



European
University
Institute

ROBERT
SCHUMAN
CENTRE FOR
ADVANCED
STUDIES

WORKING PAPERS

RSCAS 2020/05

Robert Schuman Centre for Advanced Studies
Centre for Media Pluralism and Media Freedom

Ensuring Pluralism in Social Media Markets: Some Suggestions

Maria Luisa Stasi

European University Institute

Robert Schuman Centre for Advanced Studies

Centre for Media Pluralism and Media Freedom

Ensuring Pluralism in Social Media Markets: Some Suggestions

Maria Luisa Stasi

EUI Working Paper **RSCAS** 2020/05

This text may be downloaded only for personal research purposes. Additional reproduction for other purposes, whether in hard copies or electronically, requires the consent of the author(s), editor(s). If cited or quoted, reference should be made to the full name of the author(s), editor(s), the title, the working paper, or other series, the year and the publisher.

ISSN 1028-3625

© Maria Luisa Stasi, 2020

Printed in Italy, January 2020
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu/RSCAS/Publications/
www.eui.eu
cadmus.eui.eu

Robert Schuman Centre for Advanced Studies

The Robert Schuman Centre for Advanced Studies, created in 1992 and currently directed by Professor Brigid Laffan, aims to develop inter-disciplinary and comparative research on the major issues facing the process of European integration, European societies and Europe's place in 21st century global politics.

The Centre is home to a large post-doctoral programme and hosts major research programmes, projects and data sets, in addition to a range of working groups and *ad hoc* initiatives. The research agenda is organised around a set of core themes and is continuously evolving, reflecting the changing agenda of European integration, the expanding membership of the European Union, developments in Europe's neighbourhood and the wider world.

For more information: <http://eui.eu/rscas>

The EUI and the RSCAS are not responsible for the opinion expressed by the author(s).

Centre for Media Pluralism and Media Freedom (CMPF) Working Paper Series on 'Freedom and Pluralism of the Media, Society and Markets' benefits from contributions from the CMPF's fellows as well as from leading scholars and experienced practitioners interested in and focused on the subject matter. The Working Papers Series aims at assessing theoretical issues, specific policies, and regulatory questions.

The Centre for Media Pluralism and Media Freedom is co-financed by the European Union. This initiative is a further step in the European Commission's on-going effort to improve the protection of media pluralism and media freedom in Europe and to establish what actions need to be taken at European or national levels to foster these objectives.

The aim of the EUI Centre for Media Pluralism and Media Freedom is to enhance the awareness of the importance of freedom and pluralism of the media, to contribute to its protection and promotion and to develop new ideas among academics, policy makers, regulators, market stakeholders, journalists, and all other directly involved professionals who take part in the public debate.

Abstract

Social media platforms have created unprecedented possibilities to communicate, access and share information online. However, they have also raised various challenges for both individuals and society. One is how they impact the plurality and diversity of media that each user is exposed to. This paper attempts to shed light on the current challenges concerning pluralism and content moderation in social media markets. It does so by identifying the relevant market failures and the likely theory of harm, and it suggests two possible solutions: the first is a form of regulated pluralism; the second is the unbundling of hosting and content moderation activities.

Keywords

Content moderation, media plurality and diversity, regulation, unbundling.

Introduction

Social media platforms have revolutionised the way people communicate, access and share information. In little over a decade, they have reached billions of people around the globe, enabling exciting achievements but also raising serious concerns relating to the protection of human rights. The impact they have had on society is multi-faceted, and has opened the door to many research questions, relevant for social, political and economic studies alike.

Much has been said already, both in support of and against the influence of these platforms. Recently, the major social media companies have been under attack for a number of reasons, facing proceedings before data protection and competition authorities as well as national and regional courts, and have been targeted by various legislative, regulatory and policy actions.

Notwithstanding this hype, the various challenges that social media platforms raise for competitors, users and society still require sustained scrutiny, and resolution. Among other questions, the way that Facebook, Twitter, YouTube and similar platforms moderate content remains at the centre of harsh debate. ‘Content moderation’, *per se*, is quite a vague concept, which is accused of playing a major role in phenomena like the spread of disinformation, the creation of echo chambers, the spread of hate speech online, and the manipulation of elections, among other things.

Less prominent in the discussion of content moderation on social media platforms is the issue of reduced pluralism and the diversity of information which is accessible and exposed to individual users. Access to a plurality and diversity of voices is a fundamental pillar of our democracies, and a necessary premise for an open and informed public discourse that allows citizens to form their opinions and to engage in civic space. The artificial reduction of pluralism and diversity – caused by the profit-driven business models of private actors – is a worrying evolution that must be addressed by legislators and enforcers to protect our democratic society.

This paper attempts to shed light on the current challenges concerning pluralism and content moderation in social media markets. It does so by identifying the relevant market failures, the likely theory of harm and suggests possible solutions. Part I looks at the relevant market and at content moderation practices as exercised by the dominant platforms. Part II analyses the decrease of plurality and diversity on social media platforms and shapes the theory of harm. Part III focuses on possible remedies. Part IV briefly concludes.

Part I. Content moderation

Relevant markets

Social media platforms are multi-sided markets that serve various groups of consumers. Commonly, they offer ad space and other related services to advertisers, as well as providing hosting, content moderation, instant messaging and other services to users. In addition, some platforms offer news referral services to news publishers, connecting the latter with users. Finally, they can provide application programming interfaces (APIs) to app developers, allowing them to programme and develop tools that integrate with the platform. Therefore, scholars have tentatively identified four relevant markets where social media platforms usually operate: social networking services, advertising display, news referral and platforms for apps. (Gormsen, Llanos, 2019).

For the purpose of this paper, the relevant markets are mainly two: display advertising and social networking services. On the display advertising market, platforms enter into economic transactions with advertisers, charging a monetary fee for their services. On the social networking side, platforms generally offer their services in the absence of any monetary payment: what the platforms take from

consumers – who for the purpose of this paper will be referred to as ‘users’ – is their data. This data relates to what they access, share, like, etc. While focusing on the social networking market, this paper does not ignore the other relevant markets; in fact, the business model of social media platforms cannot be explained and assessed in chunks, but it has to be seen as a continuous, complex typology of conducts that take place in all sides of the relevant multi-sided market, which are strongly interlinked to one another.

Content moderation activities

In the social networking market, platforms offer to their users several services, currently bundled together. The first is “hosting”. For example, users can create a profile on Facebook, which is hosted on the platform, where they can upload the content, pictures, and videos they want, share their activities or life events, and interact with others via likes or comments. Back in 2000, the E-Commerce Directive¹ established that a platform offering this service shall not be held liable for the content that is stored on it, unless it has “actual knowledge” of the illegality of the content.² This safeguard is still one of the fundamental principles of intermediary liability in the EU, although this may be altered in the soon-to-come review of the regulatory framework contained in the newly elected Commission’s work plan.

The second is content moderation: social media companies also moderate the content that is uploaded and circulates on their platforms. A preamble is needed here. Platforms have to deal with two main categories of content: illegal and legal. The way platforms deal with illegal content, such as child pornography or copyrighted materials, is usually dictated by national or EU rules and is outside the scope of this paper, which instead focuses on legal content – even that legal content which might be considered harmful.³

Content moderation includes various activities, such as demotion, promotion, monetization, ranking, and removal. To define each of these activities is not an easy exercise, indeed, they are performed by algorithms and guided by business strategies which we know very little about, and this extensive information asymmetry produces many of the struggles regulators currently face when dealing with content moderation.

For the purpose of this paper, content moderation activities are defined as follows. Promotion and demotion refer to increasing or decreasing the visibility of certain content, for example on Facebook News Feeds or on Twitter timelines. Some platforms explain that promotion and demotion are based on a process called ranking, which creates personalised streams of posts, tweets, and more, which are based on various criteria, including: how often the user interacts with posts from friends, groups of pages, reactions and shares to posts by the user, and how recently something was posted by the user (Facebook, 2019; Twitter, 2019). Platforms also offer various options to promote content through payment (Facebook, 2019; Twitter, 2019).

Removal can be seen as a form of content moderation too. Social media companies can decide to remove certain types of content that, albeit legal, they do not want to host on their platform. For example, a social media company might decide not to allow, and thus to remove, nudity from its platform. Removal can happen at the platform’s initiative, or can be requested by users who report the content to the platform, with the platform then making the decision whether to remove the content.

¹ Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce') *OJ L 178, 17.7.2000, p. 1–16*

² Directive on E-Commerce, cit., Article 14.

³ There is not sufficient legal clarity, nor policy consensus on what constitute legal but harmful content (see: UK DCMS and Home Department, 2019; Baines, 2019). For the purpose of this paper, I limit myself to observing that legal but harmful content falls within the broad category of legal content.

Some information about how different platforms moderate content is contained in the platforms' terms of services or community policies (hereafter, described together as Terms of Services, or ToS). ToS regulate the contractual relationship between the platform and its users, and describe the basic rules that the former imposes on the latter about the content that is admitted, or not admitted, and about how this content is treated. Nevertheless, ToS provide very little to no information about how, in practice, platforms demote or promote content (for example, based on which criteria). Nor do they provide any information about the activities the platforms put in place that constitute the essential premise for demotion and promotion (for example, users' profiling). In other words, ToS explain very little to nothing about the platforms' business model. This shows that users are not sufficiently aware of why they see what they see on their timeline and that regulators might not be able to properly assess if the service is provided in a way that complies with consumer protection and human rights standards.

I have argued in previous work (Stasi, 2019) that in order to gain a proper understanding of content moderation, one has to look at the bigger picture and consider the complex typology of conducts put in place by the platform, the final outcome of which is to decide what content each user is exposed to. In the case of an advertising-driven social media platform, this complex typology includes users' profiling and the personalisation of content.

Profiling implies the collection and use of personal data to segment people according to precise profiles. Major social media platforms use dozens of thousands of indicators to classify people's interests and personal attributes; in addition, automated profiling is capable of identifying patterns that are invisible to the human eye. Profiling requires data harvesting: the more platforms know about their users, the more they can profile them. It follows that the first step for this model to work is to gather as much personal data from users as possible, and definitely more data than would be necessary to offer them the social media service alone.

The more companies can profile their users, the more they can personalise, via automated systems, the services they offer to them, and this includes prices, advertising and content in general. Personalisation can be seen as a specific combination of promotion, demotion and removal tailored to the profile of each individual user. This paper focuses on the impact of personalisation on the diversity of content that each user is exposed to.

In theory, promotion or demotion of specific content can happen irrespective of profiling and personalisation; content moderation can be offered without profiling users. The offer of personalised content could also be based on criteria which has been provided by users, voluntarily, based on their preferences. In practice, though, the major social media platforms provide content moderation based on profiling and personalisation driven by maximisation of their monetisation channel. The principal monetisation channel for these platforms is advertising.

With regard to removal, ToS usually inform users about which content is allowed on the platform, which content is not, and how the platform will react to prohibited content. Faced with huge pressure to remove illegal speech (terrorist content, incitement to hatred, etc.), platforms may be incentivised to shape their ToS in an overly-restrictive way, in order to avoid incurring regulatory sanctions, but also societal blame. Various examples⁴ demonstrate that a grey area exists between illegal content – which

⁴ Among other cases, following the recent Turkish invasion of Kurdistan, Facebook has obscured a number of Italian pro-Kurdish pages on unclear grounds (see: https://www.repubblica.it/tecnologia/2019/10/22/news/facebook_censura_pagine_con_contenuti_pro_curdi_oscurate-239204458/); again in Italy, Facebook recently deactivated various pages linked to the political parties Forza Nuova and Casapound because they violated Facebook's policy on hate speech (see: <https://www.ilfattoquotidiano.it/2019/09/09/casapound-e-forza-nuova-offline-pagine-su-facebook-e-instagram-violano-policy-diffondono-odio/5441137/>). However, the Civil Court of Rome upheld the appeal filed by CasaPound and ordered Facebook to reactivate the page of the movement, cancelled last September 9, 2019. As explained in the judgment, the Court considered that 'it is evident the prominent importance assumed by the service of Facebook (or other social networks connected to it) with reference to the implementation of essential principles of the system such as the pluralism of political parties (art. 49 Const.), to the point that the subject who is not present on Facebook is in fact excluded (or severely limited)

social media platforms have to remove under international standards and national law – and legal content, including harmful content, which platforms regulate under their ToS.

Moreover, social media companies remain free to place boundaries on the kind of content they will host on their platforms. Nevertheless, it must be noted that when platforms are as dominant in the market as to represent the main channel of communication for individuals, their decision to include or exclude content strongly affect individuals' freedom of expression and access to information. There is a vibrant debate about whether such dominant or super dominant platforms should be considered as public spaces, public utilities or, to recur to the antitrust jargon, an essential facility, and therefore could be charged with must carry (or must view/ show) obligations. Scholars (Keller, 2019; Feld, 2019) remain hesitant, and to date domestic and regional courts have never taken this approach.

This section has tried to shed light on content moderation activities. As has been shown, ToS do not provide sufficient information about how platforms perform these activities, in particular for what concerns the criteria and the mechanisms used to select the content each user is exposed to. Therefore, for regulators dealing with content moderation and assessing its impact on individuals' freedom of expression and on media diversity in the market, ToS constitute an important element, but definitely not the only element, to be taken into account. The next section discusses the negative impact that content moderation – as currently performed by dominant social media platforms – has on plurality and the diversity of information goods as available to individual users.

Part II. Pluralism

Plurality and diversity

The concept of pluralism in the media sector is commonly used to refer to, on the one hand, the 'plurality' of suppliers – that is the plurality of media actors, types of media, and ownership – and, on the other hand, the 'diversity' of content – that is the broadest possible diversity of information, ideas and viewpoints that are communicated.

Media pluralism, together with media freedom, is an essential component of an open and free debate in society, and thus a pillar of our democracy. As the Council of Europe recently reminded us: "As the ultimate guarantors of pluralism, States have a positive obligation to put in place an appropriate legislative and policy framework to that end. This implies adopting appropriate measures to ensure sufficient variety in the overall range of media types, bearing in mind differences in terms of their purposes, functions and geographical reach" (Council of Europe, 2018). The appropriate measures, for the Council of Europe, should consist of promoting "the availability, findability and accessibility of the broadest possible diversity of media content as well as the representation of the whole diversity of society in the media" (Council of Europe, 2018).

Reduction of plurality and diversity in media is a problem both at the societal level and for individual users, because it limits their freedom of expression and information. According to the International Covenant on Civil and Political Rights (ICCPR), the right to freedom of expression "shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers..."

from the Italian political debate [...] It follows that the relationship between Facebook and the user who intends to register for the service (or with the user already enabled to the service as in the case in question) is not comparable to the relationship between any two private parties as one of the parties, precisely Facebook, holds a special position: this special position means that Facebook, in bargaining with users, must strictly adhere to the constitutional and legal principles until it is shown (with verification to be done through a fully aware phase) the violation by the user. Respect for constitutional and legal principles is for Facebook at once a condition and limit in the relationship with users who request access to its service." For reference (in Italian), see:

<https://www.ilpost.it/2019/12/12/facebook-casapound-riattivazione-pagina-tribunale-roma/>

(ICCPR, 1966). As pluralism in the media is essential for seeking and receiving information, a reduction in media pluralism constitutes also a limitation of individuals' freedom of expression.

One can look at plurality from a collective/societal perspective - e.g. how plural and diverse is the information available in a country – and from an individual one – e.g. how plural and diverse is the content each individual is exposed to. As previously mentioned, this paper focuses on the latter.

Pluralism in social media markets

From a purely quantitative perspective, today we have more plurality and diversity online than ever before. Digitalisation has made the creation and distribution of content substantially easier, cheaper and faster. Therefore, at least in principle, society has at its disposal an increasingly plural and diverse body of content.

Nevertheless, while this premise is true for online media in general, the picture looks different when we narrow the scope to social media markets, and when we look, not only at the information collectively available on which market, but rather, at the plurality and diversity of information each individual user is exposed to. Indeed, it can be argued that individuals are not exposed to the wide variety of information available online, but rather to only a tiny fraction of it. This happens for a number of reasons.

The first is excessive concentration in the market (Parcu, 2019). Social media platforms are currently one of the main channels for distribution of, and access to content. By way of example, Ofcom reports found that in 2019 about 50% of adults in the UK accessed news via social media, and the percentage is higher in people aged 16-24 (Ofcom, 2019a). However, the social media market is dominated by a few companies only, therefore what is distributed by these few companies is visible to a vast public, while what is not distributed by these few companies might not be visible to the majority of individuals. Moreover, the social media market presents high barriers to entry and it does not appear easily contestable. As a result, it can be argued that existing dominant social media companies constitute a bottleneck in the distribution of content, greatly affecting what users are exposed to.

Furthermore, social media platforms' business model is not at all neutral towards diversity. As mentioned before, the majority of these companies' business model is based on advertising. As a consequence, considering that platforms monetise users' attention, they have no incentives to expose users to all content potentially available, but only to the tiny portion of it that will keep the users more engaged. The platforms therefore shape their content moderation activities accordingly (Tufekci, 2016). Studies demonstrate that the most engaging content is typically what Mark Zuckerberg called 'borderline' (Zuckerberg, 2018), that is the content which is more sensationalist and provocative, including those stories that appeal to our baser instincts and trigger outrage and fear (Vosoughi, Roy, Aral, 2018; Tufekci, 2017).

As a result, it can be argued that, because of the way platforms perform content moderation – that is, because of the criteria and the automated system they use to promote, demote, and remove content, and to personalise users' timelines or newsfeeds – each social media user is only exposed to a subset of the vast amount of sources and viewpoints available online.

What has been described so far refers mainly to that content which users are passively exposed to. Users remain free, at least in theory, to actively look for more diverse and plural content. In practice, though, they will rarely do so. Behavioural studies have well-documented that users do not always choose the best course of action, and are generally reluctant to act outside of the status-quo (Behavioural Insights Team, 2019; ForbrukerRadet, 2018). Furthermore, platforms are able to exploit these users' bias through specific designs that promote addictive behaviours (Ofcom, 2019b). Among other practices, they do so by using default settings, dark patterns and similar forms of nudging to which users very rarely react. In other words, users will rarely, if ever, change the default settings on their profile on Facebook or Twitter, despite the fact that doing so could free them from the platforms' massive interference with the content they see, and could allow them to access more diverse information. Users

are even less likely to do so in a context of extreme asymmetry of information and lack of transparency – a context which characterises social media markets. Within these markets, users do not have sufficient awareness of the platforms’ business models nor of the incentives that determine the promotion, demotion and personalisation of content. An additional obstacle for users is the presence of high switching costs, often artificially inflated by platforms, which deprive them of alternative choices.

To conclude, this section has argued that while the plurality and diversity of content online is wider than ever before, on the contrary, when we narrow the scope to social media markets, it is possible to argue that – due to the way that platforms moderate content – the variety of content individual users are exposed to is shrinking. Among the causes of this reduction, it is possible to identify excessive market concentration, the bottleneck role played by dominant players in the distribution of content, and the advertising-driven business model that typically provides incentives which are adversarial to plurality and diversity.

For the reasons explained in the previous parts of this paper, the reduction of plurality and diversity in the social media market can be seen as both a market failure, and as a worrying constraint on people’s right to freedom of expression, which legislators and enforcers should address with care. As clearly explained by Ofcom in a recent report (Ofcom, 2019b), the complex interaction that exists between market failure and constraint to freedom of expression has to be taken into due account when looking for a solution to the problem. In essence, in order to be effective, any possible remedies to the harm must also address market failure. The next section of this paper discusses the kind of remedies available.

Part III. Possible solutions

There are arguably various ways that regulators could intervene to fix the problem at stake. *Ex ante* remedies appear to be the best fit because of the various characteristics of the social media market. First, it’s a fast moving market, where developments in technology and business models occurs rapidly, definitely more rapidly than in traditional *ex post* competition enforcement. Second, social media markets are prone to tipping. Market dynamics are as such that economies of scale and scope, once achieved, raise barriers to entry and make it difficult to reverse the situation and to make markets competitive and contestable again. Indeed, the difficulties in tackling consolidated positions of (abusive) dominance fuel extreme calls to “break up” the giants⁵.

In addition, *ex ante* remedies present a number of advantages: they can be tailored to specific situations, based on the principles of necessity and proportionality, they can be behavioural or structural in nature, and are potentially an open-ended list. Finally, the procedure to impose *ex ante* remedies could be less formal than an infringement procedure, and, depending on circumstances, it could imply a certain degree of negotiation, or at least confrontation, between the enforcer and the target.

In the case at hand – where we are confronted with a harm which is strictly linked to a market failure originating from excessive market concentration and high barriers to entry – regulators could decide to use asymmetric regulation to fix the problem. This would involve imposing specific obligations on players with significant market power (SMP), or – as distinguished scholars have argued – on those with significant media market power (SMMP; Noam, 2019).

This paper identifies two main *ex ante* regulatory solutions: (i) to impose, on SMP or SMMP players, a sort of regulated pluralism; (ii) to impose, again on SMP or SMMP players, the functional separation between hosting and content moderation activities, in order to open the market to competitors and

⁵ See, among many: The Economist, Tech giants face new threats from the government and regulators, 14 March 2019, available at: <https://www.economist.com/united-states/2019/03/14/tech-giants-face-new-threats-from-the-government-and-regulators>; The New York Times, Elizabeth Warren Proposes Breaking Up Tech Giants Like Amazon and Facebook, 8 March 2019, available at: <https://www.nytimes.com/2019/03/08/us/politics/elizabeth-warren-amazon.html>

reinstall competitive dynamics. In the following paragraphs, the pros and cons of both solutions are briefly assessed.

A. Regulated pluralism

The first option is thus to regulate content moderation in a way that guarantees pluralism. As previously mentioned, the decrease of diversity in social media markets is not linked to the availability of content, but to its distribution: social media platforms distribute content to their users via content moderation activities, which therefore act as bottlenecks for diversity. Seen from the user's perspective, the decrease in content does not concern the content they actively access, but rather, the content they are passively exposed to on a daily basis. As the market failure concerns distribution, and thus content moderation, a possible solution is to impose some form of "must view" obligation on the platforms. In other words, an obligation for the platforms to expose individual users to some degree of content diversity.

Similar obligations can be found in traditional broadcasting regulation. Many EU member States have chosen to impose positive content obligations on public service broadcasters, to protect the role that content diversity plays in the public debate and in people's engagement in society. The market failure this regulation is aimed to address is somehow similar to the one at stake in social media markets: as broadcasters need spectrum, which is a limited resource, there might be bottlenecks in the production and distribution of certain type of content, which might not be the most profitable content for broadcasters.

Moreover, "must view" obligations resemble the must carry obligations long used in the telecom sector. In that sector, the market failure to address was, and still is, the impossibility of practically or reasonably duplicating the infrastructure needed to provide the service to consumers. Therefore, incumbents have been obliged to provide alternative players with access to the infrastructure instrumental to the provision of services to consumers.

Decision-makers and regulators have started to look at this possible solution. The German broadcasting authority has adopted a framework to impose diversity obligations on media content moderation algorithms (Rundfunkkommission, 2019). The new rules impose on video platforms a number of specific duties concerning: non-discrimination in terms of access for content providers, and search and browsing access for users; priority for public service content; and the choices to be guaranteed to users with regard to content ranking and sorting. In addition, the proposal contains rules targeted at so-called 'media intermediaries', a concept that includes, but that is possibly not limited to, social media platforms. Article 53(e) of the proposal requires that media intermediaries "may not unfairly disadvantage (directly or indirectly) or treat differently providers of journalistic editorial content to the extent that the intermediary has potentially a significant influence on their visibility." However, media intermediaries are not supposed to provide users with choices concerning the ranking or sorting algorithms. Therefore, notwithstanding the additional requirements in terms of transparency, the new rules appear too mild to remedy the harm at stake. Nonetheless they could represent a step in the right direction, and it will certainly be interesting to observe how they will be implemented by the authority and what impact they will have on the market.

It should also be considered that, regulated pluralism, while providing the impression of being a targeted and well-suited remedy, brings several fundamental challenges for regulators.

First, rules would have to identify various thresholds for intervention. Which level of SMP, or SMMP would trigger regulatory obligations? How should the rules define diversity of content and viewpoints? How much diversity is enough diversity? In other words, which and how many viewpoints need to reach the individual user for the latter to be exposed to a sufficient degree of diversity? Second, would obligations be imposed on 'generalist' platforms only, or also on 'specialised' ones, like, for example, LinkedIn? Third, and possibly most importantly, to impose a certain degree of diversity while performing content moderation possibly implies an obligation for companies to monitor all content that

circulates on their platforms. This could result in a form of general monitoring of users speech by private actors, which could constitute a blatant violation of freedom of expression. Moreover, such control could open the door to censorship, or to an ever greater influence on the public debate.

In addition, as recalled in previous parts of this paper, the current EU regulatory framework provides hosting platforms with a liability exemption insofar as they are seen to not interact in any way with the content they host. This liability exemption is intended for those platforms who limit themselves to “*the technical process of operating and giving access to a communication network over which information made available by third parties is transmitted or temporarily stored, for the sole purpose of making the transmission more efficient; this activity is of a mere technical, automatic and passive nature, which implies that the information society service provider has neither knowledge of nor control over the information which is transmitted or stored*” (E-commerce Directive, 2000, Recital 42). In a scenario where companies are charged with must view obligations, it can be argued that they will necessarily interact with the content, thus losing the safe harbour, or liability exemption. One should then consider the additional consequences of this remedy, and assess if the trade-offs are compatible with the results it aims to achieve.

Finally, imposing an obligation for platforms to ensure diversity may imply that the State or the regulators will dictate the criteria through which the algorithm will perform content moderation. We may then question which safeguards should be put in place to avoid arbitrary intervention, which could result in censorship or propaganda. Some have noted that similar problems could emerge, for example, in the approach taken in the EU Code of Practice on Disinformation,⁶ which asks platforms to prioritize “trustworthy” content (Helberger, Leerssen, Van Drunen, 2019). To overcome this challenge, the Council of Europe has recently suggested a multi-stakeholder approach, calling on States to encourage social media, media, search and recommendation engines and other intermediaries which use algorithms, along with media actors, regulatory authorities, civil society, academia and other relevant stakeholders to engage in open, independent, transparent and participatory initiatives aimed at improving distribution processes in order to enhance users’ effective exposure to the broadest possible diversity of media content, among other aims (Council of Europe, 2018, section 25).

This approach is supported by various stakeholders. For example, some civil society groups have suggested the creation of ‘Social Media Councils’, a multi-stakeholder accountability mechanism for content moderation on social media. The Councils aim to provide an open, transparent, accountable and participatory forum to address content moderation issues on social media platforms on the basis of international standards on human rights (ARTICLE 19, 2019) and could help to establish best practices with regard to diversity of content.

B. (Unbundled) access to content moderation

The second option is to separate hosting activities from content moderation activities and to oblige dominant platforms to allow third parties to offer content moderation to the platforms’ users. For example, a user that creates or has a profile on Facebook should be asked by the platform whether she wants the content moderation service to be provided by Facebook itself, or by other players to be freely selected. The option to stay with the dominant platform should be presented as opt-in, rather than opt-out. Such a measure will help to defend against users’ unwillingness to change from the status-quo.

Functional separation is not a novelty in the toolbox of regulatory authorities. It has been widely used in a number of sectors, including telecoms, where it represents an effective instrument to achieve vertical separation in the market and to ensure more competition at the wholesale, as well as at the retail level. Functional separation includes a spectrum of possibilities that vary depending on the intrusiveness of the measure on the company’s freedom of economic activity and on its structure. Indeed, functional

⁶ EU Code of Practice on Disinformation, 2018, available at: <https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation>.

separation can be imposed via contractual agreement, or it can require the complete dismissal of the facility or of the infrastructure that is needed to provide a certain service. In some cases, the company is impeded from providing the service that has been separated, while in other cases it can still do so, but it has to provide competitors with the ability to provide that same service – usually by granting access to the infrastructure which is needed to do so.

The form of functional separation suggested here presents a number of advantages.

First, it appears to be capable of addressing not only the reduction of plurality and diversity on social media, but also the market failure that originates from this reduction. Functional separation opens up the content moderation market to competitors. For this to be effective, access for competitors should be provided based on fair, transparent, non-discriminatory and reasonable terms. This remedy would make the market contestable again. This competitive pressure would stimulate companies to innovate and offer better quality services, to the benefit of consumers.

Another advantage is that if competition is reinstalled in the market, plurality and diversity of content would be ensured in a sustainable way both in the long-term as well as the short-term. A healthy competitive market would deliver better quality services *per se*, without the need for further regulation.

In addition, the greater availability of choices will increase the bargaining power of consumers in relation to suppliers, and will enable users to discipline, to a certain extent, the quality and diversity of the content offered to them by switching to another supplier of content moderation if they are not satisfied by the one they are currently using.

Finally, this remedy will not necessarily imply a change in the current intermediary liability regime that social media platforms are subjected to. Indeed, separating content moderation from hosting – and mandating the possibility for third parties to provide content moderation services on dominant social media platforms – does not imply matching any quotas in the diversity of sources and viewpoints, and therefore does not require the imposition of a general monitoring obligation. Aside from eliminating the risk of violating the international standards of freedom of expression, it also has the advantage of being compatible with the European Commission's plans to review the E-Commerce Directive within the framework of its Digital Service Act initiative.

Nonetheless, this remedy also presents a number of challenges, which should be duly addressed by enforcers.

As previously mentioned, the way content is currently moderated on social media platforms is shaped by the advertising-driven business model adopted by these platforms. If the alternative operators use the same business model (advertising-based) the issue of the reduction of plurality and diversity of content might not be solved. Therefore, the key question becomes how economically sustainable other business models may be, and whether there are incentives to stimulate their adoption.

Competing providers could, for example, charge a fee for moderating content based on the parameters the consumers select. At present, rather than providing the service for free, platforms harvest as much data as possible from consumers in order to profile them and monetize the platform through targeted advertising. The sustainability of an alternative business model depends, to a large extent, on consumers' willingness to pay, which in turn depends on how much consumers value plurality and diversity of content. A way to stimulate this willingness to pay could be through media policies that increase digital literacy. More specifically, this could be achieved, firstly, through media policies that increase consumers' awareness of the key role played by media plurality and diversity, in determining their capacity to make informed decisions and engage in public life. Secondly, this could be achieved through policies that reduce the information asymmetries in social media markets, for example, by mandating or encouraging more transparency around platforms' business models and the use of algorithms in content moderation.

Another way to promote alternative business models could be by providing support via public funding for a temporary period. In this scenario, while the costs society would have to bear may be justified by the importance of the public objective to be achieved, in the long term this solution may not be sustainable.

The second major challenge concerns a number of technical aspects of functional separation. In order to be able to provide content moderation on social media platforms, third parties would need to have access to the platform's APIs, or to be able to integrate their own API on the social media platform. At present all the major social media platforms have their own APIs and provide app developers access to them according to different conditions, so it is reasonable to believe that the first option would be the easiest solution.

While access to the APIs is essential for this remedy to function in practice, a number of related issues must also be carefully addressed. An adequate system should be put in place in order to guarantee that consumers' data is collected, processed, stored and used according to GDPR⁷ rules and principles by all parties involved. Depending on the system used by the content moderator provider to deliver its services to the consumer, it will need to collect and process more or less of the consumers' data.

Indeed, automated systems used for content recommendation can work based on a vast variety of criteria, such as recency, popularity, semantic relevance and so on, and these criteria may be combined in various ways. Personalisation based on profiling is one criterion, but not the only one. It is certainly possible to build a system that does not need to be fed by a large amount of consumers' personal data, but yet delivers a variety of content based on consumers' preferences. In other words, it is possible to build a system where the personalisation or recommendation of content is based on the preferences of consumers as explicitly expressed by them in the service's settings, rather than through a, to some extent, hidden profiling exercise performed by the platform. These specific criteria can then be mixed with more common criteria such as popularity, recency or semantic relevance. Regardless, the content moderation provider will not necessarily need to process consumers' personal data.

In any case, irrespective of the amount of data that is collected in a specific case, it remains clear that behaviours and practices must be in compliance with the GDPR. The companies involved – the platforms on the one hand and the content moderation providers on the other hand – would have to agree on how to proceed via contractual agreements that should clearly allocate liability, provide adequate remedies for consumers, and establish arbitration mechanisms – all of which could help to solve problems in a more flexible and fast way. Without doubt, enforcers would have a role to play too, in providing guidance and by supporting the consolidation of best practice to the benefit of all the stakeholders involved.

Concerns similar to those highlighted for data protection could be raised with regard to security. Here as well, the introduction of an intermediary could create uncertainty about the allocation of responsibilities and liabilities, which should be dissipated either through adequate contractual agreements, or, if needed, via regulatory intervention.

More generally, the use of algorithms to make content recommendations – irrespective of whether they are used by the dominant social media platforms or by alternative operators – should be compliant with international standards on human rights. In particular, the design and deployment of algorithms should comply with the criteria suggested by various norms and recommendations that are progressively contributing to shape a regulatory framework for AI (HLEG on AI, 2019; Council of Europe Commissioner for Human Rights, 2019; UN Special Rapporteur, 2018; UN Guiding Principles, 2011). At the bare minimum, players should properly assess – both at the design phase and through the entire process – the potential impact of the algorithm on users' fundamental rights, and should put in place

⁷ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation, GDPR), OJ L 119, 04.05.2016.

efficient measures to avoid or minimise any negative impacts. In addition, players should put in place adequate remedies in case the negative impact occurs and make them easily accessible for users.

Part IV. Conclusions

The massive process of digitalisation over the past two decades has enormously increased the possibilities for individuals to create, share and access content online. While this never wider amount of information is theoretically available for all, the ability of individuals to concretely reach and enjoy such plurality and diversity is strongly limited, in social media markets, by the way content is moderated by dominant platforms.

Indeed, dominant social media provide content moderation based on a profit maximisation logic. Not only does this fail to promote diversity, but it strongly reduces it. Indeed, as the business model of dominant social media platforms is based on advertising, these platforms do not have any incentive to moderate content in a way that exposes users to a representative selection of all of the sources and viewpoints available online. On the contrary, the platforms are incentivised to favour the content that engages users the most, because they will then be able to sell more attention to advertisers. Integral to this system is the massive profiling of users, based on which the platforms strongly personalise the content they expose to each user.

As a result, social media users are exposed to only a tiny fraction of the vast diversity of content available online. This artificial reduction limits their right to freedom of expression and has a negative impact on society as well. This harm occurs also because of specific market failures present in the social media market, therefore, these market failures have to be properly assessed while looking for a solution.

This paper suggests and analyses two possible remedies to the challenge identified. The first is to regulate pluralism by imposing ‘must view’ obligations on dominant players. The second is to functionally separate hosting activities from content moderation activities and to oblige dominant platforms to provide access to those competitors that want to provide content moderation on their platforms.

The pros and cons of both remedies have been explored. One of the main differences between the two concerns is the vision behind them and the outcome to be achieved in the long term. Regulated pluralism implies a more intrusive intervention by the State, in which the State takes the responsibility to decide what and how much diversity citizens should be exposed to. In other words, the State sets the metrics and becomes the guardian of a specific dimension – media diversity – which is a fundamental constituent of our democracies.

On the contrary, with the functional separation remedy the State intervenes on market dynamics and relies on healthy competition among players to ensure media diversity. Here, the metrics are not set by the State but delivered by the market; it could be said that the State shapes the playing field but not the results of the game – which depend on the players’ behaviours. With this in mind, functional separation appears to be the best placed solution; it provides an adequate response to the market failures involved in the harm. Nevertheless, for this remedy to work properly and not to undermine users’ human rights, a few additional challenges have to be solved. This paper has tried to identify these challenges, while noting that they would benefit substantially from further research and debate in order to be properly addressed.

To that aim, a better knowledge of how algorithms for content recommendation work would certainly be beneficial for regulators to make more informed decisions. There is still scope for additional research, the main obstacle to which appears to be the lack of access to the information that is needed to perform it. Regulators could play a role here too, for example: obliging platforms and content moderators to be more transparent about the automated systems they use, and in general about their business models, and to oblige them to provide access to information and data for independent research.

Furthermore, technical standards could help to make automated systems – and therefore content moderation – more compliant with human rights. To this aim, a continuous dialogue between the industry and other interested stakeholders would likely lead to better outcomes for users as well as for society as a whole.

References

- ARTICLE 19 (2019). Social Media Councils: Consultation Paper June 2019, available at: <https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf>
- Baines, V. (2019). On Online Harms and Folk Devils: Careful Now, Medium, available at: <https://medium.com/@vicbaines/on-online-harms-and-folk-devils-careful-now-f8b63ee25584>
- Behavioural Insights Team (2019). The Behavioural Science of Online Harm and Manipulation, and What to Do about It. Report available at: <https://www.bi.team/publications/the-behavioural-science-of-online-harm-and-manipulation-and-what-to-do-about-it/>
- Council of Europe Commissioner for Human Rights (2019). Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights, available at: <https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64>
- Council of Europe (2018). Appendix to Recommendation CM/Rec(2018)1, Guidelines on media pluralism and transparency of media ownership
- EU High Level Group on Artificial Intelligence (2018). Ethics Guidelines for Trustworthy AI, available at: <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>
- Facebook (2019). Advertising policies, available at <https://www.facebook.com/policies/ads/>
- Facebook (2019). Terms of Services, available at: <https://www.facebook.com/legal/terms/update>
- Feld H. (2019). The Case for the Digital Platform Act: Market Structure and Regulation of Digital Platforms, Roosevelt Institute, Public Knowledge
- ForbrukerRadet (2018). Deceived by Design. How tech companies use dark patterns to discourage us from exercising our rights to privacy
- Helberger N., Leersen P., Van Drunen M. (2019). Germany proposes Europe’s first diversity rules for social media platforms. LSE blog, available at: <https://blogs.lse.ac.uk/medialse/2019/05/29/germany-proposes-europes-first-diversity-rules-for-social-media-platforms/>
- International Covenant on Civil and Political Rights (1966). Adopted and opened for signature, ratification and accession by General Assembly resolution 2200A (XXI) of 16 December 1966 into force 23 March 1976, in accordance with Article 49
- Keller, D (2019). Who Do You Sue? State and Platforms Hybrid Power Over Online Speech, Hoover Institution Essay, Aegis Series Paper No. 1902
- Lovdahl Gormsen, L. and Llanos, J.T., (2019). Facebook’s Anticompetitive Lean in Strategies, available at SSRN: <https://ssrn.com/abstract=3400204> or <http://dx.doi.org/10.2139/ssrn.3400204>
- Noam E, (2019). Online Video: Content Models, Emerging Market Structure, and Regulatory Policy Solutions
- Ofcom (2019). News Consumption in the UK: 2019. Main Findings, available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0027/157914/uk-news-consumption-2019-report.pdf

- Ofcom (2019). Online Market Failure and Harms. An Economic Perspective on the Challenges and Opportunities in Regulating Online Services, available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0025/174634/online-market-failures-and-harms.pdf
- Parcu P. (2019). New Digital Threats To Media Pluralism in the Information Age, RSCAS/CMPF Working Paper, RSCAS 2019/19, available at: https://cadmus.eui.eu/bitstream/handle/1814/61890/RSCAS%202019_19.pdf?sequence=1&isAllowed=y
- Rundfunkkommission, 2019
- Stasi, M. L. (2019). Social media platforms and content exposure: How to restore users' control. *Competition and Regulation in Network Industries*, 20(1), 86–110. <https://doi.org/10.1177/1783591719847545>
- Tufekci Z. (2016). The Rial Bias Built in at Facebook, New York Times, available at: <https://www.nytimes.com/2016/05/19/opinion/the-real-bias-built-in-at-facebook.html>
- Twitter, (2019). Rules and Policies, available at: <https://help.twitter.com/en/rules-and-policies>
- Twitter (2019). Twitter Ads, available at: <https://business.twitter.com/en/solutions.html>
- United Kingdom, DCMS and Home Department (2019). Online Harms White Paper, available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf
- United Nation Special Rapporteur on Freedom of Expression (2018). Report of the Special Rapporteur to the General Assembly on Artificial Intelligence technologies and implications for the information environment, A/73/348
- United Nation Special Representative of the Secretary-General on human rights and transnational corporations and other business enterprises (2011). Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework, Annex to Report A/HRC/17/31, available at: https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf
- Vosoughi S., Roy D., Aral S (2018). The spread of true and false news online. *Science* Vol. 359, Issue 6380, pp. 1146-1151, DOI: 10.1126/science.aap9559
- Zuckergeberg M. (2018). A Blueprint for Content Governance and Enforcement, available at: <https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/>

Author contacts:

Maria Luisa Stasi

ARTICLE 19

60 Farringdon Road

EC1R 3GA

London

United Kingdom

and Tilburg Law School

Email: maria@article19.org

