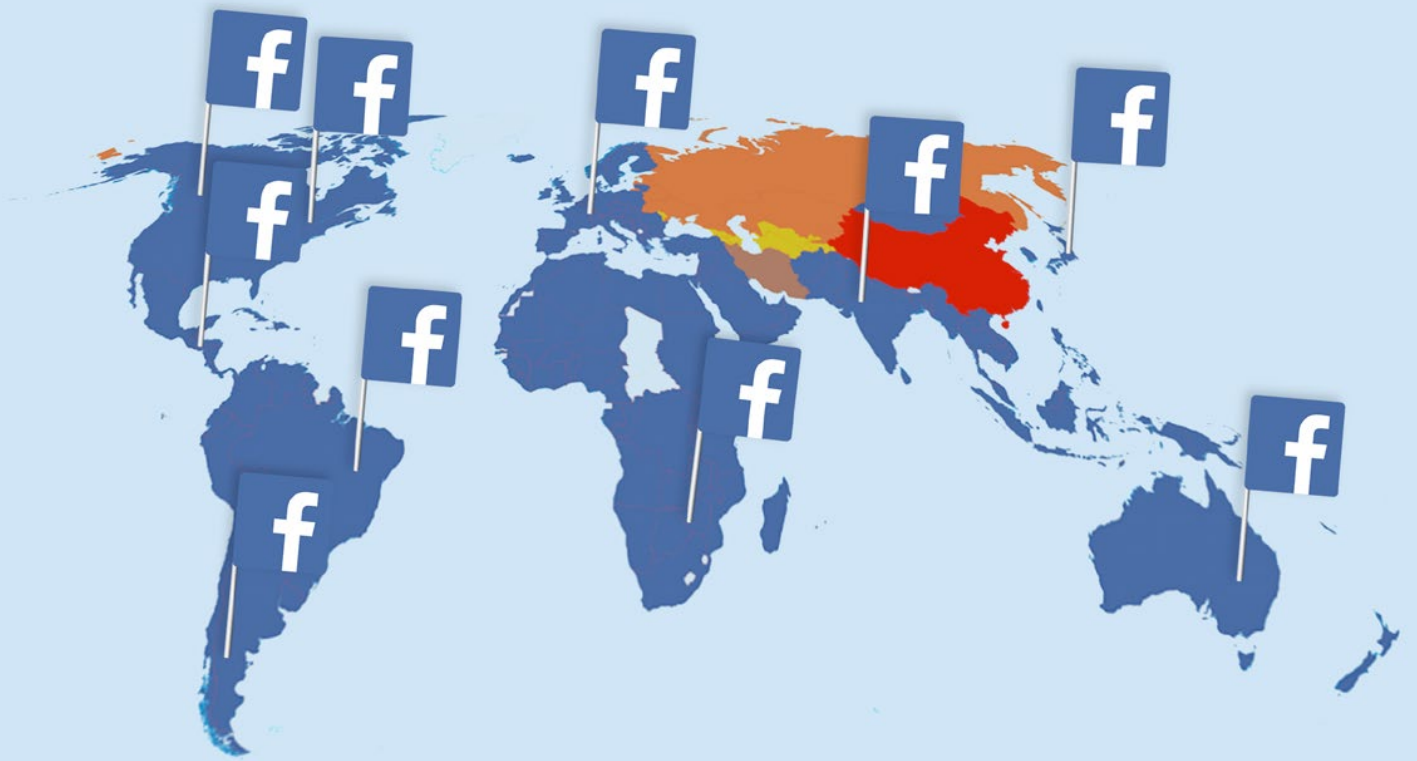
REUTERS INSTITUTE
FOR THE STUDY OF JOURNALISM

UNIVERSITY OF OXFORD

GLASNOST!
Nine ways Facebook can make itself a better forum for free speech and democracy

Timothy Garton Ash, Robert Gorwa, Danaë Metaxa

AN OXFORD-STANFORD REPORT



GLASNOST!

Nine ways Facebook can make itself a better forum for free speech and democracy

Timothy Garton Ash, Robert Gorwa, Danaë Metaxa

Dahrendorf Programme for the Study of Freedom

Stanford PACS
Center on Philanthropy and Civil Society
Project on Democracy and the Internet

An Oxford-Stanford report prepared by the Free Speech Debate project of the Dahrendorf Programme for the Study of Freedom, St. Antony's College, Oxford, in partnership with the Reuters Institute for the Study of Journalism, University of Oxford, the Project on Democracy and the Internet, Stanford University, and the Hoover Institution, Stanford University.

Contents

Summary of Recommendations

About the Authors 6

Acknowledgements 6

Introduction 7

1. Content Policy and the Moderation of Political Speech 9
   - The Problem 9
   - What Facebook Has Done 9
   - What Facebook Should Do 11

2. News Feed: Towards More Diverse, Trustworthy Political Information 14
   - The Problem 14
   - What Facebook Has Done 14
   - What Facebook Should Do 16

3. Governance: From Transparency to Accountability 18
   - The Problem 18
   - What Facebook Has Done 18
   - What Facebook Should Do 19

Conclusion 21

References 23

Summary of Recommendations

1. Tighten Community Standards wording on hate speech ... 11

2. Hire more and contextually expert content reviewers ... 11

3. Increase 'decisional transparency' ... 12

4. Expand and improve the appeals process ... 13

5. Provide meaningful News Feed controls for users ... 16

6. Expand context and fact-checking facilities ... 17

7. Establish regular auditing mechanisms ... 19

8. Create an external content policy advisory group ... 19

9. Establish an external appeals body ... 20

About the Authors

Timothy Garton Ash is Professor of European Studies at the University of Oxford, Isaiah Berlin Professorial Fellow at St Antony's College, Oxford, and a Senior Fellow at the Hoover Institution, Stanford University. He is the author of ten books of political writing or 'history of the present' including *The Magic Lantern: The Revolution of '89 Witnessed in Warsaw, Budapest, Berlin, and Prague*; *The File: A Personal History; In Europe's Name*; and *Facts are Subversive*. He directs the 13-language Oxford University research project Free Speech Debate – freespeechdebate.com – and his latest book is *Free Speech: Ten Principles for a Connected World*.

Robert Gorwa is a DPhil candidate in the Department of Politics and International Relations at the University of Oxford. Gorwa's dissertation research, supported by a SSHRC Doctoral Fellowship, examines the political role of large technology platforms, with a focus on changing notions of corporate power and private governance in the digital age. His public writing on technology and society has been published in the *Los Angeles Review of Books*, *Foreign Affairs*, *Wired Magazine (UK)*, the *Washington Post*, Quartz, and other outlets.

Danaë Metaxa is a PhD candidate in Computer Science and McCoy Center for Ethics in Society fellow at Stanford University, supported by a Stanford Graduate Fellowship. Co-advised by James Landay (Computer Science) and Jeff Hancock (Communication), Metaxa's research focuses on bias and web technologies, including the role of cognitive and psychological biases in user interfaces, and political partisanship in web search.

Acknowledgements

At Oxford, we are most grateful for comments, suggestions, and discussions around the subject matter of this report to Rasmus Kleis Nielsen, Kate O'Regan, Helen Margetts, Jacob Rowbottom, Ken MacDonald, Jonathan Bright, Carol Atack, Mark Damazer, Richard Sorabji, Alan Rusbridger, David Levy, Philip Howard, Tim Berners-Lee, Marina Jirotka, and Alexandra Borchardt. Special thanks to Alex Reid of the Reuters Institute for the Study of Journalism, who expertly saw this report through to publication. For administrative support, we give warm thanks to Maxime Dargaud-Fons, Jana Buehler, and all the staff at the European Studies Centre and St Antony's College, Oxford.

At Stanford, we owe particular thanks to Nate Persily, Rob Reich, and Eloise Duvillier of the Project on Democracy and the Internet, and to Tom Gilligan, Chris Dauer, the late, much-missed Celeste Szeto, and all the staff at the Hoover Institution. For comments, suggestions, and discussions around the subject matter of this report, we are also grateful to Daphne Keller, Francis Fukuyama, Sam Wineburg, Larry Diamond, Eileen Donahoe, Alex Stamos, Michael McFaul, and Michal Kosinski.

At Facebook, we owe particular thanks to Monika Bickert, Andy O'Connell, Parisa Zagat, Justine Isola, and Ruchika Budhraja. For other discussions and comments, we are grateful to Elliot Schrage, and other Facebook staff, too numerous to name, who took part in 'under the hood' workshops at Facebook and in Oxford. We look forward to continued conversations with Nick Clegg in his new incarnation as Vice President, Global Affairs and Communications.

No funding was received from Facebook for this report. For financial support, we are most grateful to the Project on Democracy and the Internet at Stanford University, the John S. and James L. Knight Foundation, and the foundations that have funded the Dahrendorf Programme for the Study of Freedom at St Antony's College, Oxford.

Introduction

The products deployed by firms like Facebook, Google, and Twitter have come to play a major role in the personal, political, and cultural life of billions of people (Van Dijck, Poell, and de Waal, 2018). Platforms are increasingly providing a central mechanism for news consumption, communication, and entertainment (Newman et al., 2018). Their growing influence, coupled with a series of high-profile public scandals, has led to widespread concerns about the political influence of hate speech, harassment, extremist content contributing to terrorism, polarisation, disinformation, and covert political advertising (Tucker et al., 2017), alongside mounting calls for government regulation.

A platform with more than 2.2 billion monthly active users, Facebook has found itself at the epicentre of many of the ongoing conversations about digital media, technology policy, and democracy. Following controversies such as those around the 2016 United States elections, political firm Cambridge Analytica, and inter-ethnic violence in Myanmar, Facebook no longer faces a choice between self-regulation and no regulation. Various forms of government regulation, spanning changes in intermediary liability provisions, data protection laws, antitrust measures, and more, are on the horizon, and have already been implemented with Germany's NetzDG (colloquially known as its 'Facebook law') and the EU's General Data Protection Regulation (GDPR). Mark Zuckerberg has himself recently noted that 'regulation is inevitable' (Jalonick and Ortutay, 2018). These changes will significantly affect the future of the internet and the way that people around the world use platform services.

In the meantime, platform companies are seeking to implement much-needed processes for self-regulation and governance. Responding to a flood of criticism and media commentary, including high-profile discussions of how to 'fix' Facebook (Manjoo and Roose, 2017; Osnos, 2018), the company is finally acknowledging its own public interest responsibilities as a major part of the global digital public square. Its public interest duties can be formulated in terms of international human rights norms, as outlined by UN Special Rapporteur David Kaye (2018) and the UN's so-called Ruggie Principles on business and human rights (Ruggie, 2013), and by reference to more generally formulated free-speech norms (Garton Ash, 2016). They can also be derived, at least to some extent, from Facebook's own proclaimed purposes, modestly summarised in Zuckerberg's (2017) manifesto as bringing all humankind together in a global community/ worldwide communities with five defining qualities: supportive, safe, informed, civically engaged, and inclusive.

Transparency has emerged as a key means through which Facebook has attempted to regain the trust of the public, politicians, and regulatory authorities. Facebook is now partnering with academics to create a reputable mechanism for third-party data access and independent research (King and Persily, 2018), and there is a growing body of high-quality research about the platform. Facebook has also made a number of major changes, such as introducing an appeals process for removals of most categories of content, active downranking of fact-checked and debunked news stories, and creating a public-facing 'library' of political advertisements. (For a detailed timeline of developments on Facebook and other platforms see Tow Center for Digital Journalism, 2018.)

With these developments, Facebook has entered a new era of cautious *glasnost*, inviting academics (including the authors of, and many advisers to, this report), civil society activists, journalists, and policymakers to look 'under the hood' of various aspects of its operations, understand how it formulates and implements its policies, and contribute their comments and suggestions.

This short report aims to build on this research and these interactions by (a) identifying some important specific issues concerning political information and political speech on Facebook, (b) providing an overview of the major changes that Facebook has made in recent years, and then (c) assessing these changes, and offering suggestions as to what more it should do.

We concentrate on three main areas. **Content policy**, the process which develops the rules for what speech and activity is permissible on Facebook, and then enforces those rules across the platform, is finally being recognised as a crucial issue for free speech and democracy around the globe (and has correspondingly become increasingly politicised and controversial). **News Feed**, the assemblage of algorithms that directly mediates the information to which Facebook users are exposed, is a central subject of current debates around political information, polarisation, and media manipulation. **Governance** is broader: it goes beyond transparency to the internal governance of Facebook, and the mounting concern that Facebook is not adequately accountable to its users and other stakeholders in the countries and communities in which it operates.

Needless to say, this report is not comprehensive. Its focus is tightly on some areas where we have identified things that Facebook itself can do. It does not look at the totality of issues with Facebook in relation to free speech and democracy, including around anonymity, bias and discrimination, and 'information operations' or disinformation campaigns orchestrated by foreign powers. It does not address what many have identified as the overarching problem with Facebook and other social media platforms: their business model, based on advertising and the surveillance of their users (Srnicek, 2016; Vaidhyanathan, 2018). It does not explore longstanding issues about the treatment of publishers and media organisations on Facebook, and its impact on their business models (Nielsen and Ganter, 2017; Bell et al., 2017). It looks only in passing at the important question of digital and media literacy (McGrew et al. 2017; boyd, 2017).

We understand that self-regulatory action is non-binding and can be a poor substitute for appropriate regulation, particularly in the long term. In this respect, we attach particular importance to the competition regulation being effectively advanced by the EU, and increasingly advocated in the United States (Khan, 2016). Competition is a vital way to ensure a pluralism of sources of political information and fora for political speech, and Facebook's social media platform and other wholly-owned subsidiaries (notably WhatsApp and Instagram) are so widely used as to pose real concerns about monopoly. Additionally, only comprehensive and thoughtful data protection regulation can ensure the fundamental rights of data subjects over time. The growing political influence of companies like Facebook spans industries and will require a highly creative policy approach and multiple different policy instruments. That said, the goal of this report is to focus on areas that Facebook itself can feasibly improve *now*.

To extend the metaphor from East–West relations, this report explores the possibilities of 'constructive engagement' to match Facebook's 'glasnost'. In the course of working on this report we had multiple interactions with senior staff at Facebook, and Facebook has provided comments, both on and off the record, to inform our presentation of its policies and practices. We also owe a great debt of gratitude to the multi-faceted input of many colleagues at Oxford, Stanford, and elsewhere. The final wording and judgements are, however, entirely our own.

In the following, we identify nine ways in which Facebook could incrementally make itself a better forum for free speech and democracy.

1. Content Policy and the Moderation of Political Speech

The Problem

Moderation processes have traditionally been very opaque (Roberts, 2018). On Facebook, users have historically known very little about the fundamental rules of the road: what they are permitted to post, versus what is not allowed (Gillespie, 2018). Facebook's Community Standards, which outline what sort of speech and content is not permitted, were perfunctory and unclear: for instance, while they noted that 'terrorist content' and 'sexual content' were not permitted, users did not know specifically what was considered to fall into those categories (Myers West, 2018). Users who had content removed had no possibility to appeal that decision (Klonick, 2017).

Recurrent public incidents (such as the repeated removal of the 'Terror of War', a famous Vietnam War photo of a naked, napalm-burned child, from the page of a Norwegian newspaper on child nudity grounds) (Wong, 2016), have fed a widespread perception that content policy has been ad hoc, shrouded in secrecy, and lacking in nuance. This impression has been compounded by public suspicion that content generated by public figures and Pages is subject to a different set of standards than that of ordinary users.

Investigative reporting, such as an in-depth Radiolab look at the evolution of Facebook's content policy (Adler, 2018), and a detailed Motherboard investigation (Koebler and Cox, 2018), as well as our own enquiries, strongly suggest that Facebook's policy changes and priorities are often highly reactive to outrage and scandal in the relatively narrow spectrum of Western media and politics. The concerns of non-Western users – sometimes a matter of life and death, as in Myanmar, Sri Lanka, or Libya – have been addressed much more slowly and sporadically, and often only after they have received attention from leading Western media, governments, or NGOs. Indeed, this reactive character of Facebook's policy responses – launch the product and then fix the problems afterwards, recalling the early internet motto 'Ready, Fire, Aim', or its own now-discarded motto 'Move Fast and Break Things' – seems to us a structural problem with Facebook. As a mature company, taking seriously its extraordinary global platform power and public interest responsibilities, it surely needs to work on developing more consistent and robust ways of anticipating problems in advance, rather than repeatedly scrambling to clear up the mess afterwards.

What Facebook Has Done

Among other significant changes, Facebook has published some internal guidelines for the enforcement of Community Standards, and data about the enforcement of these standards, established an appeals process, and more than doubled the number of its content reviewers.

Hard Questions – June 2017

Facebook started the Hard Questions blog to walk through key issues, such as moderation, and provide more information and transparency into how moderation is conducted, and by whom (Lyons, 2018b).

Expanded Community Standards – April 2018

An extended 30-page version of the Community Standards was released (Bikert, 2018a). This new document not only provides justification for the overarching aims of the policy, but also provides far more detail about the guidance on the basis of which types of content are removed by moderators.

Increased number of content reviewers

In response to the increased interest in this area, controversies around the use of Facebook in Myanmar and other countries, and the new German 'Facebook law' (NetzDG), Facebook has significantly increased the number of its content reviewers (most of whom remain contract workers) (Roberts, 2017). At the beginning of 2018 there were slightly more than 7500 (of whom 1500 were in Germany) (Silver, 2018). In an interview with Recode, Zuckerberg said: 'We're a profitable enough company to have 20,000 people go work on reviewing content, so I think that means that we have a responsibility to go do that.' (Swisher, 2018) However, there is some ambiguity as to how this number is used: in her autumn 2018 testimony to the Senate Intelligence Committee, COO Sheryl Sandberg said: 'We have more than doubled the number of people working on safety and security, and now have over 20,000.'[1] In Zuckerberg's November 2018 essay, 'A Blueprint for Content Governance and Enforcement', which acknowledged the many issues with Facebook's current moderation regime, he stated that 'the team responsible for enforcing these policies is made up of around 30,000 people, including content reviewers who speak almost every language widely used in the world'. Facebook tells us that at the beginning of 2019 it had some 15,000 content reviewers.

Appeals – April 2018

Simultaneously with the new Community Standards, Facebook introduced appeals for six content categories: nudity, sexual activity, hate speech, graphic violence, bullying, and harassment (Bikert, 2018a). Appeals are now also being rolled out for content that engages with dangerous organisations and individuals (including terrorist propaganda) and spam, demonstrating a move towards appeals for most if not all takedowns on Facebook. These appeals currently involve the horizontal re-moderation of a piece of content by a different content moderator; however, Facebook is extending this process, sending appeals to a more senior moderator with more context about the apparent offence (such as user information or more information about the context in which a comment was made).

Community Standards Enforcement Report – May 2018

The new Community Standards were followed by Facebook's first transparency report that illustrates different types of content takedowns, providing aggregate data into how much content is removed (Facebook, 2018). For six parts of the Community Standards (graphic violence, adult nudity and sexual activity, terrorist content, hate speech, spam, and inauthentic accounts), Facebook provides four data points: how many Community Standards violations were found, the percentage of flagged content upon which action is taken; the amount of violating content found and flagged by automated systems; and the speed with which the company takes action on these violations (York, 2018). The second report was published six months later, in November 2018, and Zuckerberg has announced that, beginning in 2019, these reports will be published quarterly and will be accompanied with a conference call discussing that quarter's content-related developments, beginning in 2019 (Zuckerberg, 2018).

Blueprint for Content Governance – November 2018

In a wide-ranging essay published under Mark Zuckerberg's name in November 2018, the company announced a major potential shift in how it approaches content moderation issues. Zuckerberg stated that he has 'increasingly come to believe that Facebook should not make so many important decisions about free expression and safety on our own', and therefore will move towards the company involving some form of external moderation oversight. This consultation period for the first possible element of this process, an external appeals 'court', which would involve outside stakeholders in the appeals process, is slated to begin in 2019 (see the section titled 'Governance: From Transparency to Accountability' below).

[1] https://www.intelligence.senate.gov/sites/default/files/documents/os-ssandberg-090518.pdf

What Facebook Should Do

1. Tighten the wording of its Community Standards on hate speech

The expanded Community Standards are an important step towards transparency, and are clear and consistent overall, but the wording on key areas, such as hate speech, remains overbroad, leading to erratic, inconsistent, and often context-insensitive takedowns. For instance, Facebook's definition of hate speech includes, in relation to people with a wide range of protected characteristics (race, ethnicity, national origin, gender, etc.): 'Expressions of contempt or their visual equivalent, including "I hate", "I don't like" … Expressions of disgust including "gross", "vile", "disgusting".'[2] A short additional section bans 'Content that describes or negatively targets people with slurs, where slurs are defined as words commonly used as insulting labels for the above-listed characteristics.' As Facebook itself acknowledges, these very broad wordings cause many problems of interpretation that are not satisfactorily resolvable by automated processes, and generate a high proportion of contested cases. **Clearer, more tightly drawn definitions** would make more consistent implementation easier, especially when combined with the measures recommended in #2, #3, and #9 below.

2. Hire more and culturally expert content reviewers

The fact that Facebook will have more than doubled the number of its content reviewers in less than two years strongly suggests that until now it has not had enough of them to meet legitimate, and sometimes urgent, public interest and human rights concerns in many different countries. The issue is quality as well as quantity. Reporting from Myanmar has shown that, despite one of the largest genocides of the 21st century, Facebook still has too little content policy capacity there (as of June 2018, only 60 Burmese-speaking contractors; as of November 2018, 100) (Stecklow, 2018). Similar problems have been reported in Sri Lanka (Goel, Kumar, and Frenkel, 2018), Libya, and the Philippines (Alba, 2018). In these cases, content that was not just hate speech but dangerous speech was left up for too long, often with disastrous consequences. In other cases, however, political speech that should have been allowed in the particular context was taken down (an amusing example (Rosenberg, 2018) was the takedown of a passage from the Declaration of Independence published in an American newspaper, on hate speech grounds), and there are also potential overblocking and freedom of expression concerns raised by the increasing external governance pressure on Facebook to remove content extremely quickly (Theil, 2018).

In both his congressional testimony and in a lengthy 'Blueprint for Content Governance' published in November 2018, Zuckerberg has argued that automated detection and flagging systems will provide an answer. The second Community Standards Enforcement Report illustrated the increasing turn towards automation for content moderation (e.g. in the latest reporting period, 67% of hate speech taken down in Myanmar was flagged by automated classifiers), which provides one answer to moderation's scale problem. However, the future impact of these systems is poorly understood, and it remains clear that AI will not resolve the issues with the deeply context-dependent judgements that need to be made in determining when, for example, hate speech becomes dangerous speech. As Zuckerberg himself ruefully observed in a call with investors, it has proved far 'easier to build an AI system to detect a nipple than what is hate speech'.

Facebook therefore needs **more human content reviewers**, with a supervising layer of **senior reviewers with relevant cultural and political expertise**. While we understand why Facebook tries to have a single, inflexible, worldwide set of standards and guidelines, this is simply unfeasible at the complex frontiers of political speech, dangerous speech, and hate speech, where context is all important. (This also relates to the unresolved ambiguity around the constantly invoked 'community': is it a single global community, or a patchwork of different communities,

[2] https://m.facebook.com/communitystandards/hate_speech/?_rdr

with somewhat different local norms?) This is emphatically not to propose a kind of cultural and moral relativism, where users in Saudi Arabia or Pakistan are not entitled to the same freedom of expression and information enjoyed by users in Rome or New York, but it is to recognise that what is merely offensive speech in one place is dangerous speech in another. **Trusted external sources (NGOs and others) should be engaged**, even more than they are already, for rapid internal review of crises as they unfold, providing an **early warning system** for Facebook to prepare more moderation capacity and better understand trouble brewing on its platform in any corner of the world.

3. Increase 'decisional transparency'

Despite the growing transparency around content moderation policies and practices, Facebook remains short on what UN Special Rapporteur David Kaye calls 'decisional transparency' (2018: 19). The public may now better understand the rules of the road, but still does not grasp how those rules are created or enforced in difficult edge cases. The 'Infowars' debacle provided a good example of inadequate transparency around decision-making (Roose, 2018).

Following an investigation by Britain's Channel 4 News, Facebook has described its Cross-Check process, whereby Pages and publishers have their content double-checked by the Community Operations team if it is flagged for removal by a human moderator (Bikert, 2018b). On occasion, Facebook will contact the publisher (e.g. if it is the *New York Times*) directly to discuss possible adjustments. In effect, there is a somewhat different standard, or at least procedure, being applied to certain established media and other Pages. Facebook has described in broad terms a set of internal guidelines, detailing the number of 'strikes' that lead accounts and Pages to be taken down (and for how long, whether merely suspended or deleted), and which vary according to kinds of content, account, and Page. Although these are obviously a work in progress, and could not be published in their entirety (not least for fear of people gaming the system) we think that **Facebook should publish more detail on these procedures**.

If we describe what platforms are doing by using the loose analogy of 'platform law', then platforms have no publicly available, systematic equivalent of case law. But they do have an internal record of multiple high-profile cases that have prompted the adjustment of policy and moderation guidelines. We think Facebook should **post and widely publicise case studies** to give users a better understanding of the basis on which content decisions are made, and therefore also potential grounds for appeal. Facebook's Hard Questions blog describes its strategies for combating false news and taking down hate speech, but seldom provides specific examples of actual hard cases. Self-regulatory organisations commonly provide such case studies in their annual reporting.

The new **Community Standards Enforcement Report** would be a good venue for more decisional transparency. As the Electronic Frontier Foundation's (EFF) Jillian York (2018) has written, the report 'deals well with how the company deals with content that violates the rules, but fails to address how the company's moderators and automated systems can get the rules wrong, taking down content that doesn't actually violate the Community Standards'. Following the publication of the second enforcement report in November 2018, civil society groups have argued that the report should provide further information about the numbers of users and pieces of content affected (in absolute terms, as opposed to partial breakdowns by violating category), how the content is being flagged (how the automated systems work, how much is being done by trusted flaggers), and provide more details about appeals (Singh, 2018). While these quantitative insights are valuable and should be expanded, these reports should go further: providing an overview of major controversies and issues of that quarter, explanations of how Facebook made the decisions it did, and what it has learned from this.

4. Expand and improve the appeals process

The new appeals process is an important step in the right direction. Facebook should continue to **expand and improve the appeals process** so that reviewers get more contextual information about the piece of content. Under the current regime, the initial internal reviewer has very limited information about the individual who posted a piece of content, despite the importance of context for adjudicating appeals. A Holocaust image has a very different significance when posted by a Holocaust survivor or by a Neo-Nazi.

Facebook shared with us, on an off-the-record basis, some interim internal statistics about appeals. These indicated that quite a large proportion of takedowns (perhaps as many as one in eight) are actually appealed, with a heavy concentration of appeals in two categories: bullying and hate speech. The highest proportion of successful appeals was on bullying, and one of the lowest on hate speech. These figures are both informative and suggestive. For example, the high proportion of appeals in the area of hate speech indicates the problem of over-breadth discussed in #1. Although we have respected Facebook's request not to give the exact percentages, since these were explicitly given on an off-the-record basis and are now out of date, we feel very strongly that this is exactly the kind of **concrete, detailed information that should be made available to analysts and users on a regular basis**.

In an open letter published in November 2018, a group of civil society organisations outlined their vision for how the appeals process could be improved (Santa Clara Principles, 2018). Noting that the appeals process will need to be functional and usable for the average user for it to be effective (it should be carefully designed from a user experience perspective), the letter advocates for a right to due process, and an ability to introduce extra evidence into the appeals system.

The process would ideally be crafted such that it is not only an evaluation of a moderator's adherence to the policies (to what extent was a Facebook policy applied correctly) but also about the important qualitative judgement of political speech/information in context (to what extent does the example show that the policy itself is incorrect). We would like to see this process developed in **global dialogue with users**, perhaps through the mechanism of the content policy advisory group suggested in #7 below, and the external appeals body discussed in #9.

2. News Feed: Towards More Diverse, Trustworthy Political Information

The Problem

The possible influence of low-quality, hyper-partisan or downright false political information (often loosely described as 'fake news') on public opinion and electoral outcomes has, especially since 2016, become the subject of multiple academic studies, journalistic reports, and governmental investigations (Allcott and Gentzkow, 2017; Guess, Nyhan and Reifler, 2018; Marwick, 2018).

Those taking advantage of Facebook to spread 'fake news' run the gamut from profit-seeking fabricators of fiction to coordinated disinformation campaigns run by foreign intelligence agencies. Facebook's advertising engine allowed Russian operatives to inject divisive content into the News Feeds of American voters (Kim et al., 2018).

A growing body of scholarship is engaging with the question of whether Facebook and other social media platforms exacerbate political polarisation, for instance by allowing users to segregate into communities of like-minded thinkers or by implementing algorithms that only show users content with which they are likely to agree. Eli Pariser and Cass Sunstein first introduced the concept of a 'filter bubble' in the early 2000s, but recent work has challenged this argument in the broader context of people's daily information consumption (Fletcher and Nielsen, 2017; Dubois and Blank 2018). The phenomenon remains vitally important but difficult to study empirically, partly due to the lack of reliable access to Facebook data for scholars.

What Facebook Has Done

Most interactions with political information happen through the News Feed, the endlessly scrolling Facebook front page composed of status updates, photos, content from Pages, and ads. While News Feed is not the only important component of Facebook (which also includes Messenger, Groups, and recently Stories), the interventions made via News Feed have the greatest impact and represent changes in the company's philosophy around social and political information more broadly.

Facebook has undertaken numerous adjustments to its News Feed with implications for free speech and democracy, including the following.

Chronological Feed – November 2011

Following a series of controversial updates to Facebook's News Feed, Facebook reinstated the option to sort the feed chronologically in the winter of 2011 (Rhee, 2011). This feature theoretically allows users to disable News Feed personalisation to display their friends' activity/stories in the order in which it happens, though it is virtually unusable in practice (as it is buried in a series of hard-to-access menus, and requires one to reset the feature every time one logs in).

Expanded News Feed Controls – November 2014

In 2014 Facebook introduced controls (Marra, 2014) for users to more easily unfollow other users, Pages, and Groups in their News Feeds in order to avoid that content while remaining connected to the entity producing it, as well as the ability to 'see less' from entities (without unfollowing altogether). However, the 'see less' functionality was opaque,[3] with users uncertain what effect it would have or how to reverse it, and since appears to have been phased out. Other Newsfeed

[3] https://www.facebook.com/help/community/question/?id=10206581533793424

controls have subsequently been introduced, including See First, Snooze, and Keyword Snooze (Fulay, 2018).

Fact-Checking Partnerships – December 2016

Following the 2016 US election, Facebook began experimenting with various fact-checking efforts. Posts flagged by several users as false, or otherwise algorithmically marked as suspicious, are reviewed by one of Facebook's fact-checking partners, who can also actively fact-check trending stories before they are flagged. (As of August 2018, Facebook had 24 partners in 17 countries;[4] as of December 2018, this number had grown to 35 in 24 countries, all of whom are members of the International Fact Checking Network organised by Poynter). Content is ranked on an eight-point scale. If it is deemed 'false', the story is downranked by News Feed to reduce visibility (Lyons, 2018a), and accompanied by a 'related articles' menu, which notes that the story is disputed and provides a link to the fact-checking organisation's analysis of the story in question. Facebook spokespeople have claimed that downranking reduces future views by up to 80% (Funke, 2018).

Context Button – October 2017

Facebook has begun adding a 'little-i' information button to news stories shared in News Feed, allowing users to easily access more context about an article, such as information about the publisher pulled from Wikipedia.[5] This feature has begun being rolled out (Hughes, Smith, and Leavitt, 2018) in the US (as of April 2018), the UK (May 2018), and several other countries: as of November 2018, these included Canada, Australia, New Zealand, Ireland, Argentina, Brazil, Colombia, Mexico, France, Germany, India, Poland, and Spain. A global roll-out was announced in December 2018.

Explore Feed – Autumn 2017 through Spring 2018

Beginning in October 2017, Facebook rolled out the Explore Feed feature in six countries, which effectively created separate feeds for users, one with posts from friends and family and the other with content from Pages (including brands, celebrities, and other organisations). This relatively small test was ended in March of 2018, with the conclusion that 'people don't want two separate feeds'. Adam Mosseri (Head of News Feed) summarised the experiment with Explore, saying that user surveys indicated that users were less satisfied with their News Feeds (Mosseri, 2018b). He also noted, however, that users indicated that Facebook 'didn't communicate the test clearly', raising the question of the degree to which users' negative responses to Explore Feed were a result of the change itself, as opposed to its problematic (opaque and involuntary) implementation.

Focusing News Feed on Friends and Family Content – January 2018

In January 2018, Facebook refocused the News Feed to prioritise 'posts that spark conversations and meaningful interactions between people' (Mosseri, 2018a). As a result, people began to see less content from Pages, including news, video, and posts from brands. Facebook stated that the proportion of 'news' in News Feed would decline from roughly 5% to roughly 4%, much to the chagrin of publishers, but argued that the quality of that news would increase. But what exactly is news according to Facebook? Facebook told us: 'The methodology used to determine the … [percentage] figure is a best estimate based on several methodologies that show the same result. One of the methods incorporates news publishers identified by ComScore, and we also use our classifiers and self-reporting on Pages.'

Multiple Security and Transparency Efforts for Political Ads – Spring/Summer 2018

In the past two years, Facebook has tested and rolled out several features with the stated aim to improve transparency around political advertising (especially in electoral periods). In October

[4] https://www.facebook.com/help/publisher/182222309230722
[5] https://www.facebook.com/facebook/videos/10156559267856729/

2017, Facebook announced that advertisers wishing to run election-related advertisements in the United States would have to register with Facebook and verify their identity, and that these ads would be accompanied by a clickable 'Paid for by...' disclosure (Goldman, 2017). In April 2018, the company announced that this verification process would be expanded to all 'issue ads', defined as those dealing with one of 20 topics, including abortion, civil rights, guns, immigration, and military (Goldman and Himel, 2018). These verification measures have currently been rolled out in the United States and Brazil: advertisers will have to provide a scanned passport or driver's licence.[6]

Beginning in November 2017, a feature called view ads was tested in Canada (Goldman, 2017), where users could navigate the dropdown menu on a page to see all of the ads that it was currently running. This was expanded to the United States and Brazil in the spring of 2018, and in June 2018 allowed users to view ads run by a page across platforms (Facebook, Instagram, and Messenger) (Leathern and Rodgers, 2018).

In May 2018, Facebook launched a public political ad archive of all political ads being run (Leathern, 2018b), and in August provided a political ad API that would allow researchers and journalists to more easily access this archive (Leathern, 2018a).

What Facebook Should Do

5. Provide meaningful News Feed controls for users

A fundamental issue with Facebook's current landing page, the News Feed, is its 'black box' user-facing model. Users who wish to get a broader, more diverse, and/or more trustworthy news diet have very few effective or intuitive controls on Facebook.

Existing options for control have largely fallen flat. The Chronological News Feed, while technically an option, is effectively unusable (every fresh reload of the page reverts the News Feed back to Facebook's preferred algorithm). Explore Feed, which separated Pages' content from content posted by friends and family, was never opt-in, and was poorly received by audiences, in part due to Facebook's lack of transparency around the feature during testing.

At the most basic level, users personalise their feeds by manually 'following' users or Pages (and thereby opting in to having those entities' content appear in the News Feed). Facebook users can later 'unfollow' anyone. Users can also choose certain friends or Pages to See First at the top of News Feed, as well as choose to Snooze certain keywords or people/Pages in News Feed (Fulay, 2018). But this level of control is still altogether inadequate.

Potential improvements would allow users to more easily understand the News Feed content to which they are exposed. Such a **News Feed Analytics feature** would provide a diagnostic for users: for example, the percentage of their News Feed content that comes from news outlets, and which ones; the percentage of their News Feed content that comes from advertisers; the breakdown of political perspectives in their News Feed (perhaps based on friends who have self-identified their political affiliation on Facebook); and how that compares with control groups of other users (e.g. national average). While many ordinary users would likely not engage with opt-in features and controls, providing a periodic visual check-up could raise awareness around these issues among a broader audience.

Going further, Facebook should allow users to have **meaningful control over their News Feed algorithm** and, therefore, the type of information they are exposed to in News Feed. This feature should clearly show what kind of information is currently being prioritised (who are the top

[6] https://www.facebook.com/business/help/167836590566506

friends, top Pages, what has been unfollowed) and provide buttons or sliders that allow users to control whether they wish to see more content that cross-cuts against their political ideology, whether they wish to see more news, and whether they wish their News Feed to be curated at all, or if it should proceed chronologically. Such features would likely be used by a fraction of all of Facebook's users, but could be powerful customisation tools for such power users.

In addition to adjusting the rankings of News Feed content, Facebook could implement the ability for users to **adopt a different point of view**, exposing them to a feed of entirely different content from the perspective they usually see; the *Wall Street Journal*'s 'Red-Feed/Blue-Feed' visualisation, which shows how left- and right-leaning users may be exposed to very different coverage on the same political issues, is a potential mock-up of such a feature (Keegan, 2016).

Facebook should conceive the provision and presentation of these controls/facilities as its own **direct contribution to improving digital and media literacy**. In doing so, it could draw on the insights of academic and NGO initiatives such as First Draft[7] and the work being done by Sam Wineburg and colleagues at Stanford (Steinmetz, 2018).

6. Expand context and fact-checking facilities

Facebook should continue to strive towards providing users with more contextual information about news stories. Research has shown that the effects of fact-checking can be complex, and some have even argued that it can be counter-productive (Nyhan and Reifler, 2015; Lazer et al., 2018). But an important piece of research has suggested that Facebook's recent fact-checking efforts have had success in reducing the amount of misinformation in the average user's feed (Allcott, Gentzkow, and Yu, 2018). We welcome the fact that at the end of 2018 **the little-i Context Button** was launched globally and believe that **significant resources should be dedicated to identifying the best, most authoritative, and trusted sources** of contextual information for each country, region, and culture.

Facebook should also continue to invest more widely in fact-checking infrastructure in countries and regions that do not have it. The current partnership under the International Fact Checking Network[8] only operates in 24 countries (Facebook operates in more than 100) and does not feature Sri Lanka, Myanmar, Libya, and other countries that have experienced recent unrest. Verificado 2018, a collaborative project organised by Mexican civil society and news organisations, and funded by Facebook's Journalism Project and Google's News Initiative ahead of the Mexican election, provides a potential model for this (Lichterman, 2018). Facebook should take responsibility for both **promoting these efforts and funding research into their effects**, especially outside of the Global North.

[7] https://firstdraftnews.org/
[8] https://www.facebook.com/help/publisher/182222309230722

3. Governance: From Transparency to Accountability

The Problem

Recent transparency initiatives are a major step forward and should be applauded. But there is still a fundamental lack of accountability to users, who have little say in important political decisions on Facebook, from what types of speech are permitted to what kind of political information is privileged.

There are broad concerns that Facebook continues to engage in deceptive behaviour when it comes to user privacy, and that it is biased against certain groups, but outsiders currently have almost no possibilities to verify these claims. Facebook remains very difficult to study, meaning that it is very difficult for policymakers to be able to formulate evidence-based policy and truly understand the scope of the relevant problems (from polarisation to disinformation, as it stands only Facebook can know the true scope of the issue).

Facebook has done much since the 2016 US election in its efforts to address some of the issues mentioned in this report. But there is little beyond goodwill, and its public relations and reputational interest in convincing the public that it is indeed taking these problems seriously, to hold it to sustained commitments to various initiatives, from combating state-run information campaigns to increasing advertising transparency. These efforts could be reversed, or have little long-term impact, if the company were (perhaps under pressure from its shareholders) to once again prioritise growth and profitability over other considerations.

What Facebook Has Done

Facebook has made fewer changes to its governance structures and practices than it has in other areas. Notable initiatives on transparency, accountability, and governance include the following.

Global Network Initiative – May 2013

In 2013, Facebook joined the Global Network Initiative (GNI), an organisation founded in 2008 in an effort to bring together technology companies and human rights groups to develop best practices for free expression and human rights online (Rusli, 2013). Companies that join the GNI voluntarily commit to implementing the GNI Principles,[9] which include a code of conduct dealing with government takedown requests, as well as a third-party assessment of how each company is adhering to the Principles conducted every two years.[10]

Collaboration with Academics – April 2018

Shortly before Mark Zuckerberg's congressional testimony, Facebook announced Social Science One, a new project with the Social Science Research Council (SSRC), spearheaded by Gary King and Nathaniel Persily (Schrage and Ginsberg, 2018). If successful, the initiative will provide academics with Facebook data for important research projects[11] focused on democracy and elections, with the additional benefit of funding this work through partnerships with philanthropic foundations. Under the suggested model, proposals will be reviewed by the SSRC, and Facebook will not have pre-publication veto power over the findings.

[9] https://globalnetworkinitiative.org/gni-principles/
[10] https://globalnetworkinitiative.org/accountability-policy-learning/
[11] https://socialscience.one/home

Civil and Human Rights Audits – May 2018

Facebook has voluntarily submitted to three rights audits: one, led by Laura Murphy, will oversee a civil rights review of Facebook's internal operations conducted by the DC law firm Relman, Dane, and Colfax (Fischer, 2018). Republican Senator Jon Kyl is leading a related investigation into the ways that Facebook may be biased against conservatives. A third, investigating Facebook's impact on human rights in Myanmar, was published in November 2018 (Warofka, 2018), finding that Facebook was not 'doing enough to help prevent [the] platform from being used to foment division and incite offline violence'.

Proposed External Oversight Mechanisms – November 2018

Zuckerberg, in an extensive November 2018 manifesto, wrote that Facebook would explore the creation of an external moderation council: 'We're planning to create a new way for people to appeal content decisions to an independent body, whose decisions would be transparent and binding.' Zuckerberg announced that the company would begin a consultation period to determine major unanswered questions with the project, and would begin beta-operations in various countries in 2019.

What Facebook Should Do

7. Establish regular auditing mechanisms

While it is laudable that the company is opening itself up to a civil rights audit of its products and practices, these initiatives remain largely focused on the US context and Global North concerns. Facebook's current privacy audits, set up following the 2011 and 2012 Federal Trade Commission rulings about Google and Facebook's deceptive privacy practices, have been shown in recent research to be based on simple self-reporting, making them 'so vague or duplicative as to be meaningless' (Gray, 2018: 4). By way of audits regarding international impact, the only public examples thus far are the Global Network Initiative Annual Reports (which include the results of the Human Rights Assessments for member companies like Facebook, but these are aggregated and provide limited insights) and the BSR audit of Facebook, WhatsApp, and Instagram's human rights impact in Myanmar. (We understand from Facebook that some other audits are being prepared.)

We believe that the model of **meaningful audits should be extended and replicated in other areas of public concern**. Global audits should involve trusted third parties that assess practices, products, and algorithms for undesirable outcomes, and identify possible improvements. They should feature academic and civil society participation, also drawing on region-specific organisations and expertise. Full access to internal data should be given, on a clear and explicit understanding about what would remain off the record. The downside of full transparency (e.g. potentially enabling bad actors to game the system) can be avoided by this model.

These audits, if acted upon, are sure to unearth other ways in which Facebook could become more responsible and accountable. For instance, the Myanmar audit, published in November 2018, makes a number of recommendations, including that the company formalise a human rights strategy with clear governance mechanisms, engage more formally on the human rights impacts of its content policy practices, and hire human rights specialists to help advise the company and engage with NGOs and academia (BSR, 2018). These are excellent suggestions, and similar efforts should be undertaken to unearth best practices around privacy, algorithmic fairness and bias, diversity, and many other areas.

8. Create an external content policy advisory group

Facebook should enlist civil society, academic experts, journalists, and other key stakeholders to **form an expert advisory group on content policy**. This body could provide ongoing feedback

on the standards and implementation of content policy, and review the record of appeals in selected areas (both thematic and geographical). Creating a body that has credibility with the extraordinarily wide geographical, cultural, and political range of Facebook users would be a major challenge, but a carefully chosen, formalised expert advisory group would be a first step. In principle, this model could be extended to other areas (e.g. News Feed) as well; bringing in external expertise and civil society engagement across Facebook's many politically salient features would be helpful.

Facebook has begun moving in this direction, bringing in experts to consult on specific policies as part of its new Community Standards Forum, a fortnightly meeting in which new moderation policies are presented and discussed. Additionally, a Data Transparency Advisory Group has been set up with Yale Law School faculty to provide external 'review [of] Facebook's measurement and transparency of content standards enforcement'.[12] These efforts should be expanded and formalised in a transparent manner. Facebook's academic partnership with the SSRC, Social Science One, provides a possible example of an organisational structure: it features substantive committees on issue areas, as well as regional committees led by academic experts.

If Facebook were to meaningfully consult with affected communities, and with academic experts in the key areas in which it forms policies, and interactively use this feedback to improve its Community Standards and general content policy practices, it would already mark a significant shift from the situation just two years ago, when these decisions and policies were crafted by a small group internally.

9. Establish a meaningful external appeals body

Since Facebook is establishing a kind of private jurisprudence, or 'platform law', for what more than two billion people can see and say, including much influential political information and political speech, there should be some form of independent, external control of both the criteria and the actual decisions made. Throughout our interactions with Facebook around this report, we have advocated for some form of external appeal procedure for content policy.

In November 2018, Zuckerberg announced the beginning of a consultation period to determine the creation of an external appellate body with power to decide top-level appeals. All the major questions remain unanswered: as Zuckerberg asks, 'How are members of the body selected? How do we ensure their independence from Facebook, but also their commitment to the principles they must uphold? How do people petition this body? How does the body pick which cases to hear from potentially millions of requests?' (Zuckerberg, 2018) Creating this appeals body will be a formidably complex task, immediately raising questions about political, geographical, and cultural representation.

Nevertheless, Zuckerberg's proposal marks a major shift in how Facebook sees its role around freedom of expression issues. If done properly, it could bring important external input and much-needed forms of due process to Facebook's private governance. However, as lawyers have begun to point out (Kadri, 2018), there are also ways in which this appellate body could fail meaningfully to devolve power from Facebook, providing granular appeals about specific pieces of content without impacting problematic policies more broadly. Facebook should strive to make this appeals body as transparent as possible (providing a publicly visible record of its decision-making) and allow it to influence broad areas of content policy (e.g. Facebook's treatment of certain populations in the Community Standards), not just rule on specific content takedowns. Civil society and digital rights advocacy groups should be key stakeholders within this process.

[12] https://law.yale.edu/yls-today/news/justice-collaboratory-lead-facebook-data-transparency-advisory-group

Conclusion

This report focuses on Facebook, which with its 2.2+ billion users is the most politically significant tech platform in the world. While we have decided to limit our scope to Facebook for analytical clarity, many of our suggestions are equally applicable to Facebook's subsidiary services WhatsApp and Instagram. WhatsApp poses unique questions around polarisation and misinformation due to its end-to-end-encrypted private and group messaging design, and Instagram poses subtly different moderation issues due to its image-based, hashtagged community organisation. However, both platforms face the same overarching governance and accountability issues as Facebook. Transparency and accountability for WhatsApp and Instagram remain rudimentary, with most public (and regulatory) attention in the past two years focused on Facebook. For example, Instagram has not implemented the same moderation overhauls as Facebook (appeals, detailed community guidelines, or content policy enforcement reports).

Many of the problems identified here are also found on other platforms, such as YouTube and Twitter. While existing accountability efforts like the Global Network Initiative provide an example of human rights groups and other civil society stakeholders joining with technology companies to increase accountability around government takedowns and surveillance, they were never designed to address the wider range of concerns facing platform companies today.

There is therefore growing interest in exploring some form of industry-wide self-regulatory body. Such a body could formulate a code of conduct for responsible, ethical behaviour. It would require meaningful civil society participation, clear governance structures, and transparent dispute resolution mechanisms. Models suggested include the press councils that create standards for media publishers (ARTICLE 19, 2018), industry-wide or company-specific ombuds who could serve as 'public editors' investigating user complaints (Kaye, 2018), and industry-specific self-governance bodies, such as the Financial Industry Regulatory Authority (FINRA), which oversees brokerage firms and the securities industry in the US.

Yet this is easier said than done. An obvious difficulty is defining the 'industry' in question, since platforms like Facebook and Google serve many different functions and span multiple services, from search to email to social networking. Smaller companies such as Reddit have expressed reservations about the idea. And what would be its geographical reach? Would there be one such body for each country, each wider region (e.g. the EU), or worldwide?

The goal of industry-wide self-regulation should be actively pursued, but attaining it, in ways that command confidence across diverse countries, cultures, and political tendencies, will be a long and complex task. In the meantime, the best should not be the enemy of the good. As we have indicated in this report, there is a great deal that a platform like Facebook can do right now to address widespread public concerns, and to do more to honour its public interest responsibilities as well as international human rights norms. Executive decisions made by Facebook have major political, social, and cultural consequences around the world. A single small change to the News Feed algorithm, or to content policy, can have an impact that is both faster and wider than that of any single piece of national (or even EU-wide) legislation.

Moreover, intrusive government regulation of political speech sets a precedent that authoritarian rulers in countries such as Russia are only too happy to follow, using their own arbitrary, self-interested definitions of 'fake news' and 'hate speech'. Alongside appropriate governmental regulation on matters such as competition, privacy, and dangerous speech, pluralistic structures of

self-regulation are in principle more desirable for political speech in a democracy. But the onus is now on platform companies such as Facebook to show that self-regulation can be effective and is not merely glorified PR. Beyond glasnost, we need perestroika.

One common thread running through our recommendations is that Facebook should offer not only more information but also more active control to its users. Ideally, the user interface and experience on Facebook should be designed to promote active, informed citizenship, and not merely clickbait addiction for the commercial benefit of Facebook, the corporation. That would be change indeed.

We recognise that Facebook employees are making difficult, complex contextual judgements every day, balancing competing interests, and not all those decisions will benefit from full transparency. But all would be better for more regular, active interchange with the worlds of academic research, investigative journalism, and civil society advocacy.

Together, we need to work towards a coherent mix of appropriate government regulation and industry-wide, platform-specific, and product-specific self-regulation, thus furnishing a credible democratic alternative to the Chinese model of authoritarian 'information sovereignty' that is gaining traction far beyond China's borders. Making Facebook a better forum for free speech and democracy is a significant part of a wider struggle to defend those values across the world.

References

Adler, Simon. 2018. 'Post No Evil.' 17 August. Radiolab. WNYC Studios. https://www.wnycstudios.org/story/post-no-evil (Accessed 31 December 2018).

Alba, Davey. 2018. 'How Duterte Used Facebook To Fuel the Philippine Drug War.' 4 September. BuzzFeed News. https://www.buzzfeednews.com/article/daveyalba/facebook-philippines-dutertes-drug-war (Accessed 31 December 2018).

Allcott, Hunt, and Matthew Gentzkow. 2017. 'Social Media and Fake News in the 2016 Election.' *Journal of Economic Perspectives* 31(2): 211–36.

Allcott, Hunt, Matthew Gentzkow, and Chuan Yu. 2018. 'Trends in the Diffusion of Misinformation on Social Media.' arXiv preprint arXiv:1809.05901.

ARTICLE 19. 2018. *Self-Regulation and 'Hate Speech' on Social Media Platforms.* London, UK. https://www.article19.org/resources/self-regulation-hate-speech-social-media-platforms/.

Bell, Emily J., Taylor Owen, Peter D. Brown, Codi Hauka, Nushin Rashidian. 2017. *The Platform Press: How Silicon Valley Reengineered Journalism.* New York: Tow Center for Digital Journalism, Columbia University. https://doi.org/10.7916/D8R216ZZ.

Bikert, Monika. 2018a. 'Publishing Our Internal Enforcement Guidelines and Expanding Our Appeals Process.' 24 April. Facebook Newsroom. https://newsroom.fb.com/news/2018/04/comprehensive-community-standards/ (Accessed 31 December 2018).

———. 2018b. 'Working to Keep Facebook Safe.' 17 July. Facebook Newsroom. https://newsroom.fb.com/news/2018/07/working-to-keep-facebook-safe/ (Accessed 31 December 2018).

boyd, danah. 2017. 'Did Media Literacy Backfire?' *Journal of Applied Youth Studies* 1(4): 83.

BSR, 2018. 'Human Rights Impact Assessment: Facebook in Myanmar.' https://fbnewsroomus.files.wordpress.com/2018/11/bsr-facebook-myanmar-hria_final.pdf (Accessed 8 January 2018).

Dubois, Elizabeth, and Grant Blank. 2018. 'The Echo Chamber Is Overstated: The Moderating Effect of Political Interest and Diverse Media.' *Information, Communication and Society* 21(5): 729–745.

Facebook. 2018. *Community Standards Enforcement Report.* https://transparency.facebook.com/community-standards-enforcement.

Fischer, Sara. 2018. 'Exclusive: Facebook Commits to Civil Rights Audit, Political Bias Review.' 2 May. Axios. https://www.axios.com/scoop-facebook-committing-to-internal-pobias-audit-1525187977-160aaa3a-3d10-4b28-a4bb-b81947bd03e4.html (Accessed 1 January 2019).

Fletcher, Richard, and Rasmus Kleis Nielsen. 2017. 'Are News Audiences Increasingly Fragmented? A Cross-National Comparative Analysis of Cross-Platform News Audience Fragmentation and Duplication.' *Journal of Communication* 67(4): 476–498.

Fulay, Amit. 2018. 'Keyword Snooze: People Turn Down the Noise.' 31 August. Facebook Newsroom. https://newsroom.fb.com/news/2018/08/inside-feed-keyword-snooze-people-turn-down-the-noise/ (Accessed 8 January 2019).

Funke, Daniel. 2018. 'Facebook Is Now Downranking Stories with False Headlines.' 24 October. Poynter. https://www.poynter.org/fact-checking/2018/facebook-is-now-downranking-stories-with-false-headlines/ (Accessed 8 January 2019).

Garton Ash, Timothy. 2016. *Free Speech: Ten Principles for a Connected World.* New Haven, CT: Yale University Press.

Gillespie, Tarleton. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media.* New Haven, CT: Yale University Press.

Goel, Vindu, Hari Kumar, and Sheera Frenkel. 2018. 'In Sri Lanka, Facebook Contends With Shutdown After Mob Violence.' *The New York Times.* https://www.nytimes.com/2018/03/08/technology/sri-lanka-facebook-shutdown.html (Accessed 31 December 2018).

Goldman, Rob. 2017. 'Update on Our Advertising Transparency and Authenticity Efforts.' 17 October. Facebook Newsroom. https://newsroom.fb.com/news/2017/10/update-on-our-advertising-transparency-and-authenticity-efforts/ (Accessed 1 January 2019).

Goldman, Rob, and Alex Himel. 2018. 'Making Ads and Pages More Transparent.' 6 April. Facebook Newsroom. https://newsroom.fb.com/news/2018/04/transparent-ads-and-pages/ (Accessed 1 January 2019).

Gray, Megan. 2018. *Understanding and Improving Privacy 'Audits' under FTC Orders.* Stanford CIS White Paper. https://cyberlaw.stanford.edu/blog/2018/04/understanding-improving-privacy-audits-under-ftc-orders.

Guess, Andy, Brendan Nyhan, and Jason Reifler. 2018. 'Selective Exposure to Disinformation: Evidence from the Consumption of Fake News During the 2016 US Presidential Campaign'. https://www.dartmouth.edu/~nyhan/fake-news-2016.pdf.

Hughes, Taylor, Jeff Smith, and Alex Leavitt. 2018. 'Helping People Better Assess the Stories They See in News Feed with the Context Button.' 3 April. Facebook Newsroom. https://newsroom.fb.com/news/2018/04/news-feed-fyi-more-context/ (Accessed 1 January 2019).

Jalonick, Mary Clare, and Barbara Ortutay. 2018. 'Zuckerberg: Regulation "inevitable" for Social Media Firms.' 12 April. Associated Press. https://apnews.com/04076e945181477cb9e08a5383528b15 (Accessed 8 January 2019).

Kadri, Thomas. 2018. 'Will Facebook Actually Give Any Power to an Independent Speech Court?' 19 November. Slate Magazine. https://slate.com/technology/2018/11/facebook-zuckerberg-independent-speech-content-appeals-court.html (Accessed 8 January 2019).

Kaye, David. 2018. *A Human Rights Approach to Platform Content Regulation.* Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression. https://freedex.org/a-human-rights-approach-to-platform-content-regulation/ (Accessed 23 June 2018).

Keegan, Jon. 2016. 'Blue Feed, Red Feed.' 18 May. *Wall Street Journal.* https://perma.cc/U94G-YZGR (Accessed 1 January 2019).

Khan, Lina M. 2016. 'Amazon's Antitrust Paradox.' *Yale Law Journal* 126: 710.

Kim, Young Mie et al. 2018. 'The Stealth Media? Groups and Targets behind Divisive Issue Campaigns on Facebook.' *Political Communication* 35(4): 515–541.

King, Gary, and Nathaniel Persily. 2018. 'A New Model for Industry-Academic Partnerships.' SSRC Working Paper. https://gking.harvard.edu/partnerships.

Klonick, Kate. 2017. 'The New Governors: The People, Rules, and Processes Governing Online Speech.' *Harvard Law Review* 131: 1598.

Koebler, Jason, and Joseph Cox. 2018. 'The Impossible Job: Inside Facebook's Struggle to Moderate Two Billion People.' 23 August. Motherboard. https://motherboard.vice.com/en_us/article/xwk9zd/how-facebook-content-moderation-works (Accessed 31 December 2018).

Lazer, David M. J., Matthew A. Baum, Yochai Benkler, et al. 2018. 'The Science of Fake News.' *Science* 359(6380): 1094–1096.

Leathern, Rob. 2018a. 'Introducing the Ad Archive API.' 22 August. Facebook Newsroom. https://newsroom.fb.com/news/2018/08/introducing-the-ad-archive-api/ (Accessed 1 January 2019).

———. 2018b. 'Shining a Light on Ads With Political Content.' 24 May. Facebook Newsroom. https://newsroom.fb.com/news/2018/05/ads-with-political-content/ (Accessed 1 January 2019).

Leathern, Rob, and Emma Rodgers. 2018. 'A New Level of Transparency for Ads and Pages.' 28 June. Facebook Newsroom. https://newsroom.fb.com/news/2018/06/transparency-for-ads-and-pages/ (Accessed 1 January 2019).

Lichterman, Joseph. 2018. 'How 90 Outlets Are Working Together to Fight Misinformation Ahead of Mexico's Elections.' 13 June. Global Investigative Journalism Network. https://gijn.org/2018/06/13/how-90-outlets-are-working-together-to-fight-misinformation-ahead-of-mexicos-elections/ (Accessed 1 January 2019).

Lyons, Tessa. 2018a. 'Hard Questions: How Is Facebook's Fact-Checking Program Working?' 14 June. Facebook Newsroom. https://newsroom.fb.com/news/2018/06/hard-questions-fact-checking/ (Accessed 1 January 2019).

Lyons, Tessa. 2018b. 'Hard Questions: Who Reviews Objectionable Content on Facebook – And Is the Company Doing Enough to Support Them?' 26 July. Facebook Newsroom. https://newsroom.fb.com/news/2018/07/hard-questions-content-reviewers/ (Accessed 1 January 2019).

Manjoo, Farhad, and Kevin Roose. 2017. 'How to Fix Facebook? We Asked 9 Experts.' 31 October. *The New York Times*. https://www.nytimes.com/2017/10/31/technology/how-to-fix-facebook-we-asked-9-experts.html (Accessed 31 December 2018).

Marra, Greg. 2014. 'More Ways to Control What You See in Your News Feed.' 7 November. Facebook Newsroom. https://newsroom.fb.com/news/2014/11/news-feed-fyi-more-ways-to-control-what-you-see-in-your-news-feed/ (Accessed 1 January 2019).

Marwick, Alice E. 2018. 'Why Do People Share Fake News? A Sociotechnical Model of Media Effects.' *Georgetown Law Technology Review* 2(2): 474–512.

McGrew, Sarah, Teresa Ortega, Joel Breakstone, and Sam Wineburg. 2017. 'The Challenge That's Bigger than Fake News: Civic Reasoning in a Social Media Environment.' *American Educator* 41(3): 4.

Mosseri, Adam. 2018a. 'Bringing People Closer Together.' 11 January. Facebook Newsroom. https://newsroom.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together/ (Accessed 8 January 2019).

Mosseri, Adam. 2018b. 'Ending the Explore Feed Test.' 1 March. Facebook Newsroom. https://newsroom.fb.com/news/2018/03/news-feed-fyi-ending-the-explore-feed-test/ (Accessed 1 January 2019).

Myers West, Sarah. 2018. 'Censored, Suspended, Shadowbanned: User Interpretations of Content Moderation on Social Media Platforms.' *New Media & Society*: 1461444818773059.

Newman, Nic, Richard Fletcher, Antonis Kalogeropoulos, David A. L. Levy, and Rasmus Kleis Nielsen. 2018. *Reuters Institute Digital News Report 2018*. Oxford, UK: Reuters Institute for the Study of Journalism. http://www.digitalnewsreport.org/.

Nielsen, Rasmus Kleis, and Sarah Anne Ganter. 2017. 'Dealing with Digital Intermediaries: A Case Study of the Relations between Publishers and Platforms.' *New Media & Society* 20(4) 1600–1617. http://journals.sagepub.com/eprint/dxNzFHygAIRHviKP9MFg/full.

Nyhan, Brendan, and Jason Reifler. 2015. 'The Effect of Fact-Checking on Elites: A Field Experiment on US State Legislators.' *American Journal of Political Science* 59(3): 628–640.

Osnos, Evan. 2018. 'Can Mark Zuckerberg Fix Facebook Before It Breaks Democracy?' 17 September. *The New Yorker*. https://www.newyorker.com/magazine/2018/09/17/can-mark-zuckerberg-fix-facebook-before-it-breaks-democracy (Accessed 31 December 2018).

Rhee, Ed. 2011. 'How to Sort Your Facebook Newsfeed in Chronological Order.' 14 November. CNET. https://www.cnet.com/how-to/how-to-sort-your-facebook-newsfeed-in-chronological-order/ (Accessed 1 January 2019).

Roberts, Sarah T. 2017. 'Social Media's Silent Filter.' 8 March. *The Atlantic*. https://www.theatlantic.com/technology/archive/2017/03/commercial-content-moderation/518796/ (Accessed 31 December 2018).

———. 2018. 'Digital Detritus: 'Error' and the Logic of Opacity in Social Media Content Moderation.' *First Monday* 23(3). http://firstmonday.org/ojs/index.php/fm/article/view/8283 (Accessed 2 March 2018).

Roose, Kevin. 2018. 'Facebook Banned Infowars. Now What?' 10 August. *The New York Times*. https://www.nytimes.com/2018/08/10/technology/facebook-banned-infowars-now-what.html (Accessed 31 December 2018).

Rosenberg, Eli. 2018. 'Facebook Censored a Post for "Hate Speech." It Was the Declaration of Independence.' 5 July. *Washington Post*. https://www.washingtonpost.com/news/the-intersect/wp/2018/07/05/facebook-censored-a-post-for-hate-speech-it-was-the-declaration-of-independence/ (Accessed 31 December 2018).

Ruggie, John Gerard. 2013. *Just Business: Multinational Corporations and Human Rights*. New York: W. W. Norton & Company.

Rusli, Evelyn M. 2013. 'Facebook Joins GNI Online Privacy-and-Freedom Group.' 22 May. *Wall Street Journal*. https://blogs.wsj.com/digits/2013/05/22/facebook-joins-gni-online-privacy-and-freedom-group/ (Accessed 1 January 2019).

Santa Clara Principles on Transparency and Accountability in Content Moderation. 2018. 'Open Letter to Mark Zuckerberg.' https://santaclaraprinciples.org/open-letter/ (Accessed 8 January 2019).

Schrage, Eliot, and David Ginsberg. 2018. 'Facebook Launches New Initiative to Help Scholars Assess Social Media's Impact on Elections.' 9 April. Facebook Newsroom. https://newsroom.fb.com/news/2018/04/new-elections-initiative/ (Accessed 1 January 2019).

Silver, Ellen. 2018. 'Hard Questions: Who Reviews Objectionable Content on Facebook – And Is the Company Doing Enough to Support Them?' 26 July. Facebook Newsroom. https://newsroom.fb.com/news/2018/07/hard-questions-content-reviewers/ (Accessed 31 December 2018).

Singh, Spandana. 2018. 'Pressing Facebook for More Transparency and Accountability Around Content Moderation.' 16 November. New America. Open Technology Institute. https://www.newamerica.org/oti/blog/pressing-facebook-more-transparency-and-accountability-around-content-moderation/ (Accessed 31 December 2018).

Srnicek, Nick. 2016. *Platform Capitalism*. Cambridge, UK: Polity Press.

Stecklow, Steve. 2018. 'Special Report: Why Facebook Is Losing the War on Hate Speech In Myanmar.' 15 August. Reuters. https://www.reuters.com/article/us-myanmar-facebook-hate-specialreport-idUSKBN1L01JY (Accessed 31 December 2018).

Steinmetz, Katy. 2018. 'How Your Brain Tricks You Into Believing Fake News.' 9 August. *Time*. http://time.com/5362183/the-real-fake-news-crisis/ (Accessed 1 January 2019).

Swisher, Kara. 2018. 'Facebook CEO Mark Zuckerberg on Recode Decode.' 18 July. Recode. https://www.recode.net/2018/7/18/17575158/mark-zuckerberg-facebook-interview-full-transcript-kara-swisher (Accessed 16 August 2018).

Theil, Stefan. 2018. 'The German NetzDG: A Risk Worth Taking?' 8 February. Verfassungsblog. https://verfassungsblog.de/the-german-netzdg-a-risk-worth-taking/ (Accessed 8 January 2019).

Tow Center for Digital Journalism. 2018. *Platforms and Publishers: A Definitive Timeline*. Columbia University. http://tow.cjr.org/platform-timeline/ (Accessed December 31, 2018).

Tucker, Joshua A, Yannis Theocharis, Margaret E. Roberts, and Pablo Barberá. 2017. 'From Liberation to Turmoil: Social Media and Democracy.' *Journal of Democracy* 28(4): 46–59.

Vaidhyanathan, Siva. 2018. *Antisocial Media: How Facebook Disconnects Us and Undermines Democracy*. Oxford, UK: Oxford University Press.

Van Dijck, José, Thomas Poell, and Martijn de Waal. 2018. *The Platform Society: Public Values in a Connective World.* New York, NY: Oxford University Press.

Warofka, Alex. 2018. 'An Independent Assessment of the Human Rights Impact of Facebook in Myanmar.' 5 November. Facebook Newsroom. https://newsroom.fb.com/news/2018/11/myanmar-hria/ (Accessed 1 January 2019).

Wong, Julia Carrie. 2016. 'Mark Zuckerberg Accused of Abusing Power after Facebook Deletes 'Napalm Girl' Post.' 9 September. The Guardian. https://www.theguardian.com/technology/2016/sep/08/facebook-mark-zuckerberg-napalm-girl-photo-vietnam-war (Accessed 8 January 2019).

York, Jillian C. 2018. 'Facebook Releases First-Ever Community Standards Enforcement Report.' 16 May. Electronic Frontier Foundation. https://www.eff.org/deeplinks/2018/05/facebook-releases-first-ever-community-standards-enforcement-report (Accessed 31 December 2018).

Zuckerberg, Mark. 2017. 'Building Global Community.' 16 February. https://www.facebook.com/notes/mark-zuckerberg/building-global-community/10103508221158471/?pnref=story.

———. 2018. 'A Blueprint for Content Governance and Enforcement.' 15 November. https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/.

Selected RISJ Publications

BOOKS

NGOs as Newsmakers: The Changing Landscape of International News
Matthew Powers (published with Columbia University Press)

Global Teamwork: The Rise of Collaboration in Investigative Journalism
Richard Sambrook (ed)

Something Old, Something New: Digital Media and the Coverage of Climate Change
James Painter et al

Journalism in an Age of Terror
John Lloyd (published with I.B.Tauris)

The Right to Be Forgotten: Privacy and the Media in the Digital Age
George Brock (published with I.B.Tauris)

The Kidnapping of Journalists: Reporting from High-Risk Conflict Zones
Robert G. Picard and Hannah Storm (published with I.B.Tauris)

Innovators in Digital News
Lucy Kueng (published with I.B.Tauris)

Local Journalism: The Decline of Newspapers and the Rise of Digital Media
Rasmus Kleis Nielsen (ed) (published with I.B.Tauris)

Journalism and PR: News Media and Public Relations in the Digital Age
John Lloyd and Laura Toogood (published with I.B.Tauris)

Reporting the EU: News, Media and the European Institutions
John Lloyd and Cristina Marconi (published with I.B.Tauris)

SELECTED RISJ REPORTS AND FACTSHEETS

More Important, But Less Robust? Five Things Everybody Needs to Know about the Future of Journalism
Rasmus Kleis Nielsen and Meera Selva

Journalism, Media, and Technology Trends and Predictions 2019
Nic Newman

Time to Step Away From the 'Bright, Shiny Things'? Towards A Sustainable Model of Journalism Innovation in an Era of Perpetual Change
Julie Posetti

An Industry-Led Debate: How UK Media Cover Artificial Intelligence
J. Scott Brennen, Philip N. Howard, and Rasmus Kleis Nielsen (Factsheet)

News You Don't Believe: Audience Perspectives on 'Fake News'
Rasmus Kleis Nielsen and Lucas Graves (Factsheet)

ISBN 978-1-907384-56-1