Paying for news diversity? A topic diversity analysis of free and paywalled online news

Authors' details

Glen Joris (corresponding author)

PhD candidate, imec-mict-UGent, Ghent University (Belgium)

E-mail: glen.joris@ugent.be

ORCID: 0000-0002-4202-2641

Jonathan Hendrickx

Postdoctoral researcher, imec-SMIT, Vrije Universiteit Brussel (Belgium)

E-mail: jonathan.hendrickx@vub.be

ORCID: 0000-0003-2802-2802

Stefaan Vercoutere

Researcher, imec-WAVES-UGent, Ghent University (Belgium)

E-mail: stefaan.vercoutere@ugent.be

ORCID: 0000-0002-2035-256X

Orphée De Clercq

Postdoctoral researcher, Language and Translation Technology Team (Belgium)

E-mail: orphee.declercq@ugent.be

ORCID: 0000-0002-6090-5552

Lieven De Marez

Researcher director, imec-mict-UGent, Ghent University (Belgium)

E-mail: lieven.demarez@ugent.be

ORCID: 0000-0001-7716-4079

Research details

1. Acknowledgements: We would like to thank DPG Media and Mediahuis for their input and support.

2. Funding: This work was supported by Ghent University under grant BOFGOA2018000601.

3. Disclosure statement: No potential conflict of interest was reported by the authors.

4. Data availability statement: The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Paying for news diversity? A topic diversity analysis of free and paywalled online news

Abstract: Media companies increasingly place (parts of) their online news content behind to compensate for shrinking digital advertising revenues. However, despite its prominence in the contemporary news environment, very few studies have assessed the effects of this practice on the diversity in the actual content provided to people. To fill this knowledge gap in scholarship, this study examines how diverse free and paid content is in several newspapers in Flanders (Belgium). Using a topic classification model and quantitative analysis of 287,856 online news articles, we differentiate between paywalled and non-paywalled articles, popular and quality news brands, two distinct media corporations and two diversity benchmarks: open and reflective diversity. Our findings reveal that in most cases, online news content behind paywalls is indeed more diverse, although the differences with freely available articles are often minimal. We also find that quality news titles do not necessarily score higher on topic diversity than their popular counterparts. These findings are contextualised within the Flemish media market as well as existing scholarship.

Keywords: Computational methods: content analysis; Flanders; IPTC; news diversity; paywalls

Introduction

News publishers are increasingly realizing that an advertising-based business model as well as an exclusive focus on user engagement is not financially sustainable (Newman et al., 2020). This realization came into place in the late 2010s when news publishers' digital advertising revenues were increasingly shrinking due to profound changes in consumption patterns and economics of advertising. Since then, news audiences have increasingly started to use ad blockers, disabling news organization's ability to track and profile users around the Internet ([AUTHOR]). Similarly, news consumption moved to smaller mobile screens and became more promiscuous in time and space (Van Damme et al., 2020), while platforms such as Facebook and Google began to rise in popularity, creating dominance in the advertising markets and leaving less room to compete (Meijer & Kormelink, 2019). As Bodó (2019) explains, all these developments made the ad-based free news business models less appealing and forced news publishers to study other revenue models such as print or digital subscriptions.

Due to this increased interest in the digital subscription model, the use and development of online paywalls became significant. While several paywall types and strategies exist and are still being optimized (i.e., via data, artificial intelligence and algorithms, see Newman et al., 2020), the widespread use of paywalls has accelerated concerns among scholars about what this implies for society in general and the public interest in specific. Sjøvaag (2016a) and Pickard (2014), for instance, argued that paywalls segregate the mass market from well-paying audiences accessing high-value content. As such, paywalls may "defy the internet principle of openness, disenfranchising people who are not able to afford such subscriptions" (Sjøvaag, 2016a, p. 7). Moreover, paywalls could also inscribe more commercial values into the news production and publication process, which, in turn, could constrict the scope of voices and views presented in the news and negatively affect the diversity that is provided to audiences (Pickard, 2014; Sjøvaag, 2016a).

Although these concerns have found their way into journalism studies by several descriptive studies on the "quality" of content behind paywalls (e.g., Benson, 2019, Myllylahti, 2017; Olsen & Solvoll, 2018; Sjøvaag, 2016a), less attention is being paid to how paywalls affect the diversity of content provided to people. This is remarkable, particularly because diversity is considered as one of the most important quality conditions for a healthy democracy and an effective public sphere. Following the Habermasian normative idea(l)s, everyone is argued to have access to the public sphere; to the various positions and

arguments people have, and to debate and exchange ideas about what the most appropriate course of action is (Habermas, 1989). Only then will people be able to find democratic consensus. Potential obstructions such as paywalls should therefore be critically assessed and examined in scholarly literature and particularly in journalism studies in which paywalls are a growing field of research.

In this article, we take the opportunity to examine how and to what degree paywalls impact the diversity in content people are exposed to. By examining how diverse content is in front of and behind paywalls, we aim to develop new insights into paywalls from a more societal perspective. We thereby also contribute to a growing field in (digital) journalism studies in which diversity is being assessed (e.g., [AUTHOR]; [AUTHOR]). However, unlike other news diversity studies in the news environment (e.g., Carpenter, 2010; van Hoof et al., 2014), we primarily focus on diversity in front of and behind paywalls. As such, we also aim to bring further nuance into existing general knowledge on news diversity, and specific knowledge on topic and content diversity.

As an appropriate case study, we use the small, highly concentrated media market of Flanders (Belgium), an affluent Dutch-speaking region with its own media market and regulatory bodies. For our assessment of content diversity in front of and behind paywalls, we use a vast data set of online news articles (n = 287,856) published by the biggest popular and quality news titles of the two most dominant media companies in Flanders. Due to shared media ownership structures with various other European countries and similar challenges news organisations around the world face when it comes to increasing subscriptions and paywalling segments of their online news output, we argue that the results of our Flemish case study are to large extents applicable as well to other media markets.

Literature review

We first discuss the historical emergence of paywalls in the news industry context. Next, we discuss how paywalls have become part of the editorial process. To understand why paywalls might affect the diversity people are confronted with, we end our literature review with a comprehensive discussion on paywalls' challenging relationship with news diversity and the public sphere.

Paywalls' historical emergence

Although paywalls are a distinctive feature of today's news environment, they are not new. Their history dates to the mid-1990s, when the first American newspapers went online and attempted to charge monthly subscription fees that ranged between 5 and 12 dollars. To do so, they developed a 'hard barrier' between the users and the publisher's content (Pickard, 2014). At the end of the 2000s, major international news titles such as The Financial Times or The New York Times incorporated the same subscription model online. Most publishers followed the 'hard' strategy popularised by the Wall Street Journal, but other types became prevalent too, including a 'softer' approach in which people could access five to ten articles before payment was required (i.e., soft paywall, see Myllylahti, 2019).

The reason why news publishers collectively switched to a digital subscription model could be found in the global financial crisis of 2007-2008 (Arrese, 2016). Back then, many news publishers faced large difficulties to sustain their financial business model. Formerly, their revenue model was primarily built on print advertising and print subscriptions, but these became inadequate when advertisers pulled out and sales revenue plummeted. As a result, many news publishers introduced paywalls to their digital channels, as these were at that time still free. Paywalls ensured that access was only permitted to those who paid and had a digital subscription.

Although paywalls were initially installed to compensate for the losses in print subscriptions and advertising revenues, they slowly became an essential part of news organizations' revenue model

towards the end of the 2010s. This trend became particularly visible from 2017 onwards, when publishers started to note the difficulties of the growing advertising competition with large platforms such as Facebook and Twitter. It forced publishers to pay more attention to the relationship with the paying customer (Meijer & Kormelink, 2019).

As a result, the use of paywalls has become commonplace in media markets globally, although the frequency and intensity of its use tends to vary per country and/or media company, each with their own characteristics. Strategies tend to shift rapidly over time, which renders comparing findings of peer-reviewed scholarly work as an intricate task. For instance, a 2020 study looking at 205 French news outlets revealed that 49% did not have a paywall at all and that a hard paywall was the least common type (Cagé et al., 2020). Meanwhile, in Flanders, and in 2022, all legacy private news brands have different types of paywalls, with the public service broadcaster an obvious yet notable exception. Even online-only news outlets have implemented subscription models with some sections of their articles placed behind a paywall.

Paywalls, editorial decisions and perceptions

A few case studies have assessed which types of news articles were typically more 'locked' behind a paywall or freely available to all online readers, at various online newspapers and at different yet recent points in time. For instance, two separate Norwegian studies assessing changes in news content before and after the introduction of a paywall revealed two similar traits (Kvalheim, 2013; Sjøvaag, 2016a). First, there were hardly any discernible changes in the ratio of different types of news articles in websites. This means that the content was not modified to accommodate the paywall, but it rather became integrated in existing content types and strategies. Second, (hyper)local news output was much more prone to being placed behind a paywall due to its exclusivity, whereas press agency and international articles tended to be freely accessible (Kvalheim, 2013; Sjøvaag, 2016a). The last point was also echoed in the case study of Australasian financial newspapers by Myllylahti (2017). However, she also found big differences in the ratio of accessible and paywalled articles, noting that the plurality in ownership structures and publication models play pivotal parts in approaching paywall strategies among news media. She also argued that this 'may enhance the commodity nature of ordinary news if the paper's content becomes over-reliant on syndicated material, or material sourced from other news sites' (p. 468).

Paywalls with the aim of enhancing online revenues were increasingly rolled out alongside with the adoption of user analytics within newsrooms. Scholarship has shown that they tend to go hand in hand in contemporary newsrooms and have become firmly integrated into news publishers' financial strategies (Cherubini & Kleis Nielsen, 2016) as well as newsroom strategies and editorial decision making based on available user data. In their study of Belgian social media editors and how they perceive and use user analytics in their daily work, Lamot & Paulussen (2019, p. 366) explain how one anonymous chief editor 'had to redirect incentives because journalists became reluctant to publish their stories behind the paywall as such stories never appear in the most-viewed list'. This example signifies that the perception of paywalls among journalists is not necessarily a positive one, which can also have ramifications for editorial decision-making. Articles behind paywalls reach smaller audiences and can therefore be considerably less impactful and shareable in today's fast-paced media environment, which is nearly by default digital and ephemeral ([AUTHOR]).

Paywalls, news diversity and the public sphere

The rise of online news and alternative business models, in which paywalls play a vital part, has fostered scholarly debate on the heightened commodification of news content and its effects on news diversity as produced and as consumed. Various scholars have discussed and presented paradigms for

media or news diversity, each focusing on other aspects of it (see [AUTHOR] for a systematic literature review). For instance, [AUTHOR] presented news diversity as a concept that operates at three different levels: micro, meso and macro. The micro level of individual news titles and their produced output (e.g. news articles, television broadcasts) have been frequently conceptualised as either vertical and horizontal diversity (Entman & Wildman, 1992), internal and external diversity (McQuail & Van Cuilenburg, 1983) and outlet and output diversity (Van Cuilenburg, 1999). Other peers took a more holistic perspective by integrating meso and macro levels, and taxonomized news diversity as comprising several dimensions and/or functions, ranging from ownership, source, content and reception diversity (Napoli, 1999) to organisational, structural, production, output and exposure diversity (Sjøvaag, 2016b). These two broader classifications later culminated, for instance, in a holistic model of news diversity which encompasses ownership, brand, production, content and consumption diversity, of which the five dimensions are argued to act as metaphorical cogs of a machine and influence each other organically in a myriad of ways ([AUTHOR]).

It is generally accepted within and beyond academia that a diverse set of news titles, sources and actors is paramount to maintain both the unique position of journalism as the fourth estate and the possibility of citizens to use their agency in political decision-making processes ([AUTHOR]; Baker, 2007; Papandrea, 2006; Sjøvaag, 2016b). In recent years, increased media ownership consolidation has caused concerns over far-reaching synergy operations of news titles (Bleyer-Simon et al., 2021; Noam, 2016), which can include sharing news content with possible negative effects on the diverse set of news content available to citizens. Quantitative content analyses executed in a number of nations have indeed established a negative relationship between ownership and content diversity ([AUTHOR]; [AUTHOR]; Badr, 2021; Beckers et al., 2017; Vogler et al., 2020), though there are also examples which have shown no deterioration following ownership structure changes, but rather stability (Sjøvaag, 2014; Skärlund, 2020).

In the Habermasian public sphere idea(l), news media mediate between politics and citizens by, put succinctly, informing the latter on the proceedings of the former. This gives the media the power to frame certain issues in specific ways, to set news agendas and to foster public accountability (Calhoun, 1992; Habermas, 2006; Hahn, 2007). As Badr (2021) argues, the public sphere then 'consists of a cluster of different messages originated and distributed by media professionals' (p. 15). The author also suggests that both media ownership and content are viable means to measure the degree by which public sphere participation can be measured. We venture that this ties in perfectly within news diversity theory as well, as the diversity of news production, content or consumption is always subject to change following ramifications in wider media or socio-political structures and systems. Following this line of thought, the introduction of paywalls makes for an interesting case as it can limit the access to diverse information to non-paying consumers and can thereby alter public sphere participation for better or for worse.

Thus far, we have shown that various studies have assessed effects of paywalls on news content, while another body of literature has focused on content analyses in the light of media ownership consolidation. To the best of our knowledge, no study has combined these two distinct viewpoints, which is precisely what we contribute to existing scholarship. Using news diversity and public sphere theories and computational analysis, we establish and explore the relationship between two recently merged Belgian media corporations' online paywall strategies across their various popular and quality news titles. We seek to investigate empirically whether paid content offers more actor and viewpoint diversity than free content.

As a starting point for this investigation, we formulate two news diversity hypotheses for this study in line with findings outlined in the literature review above. As hard news topics such as Economy and

Politics are primarily used by news publishers as 'a saleable commodity' (Myllylahti, 2017, p. 469), we expect that hard news topics are less found in free content offers and popular newspapers, which are oftentimes more focused on soft news topics such as Lifestyle and Human Interest, and subsequently have lower news diversity scores.

H1: News brands' paid content is more diverse in terms of topics than their free content offer

H2: Quality newspapers will have higher topic diversity scores than popular newspapers

Methodology

Sample and data collection procedure

To construct our sample, we exclusively looked at news brands in Flanders, the northern and Dutch-speaking part of Belgium, that use paywalls online and thus separate paid content from free content. To keep the scope of our sample focused, we concentrated on the two main news corporations and their respective quality and popular newspapers. The Flemish Media Regulator, in independent governmental body assessing media concentration, has noted how a string of mergers and acquisitions in recent years, has caused the number of Flemish companies owning legacy news outlets to decrease from seven to four in the past decade (Vlaamse Regulator voor de Media, 2020). The two main corporations, DPG Media and Mediahuis, are the results of said M&A practices, and have in recent years expanded their ownership to various other European media markets, including the ones in the Netherlands, Ireland, Denmark and Luxembourg.

For DPG Media, we looked at online news articles published by newspapers 'De Morgen' and 'Het Laatste Nieuws', while for Mediahuis, we took the titles 'De Standaard' and 'Het Nieuwsblad' into account. We base our classification of popular vs. quality titles on existing scholarship (Beckers et al., 2017; Lamot & Paulussen, 2019). For these four newspapers, we collected all online news articles published between December 2020 and August 2021. Table 1 below shows the number of articles assessed per medium and indicates that both popular titles publish considerably more articles, a finding which was already established in previous studies assessing Flemish online news outlets ([AUTHOR]). There is no specific rationale that underlies this time period, as we were not interested in the potential effect of a certain news event. However, we are aware that certain (large) news events that happened during this time period could affect our research results (e.g., COVID-19). We acknowledge this as a limitation in our study design.

Table 1. Number of articles assessed per newspaper, ranked by type and company.

Newspaper	Type	Company	n
Het Nieuwsblad	Popular	Mediahuis	112,581
De Standaard	Quality	Mediahuis	13,384
De Morgen	Quality	DPG Media	14,135
Het Laatste Nieuws	Popular	DPG Media	144,756

There were no inclusion or exclusion criteria on an article-based level, except for the fact that we removed articles of which the paywall-status (i.e., paid or free) was unclear. This resulted in a sample of 310,000 news articles before data cleaning, and 287,856 news articles after. To manage the news retrieval process, we used an online database that retrieved all news articles via RSS feeds, API's and web scraping. Articles belonging to Mediahuis were retrieved via RSS feeds and web scraping, while articles from DPG Media were gathered via APIs. To label all news articles, we operationalized a multi-label classification system developed by [AUTHOR]. This system uses deep learning and outputs topics

for each article, based on general standards that are expressed by IPTC (i.e., International Press Telecommunications Council, n.d.). The IPTC topics include 17 categories, ranging from Sports and Weather to Health and Education.

Based on the IPTC output and the highest confidence score for each of these categories, the main topic of each news article was determined (automatically). To train the classifier, we were given a historical data dump of DPG media, comprising over 200,000 news articles tagged with IPTC topics. By relying on state-of-the-art machine learning techniques, more specifically BERTje (de Vries et al. 2019), a Dutch version of the pre-trained transformer model BERT (Devlin et al. 2019), a highly accurate model was built and applied to the dataset used in this paper (see [AUTHOR]). To make this approach and its impact on our study more concrete, we added examples of articles from our data set per established IPTC news topic to the manuscript's appendix (see Table 10).

Measures and data analysis

Measuring news diversity is not an easy task, particularly on a large scale and with computational methods that are limited to certain practical boundaries ([AUTHOR]). In this article, we decided to focus on IPTC topics, as we argue, amongst others, that news audiences should be provided with a broad range of topics (Makhortykh et al., 2021). To measure topic diversity, we used a standardized measure that calculated the similarity between a news brands' topic distribution and the ideal topic distribution. The result is a score between 0 and 1, in which 1 stands for high similarity between the topic distribution and the ideal topic distribution. To determine the ideal topic distribution, we looked at two normative concepts that are concerned with the optimal outcome of diversity: reflective diversity and open diversity (McQuail & Van Cuilenburg, 1983). Open diversity claims diversity as an equal representation of all possible categories, whereas reflective diversity argues that diversity should reflect the proportions in society (or in the journalistic offer in general if the proportions in society are not known). In Figure 1 (in Appendix), a concrete example for both concepts is presented. As both concepts have other normative implications (see [AUTHOR]), we calculated two measures: (1) *measured open diversity* and (2) *measured reflective diversity*. Hence, these measures both measure topic diversity, but primarily differ in how they should be interpreted. In the former, a score of 1 means that the offer is boasting an equal representation of all possible categories, while in the latter a score of 1 means that the news brands' offer is equal to the proportions that are present in society or in the journalistic offer in general. The technical details of these measures can be found in the Appendix (see Figure 1).

Results

Our results section is divided into two main parts. First, we elaborate on our descriptive analysis of paywalled articles per newspaper based on the IPTC news topic classification. Next, we discuss the measured open and measured reflective diversity for free and paid content of all four newspapers assessed. In both sections, we group the newspapers per type (popular vs. quality) and per media company (Mediahuis vs. DPG Media).

Descriptive analysis: topic distribution

Tables 2 and 3 below respectively show the topic distribution for each of the four newspapers separately as well as in total, together with the absolute numbers and relative row percentages of all IPTC categories for the four news brands. More specifically, Table 2 reveals the share of each article type and whether it is paywalled or not, whereas Table 3 shows ratios for the latter with percentages on the accessibility of articles per IPTC category. For instance, Table 2 indicates that paywalled economic news articles make up for 18.3% of the total number of articles from 'Het Nieuwsblad'

(Mediahuis' main news brand), but that the same portion of the total data set corresponds with 62.5% of all economic-related news articles which appeared at the same news website (Table 3).

In raw numbers, sports news constitutes the biggest share of online articles in our data set, with nearly 50,000 articles on this topic. There is clearly a long tail of five outspokenly more popular IPTC categories, with Economy, Crime, Politics and Arts along with Sports as the most recurrent topics. Although the order in ranking differs per newspaper, it is noteworthy that all four different newspaper titles, in spite of their different profiles and company ownership structures, share these five main news topics on their websites between December 2020 and August 2021.

Table 3 also reveals the total shares of paywalled articles per newspaper. We denote large differences across newspapers of various types and companies, indicating a lack of comprehensive policy at the brand or overall media company level. Whereas Mediahuis' popular title 'Het Nieuwsblad' has over half of its articles paywalled, its main rival, DPG Media-owned 'Het Laatste Nieuws' offers 86.1% of its online news output free of charge. Interestingly, across company lines this situation is inverted for the quality titles, where 'De Morgen' (DPG Media) paywalls considerably more articles than 'De Standaard' (Mediahuis).

Table 2. Relative distribution percentages of all IPTC categories for the four news brands.

	Het Nieuwsblad				De Standaard				De Morgen				Het Laatste Nieuws				Total			
	Free		Paid		Free		Paid		Free		Paid		Free		Paid		Free		Paid	
	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
Economy	6536	12.3%	10911	18.3%	1144	11.5%	565	16.6%	1215	14.4%	591	10.4%	22465	17.7%	3624	17.7%	31360	15.8%	15691	17.6%
Sports	14281	27.0%	7028	11.8%	2169	21.7%	337	9.9%	877	10.4%	705	12.4%	21236	16.7%	2559	12.5%	38563	19.4%	10629	11.9%
Crime	6434	12.1%	7369	12.4%	973	9.7%	262	7.7%	834	9.9%	314	5.5%	14674	11.5%	2031	9.9%	22915	11.5%	9976	11.2%
Politics	5241	9.9%	6701	11.2%	1852	18.6%	887	26.1%	1890	22.4%	992	17.4%	13419	10.5%	1888	9.2%	22402	11.3%	10468	11.7%
Arts	4524	8.5%	6135	10.3%	834	8.4%	388	11.4%	732	8.7%	1137	20.0%	12904	10.1%	2374	11.6%	18994	9.6%	10034	11.3%
Accidents	4023	7.6%	4728	7.9%	504	5.1%	68	2.0%	293	3.5%	70	1.2%	7987	6.3%	462	2.3%	12807	6.4%	5328	6.0%
Society	1947	3.7%	3745	6.3%	288	2.9%	160	4.7%	312	3.7%	398	7.0%	7224	5.7%	1412	6.9%	9771	4.9%	5715	6.4%
Hum. interest	3341	6.3%	3134	5.3%	523	5.2%	109	3.2%	288	3.4%	362	6.4%	7180	5.6%	2151	10.5%	11332	5.7%	5756	6.5%
Health	1955	3.7%	2200	3.7%	705	7.1%	339	10.0%	1041	12.3%	455	8.0%	5292	4.2%	1170	5.7%	8993	4.5%	4164	4.7%
Education	1084	2.0%	2580	4.3%	154	1.5%	48	1.4%	142	1.7%	99	1.7%	4141	3.3%	474	2.3%	5521	2.8%	3201	3.6%
Lifestyle	1505	2.8%	2812	4.7%	91	0.9%	61	1.8%	133	1.6%	198	3.5%	5900	4.6%	1792	8.8%	7629	3.8%	4863	5.5%
Religion	339	0.6%	564	0.9%	70	0.7%	29	0.9%	67	0.8%	43	0.8%	707	0.6%	114	0.6%	1183	0.6%	750	0.8%
Environment	569	1.1%	987	1.7%	119	1.2%	44	1.3%	139	1.6%	76	1.3%	1972	1.5%	136	0.7%	2799	1.4%	1243	1.4%
Weather	499	0.9%	207	0.3%	282	2.8%	19	0.6%	165	2.0%	18	0.3%	710	0.6%	39	0.2%	1656	0.8%	283	0.3%
Science	345	0.7%	233	0.4%	159	1.6%	53	1.6%	192	2.3%	167	2.9%	776	0.6%	172	0.8%	1472	0.7%	625	0.7%
Conflicts	217	0.4%	119	0.2%	83	0.8%	15	0.4%	91	1.1%	39	0.7%	318	0.2%	30	0.1%	709	0.4%	203	0.2%
Labour	146	0.3%	142	0.2%	30	0.3%	19	0.6%	35	0.4%	25	0.4%	375	0.3%	48	0.2%	586	0.3%	234	0.3%
Total (100%)	**52986**	**100%**	**59595**	**100%**	**9980**	**100%**	**3404**	**100%**	**8446**	**100%**	**5689**	**100%**	**127280**	**100%**	**20476**	**100%**	**198692**	**100%**	**89164**	**100%**

Table 3. Absolute numbers and relative row percentages of all IPTC categories for the four news brands.

	Het Nieuwsblad					De Standaard					De Morgen					Het Laatste Nieuws				
	Free		Paid		**Total (100%)**	Free		Paid		**Total (100%)**	Free		Paid		**Total (100%)**	Free		Paid		**Total (100%)**
	#	%	#	%	**#**	#	%	#	%	**#**	#	%	#	%	**#**	#	%	#	%	**#**
Economy	6536	37.5%	10911	62.5%	**17447**	1144	66.94%	565	33.06%	**1709**	1215	67.28%	591	32.72%	**1806**	22465	86.11%	3624	13.89%	**26089**
Sports	14281	67.0%	7028	33.0%	**21309**	2169	86.55%	337	13.45%	**2506**	877	55.44%	705	44.56%	**1582**	21236	89.25%	2559	10.75%	**23795**
Crime	6434	46.6%	7369	53.4%	**13803**	973	78.80%	262	21.20%	**1235**	834	72.65%	314	27.35%	**1148**	14674	87.84%	2031	12.16%	**16705**
Politics	5241	43.9%	6701	56.1%	**11942**	1852	67.61%	887	32.39%	**2739**	1890	65.58%	992	34.42%	**2882**	13419	87.67%	1888	12.33%	**15307**
Arts	4524	42.4%	6135	57.6%	**10659**	834	68.25%	388	31.75%	**1222**	732	39.17%	1137	60.83%	**1869**	12904	84.46%	2374	15.54%	**15278**
Accidents	4023	46.0%	4728	54.0%	**8751**	504	88.13%	68	11.87%	**572**	293	80.72%	70	19.28%	**363**	7987	94.53%	462	5.47%	**8449**
Society	1947	34.2%	3745	65.8%	**5692**	288	64.28%	160	35.72%	**448**	312	43.94%	398	56.06%	**710**	7224	83.65%	1412	16.35%	**8636**
Hum. interest	3341	51.6%	3134	48.4%	**6475**	523	82.74%	109	17.26%	**632**	288	44.31%	362	55.69%	**650**	7180	76.95%	2151	23.05%	**9331**
Health	1955	47.1%	2200	52.9%	**4155**	705	67.50%	339	32.50%	**1044**	1041	69.59%	455	30.41%	**1496**	5292	81.89%	1170	18.11%	**6462**
Education	1084	29.6%	2580	70.4%	**3664**	154	76.05%	48	23.95%	**202**	142	58.92%	99	41.08%	**241**	4141	89.73%	474	10.27%	**4615**
Lifestyle	1505	34.9%	2812	65.1%	**4317**	91	60.02%	61	39.98%	**152**	133	40.18%	198	59.82%	**331**	5900	76.70%	1792	23.30%	**7692**
Religion	339	37.5%	564	62.5%	**903**	70	70.64%	29	29.36%	**99**	67	60.91%	43	39.09%	**110**	707	86.11%	114	13.89%	**821**
Environment	569	36.6%	987	63.4%	**1556**	119	73.17%	44	26.83%	**163**	139	64.65%	76	35.35%	**215**	1972	93.55%	136	6.45%	**2108**
Weather	499	70.7%	207	29.3%	**706**	282	93.56%	19	6.44%	**301**	165	90.16%	18	9.84%	**183**	710	94.79%	39	5.21%	**749**
Science	345	59.7%	233	40.3%	**578**	159	74.88%	53	25.12%	**212**	192	53.48%	167	46.52%	**359**	776	81.86%	172	18.14%	**948**
Conflicts	217	64.6%	119	35.4%	**336**	83	85.09%	15	14.91%	**98**	91	70.00%	39	30.00%	**130**	318	91.38%	30	8.62%	**348**
Labour	146	50.7%	142	49.3%	**288**	30	60.73%	19	39.27%	**49**	35	58.33%	25	41.67%	**60**	375	88.65%	48	11.35%	**423**
Total	52986	47.1%	59595	52.9%	**112581**	9980	74.57%	3404	25.43%	**13384**	8446	59.75%	5689	40.25%	**14135**	127280	86.14%	20476	13.86%	**147756**

Table 4 below summarises the above figures for the total of the four assessed newspapers and their online content offering. Our computational analysis reveals that just over 30 percent of all articles were placed behind a paywall. Articles from various IPTC categories overperformed compared to the average, most notably Lifestyle (38.9%) and Religion (38.8%). Weather-related news is mostly available freely (85.4%), although this number is 'only' 70.7% for 'Het Nieuwsblad', yet 94.8% for fellow popular title 'Het Laatste Nieuws' (see Table 4).

Table 4. Absolute numbers and relative row percentages of all IPTC categories for the four news brands (total).

			Total		
	Free		Paid		**Total**
	#	%	#	%	**#**
Economy	31360	66.65%	15691	33.35%	**47051**
Sports	38563	78.39%	10629	21.61%	**49192**
Crime	22915	69.67%	9976	30.33%	**32891**
Politics	22402	68.15%	10468	31.85%	**32870**
Arts	18994	65.43%	10034	34.57%	**29028**
Accidents	12807	70.62%	5328	29.38%	**18135**
Society	9771	63.10%	5715	36.90%	**15486**
Hum. interest	11332	66.32%	5756	33.68%	**17088**
Health	8993	68.35%	4164	31.65%	**13157**
Education	5521	63.30%	3201	36.70%	**8722**
Lifestyle	7629	61.07%	4863	38.93%	**12492**
Religion	1183	61.20%	750	38.80%	**1933**
Environment	2799	69.25%	1243	30.75%	**4042**
Weather	1656	85.39%	283	14.61%	**1939**
Science	1472	70.18%	625	29.82%	**2097**
Conflicts	709	77.78%	203	22.22%	**912**
Labour	586	71.43%	234	28.57%	**820**
Total	198692	69.02%	89164	30.98%	**287856**

To better pinpoint similarities and discrepancies across titles, we also calculated the above results per type of company and news company, as listed accordingly in Table 1 above. Below, Tables 5 and 6 shine light on the divisions of free and paid online news content per IPTC category for Mediahuis and DPG Media and for popular and quality newspapers, respectively.

For either comparison, we do not distinguish dramatic differences between the newspapers and their practices of paywalling online articles. Percentagewise when assessed from the total body of articles per company, Table 5 shows that DPG Media paywalls more articles on Arts (13.4%), Human Interest (9.6%) and Lifestyle (7.6%); for Mediahuis, this is more pronounced in the categories Economy (18.2%), Crime (12.1%), Accidents (7.6%) and Education (4.2%). The two Mediahuis-owned titles combined have a near 50/50 division of free and paid articles, whereas the overwhelming majority of DPG Media-articles was available free of charge. This is mainly caused, though, by the dominance of the two popular titles and their body of articles, as outlined earlier in Table 1.

Table 5. Absolute numbers and relative column percentages of all IPTC topics for the two main news companies: Mediahuis and DPG Media.

	Mediahuis				DPG Media			
	Free	Free	Paid	Paid	Free	Free	Paid	Paid
	#	%	#	%	#	%	#	%
Economy	7680	12.2%	11476	18.2%	23680	17.4%	4215	16.1%
Sports	16450	26.1%	7365	11.7%	22113	16.3%	3264	12.5%
Crime	7407	11.8%	7631	12.1%	15508	11.4%	2345	9.0%
Politics	7093	11.3%	7588	12.0%	15309	11.3%	2880	11.0%
Arts	5358	8.5%	6523	10.4%	13636	10.0%	3511	13.4%
Accidents	4527	7.2%	4796	7.6%	8280	6.1%	532	2.0%
Society	2235	3.5%	3905	6.2%	7536	5.6%	1810	6.9%
Human interest	3864	6.1%	3243	5.1%	7468	5.5%	2513	9.6%
Health	2660	4.2%	2539	4.0%	6333	4.7%	1625	6.2%
Education	1238	2.0%	2628	4.2%	4283	3.2%	573	2.2%
Lifestyle	1596	2.5%	2873	4.6%	6033	4.4%	1990	7.6%
Religion	409	0.6%	593	0.9%	774	0.6%	157	0.6%
Environment	688	1.1%	1031	1.6%	2111	1.6%	212	0.8%
Weather	781	1.2%	226	0.4%	875	0.6%	57	0.2%
Science	504	0.8%	286	0.5%	968	0.7%	339	1.3%
Conflicts	300	0.5%	134	0.2%	409	0.3%	69	0.3%
Labour	176	0.3%	161	0.3%	410	0.3%	73	0.3%
Total (100%)	**62966**	**100.0%**	**62999**	**100.0%**	**135726**	**100.0%**	**26165**	**100.0%**

Please note: Mediahuis newspapers include Het Nieuwsblad and De Standaard; DPG media newspapers include De Morgen and Het Laatste Nieuws

Table 6. Absolute numbers and relative column percentages of all IPTC topics for two popular newspapers and two quality newspapers.

	Popular newspapers				Quality newspapers			
	Free	Free	Paid	Paid	Free	Free	Paid	Paid
	#	%	#	%	#	%	#	%
Economy	29001	16.1%	14535	18.2%	2359	12.8%	1156	12.7%
Sports	35517	19.7%	9587	12.0%	3046	16.5%	1042	11.5%
Crime	21108	11.7%	9400	11.7%	1807	9.8%	576	6.3%
Politics	18660	10.4%	8589	10.7%	3742	20.3%	1879	20.7%
Arts	17428	9.7%	8509	10.6%	1566	8.5%	1525	16.8%
Accidents	12010	6.7%	5190	6.5%	797	4.3%	138	1.5%
Society	9171	5.1%	5157	6.4%	600	3.3%	558	6.1%
Human interest	10521	5.8%	5285	6.6%	811	4.4%	471	5.2%
Health	7247	4.0%	3370	4.2%	1746	9.5%	794	8.7%
Education	5225	2.9%	3054	3.8%	296	1.6%	147	1.6%
Lifestyle	7405	4.1%	4604	5.7%	224	1.2%	259	2.8%
Religion	1046	0.6%	678	0.8%	137	0.7%	72	0.8%
Environment	2541	1.4%	1123	1.4%	258	1.4%	120	1.3%
Weather	1209	0.7%	246	0.3%	447	2.4%	37	0.4%
Science	1121	0.6%	405	0.5%	351	1.9%	220	2.4%
Conflicts	535	0.3%	149	0.2%	174	0.9%	54	0.6%
Labour	521	0.3%	190	0.2%	65	0.4%	44	0.5%
Total (100%)	180266	100.0%	80071	100.0%	18426	100.0%	9093	100.0%

Please note: Quality newspapers include 'De Morgen' and 'De Standaard'; popular newspapers include 'Het Nieuwsblad' and 'Het Laatste Nieuws'

The division of free and paid articles per type of news outlet yields a few intriguing findings. Arguably 'popular' topics such as Crime (11.7%), Accidents (6.5%) and Lifestyle (5.7%) are indeed much more prone to being paywalled at the two assessed Flemish tabloid newspapers - but the same goes for Economy (18.2%) and Education-related (3.8%) news articles. Alternatively, the two broadsheet titles assessed charged more for Politics (20.7%), Arts (16.8%), Health (8.7%) and Science (2.4%) articles, which are arguably more 'harder' news topics which are more prevalent in quality news outlets. As such, while being unable to pinpoint differences between paywalling articles or not at the company level, we are able to identify recurring patterns at the IPTC topic level when differentiating popular and quality newspapers.

Topic diversity analysis

To start our topic diversity analysis, we first differentiate between the four separate titles and their measured degrees of open and reflective diversity for free and paid articles. Based on our results, we can only partly confirm our hypotheses. For **H1,** in which we stated that news brands' paid content is more diverse in terms of topics than their free content offer, we find that only paid 'Het Nieuwsblad' articles score higher for both measured open and reflective diversity. 'De Morgen' is tied when it comes to measured reflective diversity, while the free 'Het Laatste Nieuws' articles score an almost perfect 1 score for this specific type of diversity. Furthermore, Mediahuis' quality newspaper 'De Standaard' sees higher diversity scores on all accounts for its freely available content rather than for its paywalled articles. This means that paid articles of this newspaper are less diverse when it comes to both open diversity, boasting an equal representation of all possible categories, and to reflective diversity, which states that the proportions in society ought to be reflected accordingly in news reporting.

Table 7. Topic diversity scores for all news brands.

	Nieuwsblad		De Standaard		De Morgen		Het Laatste Nieuws	
	Free	Paid	Free	Paid	Free	Paid	Free	Paid
Measured open diversity	0.66	0.75	0.69	0.65	0.70	0.74	0.73	0.76
Measured reflective diversity	0.95	0.98	0.94	0.86	0.88	0.88	0.99	0.96

Note: The high reflective diversity scores for both 'Het Nieuwsblad' and 'Het Laatste Nieuws' can be explained by their weight – in terms of number of articles - on the used benchmark for calculating this score (see limitations in Discussion for more information).

In Tables 8 and 9 below, we again split up our results per newspaper group based on their profile (popular vs. quality) and ownership (DPG Media vs. Mediahuis). Table 8 further confirms that **H2** is proven wrong: on all accounts, popular newspapers score higher, both for measured open and reflective diversity and for their free and paid online news articles. It is noteworthy that both quality and popular newspapers score (slightly) higher for their paywalled content when it comes to measured open diversity, but that their freely available offering fairs better when measuring reflective diversity.

Table 8. Topic diversity scores for the two quality newspapers and the two popular newspapers.

	Quality newspapers		Popular newspapers	
	Free	Paid	Free	Paid
Measured open diversity	0.69	0.70	0.71	0.75
Measured reflective diversity	0.92	0.87	0.99	0.97

As indicated in Table 9, the two DPG Media-owned titles score higher on measured open diversity for free and paid content and on measured reflective diversity for free content. The latter type of measured diversity is topped, however, by the two Mediahuis newspapers and their paywalled online

news content. For all companies' newspapers, except for DPG Media's measured reflective diversity, the diversity scores for paid content are higher than for free articles. This indicates that globally, paywalled online articles of the four largest Flemish newspapers are in most cases more diverse than their free content, albeit frequently only marginally so.

Table 9. Topic diversity scores for all DPG media's newspapers and Mediahuis' newspapers.

	DPG media newspapers		Mediahuis newspapers	
	Free	Paid	Free	Paid
Measured open diversity	0.73	0.76	0.67	0.74
Measured reflective diversity	0.99	0.94	0.95	0.97

Discussion

In this study, we quantitatively assessed a dataset of 287,856 online news articles published between December 2020 and August 2021 by the four biggest newspapers in Flanders (Belgium). To do so, we used deep machine learning to automatically assign all articles a news topic derived from a typology of the International Press Telecommunications Council (IPTC). By distinguishing articles which were freely available to all or locked behind a paywall, we were able to shine light on which news topics are more prone to being paywalled than others across types of outlets and companies as well as on the diversity scores per newspaper.

According to our study, the most predominant news topics present in the online content of four leading Flemish newspapers were Economy, Sports, Crime, Politics and Arts. Interestingly, these topics were also most prevalent in one Norwegian analysis of paywalled and non-paywalled content (Sjøvaag, 2016a). Thus, regardless of the host of diverse classifications for news articles, countries and types of news outlets assessed, we find that our descriptive analysis yielded surprisingly similar results in terms of recurring news topics more prone to being paywalled. This finding not only confirms the international relevance of our Flemish-oriented case study, but also shines light on similar approaches and tactics regarding editorial decisions of (not) paywalling certain online news articles. We hold the view that in spite of different characteristics of media market structures, routines and traditions, the rise of online paywalls and the choices of charging fees for specific types of news content fits within the framework of increasingly globalised media markets, facing the same challenges and experimenting with similar possible solutions from the vantage points of both media companies as well as individual newsrooms. However, we of course also denote differences. Myllylahti (2017) concludes that hard news is among the 'most saleable commodity' (p. 469). While we too find that hard news topics such as Politics and Economy are indeed more frequently paywalled, we actually find that percentagewise, Lifestyle news is the most frequent one of all IPTC topic categories.

Furthermore, our descriptive analysis also yielded the remarkable finding that news on the Economy and Education were more prone to being paywalled among the two popular outlets analysed, in contrast to quality newspapers, along with more expected news topics such as Lifestyle, Crime and Accidents. The quality titles, meanwhile, devoted both more freely accessible and paywalled article space to news on Politics, Arts, Science and Health. Additionally, our topic diversity analysis, in which we measured the ideal topic distribution per IPTC news topic for open and reflective diversity, revealed that popular news titles consistently scored higher than their quality counterparts on both diversity types and for both freely accessible and paywalled news articles. This finding went completely against one of our hypotheses which had stated that we expected the situation to be quite the opposite. Thus, our results imply a much more nuanced and less normative perspective on practices of charging for

news content which go beyond traditional viewpoints and assumptions regarding the relationship between certain news topics and their prevalence across broadsheet and tabloid news outlets.

We are unable to link our findings to wider editorial decisions from the four newsrooms of the two media companies assessed. Both at the levels of media corporations and that of types of outlets, we mostly denote differences. Only when grouped together do we denote similarities in line with previous research, as indicated earlier in this section. If anything, the individual discrepancies in paywalling articles from certain news topics across titles reveals that each brand and its newsroom enjoy a high degree of autonomy in determining their own paywall strategies. This fits within the high position of Belgium in international rankings of press freedom (see Reporters Without Borders, 2021). We venture that these results could differ in nations with stronger hierarchical frameworks within (media) corporations and acknowledge that this dents the transferability of our study's findings to other markets. However, as the four brands assessed in this paper form part of media companies active in several other European countries, and our findings on news topics are mostly in line with existing scholarship from other markets, we hold the view that our results are still to a large extent relevant for international scholarship on paywalls and their effect on (the availability of) online news of certain topics.

Ultimately, we do not find evidence to state that paywalls at Flanders' four biggest newspapers have led to fundamental decreases in news diversity in front of the paywalls. For all four newspapers and their paid content, we find that the diversity scores are indeed higher for measured open diversity, but that it is actually lower for measured reflective diversity when compared to the newspapers' free offering. In all cases, the score differences are small, leading us to deduce that both quality and popular news outlets do not charge more for news diversity, but the situation is rather (and perhaps somewhat to our own surprise) stable.

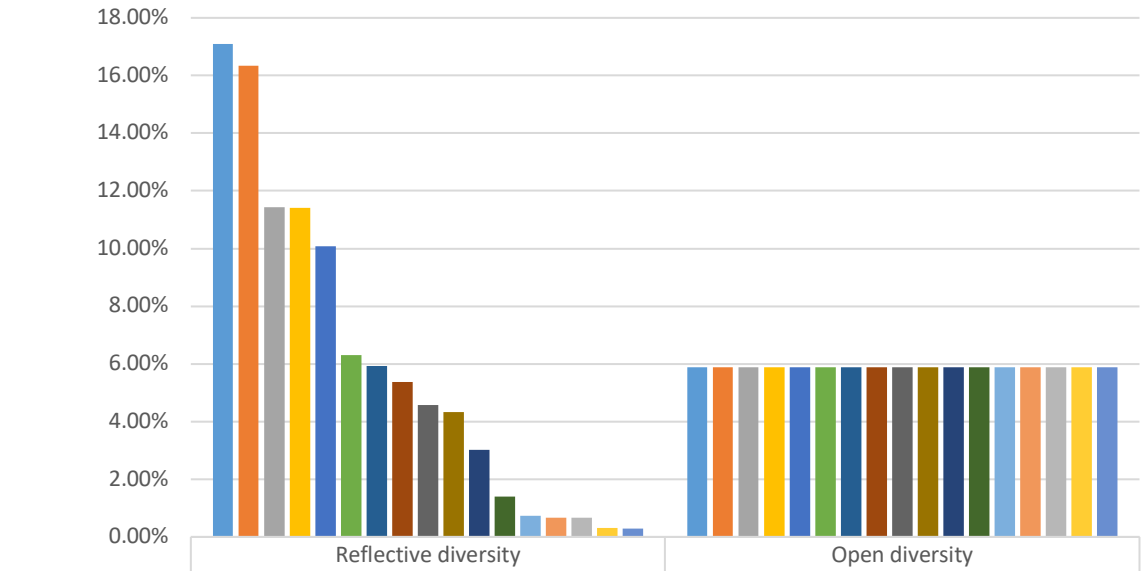
To end, we must acknowledge a few shortcomings to our study. First, several data collection techniques have been used in our sampling strategy, with varying degrees of reliability, which could possibly lead to small sampling errors. For Mediahuis' articles in particular, we used RSS feeds in combination with web scraping, which is more sensitive to technological errors in contrast to the use of API's. Although there are no concrete indications that we encountered this type of sampling error, they could not be completely excluded for the sample of Mediahuis. However, if this would be the case, we can ensure that they would have no fixed pattern and would have occurred at random occasions, making no differences in our main result or conclusion sections.

Second, in the IPTC-labelling process, we only took the news article's main topic into account, while news articles generally cover multiple topics. The rationale underlying this choice can be found in the complexity of multiple-topic analysis and interpretation. When IPTC-topics are predicted by a classification system, they do not have the same confidence score and subsequently are not in the same way related to the article as one another. Simply taking all topics into account and aggregating the news articles into one category would endanger the validity of the IPTC-categories. Moreover, multiple-topic analysis would also make it difficult to understand what certain percentages and numbers mean, as they are all in different ways related to certain IPTC-categories (e.g., some are more related to certain IPTC-categories than others).

Third, to conceptualize reflective diversity, we did not look at the distribution that is currently present in society, but at the distribution in the journalistic offer itself. Although distributions in society are preferred to use as a diversity benchmark, we had no information at our disposal about which topics are mostly discussed in society. As a result, we had to look at the journalistic offer in general and how topics were distributed among all news brands involved (see Figure 1 in Appendix). However, as the

journalistic offer is largely influenced by news brands that publish a lot, some news brands have fewer difficulties to reach the distribution level that is used as a benchmark in measured reflective diversity. This is particularly the case for ‘Het Laatste Nieuws’ and ‘Het Nieuwsblad ‘who respectively published much more than ‘De Standaard’ and ‘De Morgen’. This should be seen as a limitation of this measure, which is not an issue for the other diversity benchmark ‘open diversity’.

Appendix



	Reflective diversity	Open diversity
Sports	17.09%	5.88%
Economy	16.35%	5.88%
Crime	11.43%	5.88%
Politics	11.42%	5.88%
Arts	10.08%	5.88%
Accidents	6.30%	5.88%
Human interest	5.94%	5.88%
Society	5.38%	5.88%
Health	4.57%	5.88%
Lifestyle	4.34%	5.88%
Education	3.03%	5.88%
Environment	1.40%	5.88%
Science	0.73%	5.88%
Weather	0.67%	5.88%
Religion	0.67%	5.88%
Conflicts	0.32%	5.88%
Labour	0.29%	5.88%

Figure 1. Two topic diversity ideals used to measure topic diversity: (1) reflective diversity that follows the distribution of the journalistic offer (incl. 'Het Laatste Nieuws', 'De Morgen', 'De Standaard' and 'Het Nieuwsblad'), and (2) open diversity that gives each category equal attention.

Table 10. Examples of news articles and their automatically assigned IPTC topic

IPTC topic	Headline (in Dutch)	URL
Art	"Schouw Nieuw Zuid mag dan toch branden voor cultuurproject"	https://www.nieuwsblad.be/cnt/dmf20201127_95349307
Crime	"Inbraakpoging in Hasselt"	https://www.nieuwsblad.be/cnt/dmf20201127_95331745
Disasters	"Man veroordeeld voor dronken aanrijding: "Kindjes in andere auto gelukkig niet gewond"	https://www.nieuwsblad.be/cnt/dmf20201127_95357008
Economy	"Grootste baggercontract ooit voor Deme: containerhaven in Egypte"	https://www.nieuwsblad.be/cnt/dmf20201127_96482853
Education	"Leerlingen brengen licht en warmte in Moervaartheem"	https://www.nieuwsblad.be/cnt/dmf20201127_96371190
Environment	"Otter krijgt 7 hectare nieuw leefgebied in Polders van Kruibeke: "Heropleving van deze soort duwtje in de rug geven"	https://www.hln.be/kruibeke/otter-krijgt-7-hectare-nieuw-leefgebied-in-polders-van-kruibeke-heropleving-van-deze-soort-duwtje-in-de-rug-geven~ab8d3722/
Health	"Coronacijfers UZ Leuven: 65 patiënten met COVID-19"	www.hln.be/in-de-buurt/leuven/coronacijfers-uz-leuven-65-patienten-met-covid-19~a575dc67/
Human Interest	"André viert honderdjarige verjaardag in rusthuis in Evergem"	www.hln.be/in-de-buurt/evergem/andre-viert-honderdjarige-verjaardag-in-rusthuis-in-evergem~a4682558/
Labour	"Farma- en chemiesector blijft aanwerven, ondanks coronacrisis"	https://www.standaard.be/cnt/dmf20201026_91107828
Lifestyle	"Geef je oud brood een tweede leven als panzanella met kalkoen"	www.demorgen.be/leven-liefde/geef-je-oud-brood-een-tweede-leven-als-panzanella-met-kalkoen~b68a6be9/
Politics	"In welke staten heeft Trump al rechtszaken aangespannen? En waar wordt er opnieuw geteld?"	https://www.standaard.be/cnt/dmf20201112_92958484
Religion	"Parochies roepen op om kaarsje aan kerken en kapellen te plaatsen tijdens Allerheiligenweekend: "Troostplekken creëren om overledenen van voorbije maanden te gedenken"	www.hln.be/in-de-buurt/laakdal/parochies-roepen-op-om-kaarsje-aan-kerken-en-kapellen-te-plaatsen-tijdens-allerheiligenweekend-troostplekken-creeren-om-overledenen-van-voorbije-maanden-te-gedenken~a5924c16/
Science	"Lennikse zonnepiloot opent digitale Dag van de Wetenschap"	https://www.nieuwsblad.be/cnt/dmf20201127_95088502
Society	"Theatergarage schenkt speelgoed aan goede doelen: "De Sint voor elk kind"	https://www.nieuwsblad.be/cnt/dmf20201125_96789671
Sport	"Coronabommetje in de maak? Kapitein van Kroatië pas tijdens de rust (!) vervangen na positieve test"	https://www.nieuwsblad.be/cnt/dmf20201112_92764322
Conflict	"Ethiopische premier geeft bevel tot "eindoffensief" in Tigray"	https://www.nieuwsblad.be/cnt/dmf20201126_92676978
Weather	"Weer naar school met een boekentas en een regenjas"	https://www.nieuwsblad.be/cnt/dmf20201116_91963727

Bibliography

Arrese, Á. (2016). From Gratis to Paywalls: A brief history of a retro-innovation in the press's business. *Journalism studies, 17*(8), 1051-1067.

Badr, Z. (2021). More or More of the Same: Ownership Concentration and Media Diversity in Egypt. *The International Journal of Press/Politics, 26*(4), 774-796.

Baker, C. E. (2006). *Media concentration and democracy: Why ownership matters.* Cambridge University Press.

Beckers, K. (2017). How ordinary is the ordinary (wo) man on the street? An analysis of vox pop characteristics in television news. *Journalism Practice, 11*(8), 1026-1041.

Benson, R. (2019). Paywalls and public knowledge: How can journalism provide quality news for everyone?. *Journalism, 20*(1), 146-149.

Bleyer-Simon, K., Brogi, E., Carlini, R. M., Nenadic, I., Palmer, M., Parcu, P. L., ... & Žuffová, M. (2021). Monitoring media pluralism in the digital era: application of the Media Pluralism Monitor in the European Union, Albania, Montenegro, Republic of North Macedonia, Serbia & Turkey in the year 2020. *European University Institute.*

Bodó, B. (2019). Selling News to Audiences–A Qualitative Inquiry into the Emerging Logics of Algorithmic News Personalization in European Quality News Media. *Digital Journalism*, *7*(8), 1054-1075.

Cagé, J., Herve, N., & Mazoyer, B. (2020). Social Media and Newsroom Production Decisions. *Social Science Research Network,* available at https://doi.org/10.2139/ssrn.3663899

Carpenter, S. (2010). A study of content diversity in online citizen journalism and online newspaper articles. *New media & society*, *12*(7), 1064-1084.

Calhoun, C. J. (Ed.). (1992). *Habermas and the public sphere.* MIT press.

Cherubini, F., & Nielsen, R. K. (2016). Editorial Analytics: How News Media are Developing and Using Audience Data and Metrics. *Social Science Research Network*, available at https://doi.org/10.2139/ssrn.2739328

de Vries, W., van Cranenburgh, A., Bisazza, A., Caselli, T., van Noord, G., & Nissim, M. (2019). *Bertje: A dutch bert model.* arXiv preprint arXiv:1912.09582.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019), BERT: Pre-training of deep bidirectional transformers for language understanding, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* 4171–4186.

Dimmick, J., Feaster, J. C., & Hoplamazian, G. J. (2011). News in the interstices: The niches of mobile media in space and time. *New media & society, 13*(1), 23-39.

Entman, R. M., & Wildman, S. S. (1992). Reconciling economic and non-economic perspectives on media policy: Transcending the "marketplace of ideas". *Journal of Communication, 42*(1), 5-19.

Habermas, J. (1989). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. Polity press.

Habermas, J. (2006). Political Communication in Media Society: Does Democracy Still Enjoy an Epistemic Dimension? The Impact of Normative Theory on Empirical Research. *Communication Theory, 16*(4), 411–426.

Hahn, O. (2007). Cultures of TV News Journalism and Prospects for A Transcultural Public Sphere. In N. Sakr (Ed.), *Arab Media and Political Renewal: Community, Legitimacy and Public Life* (pp. 13–27). I.B.Tauris.

International Press Telecommunications Council. (n.d.). *Media Topics,* retrieved from https://iptc.org/standards/media-topics/

Kvalheim, N. (2013). News Behind the Wall: An Analysis of the Relationship Between the Implementation of a Paywall and News Values. *Nordicom Review, 34*(1), 25–42. https://doi.org/10.2478/nor-2013-0102

Lamot, K., & Paulussen, S. (2019). Six Uses of Analytics: Digital Editors' Perceptions of Audience Analytics in the Newsroom. *Journalism Practice, 14*(3), 358–373. https://doi.org/10.1080/17512786.2019.1617043

McQuail, D., & Van Cuilenburg, J. J. (1983). Diversity as a media policy goal: A strategy for evaluative research and a Netherlands case study. *International Communication Gazette, 31*(3), 145-162. https://doi.org/10.1177/001654928303100301

Meijer, I. C., & Kormelink, T. G. (2019). Audiences for Journalism. *The International Encyclopedia of Journalism Studies*, 1-7.

Makhortykh, M., de Vreese, C., Helberger, N., Harambam, J., & Bountouridis, D. (2021). We are what we click: Understanding time and content-based habits of online news readers. *New media & society, 23*(9), 2773-2800.

Myllylahti, M. (2017). What Content is Worth Locking Behind a Paywall? *Digital Journalism, 5*(4), 460-471.

Myllylahti, M. (2019). Paywalls. *The International Encyclopedia of Journalism Studies*, *1*-6.

Napoli, P. M. (1999). Deconstructing the diversity principle. *Journal of communication, 49*(4), 7-34.

Newman, N., Fletcher, R., Kalogeropoulos, A., & Nielsen, R. (2020). *Reuters institute digital news report*. Reuters Institute for the Study of Journalism.

Noam, E. M. (2016). *Who owns the world's media?: Media concentration and ownership around the world.* Oxford University Press.

Olsen, R. K., & Solvoll, M. K. (2018). Bouncing off the Paywall - Understanding Misalignments Between Local Newspaper Value Propositions and Audience Responses. *JMM: The International Journal on Media Management, 20*(3), 174–192.

Papandrea, F. (2006). Media Diversity and Cross-Media Regulation. *Prometheus, 24*(3), 301-322.

Pattabhiramaiah, A., Sriram, S., & Manchanda, P. (2019). Paywalls: Monetizing Online Content. *Journal of Marketin*g, 83(2), 19–36.

Pickard, V., & Williams, A. T. (2014). Salvation or folly? The promises and perils of digital paywalls. *Digital journalism, 2*(2), 195-213.

Reporters without borders (2021). *World Press Freedom Index,* retrieved from https://rsf.org/en/ranking

Simon, F. M., & Graves, L. (2019). Pay models for online news in the US and Europe: 2019 Update.

Sjøvaag, H. (2014). Homogenisation or differentiation? The effects of consolidation in the regional newspaper market. *Journalism Studies, 15*(5), 511-521.

Sjøvaag, H. (2016a). Introducing the paywall: A case study of content changes in three online newspapers. *Journalism Practice, 10*(3), 304-322.

Sjøvaag, H. (2016b). Media diversity and the global superplayers: operationalising pluralism for a digital media market. *Journal of Media Business Studies, 13*(3), 170-186.

Skärlund, S. (2020). The recycling of news in Swedish newspapers: Reused quotations and reports in articles about the crisis in the Swedish Academy in 2018. *Nordicom Review, 41*(1), 69-84.

Thurman, N., & Schifferes, S. (2012). The future of personalization at news websites: lessons from a longitudinal study. *Journalism Studies*, *13*(5-6), 775-790.

Van Cuilenburg, J. (1999). On competition, access and diversity in media, old and new: Some remarks for communications policy in the information age. *New media & society, 1*(2), 183-207.

Van Damme, K. (2020). *Transforming journalism, transforming audiences? Audience-centred research on news use in the omnipresent news environment* [Doctoral dissertation]. Biblio Ghent University.

van Hoof, A. M. J., Jacobi, C., Ruigrok, N., & van Atteveldt, W. (2014). Diverse politics, diverse news coverage? A longitudinal study of diversity in Dutch political news during two decades of election campaigns. *European Journal of Communication*, *29*(6), 668-686.

Vogler, D., Udris, L., & Eisenegger, M. (2020). Measuring media content concentration at a large scale using automated text comparisons. *Journalism Studies, 21*(11), 1459-1478.