Enduring elites in quoted sources

*Institutional alignment in Finnish media,
1999–2018*

OLLI SEURI,ᴵ ANU KOIVUNEN,ᴵᴵ HENNA LEVOLA,ᴵ & EETU MÄKELÄᴵᴵᴵ

ᴵ FACULTY OF SOCIAL SCIENCES, TAMPERE UNIVERSITY, FINLAND
ᴵᴵ SCHOOL OF HISTORY, CULTURE AND ARTS STUDIES, UNIVERSITY OF TURKU, FINLAND
ᴵᴵᴵ DEPARTMENT OF DIGITAL HUMANITIES, UNIVERSITY OF HELSINKI, FINLAND

ABSTRACT

This article addresses the question of whether and how sourcing practices of journalistic news media have been transformed by online-first publishing and the hybridisation of the media system. Focusing on four Finnish news outlets from 1999 to 2018 (a daily newspaper, a news agency, a tabloid, and a public broadcaster), we analyse the essential gatekeeping practice of giving voice to actors by quoting them directly or indirectly. According to our data, the hybrid media context does not lead to an increased diversification of sources. Instead, our analysis indicates a concentration and consolidation of sources across different news outlets: a continued elite source dominance in domestic political news. This development we term institutional alignment, suggesting that, by foregrounding elite sources, the news media seeks to maintain both its gatekeeping function and its key position in networks of power.

KEYWORDS: news media, sourcing practices, giving voice, gatekeeping, Nordic model

Introduction

In this article, we address the question of how the hybridisation of the media system (Chadwick, 2013) – the 20-year period labelled "the crisis of institutional press" (Reese, 2021) – has affected sourcing practices in journalistic news media. There is a long tradition of media research highlighting media outlets' reliance on elite sources: journalists foregrounding information and perspectives provided by individuals and institutions that hold positions of power, authority, or influence within society (see Bennett, 1990; Brown et al., 1987; Schudson, 2003; Sigal, 1973; Skogerbø et al., 2016). However, as Van Leuven and colleagues (2018) have pointed out, the breakthrough of the Internet and the changes in the media environment have drastically changed journalistic news-sourcing techniques. New practices, such as using social media, search engines, websites, wikis, and online encyclopaedias, are frequently associated with a democratisation of news access for non-elite actors, potentially indicating an increased diversity of voices among journalistic sources (Lecheler & Kruikemeier, 2016; Van Leuven et al., 2018). Furthermore, the disruption of the political public sphere might challenge the long tradition of what W. Lance Bennett (1990, 2014, 2016) has termed "indexing", the centrality of "perceived governmental power balances" for filtering news and selecting sources and, more generally, it has questioned the gatekeeping power of legacy news media (see Vos 2015; Williams & Delli Carpini, 2004).

In this article, we study news media sourcing by focusing on a particular journalistic practice: giving voice to actors by quoting them directly or indirectly, or through the more recent practice of referring to their posts on social media. Our focus is on the quotes in the news texts, which we define as visible and explicit sources (see Sundar & Nass, 2001). The assumption is that giving voice is an important function in terms of the democratic task of the media, because the cited actors are not only able to direct attention and influence, but also the flow of information (Brown et al., 1987; Richardson, 2006; Tiffen et al., 2014). When quoted by the news media, actors are recognised as legitimate interlocutors and movers of the flows of power.

In our analysis, we focus on a time period ranging from 1999 to 2018, in which the news media transitioned to online-first publishing and social media came to prominence. This longitudinal perspective allows us to investigate the thesis of shifting power balance and scrutinise the idea of digital media culture opening "virtually unlimited sources" (Williams & Delli Carpini, 2004: 1208) against the legacy of indexing and the conventional relying on elite sources (Bennett, 1990). In this article, we approach the journalistic sourcing practices and the diversity of sources in the news (see Tiffen et al., 2014) in the framework of gatekeeping theory (Shoemaker & Vos, 2009), employing quantitative analysis to answer the following empirical research questions:

RQ1. How has quotation practice developed in the last twenty years, in terms of the number of direct, indirect, and social media quotations employed?

RQ2. Has sourcing become more or less diverse overall?

RQ3. Are there differences in the extent that different societal groups get their voice heard through 1) direct, 2) indirect, and 3) social media sourcing practices? If there are, do these change over the 20-year period?

RQ4. Are there differences between different types of media with regard to any of the above?

To answer these questions, we turn to a large dataset which includes all articles published by 1) the Finnish news agency STT and 2) the main national daily newspaper *Helsingin Sanomat* between 1999 and 2018, as well as all articles published online on their sites by 3) the Finnish public service broadcaster Yle, and 4) one of the two main Finnish daily tabloids, *Iltalehti*, between 2009 and 2018. Studying a 20-year period enables us to address the question of news media's role as a political institution, identifying continuities and changes in the sourcing practices of political journalism.

Media sourcing practices and institutional power

We identify the news media practice of giving voice as a crucial gatekeeping function. There are different definitions of gatekeeping, but in media research, the concept focuses on power – that is, who is capable of influencing which information reaches people and how the social reality is constructed (Seuri & Ikäheimo, 2022). Research interest has focused on journalistic practices that define who gets a voice in the news media and how the information reaches the audience (see, e.g., Gans, 1979; Shoemaker, 1991; Shoemaker & Vos, 2009; Tuchman, 1978; White, 1950).

However, the digital transformation has challenged the centrality of gatekeeping theory, as the proliferation of alternative information channels and the growth of social media have led to the decentralisation of gatekeeping power. Legacy media institutions and industries have lost their privileged gatekeeper status to mediate between sources and recipients of information. In a high-choice media environment, "the people formerly known as the audience" (Rosen, 2012: 13) have more available communication channels to choose from and use themselves (see Van Aelst et al., 2017; Fletcher & Nielsen, 2018).

Despite the hybridisation of the public sphere, gatekeeping has not lost its importance. The news media continue to be one of the fundamental network gatekeepers (Barzilai-Nahon, 2008; DeIuliis, 2015; Wallace, 2018), who – together with social actors and platform algorithms (Karlsson et al., 2022) – cull and craft "countless bits of information into the limited number of messages that reach people every day" (Shoemaker & Vos, 2009: 1). While the news media no longer control the information space, they still act as important network nodes (Wihbey, 2019), channelling attention (Goel et al., 2016) or curating flows (Thorson & Wells, 2015). Hence, in a hybrid media environment, the questions of gatekeeping are increasingly about degree rather than inclusion and exclusion.

The gatekeeping theory and studies on media sourcing practices share a substantial overlap in their examination of how information flows through the media

ecosystem. Gatekeeping theory focuses on the role of media gatekeepers, such as editors and journalists, in selecting, filtering, and deciding which news stories or sources gain access to the public sphere (see Shoemaker & Vos, 2009). In parallel, studies on media sourcing practices delve into the process of how journalists choose and utilise sources to construct news narratives (Gans, 1979; Parmelee, 2014; Tuchman, 1978; Weaver & Elliott, 1985). Scholars have investigated source selection, the relationship between the media and sources, and the impact of sources and source diversity on news coverage in order to uncover potential biases, agenda-setting effects, and the ways in which information is framed for public consumption (see Bennett 1990; Brown et al., 1987; Gandy, 1982; Maurer & Beiler, 2018; Reich 2006; Sigal, 1973; Tiffen et al., 2014; Van Leuven et al., 2018). While previous research has identified an increased use of online sources, these appear not to have drastically changed journalistic news production (see Van Leuven et al., 2018; Thorsen & Jackson, 2018; Vliegenthart & Boukes, 2018).

Rather than investigating the power dynamics of agenda-setting, or agenda-building, our focus is on the news media as a political institution (Cook, 1998) and its relationship to other social institutions via sourcing practices (see Berkowitz, 1987; Weaver & Elliott, 1985). Our approach to sourcing and power is inspired by the work of W. Lance Bennett (1990, 2014, 2016), who has posited a theory of "indexing", whereby news coverage reflects the positions of "official" sources, or "decisive institutional power blocs" (Bennett, 1990: 125). As a theoretical framework, indexing brings together three theoretical accounts of press–state relations: 1) a "media monopoly" thesis highlighting the routines and "professional" decisions restricting diversity; 2) the organisational efficiency thesis suggesting the news are a result of "transactional" or "symbiotic" relations between journalists and officials; and 3) a "liberal reporter" thesis of press acting in a democratically responsible manner by favouring public officials, who are, in the end, representatives of the people (Bennett, 1990).

Foregrounding the institutional aspect, we wish to highlight the importance of the systemic nature of established and embedded social rules that structure social interactions (Hodgson, 2006). In a Western democracy such as Finland, the news media as a social institution has a purpose to "protect and nurture" people and the society by supporting the democratic system and its laws, regulations, and traditions (Lehdonvirta, 2022: 4, 7). In the democratic corporatist model, often associated with Finland and other Nordic countries, "political parallelism", the alignment between political parties and media outlets, is an important feature (Hallin & Mancini, 2004). To emphasise this, Syvertsen and colleagues (2014) have identified the idea of "stateness" as an important pillar of the Nordic media welfare state.

In previous research on political journalism in Finland, the peak of the democratic corporatist national media system and the political project of the welfare state has been termed as high modernity (Kantola, 2013). During this period, journalism was oriented towards the state, parties, and corporatist institutions and characterised as "subservient" towards politics and politicians, reporting the utterances of political elites. What has been identified as a shift towards

liquid modernity entailed a turn towards audiences and an increasing distance towards politicians, with journalism adopting a more critical and interpretive watchdog role (Koljonen, 2013; Väliverronen, 2022b). In an alternative account, the democratic corporatist media system geared towards a more liberal and market-oriented one (Ala-Fossi, 2020; Herkman, 2009) in the 1990s and around 2000, also featuring a generational shift among journalists.

Given these transformations, a key question concerns the role of indexing and gatekeeping power in the age of hybrid media. The narratives of the hybridisation predict and highlight disruptions, but as such, they risk overemphasising ruptures over continuities.

Research material and methods

The empirical analysis in this article is based on a research database to which four major media outlets have donated all content they have produced within either a 20-year or a 10-year timespan. From *Helsingin Sanomat*, the largest national newspaper, and STT, the Finnish news agency, the database contains all published material between 1999 and 2018 (*Helsingin Sanomat*: 1.6 million articles; STT: 2.3 million articles). Yle, the Finnish public broadcaster, and *Iltalehti*, one of the two major Finnish daily tabloids, are represented by the articles they published online between 2009 and 2018 (Yle: 930,000 articles; *Iltalehti*: 556,000 articles). Together, these four media – a news agency, a daily newspaper, a public broadcaster, and a tabloid – represent major segments of the Finnish mass media system. These outlets cover a substantial share of the Finnish text-based news media field, as each of them is the biggest actor in their sub-category. We offer our study as a case of a Nordic media system, a context characterised by strong institutions and high degree of professionalism (see Hallin & Mancini, 2004), and a national market in which the audience's trust in the news media has remained exceptionally high throughout the research period (Newman et al., 2022).

To make the data suitable for research, it has been cleaned and harmonised. For example, duplicate versions of articles have been identified algorithmically, and non-journalistic content such as television listings has been removed. For this article, we further limited our material to consist only of the core domestic, political, and economic news categories of each outlet. While lifestyle or culture sections can be viewed as important domains of the political public sphere, to make the four outlets comparable – also across time – we had to focus on stable key sections within political news journalism. In the end, this left us with a core dataset of 1.4 million articles across the four media.[1]

We approach the question of media power by focusing on sourcing practices, which we have studied as quoting practices – as literally giving voice. We analyse three types of sourcing: 1) direct quotes sourced from individuals; 2) indirect quotes (e.g., "according to X, Y", "X said that Y"); and 3) quotes where the journalist quotes an utterance originally posted on social media. This division is intended to give us a comprehensive perspective into different types of sourcing. The division into direct and indirect quotes is based on our prior research

(Koivunen et al., 2021; Seuri et al., 2021), where we identified these quotation types to fulfil partially different roles: Direct quotes were often used as flavour and to bring affect into the articles, while indirect quotes were more often used merely to attribute a reliable source for the information presented. The separating out of social media quotes, on the other hand, directly tackles a perceived shift in power, where journalists are relegated to copying and commenting on statements posted directly to followers on social media.

To enable these analyses, we first employed an automated means of identifying both direct and indirect quotations and their signified authors – both persons and non-persons.[2] This resulted in the identification of 3 million indirect and 1.2 million direct quotes in the articles. As for sourcing through social media, that is not something our automated pipeline captures. Therefore, to interrogate this question, we needed to rely on a smaller, manually curated random sample extracted from the data. To obtain this sample, we randomly extracted a total of 2,142 articles (2%) from the set of all articles (in total around 100,000) mentioning any of the words Twitter (now known as X), Facebook, or YouTube, which we identified as the major social media platforms used by Finns during our period of analysis. These were then analysed manually to first categorise the articles into instances of actual sourcing (748 or 35%) as opposed to other mentions (1,394 or 65%).

Since we are interested in the representation of different societal groups instead of individual actors, we mapped the individuals quoted in each of these three datasets to the groups they represented. This is something our method does not do fully automatically, even if it extracts mentions of roles and organisations associated with the actors. We used this information as a basis for manually categorising the actors into groups.

Given our research interest in the role of elite sources, we set out to identify different groups among traditional "decision-makers" in the Finnish context (Reunanen et al., 2010), mapping political, social, and economic elites and public administration and authorities but also non-elite voices (nongovernmental organisations, ordinary citizens). In Finland and other Nordic countries, economic interest organisations (trade unions, business associations, organisations of public service providers) have historically enjoyed privileged access to policymaking processes, and analysing the role of this "routine corporatism" (Vesa et al., 2018) in news journalism was one of our initial interests. In our study, "elites" hence refers to groups of decision-makers and representatives of routine corporatism. As a counterpoint, we set out to identify new and non-elite voices, supporting the vision of source diversification in the age of digital media.

Drawing on previous, context-specific research but also basing on qualitative analysis and evaluation of what was in the data, we ended up with 20 groups of societal actors as the target of classification (cf. Tiffen et al., 2014), which we then enacted manually.[3] While in our analysis we verified all our conclusions with these twenty groups, for presentation purposes here, these 20 have been condensed into nine larger groups based on uniformity as well as on their developing similarly in our data. Thus, the nine groups in the present analysis are: 1) political actors (including, e.g., politicians and political parties), 2) public

officials apart from 3) the police, which we keep separate to highlight its distinctly high share in the previous, 4) media outlets, 5) research organisations, 6) interest groups (mainly labour and employers' organisations), 7) other nongovernmental organisations, 8) private companies, and 9) everybody else.

Data analysis

Changes in the use of quotation (RQ1)

Before addressing the question of how quote source diversity develops in our data, it is necessary to look at how quoting practices themselves developed through the 20-year timespan. In Figure 1, this is represented by charting the number of quotes of different types identified in the news articles divided by the number of sentences in those articles. This normalisation gives a measure of quote frequency as a function of the amount of journalistic content produced, without regard to how the content is split between, for example, short- and long-form articles (this being an area where we have identified substantial changes over time).

FIGURE 1 Quoting practice from the viewpoint of production (number of quotes per ten sentences)
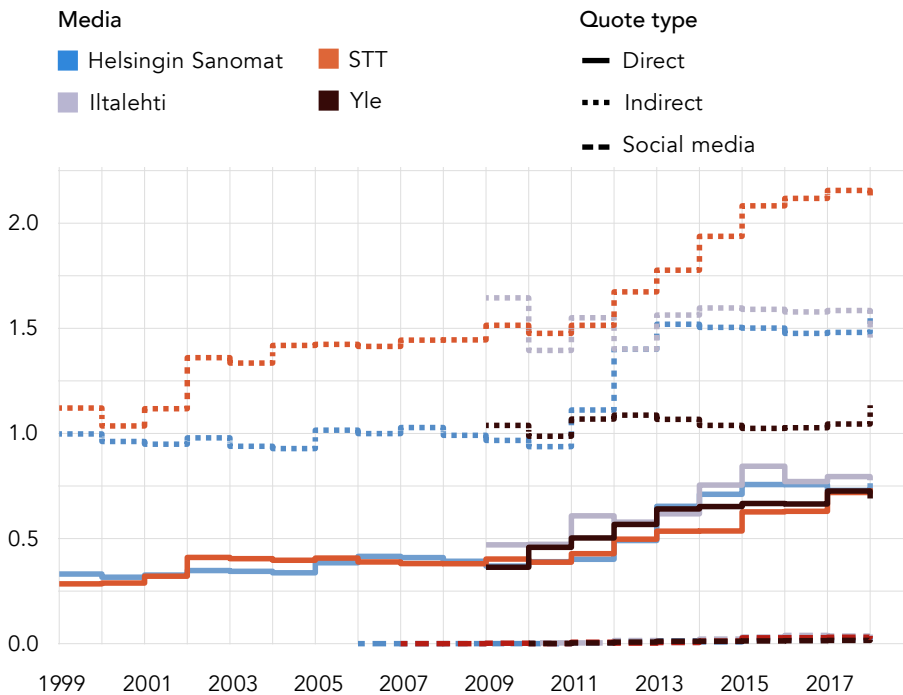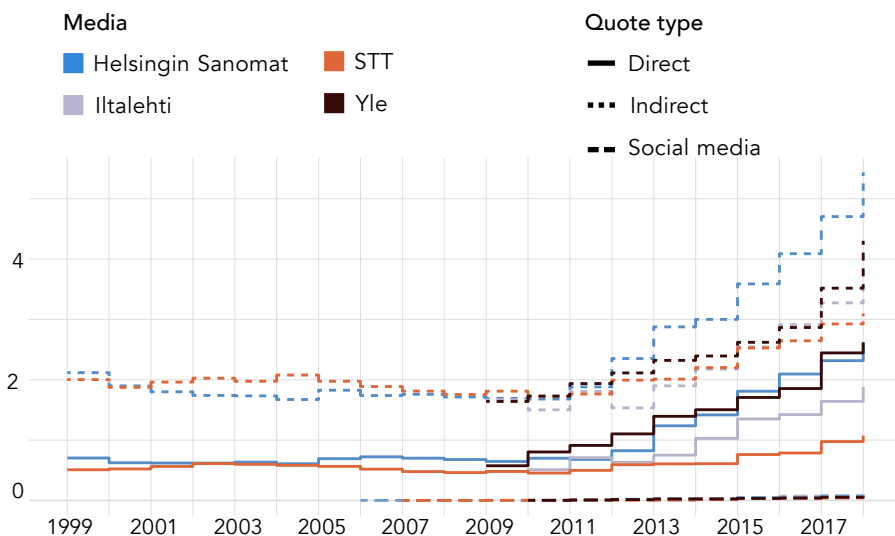

Comments: Based on domestic general/political/economic news and normalised from the viewpoint of text production.

As Figure 1 shows, in the two long-running time series we have from *Helsingin Sanomat* and STT, the use of both direct as well as indirect quotations increases over time. In particular, direct quotes have become increasingly frequent in all analysed media since 2010. In the case of indirect quotes, on the other hand, news outlets differ from one another. In *Iltalehti* and Yle (our dataset only comprising online news for these media), the amount of indirect quotes stays relatively stable. For the two sources we study in the 20-year perspective, however, we see an increased use of indirect quotes. For STT, the use of indirect quotes has grown slowly throughout the whole 20 years, while for *Helsingin Sanomat*, there is a marked increase in the use of indirect quotes coinciding with the news media turning to online-first publishing.

Finally, in terms of quotes sourced through social media, it can immediately be seen that the amount of such quotations is completely dwarfed by more traditional journalistic sourcing practices. However, if one were to zoom in to only this part of the graph, one would see an even higher proportional rate of growth as compared with the other types of quoting. In other words, the proportion of quotations that can be traced to social media increases considerably during the period of analysis. While the share of such quotes is still effectively 0 per cent in 2010, in 2018, 1.3 per cent of the total quotation pool can be traced to social media, with the proportion graph showing an exponential increase. Yet, while pointing to an ever-increasing importance moving toward the present, for the analysis conducted in this article, one must remember that we are looking at the early days of the interaction between social media and journalistic media, and the 1.3 per cent share in 2018 was still a tiny fraction of the whole.

FIGURE 2 Quoting practice from the viewpoint of consumption (number of quotes per article)

Comments: Based on domestic general/political/economic news and normalised from the viewpoint of consumption.

While the analysis above normalised how output is split into articles to make visible the changes in the journalistic process, from the viewpoint of the news consumer, it is also necessary to see how the number of quotes in each article develops. This point of view is affected by changes in typical article length, but nevertheless better reflects readers' experiences of the sourcing practices. This view, seen in Figure 2, corroborates the steady increase in the use of direct quotations already seen in Figure 1. However, from this perspective, the use of indirect quotes also rises linearly at the same speed as the use of direct quotes across all the outlets. Thus, from the viewpoint of a reader, if one in 2010 was likely to encounter approximately two indirect quotes per article and a direct one every second article, in 2018 one was likely to see two to three times that many, for both types of quotes, in each article.

Changes in overall diversity of sources (RQ2)

To get an overall high-level understanding of how the diversity of sources develops over time, a summary statistic can be used, provided that the data fed into it is representative of the whole. In this article, we are able to calculate such a statistic for the automatically detected individual sources for direct and indirect quotes. *Iltalehti* has been left out of this analysis due to the dataset issues discussed previously, and the manual social media sample is neither representative enough to be of use.

With these reservations, using the Gini coefficient over all distinct individual sources as our summary diversity measure and plotting it at different times in Figure 3, we can see that approaching the present day and disregarding short-term variability, overall source diversity generally diminishes, or at most stays the same instead of growing.

FIGURE 3 Diversity of sourcing practices over time (Gini coefficient)
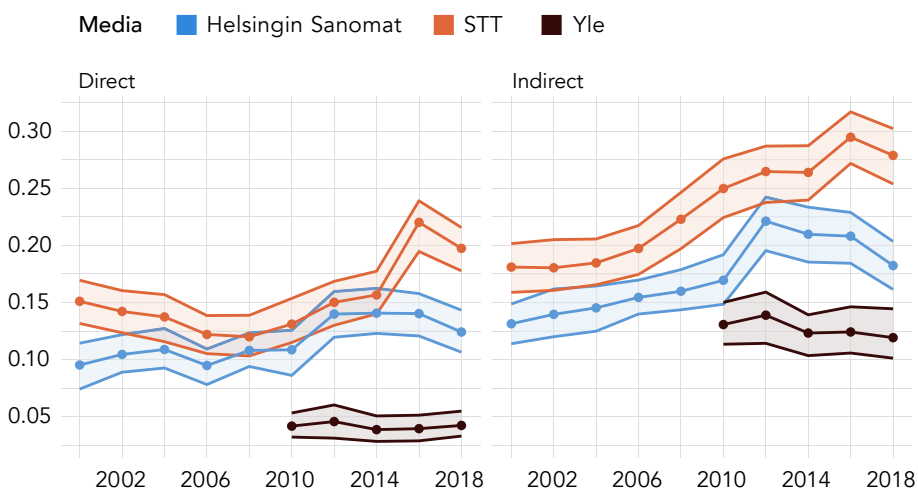
Comments: Lower values indicate more diversity. Points aggregate quotations from the 2.5 surrounding years to either side. Bands indicate 95 per cent confidence intervals for the estimates.

Examination of Lorenz curves at the different time points confirms that this process of consolidation happens on all levels, with the top 1 per cent getting ever more quotes in comparison with the 99 per cent, and the top 10 per cent likewise being more often quoted than the remaining 90 per cent, and so on. Interpreted in relation to the overall development of quoting frequency, this finding means that while there is an increasing number of quotes in the articles, the vast majority of them come from an ever-smaller group of people, particularly considering the fact that the total number of distinct persons sourced stays relatively constant through time.
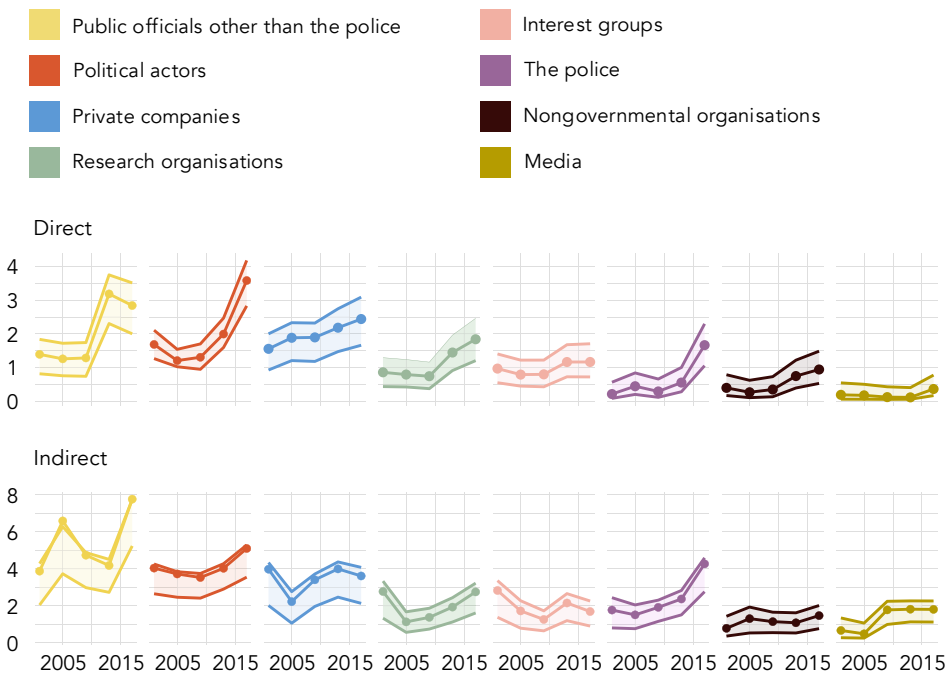
Another conclusion is that in general, there is greater diversity among directly quoted sources than indirectly quoted sources. In our earlier work (Koivunen et al., 2021), we detected that whereas indirect quotes were used to attach a source to factual information, direct quotes were often used to add affective intensity to the articles. Against this backdrop, the smaller diversity in indirectly quoted sources appears to highlight the gatekeeping function of the journalistic media: only a strictly limited, and in many ways an ever-more tightly curated pool of actors is considered authoritative and trustworthy as sources of factual information.

Finally, while a concentration and consolidation of the pool of sources characterises the development in all news outlets, there are substantial differences between the absolute levels of concentration in the different outlets at each point in time. Here, particularly the Finnish public broadcaster Yle stands out from the other news media, as it quotes both directly and indirectly from a broader list of sources. This finding corresponds with observations from other countries with strong, independent public service media (Tiffen et al., 2014).

Changes in representation of societal groups in quoted sources (RQ3)

Proceeding from the high-level overview of source consolidation to address the question of representation, we first look at traditional direct and indirect quoting practices, through both the absolute number of quotes different groups have at various points in time (see Figure 4) and the share they make of the whole body of quotes (see Figure 5). Here, Figure 4 gives us the perspective of how often a reader is likely to find a quote from a particular group when reading a news article at different points in time. Figure 5, on the other hand, normalises the fact that quoting in general becomes progressively more common after 2010, allowing us to detect changes in proportional representation – for example, how often a reader sees a quote from a political actor in *comparison* to how often they see a quote from a nongovernmental representative.

FIGURE 4 Quoted sources (quotes per ten articles)

Legend:
- Public officials other than the police
- Political actors
- Private companies
- Research organisations
- Interest groups
- The police
- Nongovernmental organisations
- Media

Direct

Indirect

Comments: Based on domestic general/political/economic news and normalised from the view-point of consumption. Points aggregate quotations from the 2.5 surrounding years to either side. Bands indicate 95 per cent confidence intervals for the estimates.

Among both direct and indirect quotes, the highest share of quotes comes from official governmental actors (notably, the police are separated from the other actors). In terms of absolute development, particularly the indirect quoting of government actors, and especially indirect but also direct quoting of the police, increased after 2010. For direct quoting of government actors other than the police, a phenomenon can be seen where direct quoting increases at the start of the online-first publication era, but not after that. In terms of proportional representation, while both quotes from the police as well as other officials increase on the whole, it is the quoting from the police that increases more, thus increasing their overall share of the quote pool at a particular time.

Political actors are the next most quoted group. In terms of the absolute number of direct quotes, they also overtake all other groups by the end of our analysis period. At the same time, this absolute increase is not substantially larger than the overall increase in quoting, thus not leading to a substantial increase in proportional representation. At the very end of our analysis period, however, this seems to be changing, with even the proportional share of direct quotes from politicians increasing substantially. Moreover, the political actors' share of quotes fluctuates the least in time, pointing to a remarkably stable representation over time, as opposed to the representation of most other groups.

The third highest quoted group is private companies, but particularly their quote share fluctuates over time. Of the other groups, the amount of absolute quotes coming from interest groups stays relatively stable, leading to their share of the whole diminishing. At the same time, both the absolute number of quotes as well as the quote share of nongovernmental organisations fluctuate in response to what is happening in society. Finally, looking at indirect quotes, an increase in sourcing from other media outlets increased between 2007 and 2015, possibly reflecting new journalistic patterns of the hybrid media environment.

Overall, while Figure 5 demonstrates some shifts between different source groups, on the whole, their shares have remained fairly constant despite the hybridisation of the media environment and the journalistic and quotation practices it entails. Considering this in light of our finding of overall consolidation and concentration of the source pool, it can only be taken to mean that while on the level of group categories the distribution of quotes has remained relatively stable over time, within the major source categories the quotations are clustering and concentrating on fewer important individual sources. Furthermore, the proportional share of quotes from sources other than the ones shown in Figure 5 has not substantially changed over time. In our data, we thus find no evidence of new societal actors substantially broadening the pool of people featured in these four Finnish media outlets.

Effects of consolidation and clustering become evident when looking at political actors as an example. Politicians and political parties have the most stable representation in our data, which means they are quoted and given voice in a relatively stable manner throughout the research period. As the overall number of quotes increases, so does the number of quotations from political actors. However, at the same time, since the distribution of quotes clusters on certain individuals with key institutional roles, certain politicians and political roles increase their visibility. For a news reader, this entails being faced with more quotes from the same politicians. Therefore, it can be argued that despite the changes in the media and the increase in quotes overall, the journalistic arena for domestic politics has not become more inclusive. Instead, the role of politicians and authorities has been foregrounded.

FIGURE 5 Sources of social media post quotes (per cent)
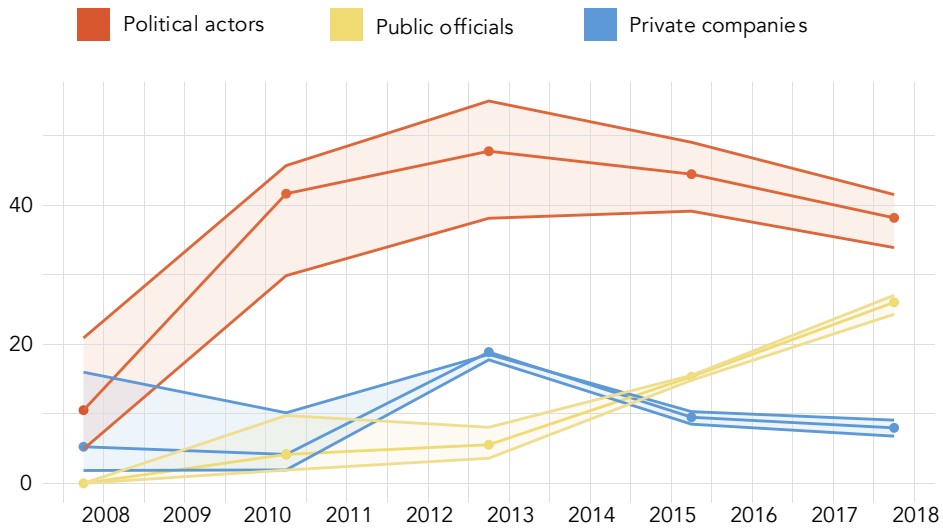
Comments: Only groups with substantial representation are included. Points aggregate quotations from the 2.5 surrounding years to either side. Bands indicate 95 per cent confidence intervals for the estimates.

We now turn to our analysis of journalists referencing posts made by actors on social media. We treat these separately to see whether the increasing share of quotes sourced this way might point to trends distinct from the more traditional sourcing practices.

Here, at the beginning of the social media era, quotes from the platforms are rare, and also non-differentiated. Social media quotes follow no clear pattern, and no group dominates. However, as the use of social media is mainstreamed in Finland, the social media quoting practices increasingly start to resemble familiar practices of quotation: The groups dominant among sources before become dominant also in social media quotes. As Figure 5 shows, while politicians are on the same level as other source categories in the beginning, they quickly become prime quoted sources also through social media. A similar rise in quotes applies to public officials who are increasingly being quoted also via social media. Their comparatively slower uptake may reflect either slower changes in quoting practices, with news outlets preferring sourcing public officials in traditional ways, or a slower uptake of social media altogether, as compared with the political elite.

Comparing sourcing from social media with sourcing in general, the share of quotes from companies is notably lower throughout the analysed period. Similarly, nongovernmental organisations and advocacy groups barely registered at all in our analyses of social media quotes at any time.

Homogenisation of Finnish media outlets in sourcing practices (RQ4)

Given the different news outlets in our dataset – a news agency, a daily newspaper, a public broadcaster, and a tabloid – one of our questions concerned potentially diverse patterns in sourcing practices. Overall, our finding is one of homogenisation. We detected no major differences in either quoting proportions or developmental trends. First, as already seen in Figure 1, overall quoting practices developed similarly in all outlets. Second, we identified few differences between the outlets in terms of proportional representation of the different societal groups. There are only two substantial differences between the outlets in this respect: *Iltalehti*, a tabloid, quoted the police slightly more often than the other outlets, while the public broadcaster, Yle, quoted somewhat fewer political actors, prioritising instead governmental actors.

Finally, in relation to the rise of social media, our findings suggest that – rather than challenging the traditional sourcing practices – this only further reinforced the voices of the political elite and the public officials in the news media.

Discussion: Institutional alignment and withering diversity

As a first outcome, to address the first research question, our analysis indicates that the hybridisation of the media system, the emergence of social media, and the transition to online-first publishing have modified news media sourcing practices. The overall development indicates an increase in both direct as well as indirect quotations over time. Especially direct quotations become distinctly more frequent towards the end of the analysed period. For the news reader, this means the news articles are becoming steadily more quote-heavy, and thus more affect-laden (Koivunen et al., 2021) and powerful in influencing news media consumers' perceptions of reality (Gibson & Zillmann, 1993).

While our quantitative analysis does not allow a comprehensive explanation to this development, several expositions can be put forward. Digital publishing is often described as a cycle, a circulation, or a flow (Chadwick, 2013), which suggests a movement or stream of information. In the hybrid media environment, content is abundant, and many actors direct their messages not only to the citizens or various target groups but also to traditional news media, who in this disrupted public sphere (Bennett & Pfetsch, 2018) act as one of the network gatekeepers, controlling attention and giving voice. The substantial rise in quotations may be partly explained by the online-first publishing tempo: an increasing need to react quickly, with outlets frequently publishing initial short stories based on information subsidies from official governmental sources and snippets from other media (see Gandy, 1982; Johnston & Forde, 2017; Seuri et al., 2021).

Also, our finding of increased quoting may be explained as indicating the persistence and continued importance of professional journalistic standards, as well as the prevalence of journalistic routines, in the hybrid media age (Bennett, 1990; Kantola, 2013; Shoemaker & Vos, 2009; Väliverronen, 2022a). Quoting sources

can be identified as a core professional practice of designating information to known and named actors, which, at the same time, legitimises the information and creates distance, if not objectivity (e.g., Tuchman, 1978). Balancing out views of different parties is a common strategy for journalists, who often lack the time or expertise to determine the veracity of a source (Dunwoody, 1999; Gans, 1979; Tuchman, 1978). The routines of news journalism can be understood simply as the means for journalists to perform the given task, which is to create a news product within a specific time limit (see Shoemaker & Vos, 2009). Thus, the professional routines of seeking a balanced position and hearing both or all sides in an argument, in other words, indicate following professional standards, but also, the same routines of making professional decisions often lead to the same sources repeatedly getting a voice in the media (Bennett, 1990).

The second outcome of our research, related to the second and third research questions, is the withering diversity among quoted sources over time. Our analysis testifies to the dominance of elite sources whose prime position among sources has persisted through the last 20 years in the Finnish news media. The hybrid media context does not change the relative visibility or shares of different categorised source groups, but instead confirms both the prominence of elite sources observed in previous studies (see Brown et al., 1987; Sigal, 1973) and the prime position of "super-influencers" (see Vesa et al., 2018; Vesa & Kantola, 2016) in the Finnish media space. Our data provides no support to predictions of the journalistic media opening to unlimited sources (Williams & Delli Carpini, 2004), or to a widening of the journalistic gate (see Bennett, 1990). Instead, our analysis corroborates the key findings of previous research: both the use of legitimised routine sources (Tuchman, 1978) and the utilisation of easily accessible sources (Gans, 1979) continue to be important in news journalism also in the age of hybrid media.

To answer the third research question, we compared how different societal groups get their voice heard through direct, indirect, and social media sourcing practices. Thus, as a third outcome, we found out how quoting social media actors follows the tradition of elite-reliance, and it seems unaffected by social media being an open platform with easily accessible and inexhaustible sources. The journalistic news media continue to give voice to the representatives of the same institutions on which they have traditionally relied on for information. After an initial period of adjustment to the social media age, all four analysed news media amplify elite voices: First, they start foregrounding the voices of politicians, after which gradually also the voices of public officials, who presumably are slower in adapting to social media platforms in their communicative practices. It seems that political tweets may be interpreted, as Parmelee (2014) has suggested, as valuable information subsidies (Gandy, 1982).

The fourth outcome of our study, addressing the fourth research question, is the substantial degree of homogenisation detected in persistent sourcing practices in our data. The four main news media outlets differ very little in terms of sourcing practices, which makes Finland stand out from many other countries with highlighted differences between different types of media (see Tiffen et al.,

2014). For instance, public broadcaster Yle and the tabloid *Iltalehti* are very similar when it comes to the relationship with institutions, sources, and giving voice. Our results corroborate previous findings (Väliverronen, 2022b) that the differences between different media outlets have diminished over the past decade. This outcome is in sharp contrast to predictions of media competition leading to diversification of news media and tabloidisation.

In this article, we have studied a period often described as a time of institutional crisis for media (Reese, 2021), as well as a period in which the relationship between news media and politics has been in flux (Väliverronen, 2022b). In most Western democracies, this period has furthermore been a time of shrinking trust in journalistic news media – Finland being one notable exception with exceptionally high trust (Newman et al., 2022). Our empirical findings offer new insights into these developments. While the news media in many countries has become polarised, or more partisan in the past decades, our data suggests that the Finnish news media rather fortified its relationships with other institutional actors: politicians and public officials. We term this development institutional alignment, and seeing it across our data of different news outlets – a news agency, a daily newspaper, a public broadcaster, and a tabloid – it may be interpreted as an industry-wide strategy of legitimation. Also, in the age of hybrid media, the importance of the news media is ever more tightly tied to its role as network gatekeepers (Barzilai-Nahon, 2008), operating between and with other institutions, curating the news flow and the national public sphere (see Thorson & Wells, 2015; Seuri et al., 2021). In this manner, journalistic news media have been able to bring together institutional power players, thereby legitimising its position as a key arena for politics.

We suggest that the institutional alignment should be interpreted as a legitimation strategy, whereby news media seeks to maintain both its gatekeeping function and its key position in networks of power. It may be partly a consequence of dwindling resources and an increased reliance of those who provide information for free, but we see it primarily as a sign of institutional cooperation in a time of disruption: traditional domains of power upholding their positions by reinforcing each other in the public realm. News journalism may be interpreted to amplify and support institutional networks of power and the tradition of seeking consensus among political, social, and economic elites (see Kunelius & Reunanen, 2012). Importantly, journalists as a professional group are, in the Finnish context, understood as members of these networks. Journalistic professionalism not only refers to journalists being educated and the field having high professional standards, but also to the social role of journalists in the networks of power. Thus, news media sourcing practices foregrounding actors exercising public power can be seen as an authority performance (Koivunen & Vuorelma, 2022) in which the news media negotiate their role and authority in network governance. This is a double-edged sword, because the news media may emphasise its integrity by making the use of sources more transparent, but by doing so, at the same time, it produces and manifests its institutional alignment, given the heavy reliance on always the same sources.

The findings of this study indicate continuities rather than ruptures. The continued sourcing practice of giving voice to the elite groups aligns with previous research findings on, for instance, Dutch economic news (Vliegenthart & Boukes, 2018) and British political news (Thorsen & Jackson, 2018). It is, however, important to note, following van Leuven and colleagues (2018), that this picture of sourcing patterns might be different if our dataset were not restricted to primary news sections. As contended, a focus on domestic, political, and economic news categories enables comparison in time and across news outlets, even if it excludes the possibility of mapping sourcing patterns across wider sections of journalism. Further, in this article, we have only *counted* quotes, without going into their content, framing, or function within the articles. While in our other work (Koivunen et al., 2021; Seuri et al., 2021), we have seen quote function correlate with quote type (indirect or direct), this is still a very blunt instrument. Moreover, we have also only looked at the sum of quotes and not the variance. Thus, while we can quite reliably show large-scale general patterns in sourcing, our approach does not allow for nuance. So, for example, while we can reliably argue for an overall trend of consolidation, there may be particular news events or article types deviating from this trend.

Conclusion: Consolidation and homogenisation

In this article, we have analysed sourcing practices, or more precisely giving voice, as a key gatekeeping function in Finnish news media between 1999 and 2018. This period has been an era in which the media environment has been substantially transformed: Digitalisation, social media, and platformisation have changed not only the way we perceive the public, societal participation, and the news media. Our findings of withering diversity in sourcing, increased institutional alignment, and overall homogenisation invoke the question of whether they indicate the persistence of the "political parallelism" which characterises the democratic corporatist media system (Hallin & Mancini, 2004) and the media welfare state model (Syvertsen et al., 2014). According to Syvertsen and colleagues, the Nordic model of the democratic corporatist media system is founded on four pillars: understanding communication as public goods, institutionalised freedom from editorial interference, media as a site of cultural policies supporting diversity and quality, as well as stateness: "a preference for consensual solutions that are durable and involve cooperation between main stakeholders: the state, media and communication industries and the public" (Syvertsen et al., 2014: 17). The elite-oriented sourcing practices identified in our analysis suggest a continued strengthening of the stateness pillar at the expense of diversity of public representation. At the same time, however, as the relationship between the state and the people in the Nordic model is seen as "a close and positive one" (Syvertsen et al., 2014: 19), the news media alignment with other public actors may, from the institutional point of view, be interpreted as including and representing the public and thereby strengthening democratic representation.

In sum, our study uncovers and makes visible a development in which the news media seek to legitimise both themselves and other key political institutions in a time of digital, and possibly also democratic, disintegration. These findings emphasise the news media's role as a political institution interacting with other institutions (see Cook, 1998), as well as the continued importance of "indexing" in press-state relations in Finland (Bennett, 1990). The relations between journalists and sources holding institutional power are evidently still very close.

While institutional alignment is the key finding of our longitudinal study, the future of this alignment is unknown. Only time will tell whether it will, normatively speaking, be good or bad for the public and the news media's role in democracy. Optimistically thinking, the cooperation of traditional institutions and their mutual reinforcement may be interpreted as democratic self-defence, a strategy for withstanding disruptive and even hostile forces. In this view, the news media, with their professional standards and conventions, act as a stabilising institution, which not only guards the good of the public but also cultivates the democratic infrastructure and the stateness pillar of the Nordic media model. From another perspective, the persistent and even growing reliance on institutional power networks will result in the news media becoming less relevant to the public, a withering, defective public discussion, with more diverse and deliberative modes of democracy moving to new arenas. In this view, the public, seizing the opportunities of a high-choice media environment, gradually move away from the news media–centred democratic model, and old institutional power networks will crumble. If this will be the case, the failure of the news media to diversify its sourcing practices and open up to the virtually unlimited sources of a hybrid media environment will look like a dramatic missed opportunity.

Acknowledgements

References

Ala-Fossi, M. (2020). Finland: Media welfare state in the digital era? *Journal of Digital Media & Policy*, *11*(2), 133–150. https://doi.org/10.1386/jdmp_00020_1

Barzilai-Nahon, K. (2008). Toward a theory of network gatekeeping: A framework for exploring information control. *Journal of the American Society for Information Science and Technology*, *59*, 1493–1512. https://doi.org/10.1002/asi.20857

Bennett, W. L. (1990). Toward a theory of press-state relations in the United States. *Journal of Communication*, *40*(2), 103–127. https://doi.org/10.1111/j.1460-2466.1990.tb02265.x

Bennett, W. L. (2014). Press–government relations in a changing media environment. In K. Kenski, & K. Hall Jamieson (Eds.), *The Oxford handbook of political communication* (pp. 249–262). Oxford Handbooks. https://doi.org/10.1093/oxfordhb/9780199793471.013.40

Bennett, W. L. (2016). *News: The politics of illusion* (10th ed.). The University of Chicago Press.

Bennett, W. L., & Pfetsch, B. (2018). Rethinking political communication in a time of disrupted public spheres. *Journal of Communication*, *68*(2), 243–253. https://doi.org/10.1093/joc/jqx017

Berkowitz, D. (1987). TV news sources and news channels: A study in agenda-building. *Journalism Quarterly*, *64*(2-3), 508–513. https://doi.org/10.1177/107769908706400231

Brown, J. D., Bybee, C. R., Wearden, S. T., & Straughn, D. M. (1987). Invisible power: Newspaper news sources and the limits of diversity. *Journalism Quarterly*, *64*(1), 45–54. https://doi.org/10.1177/107769908706400106

Chadwick, A. (2013). *The hybrid media system: Politics and power*. Oxford University Press.

Cook, T. E. (1998). *Governing with the news: The news media as a political institution*. University of Chicago Press.

DeIuliis, D. (2015). Gatekeeping theory from social fields to social networks. *Communication Research Trends*, *34*(1), 1–23. http://cscc.scu.edu/trends/v34/CRT_v34_n1_March2015.pdf

Dunwoody, S. (1999). Scientists, journalists, and the meaning of uncertainty. In S. M. Friedman, S. Dunwoody, & C. L. Rogers (Eds.). *Communicating uncertainty: Media coverage of new and controversial science* (pp. 59–79). Lawrence Erlbaum.

Fletcher, R., & Nielsen, R. K. (2018). Are people incidentally exposed to news on social media? A comparative analysis. *New Media & Society*, *20*(7), 2450–2468. https://doi.org/10.1177/1461444817724170

Gandy, O. (1982). *Beyond agenda setting: Information subsidies and public policy*. Bloomsbury Academic.

Gans, H. J. (1979). *Deciding what's news: A study of CBS Evening News, NBC Nightly News, Newsweek and Times*. Vintage Books.

Gibson, R., & Zillmann, D. (1993). The impact of quotation in news reports on issue perception. *Journalism & Mass Communication Quarterly*, *70*(4), 793–800. https://doi.org/10.1177/107769909307000405

Goel, S., Anderson, A., Hofman, J., & Watts, D. J. (2016). The structural virality of online diffusion. *Management Science*, *62*(1), 180–196. https://doi.org/10.1287/mnsc.2015.2158

Hallin, D. C., & Mancini, P. (2004). *Comparing media systems: Three models of media and politics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790867

Herkman, J. (2009). The structural transformation of the democratic corporatist model: The Case of Finland. *Javnost – The Public*, *16*(4), 73–90. https://doi.org/10.1080/13183222.2009.11009015

Hodgson, G. M. (2006). What are institutions? *Journal of Economic Issues*, *40*(1), 1–25. https://doi.org/10.1080/00213624.2006.11506879

Janicki, M., Kanner, A., & Mäkelä, E. (2023). Detection and attribution of quotes in Finnish news media: BERT vs. rule-based approach. In *Proceedings of the 24th Nordic Conference on Computational Linguistics* (NoDaLiDa) (pp. 52–59). NEALT Proceedings Series. University of Tartu Library.

Johnston, J., & Forde, S. (2017). Churnalism. *Digital Journalism*, *5*(8), 943–946. https://doi.org/10.1080/21670811.2017.1355026

Kantola, A. (2013). From gardeners to revolutionaries: The rise of the liquid ethos in political journalism. *Journalism*, *14*(5), 606–626. https://doi.org/10.1177/1464884912454504

Karlsson, M.,Van Couvering, E., & Lindell, J. (2022). Publishing, sharing, and spreading online news: A case study of gatekeeping logics in the platform era. *Nordicom Review*, *43*(2), 190–213. https://doi.org/10.2478/nor-2022-0012

Koivunen, A., Kanner, A., Janicki, M., Harju, A., Hokkanen, J., & Mäkelä, E. (2021). Emotive, evaluative, epistemic: A linguistic analysis of affectivity in news journalism. *Journalism*, 22(5), 1190–1206. https://doi.org/10.1177/1464884920985724

Koivunen, A., & Vuorelma, J. (2022). Trust and authority in the age of mediatised politics. *European Journal of Communication*, 37(4), 393–408. https://doi.org/10.1177/02673231211072653

Koljonen, K. (2013). The shift from high to liquid ideals: Making sense of journalism and its change through a multidimensional model. *Nordicom Review*, 34(S1), 141–154. https://doi.org/10.2478/nor-2013-0110

Kunelius, R., & Reunanen, E. (2012). Media in political power: A parsonian view on the differentiated mediatization of Finnish decision makers. *The International Journal of Press/Politics*, 17(1), 56–75. https://doi.org/10.1177/1940161211424207

Lecheler, S., & Kruikemeier, S. (2016). Re-evaluating journalistic routines in a digital age: A review of research on the use of online sources. *New Media & Society*, 18(1), 156–171. https://doi.org/10.1177/1461444815600412

Lehdonvirta, V. (2022). *Cloud empires: How digital platforms are overtaking the state and how we can regain control*. MIT Press.

Maurer, P., & Beiler, M. (2018). Networking and political alignment as strategies to control the news. *Journalism Studies*, 19(14), 2024–2041. https://doi.org/10.1080/1461670X.2017.1310627

Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., & Kleis Nielsen, R. (2022). *Reuters Institute digital news report 2022*. Reuters Institute for the Study of Journalism. University of Oxford. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022

Parmelee, J. H. (2014). The agenda-building function of political tweets. *New Media & Society*, 16(3), 434–450. https://doi.org/10.1177/1461444813487955

Reese, S. D. (2021). *The crisis of the institutional press*. Polity.

Reich, Z. (2006). The process model of news initiative: Sources lead first, reporters thereafter. *Journalism Studies*, 7(4), 497–514. https://doi.org/10.1080/14616700600757928

Reunanen, E., Kunelius, R., & Noppari, E. (2010). Mediatization in context: Consensus culture, media and decision making in the 21st century, the case of Finland. *Communications*, 35(3), 287–307. https://doi.org/10.1515/comm.2010.016

Richardson, J. (2006). Who gets to speak? A study of sources in the broadsheet press. In E. Poole, & J. Richardson (Eds.). *Muslims and the news media* (pp. 103–115). I. B. Tauris.

Rosen, J. (2012). The people formerly known as the audience. In M. Mandiberg (Ed.), *The social media reader* (pp. 13–16). NYU Press.

Schudson, M. (2003). *The sociology of news*. W. W. Norton.

Seuri, O., Era, R., Koivunen, A., Janicki, M., Toivanen, P., Hokkanen, J., & Mäkelä, E. (2021). Uutisvuon hallitsija: Uutismedia kiky-kamppailussa 2015–2016 [Ruler of the news flow: Traditional media in the competitiveness pact negotiations 2015–2016]. *Politiikka, 63*(3). https://doi.org/10.37452/politiikka.99432

Seuri, O., & Ikäheimo, H.-P. (2022, November 25). *Gatekeeping in the digital age* [Sitra Working Paper]. Sitra.

Shoemaker, P. (1991). *Gatekeeping*. Sage.

Shoemaker, P., & Vos, T. P. (2009). *Gatekeeping theory*. Routledge. https://doi.org/10.4324/9780203931653

Sigal, L. V. (1973). *Reporters & officials: The organisation & politics of newsmaking*. D. C. Heath.

Skogerbø, E., Bruns, A., Quodling, A., & Ingebretsen, T. (2016). Agenda-setting revisited: Social media in mainstream journalism. In G. Enli, A. Bruns, A. O. Larsson, E. Skogerbo, & C. Christensen (Eds.), *The Routledge companion to social media and politics* (pp. 104–120). Routledge. https://doi.org/10.4324/9781315716299

Sundar, S. S., & Nass, C. (2001). Conceptualizing sources in online news. *Journal of Communication*, 51, 52–72. https://doi.org/10.1111/j.1460-2466.2001.tb02872.x

Syvertsen, T., Gunn, E., Mjøs, O., & Moe, H. (2014). *The media welfare state: Nordic media in the digital era*. The University of Michigan Press. https://doi.org/10.3998/nmw.12367206.0001.001

Thorsen, E., & Jackson, D. (2018). Seven characteristics defining online news formats. *Digital Journalism*, 6(7), 847–868. https://doi.org/10.1080/21670811.2018.1468722

Thorson K., & Wells, C. (2015). How gatekeeping still matters: Understanding media effects in an era of curated flows. In T. P. Vos, & F. Heinderyckx (Eds.), *Gatekeeping in Transition* (pp. 25–44). Routledge. https://doi.org/10.4324/9781315849652

Tiffen, R., Jones, P. K., Rowe, D., Aalberg, T., Coen, S., Curran, J., Hayashi, K., Iyengar, S., Mazzoleni, G., Papathanassopoulos, S., Rojas, H., & Soroka, S. (2014). Sources in the news: A comparative study. *Journalism studies*, *15*(4), 374–391. https://doi.org/10.1080/1461670X.2013.831239

Tuchman, G. (1978). *Making news: A study in the construction of reality*. The Free Press.

Van Aelst, P., Strömbäck, J., Aalberg, T., Esser, F., de Vreese, C., Matthes, J., Hopmann, D., Salgado, S., Hubé, N., Stępińska, A., Papathanassopoulos, S., Berganza, R., Legnante, G., Reinemann, C., Sheafer, T., & Stanyer, J. (2017) Political communication in a high-choice media environment: A challenge for democracy? *Annals of the International Communication Association*, *41*(1), 3–27. https://doi.org/10.1080/23808985.2017.1288551

Van Leuven, S., Kruikemeier, S., Lecheler, S., & Hermans, L. (2018). Online and newsworthy. *Digital Journalism*, *6*(7), 798–806. https://doi.org/10.1080/21670811.2018.1498747

Vesa, J., & Kantola, A. (2016). Kuka pääsee mukaan? Miten järjestöjen ääni kuuluu lakien valmistelussa [Who can join? How the voice of organisations is heard in the preparation of laws]. Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 59/2016 [Government investigation and research publication series 59/2016.]. Prime Minister's Office Finland.

Vesa, J., Kantola, A., & Binderkrantz, A. (2018). A stronghold of routine corporatism? The involvement of interest groups in policy making in Finland. *Scandinavian Political Studies*, *41*(4), 239–262. https://doi.org/10.1111/1467-9477.12128

Vliegenthart, R., & Boukes, M. (2018) On the street and/or on Twitter? *Digital Journalism*, *6*(7), 829–846. https://doi.org/10.1080/21670811.2018.1497449

Vos, T. P. (2015). Revisiting gatekeeping theory during a time of transition. In T. P. Vos, & F. Heinderyckx (Eds.), *Gatekeeping in transition* (pp. 3–25). Routledge. https://doi.org/10.4324/9781315849652

Väliverronen, J. (2022a). Boundaries in motion? Finnish political journalists' external and internal boundary work in a time of change. *Journalism Studies*, *23*(9), 1037–1055. https://doi.org/10.1080/1461670X.2022.2065340

Väliverronen, J. (2022b). Political journalism content in a new era: The case of Finnish newspapers, 1995–2015. *European Journal of Communication*, *37*(4), 373–392. https://doi.org/10.1177/02673231211072662

Wallace, J. (2018). Modelling contemporary gatekeeping. *Digital Journalism*, *6*(3), 274–293, https://doi.org/10.1080/21670811.2017.1343648

Weaver, D., & Elliott, S. N. (1985). Who sets the agenda for the media? A study of local agenda-building. *Journalism Quarterly*, *62*(1), 87–94. https://doi.org/10.1177/107769908506200113

White, D. M. (1950). The "gate keeper": A case study in the selection of news. *Journalism Quarterly*, *27*(4), 383–390. https://doi.org/10.1177/107769905002700403

Wihbey, J. P. (2019). *The social fact: News and knowledge in a networked world*. MIT Press.

Williams, B. A., & Delli Carpini, M. X. (2004). Monica and Bill all the time and everywhere: The collapse of gatekeeping and agenda setting in the new media environment. *American Behavioral Scientist*, *47*(9), 1208–1230. https://doi.org/10.1177/00027642032623

Endnotes

1 While the data used for this article have been extensively tested and evaluated as well as used in multiple prior articles, one of the analyses conducted for this article surfaced an additional error that only affects a particular type of analysis first conducted here. This anomaly was that the dataset given to us by *Iltalehti* had many domestic news articles missing from individual months from 2009–2013. These are also missing from the website of the paper itself, so were probably lost at the source. Because we mostly calculate metrics either on a yearly basis or proportionally, this skew has not affected our calculations in previous articles, nor even the calculations in the current article, apart from the one exception that surfaced the problem. That one exception is the general diversity metric we use, which is affected by the monthly pattern of what is and isn't there for this source and period. As a consequence, we have left *Iltalehti* out of our general diversity analysis. This is noted also explicitly in the relevant section.

2 Based on an evaluation (Janicki et al., 2023), our approach captures 86 per cent of direct quotes in the material (recall), and 93 per cent of the passages it detects are actual direct quotes (precision). Indirect quotes are harder to detect, and relating to them, in our evaluation we were able to capture only 66 per cent of these indirect quotes, and our precision was also substantially worse at 75 per cent. Correct attribution is harder still. Here, while our method obtains a precision of 91 per cent, we are able to capture only 58 per cent of quote attributions. However, because in this article we are more interested in proportions and their changes through time, absolute recall numbers are not actually important. Instead, our requirement is that what is recalled should be a representative sample of the whole. Based on the high precision, as well as the fact that a more in-depth error-analysis yielded no apparent patterns of bias, this does seem to hold. Thus, we conclude that our method should bring us an unbiased view of both the quotes as well as their attributions in the material.

3 Of the manual social media sample, we categorised all actors. Of the automatically extracted direct and indirect quotes, to ensure good coverage and statistical power of our analyses, we categorised, for each year in our 20-year timespan, all the top-100 most often appearing actors in both direct and indirect quotes (718 unique actors). At the same time, to ensure that our data overall has a good representation of also the long tail of rarely quoted actors, we also categorised a random sample of altogether 1,200 quotes, filtered to not include quotes from the already categorised 718 actors, as well as stratified by quote type (direct/indirect), media outlet, and coarse time period (before/after 2009). Taken together, these three samples finally provided us with a dataset of group attributions that is both representative as well as large enough to show differences in a statistically robust manner. More specifically, due to the large datasets used, all results reported in this article pass a test for statistical significance at least to the level that equal differences could appear randomly at most one time in a thousand ($p \leq 0.001$). Differences remain between the semiautomatically derived set of indirect and direct quotes and the fully manually categorised social media dataset in terms of analysis resolution, however. With the larger dataset, we are able to reliably talk about differences in quoting between outlets as well as track the representation

of smaller groups. For the social media quote dataset, our resolution is limited to an aggregation of all outlets and only the three largest of the societal groups and an aggregation of all the others.