multi-home between the platform and the publisher's website, and direct traffic is more responsive to quality improvements than without content bundling. Indeed, because the platform shows some news content, increasing one's consumption of news by one unit of time requires reallocating *more* than one unit of time from the platform's website to the publisher's.

These results obtained in the case of a monopolist publisher who cannot prevent the platform from using its content extend to a setup with many differentiated publishers who can opt out of content bundling. Indeed, the existence of alternative news sources that the platform could link to deters all but the most efficient newspapers from opting out. Content bundling has the same effect on opting-in newspapers as in the baseline model, and weakly increases incentives to invest of the opting-out ones. Similarly, we also find that, if we allow newspapers to erect a paywall and collect subscription revenue from readers, our results pertaining to quality remain the same.

These core findings on the effect of content bundling on newspapers incentive to invest in quality are important in a context of general concern for the quality of news in the age of social media. Our results indicate that the effect of social networks on news providers is not uniform, and instead of leading to an unambiguous decline of content quality it could result in an increase in the *dispersion* of quality with some newspapers raising their

quality investment.

Motivated by European and Australian attempts to force large digital platforms to compensate newspapers, we also consider two policy initiatives: the imposition of link tax and mandatory bargaining code. While both policies unsurprisingly make newspapers better-off, we find that their effect on equilibrium quality is to *reduce* the dispersion of quality, by inducing low quality outlets to invest more and high quality ones to invest less, leading to ambiguous effects on the overall quality of news.

Finally, we enrich the baseline model by introducing consumer heterogeneity and (imperfect) personalization of the newsfeed by the platform. Under this more realistic setup, unilateral content bundling still increases the dispersion of quality. Additionally, it may now benefit the publisher, as it increases news consumption (a market expansion effect). Moreover, we also show that an increase in the platform’s personalization capabilities induces the publisher to invest *more* in the quality of its content. Interestingly, this suggests that a strategic platform may commit *not* to fully personalize consumers’ newsfeed so as to limit the intensity of competition with the publisher.

The next section summarizes the relevant literature. This is followed by the description (Section 3) and detailed analysis (Section 4) of a baseline model with unilateral content bundling and homogeneous consumers. In Section 5, we analyze the case when a publisher can erect a paywall and the case with multiple publishers that can each opt out of content bundling. Section 6 then looks at what happens when the platform and the publisher negotiate over content bundling. Finally, Section 7 introduces consumer heterogeneity and imperfect personalization by the platform. Appendix A contains omitted proofs, while the online appendix extends the baseline model to the case of competing publishers.

2 Relevant literature

Internet and the news The Internet has had a dramatic effect on the news industry, with a joint decrease of circulation and advertising revenue for the printed press (see Peitz and Reisinger, 2015, for an overview). In their online transition, newspapers have also experienced various challenges, among others self-cannibalization (Gentzkow, 2007), consumer switching behavior (Athey, Calvano, and Gans, 2018),⁴ or copyright violation (Cagé, Hervé, and Viaud, 2017).

An issue which is particularly relevant to this paper is the emergence of news aggregators, such as Yahoo News or Google News (see Jeon, 2018, for a survey). As in our paper, the central question is how these intermediaries impact the consumption of news as well as the quality of content produced. On the theory side, Jeon and Nasr (2016) and Dellarocas,

⁴see also Ambrus, Calvano, and Reisinger (2016) and Anderson, Foros, and Kind (2018) for the study of consumer multi-homing.

Katona, and Rand (2013) model aggregators as enabling consumers to find high-quality news more easily. They find that the entry of an aggregator tends to increase competition among websites, leading to higher quality. The impact on newspapers' profit depends on which effect is stronger: business stealing or market expansion. Rutt (2011) studies how the presence of an aggregator affects newspapers' choice of business model, and shows that it has different effects on the quality provided by free versus paying outlets. In George and Hogendorn (2012), the aggregator reduces the cost of multi-homing for consumers. In contrast to the way we model the platform's behaviour, in these papers, aggregators are non-strategic and do not produce their own content, but merely replicate the experience of a newspaper.

A recent series of empirical papers examine the impact of aggregators on the news industry. Using disputes between Google News and Spanish publishers (Athey, Mobius, and Pal, 2017; Calzada and Gil, 2016) or the Associated Press (Chiou and Tucker, 2017), empirical research finds that Google News increases overall news consumption. In particular, Athey, Mobius, and Pal (2017) document that this effect is mostly present for small publishers, who cannot rely on brand recognition to attract users and therefore benefit most from the aggregator. In relation to the theoretical work on aggregators, these papers suggest that the demand-expansion effect of aggregators dominates. George and Hogendorn (2013) study the consequences of a redesign of Google News, and find that news aggregators can potentially also change the composition of news consumption. Sismeiro and Mahmood (2018) study the impact of social media on news consumption using a global outage of Facebook in 2013, and show that the outage reduces traffic to newspapers' websites and changes the types of articles that consumers read.

In our model, the platform allocates consumers' attention by choosing the mix of content that it displays. In this respect it is similar to a search engine, which allocates traffic through its ranking and design (see de Cornière and Taylor, 2014; Burguet, Caminal, and Ellman, 2015). However, in these papers the intermediary enjoys an exogenous bottleneck position: consumers have to use the search engine to find content. In contrast, our mechanism is one where the allocation of attention while on the platform (i.e. content bundling) determines how consumers allocate their attention between the platform and the newspaper. The gatekeeping role of the platform thus emerges endogenously.

A few recent papers explicitly examine the relationship between social media and the news industry. Social platforms have been accused of fostering echo chambers and spreading fake news. Some critics argue that platforms should be held responsible for the content displayed on their websites (see Rolnik et al., 2019, for a recent overview of the issues). Abreu and Jeon (2020) and Campbell, Leister, and Zenou (2019) study how the distribution of news through a social network can lead to more polarized content being produced. Berman and Katona (2020) look at the effect of various curation algorithms on content quality and link formation in a social network. Our work is related as it focuses

on the fundamental effect of content bundling on publishers’ incentive to invest in content quality.

As Alaoui and Germano (2020), we also assume that consumers are time constrained in their consumption of media and our results resonate with theirs in that competition between content suppliers (including the social network) may distort consumers’ media consumption (in our case, when there is imperfect personalization). However, we focus on consumers’ time allocation across qualitatively different content providers and we abstract away from the editorial process of publishers when multiple topics are present.

News and news quality In our modeling approach, we do not separate news narrowly defined (facts and events) and the actual ‘news articles’ reporting the news. The model is based on the broadly accepted assumption that news publishers “curate” or “package” the available news items (facts) around a specific narrative that itself is an integral part of their value proposition (brand). In other words, updates (facts) are not considered separate from , so-called ‘narratives’. This view is largely consistent with the general theory of news media as seen by economics (see Mullainathan and Shleifer, 2005, for example).

Similarly, in this paper, we define quality as a demand-enhancing investment, which applies to the content itself (e.g. better investigative journalism). In practice, what constitutes news quality is a complex issue (Lacy and Rosenstiel, 2015), with many relevant dimensions, such as the accuracy of the reporting or variety of content.⁵ Empirical researchers have used the number of pages, size of the newsroom and the number of Pulitzer prize winners per staff member as measures of quality (see, e.g. Berry and Waldfogel (2010) and Angelucci and Cagé (2016)).

Notice that this approach does not encompass situations where consumers have heterogeneous views on what constitutes high quality as, for example, in the literature on media bias (see Gentzkow, Shapiro, and Stone, 2015; Puglisi and Snyder Jr, 2015, for recent surveys of media bias).⁶

3 Baseline model

Content and preferences We consider a model where consumers can consume two kinds of content: news and user-generated content (UGC). UGC is produced by a monopolist social platform (indexed by 0), at no cost, and its quality is exogenous. News are produced by a monopolist publisher, indexed by 1. The quality of news is denoted by q . The associated fixed cost for the newspaper is $C(q) = \frac{cq^2}{2}$. The newspaper’s marginal

⁵Bogart, 2004 cites characteristics such as “integrity, fairness, balance, accuracy, comprehensiveness, diligence of discovery, authority, breadth of coverage, variety of content, reflection of the entire home community, vivid writing, attractive makeup, packaging or appearance, and easy navigability”.

⁶See also Angelucci and Cagé (2016) for a model where newspapers can invest more in “hard news”, which is a horizontal feature.

cost is zero. Note that what we call “news” may be interpreted more broadly as any professional content produced by a third party.

Consumers are homogeneous in their relative preferences between news and UGC. If we denote by x_U the consumption of UGC and x_N that of news, consumers’ utility function is

$$U(x_U, x_N) = x_U + (1 + q)x_N - \frac{(x_N)^2}{2}. \quad (1)$$

Equation 1 captures the ideas that consumers enjoy both kinds of content and that their relative preference for news is increasing in its quality.⁷

Consumers are time constrained and we assume that they have one unit of time to allocate between consumption of news and UGC. Maximizing U under the constraint that $x_U + x_N \leq 1$, we obtain the demand for news of quality q : $D_N(q) = q$.

While unrealistic, the assumptions of consumer homogeneity and of a monopolist newspaper allow us to isolate and explain the main results’ driving mechanism. We extend the analysis to a setup with multiple heterogenous newspapers in Section 5.2,⁸ and to the case of heterogenous consumers in Section 7.

Content bundling Consumers divide their time between the platform and the publisher’s website. While the publisher’s website only shows news, the platform has a content curating role and may offer a mix of UGC and news, a strategy we refer to as *content bundling*. More precisely, for each unit of time spent on the platform’s website, consumers are exposed to a share λ of news content, and a share $1 - \lambda$ of UGC, where λ is chosen by the platform. Note that λ does not have to be literally interpreted as the share of news content that is shown to consumers, but rather as the share of news content that consumers read while on the platform. The model thus captures the idea that the platform may influence consumers’ news diet through the design of its algorithm.

We assume that the platform is restricted to choosing $\lambda \leq \Lambda$, where $\Lambda \in [0, 1/2]$ is an exogenous bound on the amount of news it can induce consumers to read while on the platform. The parameter Λ captures, in a reduced-form way, the idea that the platform needs to have a significant amount of UGC to be viable as a social platform and cannot show more content produced by a specific publisher than its own social content. Alternatively, one could interpret Λ as the maximal quantity of news that consumers are willing to read through the platform. Therefore, if a consumer spends T_0 units of time on the platform and $T_1 = 1 - T_0$ units of time on the publisher’s website, his consumption of news equals $x_N = T_1 + \lambda T_0$. We call T_1 and λT_0 respectively the direct and indirect

⁷Equation 1 also implies that there is a decreasing marginal utility for news only. This property is not essential for our analysis. Other utility functions, such as $U(x_U, x_N, q) = \ln x_U + \frac{q}{1-q} \ln x_N$ or $U(x_U, x_N, q) = x_U + 2qx_N - (x_U^2 + x_N^2)/2$, lead to the same demand for news and, as a result, to the same equilibria, with diminishing marginal utility for x_U .

⁸See also the online appendix.

consumption of news.

We start by considering the case of what we call “unilateral content bundling”, in the sense that the publisher has no control on how much of its content the platform bundles in its newsfeed. Later, we examine the case when the publisher can deny its content to the platform (can opt out of content bundling) and show that, with multiple publishers, “voluntary content bundling” emerges endogenously. Finally, we also explore the case where the two parties engage in bargaining about content sharing (“negotiated content bundling”).

Revenues We assume that both the platform and the publisher are purely advertising supported.⁹ Moreover, we assume the following: (i) the publisher generates more revenues from direct consumption than from indirect consumption of news (for instance because of “brand dilution”, or because indirect news consumption does not allow publishers to obtain data about consumers); (ii) the platform generates more advertising revenues (per unit of time) from showing UGC than from showing news, either because of revenue sharing or because of the risk that a consumer who clicks a link to the publisher does not come back to the platform.

Formally, we normalize to 1 the platform’s revenue when it shows UGC (all revenues are expressed per unit of time). Its revenue when it shows news is $\phi_0 \leq 1$, as per assumption (ii) above. We denote the publisher’s revenue from direct traffic as β and its revenue from indirect news consumption is denoted $\phi_1 < \beta$, as per assumption (i).

Overall, for given T_0 , T_1 and λ , the platform’s revenue equals $R_0 = (1 - \lambda + \lambda\phi_0)T_0$, while the publisher’s revenue is $R_1 = \beta T_1 + \phi_1 \lambda T_0$. The platform’s profit equals its revenue ($\pi_0 = R_0$), and the publisher’s profit is $\pi_1 = R_1 - \frac{cq^2}{2}$.

Timing and equilibrium In the first stage of the game, the publisher chooses the quality of its content q . In the second stage, the platform chooses how much news to incorporate to its own content, λ . In the third stage, consumers choose how to allocate their attention across the two websites. We look for subgame-perfect equilibria.

Table 1 presents a summary of the model variables. In order to focus on interior solutions (i.e. such that $q < 1$) we assume that $c > 2\beta$.

⁹Our results continue to hold if the publisher can charge a subscription price using a paywall (see Section 5.1).

Table 1: Summary of important variables and parameters in the model.

Variable	Meaning
q	Quality of news.
$\frac{cq^2}{2}$	Cost of news production.
T_i	Time spent on website $i \in \{0, 1\}$.
λ	Share of news on the platform.
Λ	Maximal share of news on the platform.
x_N	News consumption. $x_N = T_1 + \lambda T_0$
β	Newspaper's revenue from direct traffic (per unit of time).
ϕ_1	Newspaper's revenue from indirect traffic (per unit of time).
ϕ_0	Platform's revenue when showing news (per unit of time).
1	Platform's revenue when showing UGC (per unit of time).
R_i	Firm i 's total revenue.
Assumptions	
$\phi_0, \phi_1 \in (0, 1)$	$\beta > \phi_1, c > 2\beta, \Lambda \in (0, 1/2)$

4 Equilibrium

4.1 Benchmark without content bundling

We start by providing the analysis of the case where the platform cannot display content produced by the publisher on its own website. Formally, this corresponds to $\Lambda = 0$. We denote by \hat{q} the equilibrium quality.

In the last stage of the game, given a quality q , consumers allocate their attention so as to satisfy their demand for each type of content, $D_N = q$ and $D_U = 1 - q$, choosing $T_1 = q$ and $T_0 = 1 - q$. The publisher's profit is $\beta q - \frac{cq^2}{2}$. Maximizing this profit over q , we obtain:

Lemma 1. *In the benchmark case without content bundling, the equilibrium quality of news and the publisher's profit are*

$$\hat{q} = \frac{\beta}{c}, \quad \text{and} \quad \hat{\pi}_1 = \frac{\beta^2}{2c}.$$

Note that the total cost, $c\hat{q}^2/2$, is decreasing in c . This is consistent with the interpretation of newspapers with a low c as being "high quality" newspapers, which usually have a larger newsroom and thus higher total costs.

4.2 Equilibrium with unilateral content bundling

Let us now solve for the equilibrium of the game with content bundling through backward induction.

Stage 3: Consumers Suppose that quality of news is q and that the platform has chosen λ . Then, as we have seen, consumers' desired share of news consumption is also q . We need to distinguish two cases: (i) If $\lambda \geq q$, the optimal allocation of attention is $T_0 = 1$ because by spending all their time on the platform, consumers already obtain (weakly) more news than they would like. (ii) If $q > \lambda$, consumers must spend a positive amount of time on each website to consume their optimal mix of content. The optimal allocation of attention $(T_0, T_1) = (1 - T_1, T_1)$ solves

$$q = T_1 + \lambda T_0 = \lambda + T_1(1 - \lambda).$$

Lemma 2. Allocation of consumers' attention *Given a quality q and a share of news shown on the platform λ , the share of time that consumers spend on the publisher's website is $T_1(q, \lambda) = \max\{\frac{q-\lambda}{1-\lambda}, 0\}$.*

Stage 2: Platform The platform chooses λ to maximize its revenue

$$R_0 = (1 - \lambda(1 - \phi_0)) T_0(q, \lambda)$$

where $T_0(q, \lambda) = \min\{\frac{1-q}{1-\lambda}, 1\}$. Here again we need to distinguish two cases:

If $q < \Lambda$, notice first that it is never optimal to set $\lambda > q$: indeed, such a strategy is dominated by setting $\lambda = q$, as this would still maximize T_0 while reducing the quantity of news showed on the platform. Second, setting $\lambda < q$ is also suboptimal: in that case, consumers would adjust their allocation of attention so as to consume a share q of news and $1 - q$ of UGC, but some of the news consumption would occur on the publisher's website. Setting $\lambda = q$ leads to consumers receiving a share q of news, exclusively through the platform.

If $q > \Lambda$, the platform cannot offer the optimal mix on its website. The best it can do is to set $\lambda = \Lambda$. Any lower λ would lead to the same mix of content being consumed, with more time spent on the publisher's site.

Lemma 3. Platform's strategy *The optimal strategy for the platform is to set $\lambda = \min\{q, \Lambda\}$.*

Stage 1: Publisher Given the above analysis, the publisher's profit can be written as follows:

$$\pi_1(q) = \begin{cases} \phi_1 q - \frac{cq^2}{2} & \text{if } q \leq \Lambda, \\ \phi_1 \Lambda \frac{1-q}{1-\Lambda} + \beta \frac{q-\Lambda}{1-\Lambda} - \frac{cq^2}{2} & \text{otherwise.} \end{cases} \quad (2)$$

Broadly speaking, the publisher has two strategies. It can either choose a relatively low quality ($q \leq \Lambda$), in which case its content will be consumed exclusively through the platform. Such a strategy therefore induces consumers to single-home on the platform

and it resonates with the emergence of so-called “social native” publishers whose primary distribution is via social networks. Alternatively, it can choose a relatively high quality ($q > \Lambda$) in order to attract some direct traffic (equal to $\frac{q-\Lambda}{1-\Lambda}$), thereby leading to consumers multihoming, which is the typical strategy of established news publishers. The relative profitability of these strategies depends on the cost of producing quality, captured by the parameter c .

Let us define $\pi_{1L}(q, c) \equiv \phi_1 q - \frac{cq^2}{2}$, $\pi_{1H}(q, c) \equiv \phi_1 \Lambda \frac{1-q}{1-\Lambda} + \beta \frac{q-\Lambda}{1-\Lambda} - \frac{cq^2}{2}$, $q_L(c) \equiv \operatorname{argmax}_q \pi_{1L}(q, c) = \frac{\phi_1}{c}$, and $q_H(c) \equiv \operatorname{argmax}_q \pi_{1H}(q, c) = \frac{\beta - \Lambda \phi_1}{c(1-\Lambda)}$. One can readily check that $q_H(c) > q_L(c)$.

Let us also define c^\dagger such that $\pi_{1L}(q_L(c^\dagger), c^\dagger) = \pi_{1H}(q_H(c^\dagger), c^\dagger)$, i.e. $c^\dagger = \frac{\beta + \phi_1(1-2\Lambda)}{2\Lambda(1-\Lambda)}$.

Maximizing (2) leads to the following result, whose proof is in Appendix A:

Lemma 4. *Equilibrium quality* *The equilibrium quality is $q^* = q_L(c)$ if $c \geq c^\dagger$, and $q^* = q_H(c)$ otherwise.*

When c is very large, competing to attract direct traffic by setting $q > \Lambda$ is too costly for the newspaper. The optimal strategy therefore consists in choosing $q = q_L(c)$ and focus on indirect traffic. If the newspaper is more efficient, or if the platform cannot show too much of its content, the newspaper chooses a quality $q_H(c)$ that enables it to attract some direct traffic.

4.3 Effects of content bundling and comparative statics

We are now ready to evaluate the effect of content bundling on newspapers’ content quality. Comparing Lemmas 1 and 4, we can state the following result:¹⁰

Proposition 1. *Compared to a benchmark without content bundling:*

- (i) *Content bundling lowers the publisher’s profit.*
- (ii) *The quality of the publisher’s content and consumer surplus increase if and only if $c < c^\dagger$.*

The first part of Proposition 1 is fairly intuitive: because content bundling does not increase news consumption but diverts some consumption onto the platform, it cannot benefit the publisher. Note that this result does not always hold with heterogeneous consumers, as we will see in Section 7.

The intuition for the second part of the Proposition is as follows: when c is large, content bundling leads the publisher to adopt a low-quality strategy leading consumers to single-home on the platform, where the marginal revenue is ϕ_1 , which is less than the marginal revenue from direct traffic (β). Publishers who rely exclusively on the platform therefore invest less than they would absent content bundling.

¹⁰The proof is in Appendix A.

The story is different for publishers with a low cost of quality. Under the benchmark, a unit increase in quality leads to a β increase in R_1 . Under content bundling, the same increase in quality leads to a revenue increase of $\beta + \frac{\Lambda}{1-\Lambda}(\beta - \phi_1)$. Indeed, in order to achieve a unit increase of news consumption, consumers need to increase their direct traffic to the newspaper by $\frac{1}{1-\Lambda} > 1$. Part of this extra direct traffic comes at the expense of indirect revenue from the platform (an effect captured by the second term), but the overall effect is unambiguous: the newspaper’s revenue is more sensitive to quality in the regime where both websites get direct traffic.

Using a broader interpretation of our model, namely that there are multiple publishers that differ in their cost of producing quality (see Section 5.2 for a formal treatment of that case), Proposition 1 suggests that content bundling has a heterogeneous effect on newspapers: for high-quality newspapers (with low c), content bundling increases the incentive to invest in quality, while for low-quality newspapers it reduces this incentive. A testable prediction of the model is then that content bundling should increase the variance of the distribution of quality.

A note on total welfare In this model, if the mass of consumers is μ and if they consume the optimal mix $x_N = q$, total welfare is $\mu + (\mu - c)\frac{q^2}{2}$, so that the socially optimal quality is either 0 (if $\mu < c$) or 1 (assuming we cannot produce higher quality). Focusing on the more plausible scenario where news quality is socially desirable would therefore deliver the same normative conclusions as those based on consumer surplus.

Comparative statics Let us now examine the effect of the parameters of the model on the equilibrium under content bundling.

First, let us consider the effects of ϕ_1 , the indirect monetization parameter. An increase in ϕ_1 can be viewed as a way to model a “link tax”, forcing the platform to pay the publisher as is considered in recently discussed policy proposals.¹¹ If the publisher has a low cost ($c < c^\dagger$) and therefore a higher quality, its incentive to invest in quality goes down with ϕ_1 (i.e. $\frac{\partial q_H}{\partial \phi_1} < 0$) because of what we call the “softening effect”: as indirect monetization improves, attracting direct traffic becomes relatively less profitable. On the other hand, the incentives to produce higher quality increase with ϕ_1 if the publisher has a high cost ($c > c^\dagger$) because in this case all news is consumed on the platform and a higher quality increases indirect traffic only. In summary, a link tax does *not* lead to a general increase in quality. Rather, it leads to a reduction in newspapers’ quality dispersion.

Next, an increase in the advertising revenue from direct traffic β makes it more profitable for the publisher to adopt a “high quality” strategy ($\frac{\partial c^\dagger}{\partial \beta} > 0$), and the corresponding quality also increases ($\frac{\partial q_H}{\partial \beta} > 0$). If the publisher adopts the “low quality” strategy, its

¹¹See for instance the recent EU Copyright Directive: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32019L0790&from=EN>.

quality choice is unaffected by β .

Finally, one may look at the effects of a change in Λ , which one could interpret as a strategic choice by the platform to give relatively more or less prominence to externally produced content, as for instance when Facebook announced that it would reduce the share of news in the newsfeed in 2019. In the model, a reduction in Λ would lead to a lower $q_H(c)$, because the reduced sensitivity of direct traffic would more than offset the increased relative value of such traffic.¹² Notice that, starting from $\Lambda = 0$, a small increase in Λ increases quality for all types of publishers (because $c^\dagger \rightarrow +\infty$ when $\Lambda \rightarrow 0$).

These results are summarized below:

Proposition 2. *Under unilateral content bundling, comparative statics results are given by the following table:*

	$q_H(c)$	c^\dagger	$q_L(c)$
$\beta \uparrow$	\uparrow	\uparrow	$=$
$\phi_1 \uparrow$	\downarrow	\uparrow	\uparrow
$\Lambda \uparrow$	\uparrow	\downarrow	$=$

5 Paywalls, newspaper opt-out and voluntary participation

The baseline model presented above relies on some stark assumptions, in particular that the newspaper derives all its revenues from advertising and cannot erect a “paywall”, and that it has no way of preventing the platform from using its content. In this section, we relax these assumptions. We first look at a variant of the model where the newspaper can charge a price for content to consumers using a paywall (i.e. preventing the platform from displaying its content to non-subscribers). Second, we study a model where the platform can directly block the platform from using its content. For the analysis of the latter case we also enrich the model by introducing multiple newspapers that are imperfect substitutes. The main takeaway from this section is that, broadly speaking, the results from the baseline model continue to hold.

5.1 Subscription pricing by the publisher

Suppose that the newspaper can erect a paywall and charge a price p for consumers to access its content. Subscribers can read the newspaper’s content either on its website or through the platform, while non-subscribers can only consume UGC. We maintain the assumption that the platform cannot show more than a share Λ of news, so that subscribers

¹²The logic is similar to the one explaining that quality increases under content bundling for $c < c^\dagger$.

of a newspaper with quality $q > \Lambda$ will read their news both on the newspaper’s website and on the platform.

Proposition 3. *When the newspaper can charge consumers a subscription price:*

1. *For a given quality, the subscription price is the same with or without content bundling, equal to $p(q) = \frac{q^2}{2}$.*
2. *The effect of content bundling on equilibrium quality is the same as in Proposition 1.*

Proof. (1) Suppose that the publisher can set a price p that consumers must pay in order to access its content. A consumer who does not pay p only consumes UGC and gets a surplus of $U(1,0) = 1$ (by Equation 1). If the consumer pays the price, his news consumption (with or without content bundling) becomes $x_N = q$, and his utility is $U(1 - q, q) - p = 1 + \frac{q^2}{2} - p$. For a given quality q , the optimal price is thus $p(q) = \frac{q^2}{2}$.

(2) The publisher maximizes $p(q) + R_1(q) - C(q)$, where $R_1(q)$ is advertising revenue, equal to $\widehat{R}_1(q) = \beta q$ without content bundling, and to $R_1^{CB}(q) = \beta T_1(q, \lambda^*(q)) + \phi_1 \lambda T_0(q, \lambda^*(q))$ with content bundling. Because the price $p(q)$ is the same with and without content bundling, quality increases with content bundling, if and only if, $\widehat{R}_1'(q) < R_1^{CB'}(q)$, which is the same condition as in the model without subscription pricing. ■

This analysis suggests that our results do not critically depend on the assumption that newspapers are purely advertising-driven, which is important given the increased prevalence of subscription models. One important assumption for this “neutrality” result is that of homogeneous consumers. With heterogeneous consumers (as in Section 7), we can show (analysis available upon request) that content bundling is more likely to increase quality with subscription models than with pure advertising-funded newspapers, because it induces the newspaper to focus more on surplus extraction than on readership expansion (extra eyeballs are less valuable with content bundling). However, a result reminiscent of Proposition 1 is that quality increases more for low values of c than for high values. That is, we still expect the dispersion of quality to increase.

Note that the choice of a subscription model and the associated paywall by a publisher is distinct from that of preventing the platform from linking to its content. Indeed, many outlets with a subscription model are present on social networks. The next subsection examines the publisher’s decision of whether to be present on the platform at all.

5.2 Endogenous newspaper participation

In practice many newspapers are voluntary participants to social networks, for instance by maintaining a Facebook page with regular posting of articles, or by displaying various options for their readers to share articles on social networks. This fact seems hard to reconcile with Proposition 1’s result on profit. Indeed, in the baseline model, the

newspaper would never choose to be an active participant to the platform. However, another important force missing from the simple baseline model is the competitive pressure faced by newspapers: a newspaper who would choose not to be present on social media would risk losing market share to other outlets who have opted in. We formalize this argument below, and show that the baseline model's predictions regarding quality hold in this context.

Suppose that there is a continuum of newspapers, indexed by $i \in [0, 1]$. Each newspaper has a cost function $c_i q_i^2/2$, where the parameters c_i are heterogeneous across newspapers, and distributed over $[\underline{c}, \infty)$, with $\underline{c} \geq 2\beta$. Each newspaper i has a mass 1 of regular readers, who allocate their time between newspaper i and the platform.¹³ Suppose that newspapers control whether their content is available on the platform. If newspaper i chooses to allow the platform to show its content (opt-in), the situation is similar to our baseline model, in that the platform shows newspaper i 's regular readers a mix of UGC and news from i (say, because they follow the newspaper on the platform). If newspaper i opts out, the platform can still show its regular readers news content from other newspapers who have opted in. However, we assume that i 's regular readers do not value news from other newspapers as much. Formally, the perceived quality of newspaper j by regular readers of i is $\min\{\underline{q}, q_j\}$. The parameter \underline{q} is an inverse measure of brand loyalty. When \underline{q} is small i 's readers put little value on j 's quality, whereas $\underline{q} \rightarrow \infty$ corresponds to the case where j is evaluated the same way i is. One interpretation is that reading news from another source than i would allow a regular reader of i to learn about the main events of the day, but does not bring as much benefit as the coverage of i . We focus on cases where \underline{q} is relatively small (in particular, $\underline{q} < \Lambda$). Suppose that, in the case where i opts out, the platform shows content from the opting-in newspapers that provide the largest utility, i.e. whose quality is larger than \underline{q} (we show in the appendix that it is an equilibrium for some of these newspapers to opt in). We index these newspapers with the subscript $-i$.

The utility of a regular reader of newspaper i who spends $x_{N,i}$ units of time reading i , $x_{N,-i}$ units of time reading news from other sources, and x_U units of time reading UGC, is then:

$$U(x_{N,i}, x_{N,-i}, x_U) = (1 + q_i)x_{N,i} + (1 + \underline{q})x_{N,-i} + x_U - \frac{(x_{N,i} + x_{N,-i})^2}{2},$$

For $q_i \geq \underline{q}$, the optimal content mix for a regular reader of i is $x_{N,i} = q_i$, $x_U = 1 - q_i$ and $x_{N,-i} = 0$.¹⁴

The timing of the game is the following: (1) Publishers choose their quality, and choose whether to opt in or out of the platform; (2) The platform observes the quality of the publishers who opt in, and chooses its content mix; (3) Consumers allocate their attention

¹³One could easily allow heterogeneity in the mass of regular readers, but this would slightly complicate notations without much additional insight.

¹⁴When $q_i < \underline{q}$, it would be optimal for consumers to only read news from $-i$.

between the platform and their favorite newspaper.

We denote by λ_i the share of content from newspaper i that the platform shows to its regular readers if i has opted in, and by λ_{-i} the share of news (from outlets other than i who have opted in) that the platform shows to i 's regular readers if i has opted out.

If publisher i opts in, then the platform will choose $\lambda_i = \min\{\Lambda, q_i\}$, so that the equilibrium quality is the same as in Lemma 4.

Suppose that publisher i opts out, and that at least one other publisher, say j , opts in. Then the platform can show news from j to i 's regular readers. If it shows a share λ_{-i} of news from j , then in stage (3) a regular reader of i chooses the time he spends on the platform, $T_{0,i}$, so as to solve

$$\max_{T_{0,i}} U \left(\underbrace{1 - T_{0,i}}_{x_{N,i}}, \underbrace{\lambda_{-i} T_{0,i}}_{x_{N,-i}}, \underbrace{(1 - \lambda_{-i}) T_{0,i}}_{x_U} \right).$$

The solution is $T_{0,i} = \max \left\{ \min \left\{ \frac{1 - q_i - (1 - q) \lambda_{-i}}{(1 - \lambda_{-i})^2}, 1 \right\}, 0 \right\}$.

In stage (2), if i has opted out, the platform does not observe q_i but forms rational expectations, and chooses λ_{-i} so as to maximize $(1 - \lambda_{-i} + \lambda_{-i} \phi_0) T_{0,i}$, under the constraint that $\lambda_{-i} \leq \Lambda$. We provide the formal analysis of this problem in the proof of Proposition 4 below. Qualitatively, we find that λ_{-i} is an inverted U-shaped function of q_i . For small values of q_i , the platform's optimal strategy consists in inducing consumers to read all their news on the platform, albeit not from their favorite newspaper. To do so, it shows news from a "generic" newspaper who has opted in, and this reduces consumers' marginal utility of news consumption, thereby leading consumers not to visit i 's website. As q_i increases, such a strategy ceases to be profitable. Then λ_{-i} becomes a decreasing function of q_i : as q_i increases, it becomes more costly for the platform to retain consumers by showing them news of relatively low quality, and so the platform focuses more on UGC.

The following Proposition summarizes the characteristics of this equilibrium. Its proof is provided in Appendix A.

Proposition 4. *There exists an equilibrium such that:*

1. *Newspapers with $c_i \leq \frac{\beta(1+\phi_0)}{q+\phi_0}$ opt out and choose the same quality as without content bundling. The platform does not show news to these newspapers' regular readers ($\lambda_{-i} = 0$)*
2. *Newspapers with c_i slightly above $\frac{\beta(1+\phi_0)}{q+\phi_0}$ also opt out, but choose a higher quality than without content bundling. The platform chooses $\lambda_{-i} > 0$.*
3. *For larger values of c_i , newspapers opt in. Then, they choose a higher quality than without content bundling if and only if $c < c^\dagger$ (see Proposition 1).*

For very small values of c_i , a publisher can ignore the platform and choose the same quality as in the benchmark: its quality advantage over “generic” news is so large that in case of opt-out the platform would not display any news. The relevant region is not empty as long as $\underline{q} < \frac{1-\phi_0}{2}$. Note that, as \underline{q} increases, i.e. as brand loyalty decreases, it is less likely that a publisher with a given c_i will be able to opt out and ignore competition from other publishers that opt in. For values of c_i slightly above the previous threshold, if publisher i opts out, the platform finds it optimal to show some content from another publisher. This makes publisher i ’s demand more sensitive to quality, thereby leading to an increase in q_i . This strategy is more profitable for publisher i than opting in if c_i is small enough. For larger values of c_i , opting in dominates. In that case, the outcome is identical to the one studied in Section 3: if c_i is relatively small publisher i invests more than under the benchmark and consumers spend some time on its website and some time on the platform. For larger values of c_i , the best strategy for publisher i is to rely exclusively on the platform for the distribution of its content, in which case the equilibrium quality is lower than under the benchmark.

Discussion The pattern revealed in Proposition 4 shows that the existence of alternative sources of news can lead some publishers to opt-in, for fear of losing too much traffic. Under this outcome our previous analysis in the baseline model applies: under a certain cost-of-quality threshold a publisher will increase its equilibrium content quality compared to the benchmark without content bundling by the platform. The novel finding of this extension is that a publisher with a very small cost of quality will opt out: its quality advantage is large enough not to have to make its content available on the platform. However, such opt-out publisher chooses a (weakly) higher quality. Indeed, the threat of losing readers to other publishers on the platform provides an incentive to increase quality.

6 Negotiated content bundling

As the previous section showed, enabling newspapers to block the platform from using their content may not achieve much, either in terms of quality or profits. One policy approach recently put in place by regulators in Australia and France consists in requiring dominant platforms to negotiate revenue-sharing agreements with publishers.

To shed light on the economic implications of such a policy we go back to the baseline model with a single newspaper, and introduce Nash bargaining. Suppose that the timing is the following: in the first stage, the publisher chooses its quality q . In the second stage, the publisher and the platform engage in Nash bargaining (with equal bargaining power) to determine a lump-sum payment to the publisher.¹⁵ In the third stage, the platform chooses

¹⁵Regarding the assumption that investment takes place before bargaining, the alternative game where bargaining takes place before the newspaper chooses quality would result in the platform paying the

the quantity of news it shows to its users subject to the constraint $\lambda \leq \Lambda$, and gets all the advertising revenue when consumers use its website.¹⁶ In stage 4, consumers allocate their attention among the two websites so as to consume their optimal mix. Direct traffic to the publisher's website generates revenues of β per unit of time. If $\beta > 1$, there would be no agreement in equilibrium as content bundling would not create surplus. Accordingly, we assume $\beta \leq 1$. To rule out corner solutions we focus on cases where $\beta > 1/3$.

In practice, the game we are studying could be implemented by having the publisher's content hosted on the platform, and the advertising slots sold by the platform to advertisers. Whether revenues are shared in a lump-sum way or on a per-unit basis would be equivalent in this set-up, as the quantity of news shown to consumers is always $\min\{q, \Lambda\}$ and would not depend on the fee paid to the publisher.¹⁷ The constraint $\lambda \leq \Lambda$ captures the idea that, even with a formal agreement, the platform must show enough original content to be viable. We assume that when the platform is indifferent between several values of λ it chooses the smallest one. In case bargaining fails, the outcome is the one without content bundling, described in Lemma 1.¹⁸

Equilibrium We solve the game by backward induction. In Stages 3 and 4, consumers face the same decision as in the baseline analysis, and their optimal behavior is given by Lemma 2: If $\lambda \geq q$, $T_1(q, \lambda) = 0$. If $q > \lambda$, $T_1(q, \lambda) = \frac{q-\lambda}{1-\lambda}$.

Given that $\phi_0 = 1$, the platform simply seeks to maximize the amount of time users spend on its website. If $q \leq \Lambda$, this entails setting $\lambda = q$, which leads to $T_0(q, q) = 1$. If $q > \Lambda$, the platform sets $\lambda = \Lambda$ and $T_0(q, \Lambda) = \frac{1-q}{1-\Lambda}$.

At stage 2 (the bargaining stage), the cost $C(q)$ is sunk and does not affect the bargaining outcome. Given a quality q , the outside option payoffs are, respectively, βq for the publisher and $1 - q$ for the platform.

Let F be the lump sum payment to the publisher in case of an agreement. If the publisher has chosen $q < \Lambda$, we know that the platform will choose $\lambda = q$ and obtain advertising revenues equal to 1, while the publisher will get zero revenues. By the logic of Nash bargaining, the negotiated payment F must then be the solution to

$$\max_F (F - \beta q) (1 - F - (1 - q)), \quad (3)$$

that is

$$F = \frac{q(1 + \beta)}{2} \equiv F_1(q). \quad (4)$$

newspaper not to invest in quality.

¹⁶To use notations from the previous section, we then have $\phi_0 = 1, \phi_1 = 0$.

¹⁷This feature is due to our assumption of homogeneous consumers.

¹⁸This framework also works when the publisher uses a paywall and collects subscription revenues. The difference is that in this case its fallback revenue is higher.

If the publisher has chosen $q > \Lambda$, the parameters of the negotiation are different. Indeed, even with an agreement, the publisher will receive a share of traffic $T_1 = \frac{q-\Lambda}{1-\Lambda}$, and a revenue βT_1 . The negotiated payment F must then be the solution to

$$\max_F \left(\beta \frac{q-\Lambda}{1-\Lambda} + F - \beta q \right) \left(\frac{1-q}{1-\Lambda} - F - (1-q) \right), \quad (5)$$

that is

$$F = \frac{(1-q)(1+\beta)\Lambda}{2(1-\Lambda)} \equiv F_2(q). \quad (6)$$

At stage 1, anticipating the outcome of the bargaining stage, the publisher maximizes

$$\pi_1(q) = \begin{cases} F_1(q) - \frac{cq^2}{2} & \text{if } q \leq \Lambda \\ F_2(q) + \beta \frac{q-\Lambda}{1-\Lambda} - \frac{cq^2}{2} & \text{otherwise.} \end{cases} \quad (7)$$

Let us denote $q_1(c) \equiv \operatorname{argmax}_q F_1(q) - \frac{cq^2}{2} = \frac{1+\beta}{2c}$ and $q_2(c) \equiv \operatorname{argmax}_q F_2(q) + \beta \frac{q-\Lambda}{1-\Lambda} - \frac{cq^2}{2} = \frac{\beta(2-\Lambda)-\Lambda}{2(1-\Lambda)c}$. Maximizing (7) leads to the following result.¹⁹

Lemma 5. *Under negotiated content bundling, the equilibrium quality is:*

$$\tilde{q} = \begin{cases} q_2(c) & \text{if } c < \frac{\beta(2-\Lambda)-\Lambda}{2(1-\Lambda)\Lambda} \\ \Lambda & \text{if } c \in \left[\frac{\beta(2-\Lambda)-\Lambda}{2(1-\Lambda)\Lambda}, \frac{1+\beta}{2\Lambda} \right] \\ q_1(c) & \text{if } c > \frac{1+\beta}{2\Lambda}. \end{cases}$$

Intuitively, when the cost of producing quality is small, the publisher operates in the parameter region such that $q > \Lambda$ and therefore chooses $q_2(c)$. Conversely, when c is large, it is more profitable for the publisher to choose a strategy of exclusive distribution through the platform, and then $\tilde{q} = q_1(c)$.

Comparison with benchmark and unilateral content bundling Comparing the equilibrium under bargaining to the equilibrium under the benchmark (Lemma 1) and under unilateral content bundling (Lemma 4), we have the following:

Proposition 5. *1. Negotiated content bundling increases quality compared to the benchmark if and only if $c > \frac{\beta}{\Lambda}$.*

2. Negotiated content bundling increases quality compared to unilateral content bundling if and only if $c > c^\dagger$.

The intuition for the result is as follows. Suppose first that c and Λ are large enough so that we are in the region where $q < \Lambda$. In this case the newspaper's content is exclusively distributed through the platform under negotiated content bundling (low-quality content

¹⁹See Appendix A for the proof.

with single-homing on the platform). Then, the incentive to invest comes solely from the effect of quality on each party’s outside option. A quality increase of Δq improves the publisher’s outside option by $\beta\Delta q$ and reduces the platform’s one by Δq . Under Nash bargaining, this results in an increase of $\frac{1+\beta}{2}\Delta q$ of the publisher’s payoff, which is larger than both the corresponding increase without content bundling ($\beta\Delta q$) and the one under unilateral content bundling ($\phi_1\Delta q$).

When c is small and we are in the region such that $q > \Lambda$, an increase in q still affects the parties’ outside options, but it also entails an inefficiency from the parties’ joint profit perspective, as it attracts more traffic to the publisher’s website, where advertising revenues are lower than on the platform.²⁰ The joint profit is indeed $1 - \frac{(1-\beta)(q-\Lambda)}{1-\Lambda}$, decreasing in q . This inefficiency dampens the publisher’s incentives to invest. Because the newspaper does not internalize the platform’s revenues under the benchmark or with unilateral content bundling, such an effect is absent in these cases.

If one worries about the negative effect of social platforms’ content bundling on newspapers’ profitability, putting their survival at risk, a solution based on bargaining appears promising as it ensures that publishers are better-off than under unilateral content bundling. However such a regime is not neutral from a quality perspective. Indeed, the model predicts that established newspapers (with larger newsrooms, and a higher ability to cover many topics) are more likely to invest less under negotiated content bundling, while less established ones will tend to invest more. In sum, similar to the effect of a link tax, forcing negotiations between the platform and newspapers may only lead to a reduction in the variation across newspapers’ qualities and not to a general increase of quality. Note that a particularly undesirable outcome, from our model’s perspective, would be for established newspapers to negotiate with the platform while smaller players (with a higher c) operate under the unilateral content bundling framework (even with possibility to opt-out), as in this case quality would decrease for most outlets.

7 Heterogeneous consumers and platform analytics

So far, we have assumed homogeneous consumers and as a result, the social network’s content strategy was reflected by a single number λ and each consumer’s newsfeed contained the same proportions of news and UGC. In reality however, there is a large variation of preferences with respect to content across members on a social network. More importantly, a fundamental appeal of social platforms is that by observing their members’ behavior the platform can personalize each member’s newsfeed. In this section we extend our basic model to capture these important characteristics.

To introduce consumer heterogeneity in a tractable way, we modify consumers’ utility

²⁰Recall that if this was not the case there would be no agreement.

function in (1) by assuming that a consumer of type θ who consumes a quantity x_N of news (of quality q) and x_U of UGC derives utility:

$$U(x_U, x_N, q, \theta) = x_U + (1 + \theta + q)x_N - \frac{(x_N)^2}{2}, \quad (8)$$

where θ is a parameter that measures a consumer's relative preference for news compared to UGC, and is uniformly distributed on $[0, 1]$. Given the constraint that $x_U + x_N \leq 1$, the demand for news of a consumer of type θ is $\min\{\theta + q, 1\}$.²¹ As in the baseline model, consumers allocate their unit of attention across two websites: one operated by the publisher, and one by the platform.

We assume that the platform is imperfectly informed about consumers' types. More specifically, the platform observes a consumer's type θ with probability α , and observes nothing with probability $1 - \alpha$. We interpret the parameter α as the exogenous quality of the platform's analytics technology, enabling it to infer consumers' preferences.²²

The platform can adapt the content it shows to a consumer depending on his type, showing a higher share of news (λ) to higher types. Consumers whose type the platform does not observe all face the same λ . We assume that the maximal amount of news that the platform can show, Λ , is arbitrarily close but strictly below one, meaning that the platform can show as much news as it wants to high types (when it identifies them), but consumers who only desire to read news visit the publisher directly (tie-breaking rule).²³ For notational simplicity, we also assume that the platform shares with the newspaper the total advertising revenue when advertising on behalf of the newspaper, i.e. $\phi_1 + \phi_0 = 1$. In this case, we simply denote by ϕ the share of ad revenues that the platform keeps for itself. Given our previous assumptions on the ϕ -s and β , this means that $1 - \phi < \beta$.

The timing is as follows: at $\tau = 1$, the newspaper chooses a quality q , publicly observed, and incurs cost $C(q)$. At $\tau = 2$, the platform observes the type of a share α of consumers, and chooses the share of news $\lambda(\theta, q)$ it shows to its users (for consumers whose type it does not observe, the platform's content bundling strategy is $\lambda(q)$). At $\tau = 3$, consumers observe the λ they face, and choose $t_0(\theta, q, \lambda)$, the time they spend on the platform as a function of their type, of the quality of news and of the platform's content mix. The resulting news consumption is $x_N = \lambda t_0 + t_1$. We look for subgame-perfect equilibria.

²¹Note that, even though θ is not a measure of a consumer's taste for quality, consumers with high θ derive more benefits from an increase in quality, given that they consume more news. We have also explored a model where $U(x_U, x_N, q, \theta) = x_U + (1 + \theta q)x_N - \frac{(x_N)^2}{2}$. It is harder to obtain closed-form solutions there, but the numerical results we obtained were consistent with those that are presented in the paper.

²²An alternative, maybe broader way to define α is that, while the platform knows every consumers' θ , each consumer deviates from their θ with probability $1 - \alpha$. In this interpretation, α measures the platform's capability to predict consumers' preferences.

²³We make this assumption for notational simplicity, by removing Λ from the picture.

7.1 Benchmark: UGC-only newsfeed

When the platform cannot bundle news content alongside UGC consumers can consume their desired mix of content: a consumer of type θ then spends $t_1(\theta, q, 0) = \min\{\theta + q, 1\}$ units of time on the newspaper's website, and $t_1(\theta, q, 0) = 1 - t_1(\theta, q, 0)$ on the platform. Thus, the total time spent on the newspaper is

$$T_1(q, 0) = \int_0^1 \min\{\theta + q, 1\} d\theta = \int_0^{1-q} (\theta + q) d\theta + \int_{1-q}^1 1 d\theta = \frac{1 + 2q - q^2}{2}.$$

The profit of the newspaper is $\beta T_1(q, 0) - cq^2/2$, which leads to the following equilibrium quality and profit in the benchmark:

$$\tilde{q} = \frac{\beta}{\beta + c} \quad \text{and} \quad \tilde{\pi}_1 = \frac{\beta(2\beta + c)}{2(\beta + c)}. \quad (9)$$

As in the base model, quality increases in β and decreases with c .

7.2 Equilibrium with content bundling

Under content bundling consumers' time allocation across platforms depends on θ . A consumer of type θ to whom the platform shows a share λ of news will allocate his attention so as to come as close as possible to his desired consumption mix (see Anderson and Neven, 1989, for a related analysis of combinable goods). The optimal allocation of attention mirrors that of Lemma 2, where we replace q by $\theta + q$. Following the tie-breaking rule we also observe that consumers such that $\theta + q \geq 1$ do not visit the platform (i.e. $t_0(\theta, q, \lambda) = 0$).

The platform's problem consists in choosing a share of news $\lambda(\theta, q)$ to show to consumers of type θ , and a share $\bar{\lambda}(q)$ to show to consumers whose type it does not observe. First, the platform's decision regarding consumers whose type it can observe is straightforward: $\lambda(\theta, q) = \min\{\theta + q, 1\}$. By showing each consumer his ideal content mix, the platform ensures that consumers with a positive demand for UGC (such that $\theta < 1 - q$) allocate all their attention to its website: $t_0(\theta, q, \lambda(\theta, q)) = 1$ for all $\theta < 1 - q$ (consumers such that $\theta \geq 1 - q$ only visit the newspaper).

For the $1 - \alpha$ unidentified consumers, whose θ is unknown to the platform, the problem is more subtle. If the platform displays a share $\bar{\lambda} \geq q$ of news content, the total amount of attention that it receives from this segment is

$$T_0(\bar{\lambda}, q) = \int_0^{\bar{\lambda}-q} 1 d\theta + \int_{\bar{\lambda}-q}^{1-q} \frac{1 - (\theta + q)}{1 - \bar{\lambda}} d\theta = \frac{1 + \bar{\lambda} - 2q}{2}. \quad (10)$$

The platform's profit is then $(1 - \bar{\lambda} + \bar{\lambda}\phi_0)T_0(\bar{\lambda}, q)$, which is maximized for $\bar{\lambda}(q) = q + \frac{\phi}{2(1-\phi)}$. The first term, q , corresponds to the demand for news of the lowest type ($\theta = 0$). Beyond

this level, the platform’s optimal strategy depends on the share ϕ of revenues it captures when it shows news: for large values of ϕ the platform has an incentive to show a lot of news content to its users.

At $\tau = 1$, the publisher can anticipate the platform’s strategy and consumers’ time allocation across sites and chooses its quality so as to maximize its profit. Adding revenues from consumers across sites, the publisher’s optimal choice of quality is as follows (see the proof in the Appendix):

Lemma 6. *With heterogeneous consumers, under content bundling by the platform, the profit-maximizing quality level for the newspaper is*

$$q^* = \frac{\beta + 1 - \phi + \alpha(\beta - (1 - \phi))}{2(1 - \phi + c)}. \quad (11)$$

As in the homogeneous case, quality is increasing in β and decreasing in c and ϕ . Interestingly, it is also increasing in α (see below for a discussion).

7.3 Effects of content bundling

Now, we can compare the equilibrium outcomes with and without content bundling, from the two previous subsections.²⁴ We start with the publisher’s profit:

Proposition 6. *There exists $\bar{\beta} \in [1 - \phi, 1)$ such that the newspaper is better-off under content bundling, if and only if, $\beta < \bar{\beta}$.*

Content bundling has two effects on the newspaper: first, as in the baseline model, content bundling diverts news consumption onto the platform, thereby reducing the newspaper’s revenue per-unit of attention. This effect is negative. Second, unlike the baseline model with homogeneous consumers, content bundling increases news consumption (as long as $\alpha < 1$): by setting $\bar{\lambda}(q) > 0$, the platform induces some low-types to read more news than they would otherwise, while higher types still consume their optimal mix. This is a market expansion effect. If the newspaper’s revenue from direct traffic, β is close to its revenue from indirect traffic ($1 - \phi$), the first effect dominates, while the second effect dominates for β close to 1.

Of central interest is the newspaper’s incentive to invest in quality under content bundling compared to the benchmark. Simple algebra yields the following result.

Proposition 7. *The equilibrium quality of news increases under content bundling compared to the benchmark if and only if $c \leq \frac{1+\alpha}{1-\alpha}\beta$.*

²⁴The proof is in the Appendix. Furthermore, there we show additional analysis, where we explore and discuss the impact of a “link tax”, as proposed by European legislators.

This result mirrors Proposition 1 in the case with heterogeneous consumers: the effects of content bundling on quality depend on the publisher’s cost of quality, as a result of two opposite effects: (i) increasing demand for news of low-types is less valuable, as this consumption takes place on the platform; (ii) direct traffic by relatively higher types is more sensitive to quality under content bundling.

Lastly, we examine the impact of the platform’s analytic capabilities by asking: suppose that the platform were to use the optimal content bundling strategy, given its analytic capability α , how would an increase in said ability affect the equilibrium? From (11) we directly see the following:

Proposition 8. *The equilibrium quality of news is an increasing function of the platform’s analytics capability, α .*

With personalization, competition between the platform and the publisher is more intense. In particular, turning a consumer into an exclusive news reader ($\theta + q \geq 1$) is much more valuable to the publisher when the platform offers personalized content, because non-exclusive news readers spend all their time on the platform, even if the share of news they read is close to one.²⁵

The result that personalization leads the publisher to invest more in the quality of its content has some interesting implications. First, it highlights a potential positive side-effect of platforms’ increasing ability to tailor content to users’ preferences, which so far has not been studied in the literature (see de Cornière and Taylor, 2021, for a discussion of improved analytics on competition).

Second, one intriguing result is that the platform’s profit is not always increasing in α : under some parameters, even if the platform had the ability to offer a fully personalized content mix to all its users, it would not be optimal to do so, in order to deter the publisher from investing too much in quality. We can show that the platform tends to optimally limit the extent of personalization when $\beta/(1 - \phi)$ is large, i.e. when the publisher’s incentive to compete fiercely to attract attention on its own website is large. This result illustrates, in a very stark way, how the returns on investment in analytics technology may decrease due to the strategic reaction of third party content producers.

Finally, with respect to consumer surplus, our findings are ambiguous. When we compare content bundling to the benchmark, two effects must be taken into account, namely the impact of bundling on quality, and its impact on the equilibrium mix of content. While the first effect can go either way, as discussed previously, the second one can only be negative, with low-type consumers being offered too much news compared to their optimal level.

Importantly, note that the effect of more personalization (α) is unambiguously positive: we saw that quality goes up with α (Proposition 8), and personalization reduces the

²⁵A similar logic would work if there was a limit to how much news the platform can show.

content mix distortion.

8 Discussion and concluding remarks

Social networks have gained tremendous importance in the last decade, claiming a significant share of consumer attention. They have achieved such prominence by leveraging network effects *and*, more recently, by successful content bundling, whereby third party content is presented in their users’ “newsfeed”. This strategy, in turn, has started to fundamentally transform media production and consumption. For some content domains with so-called public calling, such as news for example, concerns have been raised about the overall health of the industry. In particular, given the market power of social platforms, it is important to understand whether news providers’ maintain adequate incentives to invest in content quality. Accordingly, policy makers have pushed for ways to encourage transfers between social platforms and news providers.

In this paper we develop a model of competition for attention between a social platform and third-party content providers, allowing us to shed light both on the strategic motives for content bundling and on its implications for publishers and their incentives to invest in quality. By bundling news content with UGC, the platform increases the share of attention it receives from consumers: part of their demand for news is satisfied by the platform, which induces them to spend less time on the newspaper’s website. We show that the practice of unilateral content bundling tends to harm publishers, but has a heterogeneous impact on their incentives to invest in the quality of their content: initially low quality outlets invest less under content bundling, while initially high quality ones invest more.

Importantly, the above core result continues to hold when publishers can opt out of content bundling as long as they face significant competition. It also holds under the more realistic case with consumer heterogeneity when the platform can (partially) customize consumers’ newsfeed.

The paper also sheds light on the efficacy of recent policy initiatives to address the newspapers’ concerns about platforms’ increased dominance. We show that when the platform and the publisher negotiate for the former to use the latter’s content, the results are reversed: the publisher is better-off than under the benchmark, but the dispersion of quality goes down, as low quality outlets invest more and high quality ones invest less. This highlights a potential trade-off associated with current proposals (e.g. in Australia and the European Union) to regulate relations between publishers and platforms by forcing platforms to negotiate with publishers.

In order to keep the model parsimonious, we have abstracted away from several interesting considerations. In particular, the structure of the social network is notably absent from the analysis. An interesting avenue for future research would be to study environments where consumers have heterogeneous preferences and where the social

network exhibits homophily. We have also mostly abstracted away from users' behavior on the platform, regarding, for instance, their decision to share third-party content or to produce UGC. Finally, we have abstracted away from an in-depth interpretation of content quality. Recent research in the context of news (see e.g. Garz et al. (2018)) shows that the consumption of news is driven by a complex web of psychological factors. These considerations may shed different light on our results with respect to news quality in particular.

Our analysis focuses on the impact of a social network on publishers. We have illustrated our results for a particular content domain, news, because of its special importance for public life. However, our model readily applies to third-party publishers in other content domains (e.g. games, videos) who also seek to be present in consumers' 'newsfeed' on social media. Beyond social networks narrowly defined, the modeling framework also seems to be applicable to a broader set of interactions between multi-sided platforms and third-party 'content' providers.

For example, video distribution platforms such as Netflix, Hulu or Amazon Prime Video all offer third-party content alongside shows they produce themselves. Here, the role of newspapers is played by movie studios or TV networks who can monetize their content independently but are attracted by the platforms' customer base. In the e-commerce sector, Amazon.com also offers consumers the possibility to buy some products from third-party merchants or from Amazon itself. In the search engine market, Google often displays a result box that contains information produced by a website (e.g. a definition, the rating of a merchant or of a movie), with a small link to that website. While some consumers may click that link, many will simply stop there, having obtained the relevant information. This practice has been denounced by some websites (e.g. Yelp) who argue that it makes it difficult for them to attract visitors. This case raises issues of copyright protection, and also of efficiency gains.

These examples share some of the characteristics of our framework, and some of our insights might apply there as well. However, they also have some specific features, which are not captured by our model. This calls for future research in this area.²⁶

References

- Abreu, Luis and Doh-Shin Jeon (2020). "Homophily in Social Media and News Polarization". *TSE Working Paper*.
- Alaoui, Larbi and Fabrizio Germano (2020). "Time scarcity and the market for news". *Journal of Economic Behavior & Organization* 174, pp. 173–195.

²⁶In quite different contexts, Moorthy, Chen, and Tehrani (2018) and Hagiu, Jullien, and Wright (2020) also investigate hybrid forms of competition, where a firm distributes the product of its rivals.

- Ambrus, Attila, Emilio Calvano, and Markus Reisinger (2016). “Either or both competition: A” two-sided” theory of advertising with overlapping viewerships”. *American Economic Journal: Microeconomics* 8.3, pp. 189–222.
- Anderson, Simon P, Øystein Foros, and Hans Jarle Kind (2018). “Competition for advertisers and for viewers in media markets”. *The Economic Journal* 128.608, pp. 34–54.
- Anderson, Simon P and Damien J Neven (1989). “Market efficiency with combinable products”. *European Economic Review* 33.4, pp. 707–719.
- Angelucci, Charles and Julia Cagé (2016). “Newspapers in times of low advertising revenues”.
- Athey, Susan, Emilio Calvano, and Joshua S Gans (2018). “The impact of consumer multi-homing on advertising markets and media competition”. *Management Science* 64.3, pp. 1574–1590.
- Athey, Susan, Markus Mobius, and Jenő Pal (2017). “The Impact of Aggregators on Internet News Consumption”. *Working Paper, Stanford University* January.
- Berman, Ron and Zsolt Katona (2020). “Curation algorithms and filter bubbles in social networks”. *Marketing Science* 39.2, pp. 296–316.
- Berry, Steven and Joel Waldfogel (2010). “Product quality and market size”. *The Journal of Industrial Economics* 58.1, pp. 1–31.
- Bogart, Leo (2004). “Reflections on content quality in newspapers”. *Newspaper research journal* 25.1, pp. 40–53.
- Burguet, Roberto, Ramon Caminal, and Matthew Ellman (2015). “In Google we trust?” *International Journal of Industrial Organization* 39, pp. 44–55.
- Cagé, Julia (2016). *Saving the Media*. The Belknap Press of Harvard University Press, Cambridge, MA.
- Cagé, Julia, Nicolas Hervé, and Marie-Luce Viaud (2017). “The Production of Information in an Online World: Is Copy Right?”
- Calzada, Joan and Ricard Gil (2016). “What Do News Aggregators Do? Evidence from Google News in Spain and Germany”.
- Campbell, Arthur, C Matthew Leister, and Yves Zenou (2019). “Social media and polarization”. *Available at SSRN 3419073*.
- Chiou, Lesley and Catherine Tucker (2017). “Content aggregation by platforms: The case of the news media”. *Journal of Economics & Management Strategy* 26.4, pp. 782–805.
- Competition and Markets Authority (2020). “Online platforms and digital advertising market study”.
- De Cornière, Alexandre and Greg Taylor (2014). “Integration and search engine bias”. *RAND Journal of Economics* 45.3, pp. 576–97.
- De Cornière, Alexandre and Greg Taylor (2021). “Data and competition: a simple framework with applications to mergers and market structure”.

- Dellarocas, Chrysanthos, Zsolt Katona, and William Rand (2013). “Media, Aggregators, and the Link Economy: Strategic Hyperlink Formation in Content Networks”. *Management Science* 59.10, pp. 2360–79.
- Ershov, Daniel and Juan S. Morales (2021). “Sharing News Left and Right: The Effects of Policies Targeting Misinformation on Social Media”. *Working paper*.
- Fan, Ying. (2013). “Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market”. *American Economic Review* 103.5, pp. 1598–1628.
- Garz, Marcel, Sood Gaurav, Daniel F. Stone, and Justin Wallace (2018). “What drives demand for media slant?” *Working Paper*.
- Gentzkow, Matthew (2007). “Valuing new goods in a model with complementarity: Online newspapers”. *American Economic Review* 97.3, pp. 713–744.
- Gentzkow, Matthew, Jesse M Shapiro, and Daniel F Stone (2015). “Media bias in the marketplace: Theory”. *Handbook of media economics*. Vol. 1. Elsevier, pp. 623–645.
- George, Lisa and Christiaan Hogendorn (2012). “Aggregators, search and the economics of new media institutions”. *Information Economics and Policy* 24.1, pp. 40–51.
- (2013). “Local News Online: Aggregators, Geo-Targeting and the Market for Local News”.
- Hagiu, Andrei, Bruno Jullien, and Julian Wright (2020). “Creating platforms by hosting rivals”. *Management Science* 66.7, pp. 3234–3248.
- Jeon, Doh-Shin (2018). “Economics of News Aggregators”.
- Jeon, Doh-Shin and Nikrooz Nasr (2016). “News Aggregators and Competition among Newspapers on the Internet”. *American Economic Journal: Microeconomics* 8.4, pp. 91–114.
- Lacy, Stephen and Tom Rosenstiel (2015). “Defining and measuring quality journalism”. *New Brunswick, Rutgers*.
- Moorthy, Sridhar, Yongmin Chen, and Shervin Shahrokhi Tehrani (2018). “Selling Your Product Through Competitors’ Outlets: Channel Strategy When Consumers Comparison Shop”. *Marketing Science* 37.1, pp. 138–152.
- Mullainathan, Sendhil and Andrei Shleifer (2005). “The market for news”. *American Economic Review* 95.4, pp. 1031–1053.
- Peitz, Martin and Markus Reisinger (2015). “The Economics of Internet Media”. *Handbook of Media Economics* Volume 1A.Chapter 10, pp. 445–530.
- Pew Research Center (2019). “Americans are Wary of the Role Social Media Sites Play in Delivering the News”.
- Puglisi, Riccardo and James M Snyder Jr (2015). “Empirical studies of media bias”. *Handbook of media economics*. Vol. 1. Elsevier, pp. 647–667.
- Reuters (2017). “Reuters Institute Digital News Report”.
- (2020). “Reuters Institute Digital News Report”.

- Rolnik, Guy, Julia Cagé, Joshua Gans, Ellen Goodman, Brian Knight, Andrea Prat, Anya Schiffrin, and Prateek Raj (2019). “Protecting Journalism in the Age of Digital Platforms”. *George J. Stigler Center for the Study of the Economy and the State*.
- Roose, Kevin and Gabriel J.X. Dance (2020). “The Coronavirus Revives Facebook as a News Powerhouse”. *The New York Times*. URL: <https://www.nytimes.com/2020/03/23/technology/coronavirus-facebook-news.html>.
- Rutt, James (2011). “Aggregators and the News Industry: Charging for Access to Content”. *Net Institute Working Paper*. October, #11-19.
- Sismeiro, Catarina and Ammara Mahmood (2018). “Competitive vs. Complementary Effects in Online Social Networks and News Consumption: A Natural Experiment”. *Management Science*.

A Appendix - Omitted proofs

Proof of Lemma 4. For any $c < c^\dagger$, we have $\pi_{1L}(q_L(c), c) < \pi_{1H}(q_H(c), c)$ while the inequality is reversed for $c > c^\dagger$.

Because the actual profit of the newspaper is equal to π_{1L} for $q < \Lambda$ and to π_{1H} otherwise, we need to check whether $q_L(c) < \Lambda$ when $c > c^\dagger$ and whether $q_H(c) > \Lambda$ when $c < c^\dagger$. Moreover, we need to check that $q_H(c) \leq 1$, because the profit expression is only valid for $q \leq 1$ (consumers cannot consume more news than $x_N = 1$).

Notice that c^\dagger is non-monotonic in Λ , and that $c^\dagger > 2\beta$ for $\Lambda < \frac{1}{2}$.

Consider $c \in (2\beta, c^\dagger)$. Straightforward algebra reveals that $q_H(c)$ is then in $(\Lambda, 1)$, so that it is the optimal quality. Consider now $c > c^\dagger$. The solution is then $q_L(c) = \phi/c$, provided that it is smaller than Λ . This is true because we have $c > c^\dagger$ and $c^\dagger > \phi/\Lambda$. ■

Proof of Proposition 1. (i) With or without content bundling, news consumption is equal to q . Because under content bundling part of the consumption occurs through the platform, revenue is always lower under content bundling *for a given* q . Let $\widehat{R}_1(q)$ and $R_1^*(q)$ be the respective revenue functions in the benchmark and with content bundling for a given q . We then have, for $q = q^*$, $\widehat{R}_1(q^*) > R_1^*(q^*)$. Subtracting $C(q^*)$, we obtain $\widehat{R}_1(q^*) - C(q^*) > \pi_1^*$. By revealed preferences, $\widehat{\pi}_1 \geq \widehat{R}_1(q^*) - C(q^*)$, which implies that $\widehat{\pi}_1 > \pi_1^*$.

(ii) Because $\beta > \phi_1$, it is straightforward to check that $q_L(c) = \frac{\phi_1}{c} < \hat{q} = \frac{\beta}{c}$ and that $q_H(c) = \frac{\beta - \Lambda\phi_1}{c(1-\Lambda)} > \hat{q}$. Recall from the proof of Lemma 4 that when $\Lambda > \frac{\beta + \phi_1}{\beta + \phi_1 + \sqrt{\beta^2 - \phi_1^2}}$, $c^\dagger < 2\beta$, so that c is necessarily larger than c^\dagger and $q^* = q_L(c)$. When $\Lambda \leq \frac{\beta + \phi_1}{\beta + \phi_1 + \sqrt{\beta^2 - \phi_1^2}}$ we have $c^\dagger \geq 2\beta$ and $q^* = q_L(c)$ if $c \geq c^\dagger$ and $q^* = q_H(c)$ otherwise. Because consumers always get their optimal content mix, consumer surplus moves in the same direction as quality of news. ■

Proof of Proposition 4. If a publisher opts in, the analysis is identical to our baseline model. Let us thus look at the case where publisher i opts out. If the platform chooses λ_{-i} and the quality is q_i , maximizing utility leads i 's regular readers to spend $T_0(\lambda_{-i}, q) = \max\{\min\{\frac{1-q-(1-q)\lambda_{-i}}{(1-\lambda_{-i})^2}, 1\}, 0\}$ on the platform and $1 - T_0(\lambda_{-i}, q_i)$ on the i 's site.

The platform's profit with respect to these consumers is $\pi_0(\lambda_{-i}, q_i) = (1 - \lambda_{-i} + \lambda_{-i}\phi_0)T_0(\lambda_{-i}, q_i)$. If $q_i > (1 + \underline{q})^2/4$ then $\frac{1-q_i-(1-q)\lambda_{-i}}{(1-\lambda_{-i})^2} < 1$, which means that $T_0(\lambda_{-i}, q_i) = \max\{\frac{1-q_i-(1-q)\lambda_{-i}}{(1-\lambda_{-i})^2}, 0\}$. Because $T_0 = 0$ cannot be optimal for the platform, its best-response either solves $\frac{\partial \pi_0(\lambda_{-i}, q_i)}{\partial \lambda_{-i}} = 0$ or hits the constraint $q_i \leq \Lambda$, i.e. $\lambda_{-i}(q_i) = \min\{\frac{\underline{q} + \phi_0 - q_i(1 + \phi_0)}{\underline{q}(1 - 2\phi_0) - q_i(1 - \phi_0) + \phi_0}, \Lambda\}$.

If $q_i \leq (1 + \underline{q})^2/4$, the platform can induce $T_0 = 1$, by choosing $\lambda_{-i} = \frac{1 - \sqrt{(1 + \underline{q})^2 - 4q_i}}{2}$. It never wants to choose a higher λ_{-i} . It may want to choose a lower level if profit is

maximized at $T_0 < 1$ (as in the case above), or if the constraint $\lambda_{-i} \leq \Lambda$ binds. So $\lambda_{-i}(q_i) = \max\{\min\{\frac{\underline{q} + \phi_0 - q_i(1 + \phi_0)}{q(1 - 2\phi_0) - q_i(1 - \phi_0) + \phi_0}, \Lambda, \frac{1 - \sqrt{(1 + \underline{q})^2 - 4q_i}}{2}\}, 0\}$.

Inspecting the two expressions for the platform's best-response, one can see that, if $q_i > \frac{\underline{q} + \phi_0}{(1 + \phi_0)}$, $\lambda_{-i}(q_i) = 0$. Therefore, if the benchmark quality $\hat{q} \geq \frac{\underline{q} + \phi_0}{(1 + \phi_0)}$, i.e. if $c_i \leq \frac{\beta(1 + \phi_0)}{\underline{q} + \phi_0}$, it is an equilibrium for publisher i to opt out and set $q_i = \hat{q}$, while $\lambda_{-i} = 0$. This proves part 1.

Now suppose that c_i is such that the benchmark quality, $\hat{q} \equiv \frac{\beta}{c_i}$, is equal to $\frac{\underline{q} + \phi_0}{(1 + \phi_0)} - \epsilon$, where $\epsilon > 0$ is very small. In other words, suppose that c_i is slightly above $\frac{\beta(1 + \phi_0)}{\underline{q} + \phi_0}$. $\lambda_{-i} = 0$ cannot be an equilibrium, because then the publisher would choose $q_i = \frac{\beta}{c_i}$, but $\lambda_{-i}(\frac{\beta}{c_i}) > 0$. Suppose that the publisher expects λ_{-i} to be positive but small, so that consumers spend time on both the platform and the publisher's site. Then we have $T_1(\lambda_{-i}, q_i) = 1 - \frac{1 - q_i - (1 - \underline{q})\lambda_{-i}}{(1 - \lambda_{-i})^2}$, and the publisher's best-response is $\operatorname{argmax}_{q_i} \{\beta T_1(\lambda_{-i}, q_i) - c_i \frac{q_i^2}{2}\} = \frac{\beta}{c_i(1 - \lambda_{-i})^2}$. For ϵ sufficiently small, the equilibrium of the case with opt-out is thus the solution to the system

$$\lambda_{-i}^* = \frac{\underline{q} + \phi_0 - q_i^*(1 + \phi_0)}{q(1 - 2\phi_0) - q_i^*(1 - \phi_0) + \phi_0} \quad \text{and} \quad q_i^* = \frac{\beta}{c_i(1 - \lambda_{-i}^*)^2}$$

As ϵ goes to zero, notice that the profit converges to the benchmark profit, which is strictly higher than the profit under opt-in. Therefore, by continuity, there exists a \tilde{c} such that, for any c in $(\frac{\beta(1 + \underline{q})(1 + \phi_0)}{\phi_0}, \tilde{c})$, the premium publisher prefers to opt out. Notice that, because $\lambda_{-i}^* > 0$, we have $q_i^* > \frac{\beta}{c_i} = \hat{q}$. This proves part 2.

If publisher i opts out and chooses a quality too close to \underline{q} , the presence of other publishers on the platform is enough to induce $T_0 = 1$, so that the publisher's profit is negative. More precisely, this happens when q_i is below the level \tilde{q} such that $\frac{1 - \sqrt{(1 + \underline{q})^2 - 4\tilde{q}}}{2} = \frac{\underline{q} + \phi_0 - \tilde{q}(1 + \phi_0)}{q(1 - 2\phi_0) - \tilde{q}(1 - \phi_0) + \phi_0}$ (for any $q \leq \tilde{q}$ the platform sets λ_{-i} such that $T_0 = 1$). One can show that $\tilde{q} > \underline{q}$. Take c_i such that, for any $q_i \geq \tilde{q}$, $\beta q_i - c_i q_i^2/2 \leq 0$. Publisher i is then necessarily better-off opting in onto the platform. This proves part 3.

Finally, we need to prove that some of the newspapers who opt in have $q_i > \underline{q}$. If this was not the case, and if the maximal quality among opting-in newspapers was \hat{q} , a newspaper could opt-in and chose $q > \hat{q}$ and become the sole provider of news through the platform, which would allow it to reach all the regular readers of opting-out newspapers, which would be profitable. We checked numerically that there exist combinations of parameters such that the four regions described are non-empty. For other parameters, the opt-in threshold is larger than c^\dagger so that all newspapers who opt in reduce their quality. ■

Proof of Lemma 5. First, one can check that $q_1(c) > q_2(c)$ and that profit is continuous. If $\Lambda < q_2(c)$, profit is increasing over $[0, q_2(c)]$ and decreasing afterwards. The condition $\Lambda < q_2(c)$ can be rewritten as $c < \frac{\beta(2 - \Lambda) - \Lambda}{2(1 - \Lambda)\Lambda}$. If $\Lambda > q_1(c)$ (i.e. $c > \frac{1 + \beta}{2\Lambda}$), profit is increasing

over $[0, q_1(c)]$ and decreasing afterwards. Finally, for $\Lambda \in [q_2(c), q_1(c)]$, profit is increasing over $[0, \Lambda]$ and decreasing afterwards.

Note that $q_1(c) < 1$ for all $c > 2\beta$ if $\beta > 1/3$, which we have assumed, and that $q_2(c) < q_1(c)$ for all c , so that, in equilibrium, quality is indeed between 0 and 1. ■

Proof of Proposition 5. Part 1: Negotiated content bundling (NCB) versus benchmark.

As a preamble, one can readily check that $q_2(c) < \hat{q}(c) \equiv \beta/c < q_1(c)$.

The only possibilities for quality to be higher under NCD are (1) that $\tilde{q} = q_1(c)$ or (2) that $\tilde{q} = \Lambda$ and $\Lambda > \hat{q}$. The second case happens when $c > \frac{\beta}{\Lambda}$, while the first happens when $c > \frac{1+\beta}{2\Lambda}$, the latter term being larger than $\frac{\beta}{\Lambda}$. So the relevant threshold is $\frac{\beta}{\Lambda}$.

Part 2: Negotiated versus unilateral content bundling (UCB).

We need to distinguish cases depending on whether c^\dagger is below $\frac{\beta(2-\Lambda)-\Lambda}{2(1-\Lambda)\Lambda}$, between $\frac{\beta(2-\Lambda)-\Lambda}{2(1-\Lambda)\Lambda}$ and $\frac{1+\beta}{2\Lambda}$, or above $\frac{1+\beta}{2\Lambda}$. In each case, the relevant comparisons reveal that quality is higher under NCB if and only if $c > c^\dagger$. ■

Proof of Lemma 6. Adding visitors across sites, the publisher's profit reads as follows:

$$\begin{aligned} \pi_1(q) = & \underbrace{\alpha(1-\phi) \int_0^{1-q} \hat{x}(\theta, q) d\theta}_{(i)} + \underbrace{(1-\phi)(1-\alpha) \int_0^{\bar{\lambda}(q)-q} \bar{\lambda}(q) d\theta}_{(ii)} \\ & + \underbrace{(1-\alpha) \int_{\bar{\lambda}(q)-q}^{1-q} (\bar{\lambda}(q)(1-\phi)t_0(\theta, q, \bar{\lambda}(q)) + \beta t_1(\theta, q, \bar{\lambda}(q))) d\theta}_{(iii)} + \underbrace{\beta \int_{1-q}^1 d\theta}_{(iv)} \quad (12) \\ & - C(q). \end{aligned}$$

Term (i) in equation (12) corresponds to the consumers that the platform identifies and to whom it shows a tailored mix of content ($\lambda(\theta, q) = \theta + q$), inducing them to spend all their time on the platform. The newspaper's revenue over these consumers is $(1-\phi)$ per unit of time. Term (ii) represents the consumers that the platform does not identify and who have a relatively low type ($\theta < \bar{\lambda}(q) - q$): these consumers spend all their time on the platform, are exposed to a share $\bar{\lambda}(q)$ of news (which is more than enough to satisfy their demand for news), over which the newspaper gets a revenue $(1-\phi)$ per-unit of time. Term (iii) corresponds to the consumers who are not identified by the platform and who wish to complement the news they get from the platform by direct traffic to the newspaper. They spend $t_0(\theta, q, \bar{\lambda}(q)) = \frac{1-\theta-q}{1-\bar{\lambda}(q)}$ units of time on the platform's website, generating revenue $\bar{\lambda}(q)(1-\phi)$ per unit of time for the newspaper, and $t_1(\theta, q, \bar{\lambda}(q)) = 1 - t_0(\theta, q, \bar{\lambda}(q))$ units of time on the newspaper's website, generating a revenue of β per unit of time. Finally, term (iv) represents the consumers who only visit the newspaper's website.

Given that $\bar{\lambda}(q) = q + \frac{\phi}{2(1-\phi)}$, after some algebra we obtain the Lemma. ■

Proof of Proposition 6. When $\beta = 1 - \phi$, $\pi_1(q^*) - \tilde{\pi}_1 = \frac{(1-\alpha)\phi^2}{8(1-\phi)} \geq 0$.

When $\beta = 1$ the newspaper is necessarily worse-off under content bundling. Indeed, in such a case the industry revenue is constant and equal to 1. Let $\tilde{R}_0(q)$ and $\tilde{R}_1(q)$ be the platform's and the newspaper's revenue without content bundling if news quality is q , and $\hat{R}_0(q)$ and $\hat{R}_1(q)$ their revenues under content bundling with quality q . Because $R_0 + R_1 = 1$ and $\hat{R}_0(q) > \tilde{R}_0(q)$ for any q (optimal content bundling necessarily increases the platform's revenue), we have

$$\hat{\pi}_1(q^*) \equiv \hat{R}_1(q^*) - c(q^*) < \tilde{R}_1(q^*) - c(q^*) \leq \tilde{R}_1(\tilde{q}) - c(\tilde{q})$$

where the last inequality comes from a revealed-preferences argument.

Last, the difference between the newspaper's profit with and without content bundling is decreasing in β . ■