

An illustration of a diverse crowd of people, mostly men, looking down at their mobile devices. The style is a flat, colorful illustration with a warm color palette of blues, yellows, and browns. The people are wearing various clothing, including jackets, shirts, and suits. Some are wearing glasses. The overall scene suggests a busy, technology-driven environment.

**CEPR PRESS**

Edited by Filipe R. Campante, Ruben Durante  
and Andrea Tesei

# The Political Economy of Social Media

**CENTRE FOR  
ECONOMIC  
POLICY  
RESEARCH**

# The Political Economy of Social Media

CEPR

WITH SUPPORT OF CEPR PARIS FOUNDING PARTNERS



Research Fund



Direction générale  
du Trésor



SciencesPo

**CEPR PRESS**

Centre for Economic Policy Research  
187 boulevard Saint-Germain  
75007, Paris, France

33 Great Sutton Street  
London, EC1V 0DX, UK

Tel: +44 (0)20 7183 8801

Email: [cepr@cepr.org](mailto:cepr@cepr.org)

Web: [www.cepr.org](http://www.cepr.org)

ISBN: 978-1-912179-80-0

Copyright © CEPR Press

# The Political Economy of Social Media

Edited by Filipe R. Campante,  
Ruben Durante and Andrea Tesei

**CEPR** PRESS

## **CENTRE FOR ECONOMIC POLICY RESEARCH (CEPR)**

The Centre for Economic Policy Research (CEPR) is a network of over 1,700 research economists based mostly in European universities. The Centre's goal is twofold: to promote world-class research, and to get the policy-relevant results into the hands of key decision-makers. CEPR's guiding principle is 'Research excellence with policy relevance'. It was founded in the UK in 1983, where it is a Charity, and in November 2019 CEPR initiated the creation of an Association under French law, in order to provide a vehicle for an expansion in France. The members of the Conseil d'Administration of the Association are identical to the UK Board of Trustees.

CEPR is independent of all public and private interest groups. It takes no institutional stand on economic policy matters and its core funding comes from its Institutional Members, projects that it runs and sales of publications. Because it draws on such a large network of researchers, its output reflects a broad spectrum of individual viewpoints as well as perspectives drawn from civil society. CEPR research may include views on policy, but the Trustees/members of the Conseil d'Administration of the Association do not give prior review to its publications. The opinions expressed in this report are those of the authors and not those of CEPR.

Chair of the Board	Sir Charlie Bean
Founder and Honorary President	Richard Portes
President	Beatrice Weder di Mauro
Vice Presidents	Maristella Botticini
	Philippe Martin
	Ugo Panizza
	Mar Reguant
	Hélène Rey
Chief Executive Officer	Tessa Ogden

# Contents

<i>Foreword</i>	<i>vii</i>
<b>Introduction</b>	<b>9</b>
Filipe R. Campante, Ruben Durante and Andrea Tesei	
<b>Section 1 Welfare effects of social media</b>	
<b>1 The welfare effects of social media</b>	<b>25</b>
Hunt Allcott, Luca Braghieri, Sarah Eichmeyer and Matthew Gentzkow	
<b>2 Social media and mental health</b>	<b>31</b>
Luca Braghieri, Ro'ee Levy and Alexey Makarin	
<b>Section 2 Social media and voters</b>	
<b>3 Social media, news consumption and polarisation</b>	<b>47</b>
Ro'ee Levy	
<b>4 Homophily, group size and the diffusion of political information in social networks</b>	<b>59</b>
Yosh Halberstam and Brian Knight	
<b>5 Political implications of the rise of mobile broadband internet</b>	<b>67</b>
Sergei Guriev, Nikita Melnikov and Ekaterina Zhuravskaya	
<b>6 Mobile internet and the rise of communitarian politics</b>	<b>77</b>
Marco Manacorda, Guido Tabellini and Andrea Tesei	
<b>7 The effect of social media on elections: Evidence from the United States</b>	<b>89</b>
Thomas Fujiwara, Karsten Müller and Carlo Schwarz	
<b>Section 3 Social media and politicians</b>	
<b>8 Politics 2.0: The multifaceted effect of broadband internet on political participation</b>	<b>101</b>
Filipe Campante, Ruben Durante and Francesco Sobbrío	
<b>9 New technologies and political competition: The impact of social media communication on political contributions</b>	<b>111</b>
Maria Petrova, Ananya Sen and Pinar Yildirim	

## **Section 4 Social media and mobilisation**

- 10 Social media and protest participation: Evidence from Russia** 121  
Ruben Enikolopov, Alexey Makarin and Maria Petrova
- 11 Social media and mobilisation** 131  
Leopoldo Fergusson and Carlos Molina
- 12 Liberation technology: Mobile phones and political mobilisation in Africa** 143  
Marco Manacorda and Andrea Tesei

## **Section 5 Social media and hatred**

- 13 Social media and Xenophobia: Evidence from Russia** 153  
Leonardo Bursztyn, Georgy Egorov, Ruben Enikolopov and Maria Petrova
- 14 Can social media spur offline hatred?** 163  
Karsten Müller and Carlo Schwarz

## **Section 6 Social media in autocracies**

- 15 The political economy of social media in China** 175  
Bei Qin, David Strömberg and Yanhui Wu
- 16 Social media in autocracies** 185  
David Y. Yang

## **Section 7 Social media and legacy media**

- 17 Social media and legacy media** 195  
Sophie Hatte and Ekaterina Zhuravskaya
- 18 Contagion from social media to mainstream media** 203  
Julia Cagé, Nicolas Hervé and Béatrice Mazoyer

# Foreword

The rise of social media has profoundly transformed society, reshaping communication and information consumption. Its widespread use has placed unprecedented pressure on norms and institutions, challenging them to adapt to a rapidly changing social and political landscape. Its legacy, meanwhile, is already being fiercely contested by academics.

This eBook brings together a diverse array of contributions focusing on the political economics of social media, providing a comprehensive exploration of the impact of the internet and social media on the global political landscape. The chapters contribute valuable insights into the welfare effects of social media, highlighting its addictive nature and negative ramifications on mental health. The research also explores the influence of social media on voter behaviour, including its role in exacerbating recent polarisation trends through the creation of like-minded 'echo chambers' and the selective dissemination of political information, which can both enhance transparency and contribute to political discontent.

Other chapters explore the dynamic relationship between politicians and social media, revealing how these platforms enable politicians to effectively raise resources and mobilise voters, at a fraction of the cost of traditional methods. The authors uncover the influential role of social media in shaping elections, but not always in a consistent direction. The research also addresses the complex interplay between social media and protests, tracing its emergence as a highly effective tool for coordination and collective action. Finally, the authors explore the unique impact of social media in autocratic regimes and investigates their impact on the global legacy media.

Overall, this eBook provides compelling insights into social media's nuanced and complicated impact on the political and social landscape. Rather than providing judgements on whether social media is 'good' or 'bad', the chapters underscore the variability of consequences based on diverse outcomes and contexts. These findings will help inform policymakers on the future direction of social media regulation and moderation, recognising its enduring significance in society, particularly in the realm of politics.

CEPR is grateful to Filipe R. Campante, Ruban Durante, and Andrea Tesei for their expert editorship of the eBook. Our thanks also go to Anil Shamdasani for his skilled handling of its production.

CEPR, which takes no institutional positions on economic policy matters, is delighted to provide a platform for an exchange of views on this important topic.

Tessa Ogden  
Chief Executive Officer, CEPR  
November 2023





# Introduction

**Filipe R. Campante,<sup>a</sup> Ruben Durante<sup>bce</sup> and Andrea Tesei<sup>de</sup>**

<sup>a</sup>Johns Hopkins University; <sup>b</sup>National University of Singapore; <sup>c</sup>ICREA-Universitat Pompeu Fabra; <sup>d</sup>Queen Mary University of London; <sup>e</sup>CEPR

9

The advent of the internet and social media represents one of the most important social transformations of our time. Their ubiquitous presence in our daily lives has reshaped the way humans communicate, interact, and coordinate with each other. This has far-reaching consequences for norms and institutions, which are subject to unprecedented pressures to cope with a rapidly changing social and political environment. A well-established body of evidence has shown that traditional media, such as radio and TV, have had a substantial impact on political behaviour and outcomes.<sup>1</sup> Due to its two-way nature and low barriers to entry into the production and dissemination of content, social media has the potential to stir politics even further.

Looking at the public debate on the impact of technology, there seems to be a consensus that the political implications of social media are huge. And yet, there is substantial disagreement as to the nature of those implications. Some have argued that platforms such as Facebook helped make a massive blow for democracy and citizen participation, and against the power of autocrats (Ghonim 2012). More recently, however, the view has turned decidedly more negative, with some laying the blame on social media for a host of negative political developments, including the rise of fake news and hate speech (Haidt 2022). This is a call for empirical research, if there has ever been one.

Fortunately, social scientists have answered that call. At this point, in 2023, we can confidently state that a body of literature has emerged that provides credible evidence on many aspects of the impact of social media on politics. While we are far from having all the answers we would like, we now know way more than we did many years ago.

This volume is an attempt to capture a snapshot of that effort. It is by necessity a partial one, given how voluminous the literature has become. It focuses on the political economics of social media, meaning that it compiles contributions from the field of economics. Even within that field, it is meant to be more illustrative than comprehensive – there is simply too much quality research for any effort of this kind to aspire to be encyclopaedic. Yet it aims to paint a coherent picture of what the internet and social media have wrought to our political landscape and beyond.

<sup>1</sup> For instance, see Stromberg (2004), Yanagizawa-Drott (2014) and Adena et al (2015) on the impact of the radio, Gentzkow (2006), Prior (2007), Campante and Hojman (2013), and Durante et al. (2019) on broadcast TV, Della et al. (2007) and Martin and Yurukoglu (2017) on cable TV, among others. For a survey, see Prat and Stromberg (2015).

That picture shows that social media has had a unique kind of impact, because of specific features that set it apart from previous media technologies. First, while some of these technologies were portable (e.g. radio) and arguably had addictive properties (e.g. television), social media can be carried around and exert its pull on consumers literally everywhere and at all times, via mobile phones. This ubiquity entails a broader scope for impact on everyday life and human welfare.

Second, social media is prone to forming homophilic networks through which like-minded, pro-attitudinal content is more likely to spread. This gives rise to a specific type of political impact, with social media being linked to increased polarisation and the diffusion of political content that capitalises on distrust of others, strengthening in-group biases and animosity toward outsiders, as in the case of the recent rise of populism.

Third, and closely related to the previous feature, social media is characterised by uniquely low barriers to entry into the production and dissemination of content – political or otherwise. While previous technologies such as radio or TV were largely one-way avenues, with a (relatively) small number of outlets broadcasting content for a mass audience, social media allows everyone to be a content provider. This makes a huge difference to the possibilities they offer to citizens – in terms of organising and coordinating for collective action – but also to political entrepreneurs, who can use social media to spread their own messages, for their own strategic purposes.

## THE WELFARE EFFECTS OF SOCIAL MEDIA

In the broadest sense, and as with every technology, we care about the impact of social media on wellbeing. That is where this volume starts, in Part 1. As challenging as measuring wellbeing can be, the literature has come up with ingenious approaches. In Chapter 1, Hunt Allcott, Luca Braghieri, Sarah Eichmeyer and Matthew Gentzkow implement a randomised evaluation incentivising Facebook users to drop off the platform for one month. The outcome is greater offline interaction and, perhaps relatedly, a small but significant increase in self-reported subjective wellbeing. In fact, the treatment also increased the likelihood of experimental subjects being interested in tools to limit social media usage after the experiment. This highlights the idea that social media has addictive features (Allcott et al. 2022), suggesting the limitations of ‘revealed preference’-type arguments whereby the widespread use of social media implies that it has ipso facto net positive effects on users.

The evidence of negative effects of social media on individual wellbeing is reinforced by the observational evidence in Chapter 2, in which Luca Braghieri, Ro’ee Levy and Alexey Makarin exploit the variation induced by the initial spread of Facebook through different college campuses, using a differences-in-differences approach. They show that the introduction of Facebook at a college had a negative impact on students’ self-reported mental health, and especially for individuals already susceptible to mental illness. This underscores that the effects of social media may differ across different groups.

In sum, social media can have negative effects on subjective wellbeing, and the fact that people use it cannot be taken by itself as evidence that it increases their welfare. This suggests that there is scope for considering regulation of social media usage and for encouraging content moderation in online platforms.

This conclusion, based on the *individual* consequences of social media usage, is further strengthened when their *aggregate* implications are considered. In particular, as with other media technologies, social media could have important implications in the political arena, which obviously distinguishes the media industry from other markets. That is where we turn to next.

## SOCIAL MEDIA AND VOTERS

We start in the specific realm of individual voter behaviour, in Part 2. In fact, the same work that identifies the negative effects of social media on individual wellbeing already points to possible trade-offs, from a societal perspective, when it comes to political engagement. After all, Chapter 1 indicates that individuals who dropped off Facebook end up less informed about politics, though the experimental variation does not seem enough to affect turnout. It is enough, however, to reduce political polarisation, suggesting that social media may have indeed played a part in increasing polarisation in recent years, as many have accused it of doing (Sunstein 2017, Haidt 2022).

The polarising effect of social media is explored further in Chapter 3. In it, Ro'ee Levy presents another experimental intervention, manipulating not presence on Facebook but rather the content users are exposed to. The key result is that exposure to social media content indeed increases affective polarisation, that is, the extent to which partisans dislike the other party. Importantly, the experimental intervention randomly assigns pro- versus counter-attitudinal content, and in doing so – and by following social media usage behaviour by the experimental subjects – it is able to adjudicate between different possible mechanisms. In particular, Levy shows that the effect on polarisation works through the fact that social media exposes users to disproportionately pro-attitudinal content, and is able to attribute a substantial part of that to the algorithms they employ. This substantiates the common intuition that the business models of existing social media platforms may have consequences in terms of the political environment.

The interplay between social media and polarisation is also the theme of Chapter 4, in which Yosh Halberstam and Brian Knight turn to another major social media platform that has been widely accused of fostering polarisation – Twitter (now rebranded as X). The authors start with a model that predicts that social media users will tend to be disproportionately exposed to like-minded content, through endogenous homophily in their networks. They then look at the Twitter presence of candidates from the two US major political parties, and the network formed by their followers, which allows them to proxy for the ideology of a subset of Twitter users based on the partisan affiliation of the politicians they follow. They find strong homophily: liberals (resp. conservatives)

are disproportionately likely to have liberal (resp. conservative) followers themselves, and much more so than with other types of social interactions. This in turn translates into disproportionate consumption of pro-attitudinal information, especially in the political realm.

The evidence that social media can indeed create ‘echo chambers’ that amplify polarisation feeds naturally into the possibility that the content flowing through those chambers could have implications for political outcomes. This is the theme explored in Chapter 5, in which Sergei Guriev, Nikita Melnikov and Ekaterina Zhuravskaya take a global look at the impact of the spread of social media. They do so by taking advantage of the fact that, at the global scale, the spread was intimately connected to the expansion of 3G mobile networks – the first technology that made mobile internet access a reality. Using a differences-in-differences approach, their key finding is that increased mobile internet coverage led to reduced levels of trust in government. This seems to be driven by countries where traditional media were not free, and where the arrival of social media thereby facilitates access to information on government misconduct. While this is arguably a welcome development of increased transparency, the authors also find that things are different in the European context: there, it is populist politicians who seem to have benefited from the new technology.

This is precisely the theme developed further by Marco Manacorda, Guido Tabellini and Andrea Tesei in Chapter 6. Their key finding is that, in Europe, the expansion of 3G and 4G mobile technology was associated with increased support for ‘communitarian’ parties, that is, those advocating for the interests of insiders (e.g. native-born populations, majoritarian groups) against outsiders (e.g. immigrants, minorities). They exploit granular geographical variation in the expansion of the mobile networks, showing that their arrival was followed by increased voting for communitarian parties. To establish a causal impact, they use the variation in network access induced by proximity to the birthplace of a manager in charge of decisions by telecommunications firms. They argue that the effect is being driven by online content favouring messages that strengthen bias towards in-groups and against out-groups, and provide survey evidence showing voters’ opinions shifting in that direction. This benefited communitarian parties in general, and the subgroup classified as populist as part of that broader pattern.

While the conclusion that social media works to the electoral advantage of populist politicians seems tempting, a word of caution comes from Chapter 7. In it, Thomas Fujiwara, Karsten Müller and Carlo Schwarz look at the 2016 US election, where many observers partly attributed the unexpected victory of Donald Trump to his social media presence, especially on Twitter. They exploit the geographical variation in early Twitter adoption generated by attendance at the 2007 South by Southwest (SXSW) festival, which had a strong impact on the platform’s spread. They find that exposure to Twitter induced by that variation was negatively associated with Trump’s vote share in 2016 and

2020 (but not with the Republican congressional vote share). This underscores the power of social media to affect elections, but not always in a consistent direction – instead, it can be contingent on specific candidate characteristics.

In sum, the evidence strongly points in the direction of social media having significant effects on the behaviour of voters. This starts from polarisation, reinforced by the tendency of social media to create homophilic ‘echo chambers’ disseminating pro-attitudinal content. Social media also spreads political information to voters in general, which can increase transparency, but the selective nature of the content that is spread potentially opens the way for political discontent.<sup>2</sup>

## SOCIAL MEDIA AND POLITICIANS

Having gone over the impact of social media on voters, Part 3 then turns to what it means for the behaviour of the supply side of the political ‘market’, namely, politicians. While there is less written on that topic, some evidence has emerged that the ability to strategically use these technologies, given the low barriers to entry, may well advantage politicians, possibly at the expense of voters.

Chapter 8, by Filipe Campante, Ruben Durante and Francesco Sobbrío, illustrates how the introduction of social media changes the possibilities facing political entrepreneurs, in ways that can undo the initial impact of the pre-social media internet. They exploit variation in the expansion of broadband technology in Italy and show that this expansion initially translated into reduced voter turnout, as supporters of more extreme parties became less likely to vote. However, the arrival of a new political movement, created around the early social media platform Meetup.com, helped revert that trend: the Five Star Movement expanded more rapidly in places with greater broadband access, and attracted more voters in those places when it eventually started competing in elections. In short, the ability to mobilise voters online, at very low cost, offers a possibility for new political actors to enter the fray and reverse the initial demobilising effect of the internet.

But what do incumbent politicians do, as they can also take advantage of the new technologies?

In Chapter 9, Maria Petrova, Ananya Sen and Pinar Yildirim study the role of social media in helping politicians raise resources. They analyse how contributions to US congressional campaigns change in response to the candidates’ opening Twitter accounts, exploiting variation in Twitter penetration across different states. They show that, in the month where a politician opens their account, there is a positive jump in the individual contributions they receive, but only in states with high Twitter penetration.

<sup>2</sup> On the importance of the content of the information spread by social media, a particularly prominent instance is that of disinformation, as illustrated by the phenomenon of ‘fake news’. Allcott and Gentzkow (2017) and Lazer et al (2018) discuss the reach of this particular problem.

Their evidence suggests that this happens mostly because of the platform's impact in helping politicians become better known by a previously untapped set of individuals, as opposed to reminding prior contributors to give more.

In sum, the early evidence from this literature suggests that social media empowers politicians, enabling them to mobilise their supporters in more effective ways, which are available at a very low cost – and hence to all politicians rather than just a few. This offers new strategic possibilities, which can empower new actors. Much remains to be learned on what these new actors bring to the picture.

## **SOCIAL MEDIA AND PROTEST**

The literature has also addressed the impact of social media beyond electoral politics. The role of social media as an effective vehicle for coordination and collective action (Campante et al. 2022) naturally makes it a potential facilitator of costly political activities, perhaps best exemplified by political protest. This is the topic of Part 4.

In Chapter 10, Ruben Enikolopov, Alexey Makarin and Maria Petrova take advantage of the idiosyncratic role played by one specific university – St Petersburg State University (SPSU) – in the creation and expansion of VKontakte (VK), a prominent Russian social media platform. They show that cities with a large number of SPSU students in the cohort to which VK's founder belonged, relative to the numbers in other cohorts, had stronger VK penetration. They were also more likely to have demonstrations (and larger ones at that) against the government following the 2011 elections, which were widely perceived as manipulated. In contrast, the effect in terms of electoral results (in less disputed elections) seemed to be in favour of the government's party. This indicates that social media played an important role both through the content shared on it and by facilitating coordination for collective action.

In Chapter 11, Leopoldo Fergusson and Carlos Molina look at the problem through a broader lens. They study the impact of Facebook across different countries, taking advantage of the fact that versions of the platform in different languages were introduced at different points in time. As it turns out, the incidence of protest increased after the introduction of a Facebook platform in a given language in countries with a greater presence of speakers of that language relative to those with fewer speakers. Given that no similar pattern occurs for the prevalence of different political views, the authors conclude that the mechanism is via the facilitation of coordination towards collective action.

The role of coordination is underscored, in a different context, in Chapter 12. Marco Manacorda and Andrea Tesei focus on 2G mobile phone technology, which does not lend itself to web use but facilitates communication via voice and text. Using detailed georeferenced data on the rollout of 2G technology and on the incidence of protests across Africa over 15 years, they show that mobile phones were instrumental to mass political mobilisation across local areas within countries, but only during periods of economic

downturn, when reasons for grievance emerge and the cost of participation falls. Their results highlight the interplay between the online and offline worlds, suggesting that communication technologies can act as a tool for coordinating pre-existing offline grievances, leading individuals to mobilise and amplifying their discontent.

The evidence thus suggests that the low barriers to entry and the effectiveness of social media in allowing for communication do translate into greater mobilisation for collective action by citizens. In conjunction with its effect on electoral politics, it becomes clear that the political impact of social media is deep and wide-ranging: it can empower all sorts of different political actors to pursue many sorts of political activities.

## **SOCIAL MEDIA AND HATRED**

One particularly important kind of political action that one may be concerned about due to its potential negative impact is what we may call ‘hatred’: collective action directed against specific groups. This is the focus of Part 5.

Chapter 13, by Leonardo Bursztyn, Georgy Egorov, Ruben Enikolopov and Maria Petrova, looks once again at the Russian context and investigates the impact of social media on the prevalence of xenophobic attitudes and ethnic hate crimes. Using the same variation exploited in Chapter 10, they show that the expansion of the VK social media platform increased the share of individuals holding extreme xenophobic views (based on an online survey experiment) and the occurrence of hate crimes against ethnic minorities. Moreover, they show that the two are connected: the increase in crimes is especially strong in cities with higher pre-existing levels of social media and for crimes with multiple perpetrators. This underscores the ‘dark side’ of social media’s ability to facilitate collective action, by allowing like-minded individuals to find one another and to coordinate towards action.

Important recent episodes in Western Europe and in the United States are the focus of Chapter 14, by Karsten Müller and Carlo Schwarz. They show first that the large influx of refugees into Germany in 2015, as a result of the civil war in Syria, was associated with an increase in anti-refugee rhetoric in social media by the far-right Alternative für Deutschland (AfD) party and also in violent attacks against refugees. Using variation in internet and Facebook outages over time and across German towns, they show that the correlation between online hatred and offline hate crimes weakens when the internet or Facebook goes offline. Similarly, anti-Muslim sentiment and hate crimes increased in the United States, especially after Donald Trump initiated his presidential campaign, and those increases were stronger in places with greater Twitter penetration. Using the variation induced by the SXSW festival, discussed in Chapter 7, they argue that Twitter had a causal impact, mediated by Trump’s tweeting. Importantly, the effect was concentrated in places with substantial pre-existing activity by hate groups.



The evidence thus suggests that offline and online activities are also inextricably linked by social media when it comes to hatred against specific groups. These technologies facilitate the spread of hateful content as well as coordination around translating animosity into action.

## **SOCIAL MEDIA IN AUTOCRACIES**

The role of social media in autocracies has been another prominent topic of study. The case of China has been of special interest, in light not only of the country's size and importance, but also of the well-documented efforts by the Chinese regime in trying to harness new media and communications technologies towards its own goals. This is the focus of Part 6.

In Chapter 15, Bei Qin, David Strömberg and Yanhui Wu sketch out the social media landscape in China. They start by estimating the Chinese government's social media presence, looking at the Sina Weibo platform, and show that it is much more extensive than commonly assessed, with some 600,000 government-affiliated Weibo accounts, with posting particularly concentrated on politically sensitive topics. In addition, they find evidence that the government uses social media to monitor local officials, as well as to exert surveillance over collective action by citizens. Yet the spread of information on such action can take place over social media, which highlights a key tension facing an autocratic government such as that of China: the value of information versus the risk of collective action against the regime.

This tension is further exemplified in Chapter 16, by David Yang. The chapter reports on a field experiment conducted in China randomly assigning a group of university students to free access to a virtual proxy network (VPN) tool, which allows access to internet content from outside China, thereby bypassing censorship. The key finding is that access to uncensored content has little impact on actual acquisition of politically sensitive information. On the other hand, additional incentives to subscribe to a Western news outlet did increase that acquisition, and furthermore, this brought important and persistent changes to knowledge, beliefs and (intended) behaviour, in ways that run against the Chinese government. This pattern suggests that, while censorship may have managed to create an apathetic environment low demand for politically sensitive information, this does not necessarily imply fear. Sufficient incentives towards the acquisition of information could actually lead to an increased propensity for political action.

In sum, the role of social media in an autocratic regime once again highlights how its effects can go in different directions, as it can empower citizens but also governments.

## **SOCIAL MEDIA AND LEGACY MEDIA**

Last but not least, understanding the broader impact of social media requires not only considering their direct effects, but also how they shape and affect other media. After all, social media never exist in isolation, but as a part of a broader ecosystem. As such, Part 7 looks at the impact on so-called “legacy” media.

Chapter 17, by Sophie Hatte and Ekaterina Zhuravskaya, focuses on the interplay between social media and TV news coverage in the context of the Israeli-Palestinian conflict. Once again, the low barriers to entry in the dissemination of content comes to the forefront, in the example of ‘citizen-journalists’: social media content documenting citizens’ experience of conflict can affect how the conflict gets covered on TV. Using variation in social media access driven by internet outages in Israel and Palestine, they show that more citizen-generated content leads to more TV coverage of the conflict. What is more, it changes the tone of that coverage, bringing greater emotional intensity and focus on the impact on civilians. As a result of that influence, coverage becomes more similar across different TV news channels.

The direct influence of social media content on news coverage is also the topic of Chapter 18, by Julia Cagé, Nicolas Hervé and Béatrice Mazoyer. Using a large sample of French-language tweets, they identify instances of events covered in both social and legacy media. They then rely on variation in the impact of Twitter content that is driven by the network centrality of users, as opposed to the intrinsic newsworthiness of the event they address, interacted with variation in the presence of competing stories at the time of the event. They find that an increase in the number of tweets about an event has a causal impact on the news coverage by legacy media, which is particularly strong for legacy media outlets that rely more on advertising revenues. The findings also suggest that this may bias traditional news coverage, since Twitter users are not a representative sample of the broader audience, and Twitter-pushed stories do not seem to generate greater interest among readers.

The evidence clearly indicates that social media influences the behaviour of legacy media, both in terms of the issues they cover and how they cover them. This suggests that the impact of social media can go well beyond the direct effect that is often more easily measured and assessed.

## **WHAT TO MAKE OF THIS?**

The unique features of social media, compared to pre-existing media technologies – its pervasiveness, the kind of content it favours, the low barriers to entry, and the multi-way mass communication it facilitates – imply that its effects on political life can be complex and nuanced. Social media can empower voters and politicians, citizens and autocrats, increase participation or decrease it, bring people together or push them apart. Far from

simple answers regarding whether social media is ‘good’ or ‘bad’, the work compiled in this volume establishes that the implications can push in different directions, depending on different outcomes and contexts.

One thing that remains a crucial, open topic for future research is what to do about social media from a policy point of view. How can we harness its potential and minimise its harm? In contrast with many addictive goods that bring little social upside beyond individual enjoyment – and whose consumption, for the most part, we would therefore like to discourage – social media can also have positive effects that are not fully internalised by individuals. This makes it difficult to figure out what optimal regulatory policy should be, and how to achieve it.

Some possible directions are already apparent. To give but one example, the evidence in Chapter 3 indicates that the algorithms employed by social media companies play an important role in the impact of social media usage on polarisation. Regulating them, or changing the incentives that underpin their design, seems worthy of consideration by policymakers. By the same token, the chapters in Part 7 underscore that social media strongly conditions what happens on traditional media platforms, and this interplay should affect how policy approaches the regulation of those platforms.

In any case, more research is needed to guide policy. Yet there are substantial challenges to pushing research in this direction. For one, social media has become ubiquitous, and the kind of exogenous variation in access exploited by some of the contributions covered here – the spread of broadband or 3G access, or idiosyncrasies in the initial expansion of certain platforms – is, by now, hard to come by. The experimental approach that other contributions have used is a natural alternative, but they generally preclude the kind of large-scale, general-equilibrium impact that is often the object of policy interest.

Yet the importance of the questions associated with these issues is bound to keep drawing intense attention from researchers going forward. One recent example, published too late to be included in this book, is the series of collaborations between teams of social scientists and social media companies for large-scale experiments (González-Bailón et al. 2023, Guess et al. 2023a, 2023b, Nyhan et al. 2023). The results shed light on the power and limitations of algorithms and user features, the degree of segregation in news consumption, and the limits of its effects on polarisation. At the same time, they are bound to raise further questions that will continue to stimulate research going forward.

Another important avenue to explore, which much of the work here exemplifies, is considering social media in its role as an arena for political action, which means that social media provides a rich source of data for answering political economy questions beyond those related to the impact of the technology itself. The continual improvement of language processing tools – possibly enhanced by artificial intelligence – will ensure that this will be a vibrant area of research going forward.

Towards that goal, one possible type of policy intervention would be to facilitate access to data for research purposes. Data are both a key by-product of social media usage and a resource that is crucial to the business models of the companies operating in this space. While companies have an interest in allowing access to their data for research whose results might be of use to them, they also have a strong incentive to try and control the dissemination of those results. There seems to be scope for policy to induce companies to increase ‘no strings attached’ opportunities for data access.

What seems certain, however, is that social media will endure as a key force in society, and in politics in particular. We will keep studying it also, to paraphrase Sir Edmund Hillary, because it’s there.

## REFERENCES

Adena, M, R Enikolopov, M Petrova, V Santarosa and E Zhuravskaya (2015), “Radio and the Rise of the Nazis in Pre-war Germany”, *The Quarterly Journal of Economics* 130(4): 1885-1939.

Allcott, H and M Gentzkow (2017), “Social Media and Fake News in the 2016 election”, *Journal of Economic Perspectives* 31(2): 211-236.

Allcott, H, M Gentzkow, and L Song (2022), “Digital Addiction”, *American Economic Review* 112(7): 2424-2463.

Campante, F, R Durante, and A Tesei (2022), “Media and Social Capital”, *Annual Review of Economics* 14: 69-91.

Campante, F and D A Hojman (2013), “Media and Polarization: Evidence from the Introduction of Broadcast TV in the United States”, *Journal of Public Economics* 100: 79-92.

Della Vigna, S and E Kaplan (2007), “The Political Impact of Media Bias”, mimeo, UC Berkeley and Stockholm University.

Durante, R, P Pinotti, and A Tesei (2019), “The Political Legacy of Entertainment TV”, *American Economic Review* 109(7): 2497-2530.

Gentzkow, M (2006), “Television and Voter Turnout”, *The Quarterly Journal of Economics* 121(3): 931-972.

Ghoni, W (2012), *Revolution 2.0: The Power of the People is Greater than the People in Power: A Memoir*, Houghton Mifflin Harcourt.

González-Bailón, D, D Lazer, P Barberá et al. (2023), “Asymmetric Ideological Segregation in Exposure to Political News on Facebook”, *Science* 381.6656, 392-398.

Guess, A M, N Malhotra, J Pan et al. (2023), “How do Social Media Feed Algorithms affect Attitudes and Behavior in an Election Campaign?”, *Science* 381(6656): 398-404.

Guess, A M, N Malhotra, J Pan et al. (2023), “Reshares on Social Media amplify Political News but do not detectably affect Beliefs or Opinions”, *Science* 381(6656): 404-408.

Haidt, J (2022), “Why the Past 10 years of American life have been uniquely stupid”, *The Atlantic*, 11 April.

Lazer, D M J, M A Baum, Y Benkler et al. (2018), “The Science of Fake News”, *Science* 359(6380): 1094-1096.

Martin, G J and A Yurukoglu (2017), “Bias in Cable News: Persuasion and Polarization”, *American Economic Review* 107(9): 2565-2599.

Nyhan, Brendan, J Settle, E Thorson et al. (2023), “Like-minded Sources on Facebook are Prevalent but not Polarizing”, *Nature* 620(7972): 137-144.

Prat, A, and D Strömberg (2013), “The Political Economy of Mass Media”, *Advances in Economics and Econometrics* 2: 135-187.

Prior, M (2007), *Post-broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections*, Cambridge University Press.

Strömberg, D (2004), “Mass Media Competition, Political Competition, and Public Policy”, *The Review of Economic Studies* 71(1): 265-284.

Sunstein, C R, (2017), “A Prison of our own Design: Divided Democracy in the Age of Social Media”, Democratic Audit UK.

Yanagizawa-Drott, D (2014), “Propaganda and Conflict: Evidence from the Rwandan Genocide”, *The Quarterly Journal of Economics* 129(4): 1947-1994.

## ABOUT THE AUTHORS

**Filipe Campante** is a Bloomberg Distinguished Professor of International Economics at Johns Hopkins University, and a Research Associate at the National Bureau of Economic Research. He is interested in political economy, development economics, and urban/regional issues. His research looks at what constrains politicians and policy makers beyond formal checks and balances: media, cultural norms, institutions, political protest. His work has appeared in leading academic journals, and received numerous mentions in media outlets all over the world. Born and raised in Rio de Janeiro, Brazil, he holds a PhD from Harvard University, an MA from the Pontifical Catholic University of Rio de Janeiro, and a BA from the Federal University of Rio de Janeiro, all in economics.

**Ruben Durante** is Professor of Economics at the National University of Singapore and ICREA researcher at Pompeu Fabra University. He is also affiliated faculty to the Barcelona School of Economics and IPEG, Research Advisor at Gallup US, and Research Fellow of CEPR, CESifo, and IZA. He has held visiting positions at Yale University, PSE, and INSEAD. He studies political economy with a focus on the functioning and impact

of traditional and new media in democratic societies. His work has been published in top journals in economics, political science, and management. His research has been supported by several funding agencies including the European Research Council through a Starting Grant, a Consolidator Grant, and a Proof of Concept Grant. He holds a BA in Economics from the University of Messina, a Master's in Political Economy from Sorbonne University, and a Master's and a PhD in Economics from Brown University.

**Andrea Tesei** is Associate Professor of Economics at Queen Mary University of London. He is also a Research Affiliate of CEPR, a Research Associate of the CEP at the London School of Economics and a Research Fellow of CESifo. He has been a visiting scholar at Northwestern University. His main research interests are in political economy and development economics, with a focus on the political, economic and social impact of traditional and new media in both developed and developing countries. His research has been published in leading scientific journals, including the *American Economic Review*, *Econometrica* and the *Review of Economics and Statistics*. Andrea obtained his PhD from Universitat Pompeu Fabra in 2012.



## **SECTION 1**

# **WELFARE EFFECTS OF SOCIAL MEDIA**





# CHAPTER 1

## The welfare effects of social media

25

Hunt Allcott,<sup>a</sup> Luca Braghieri,<sup>b</sup> Sarah Eichmeyer<sup>b</sup> and Matthew Gentzkow<sup>a</sup>

<sup>a</sup>Stanford University; <sup>b</sup>Bocconi University

In the last decade, social media has woven its way deep into our lives. Facebook has 2.3 billion monthly active users, and by 2016 the average user was spending nearly an hour per day on it and its sister platforms. There may be no technology since television that has so dramatically reshaped the way we get information and spend our time.

Early on, platforms like Facebook, Twitter and Instagram were hailed for their potential to make communication and the sharing of information easier. Now, the conversation is dominated by potential harms, from addiction to depression to political polarisation. Despite the abundance of speculation about the potential effects of social media, hard evidence remains scarce.

In a recent paper, we provide a large-scale randomised evaluation of the welfare impacts of Facebook, the largest social media platform (Allcott et al. 2020). This provided the largest-scale experimental evidence to date on Facebook's impact on a range of outcomes.

We find that deactivating Facebook for one month leads people to spend more time with friends and family. It also leaves them less informed about the news, less polarised in their political opinions, and a little happier and more satisfied with their lives. We find that after the time off Facebook, users want it back, but they use it significantly less than before. Our findings are in line with other important work on the same topic (e.g. Mosquera et al. 2020, Müller and Schwarz 2021, Braghieri et al. 2022).

### STUDY DESIGN

We recruited 1,600 US Facebook users online and randomised them into a 'deactivation' (or 'treatment') group and a 'control' group. The deactivation group received US\$102 in exchange for staying off Facebook for the four weeks leading up to the US midterm election in November 2018; the control group kept using Facebook as usual.

We measured a suite of outcomes using text messages, surveys, emails and administrative voting records. We recorded key measures twice – once in October, before the beginning of the deactivation period ('baseline'); and once in November, after the deactivation period had concluded ('endline'). We then compared the changes in those outcomes in the deactivation group to those in the control group. Our surveys had very high response

rates: of the 580 people in the deactivation group, only seven failed to complete the endline survey. Of the 1,081 people in the control group, only 17 failed to complete the endline.

To verify deactivation, we repeatedly pinged the URLs of participants' public Facebook profiles. While a user can limit how much content other people can see in their profiles, they cannot hide their public profile page. The public profile page returns a valid page when an account is active but returns an error message when an account is deactivated. Overall, 90% of users in the deactivation group followed our instructions and deactivated their accounts. For our impact evaluation, we estimate the local average treatment effect of deactivation. That is, we use the treatment indicator to instrument for the percentage of deactivation checks in which a person is observed to be deactivated.

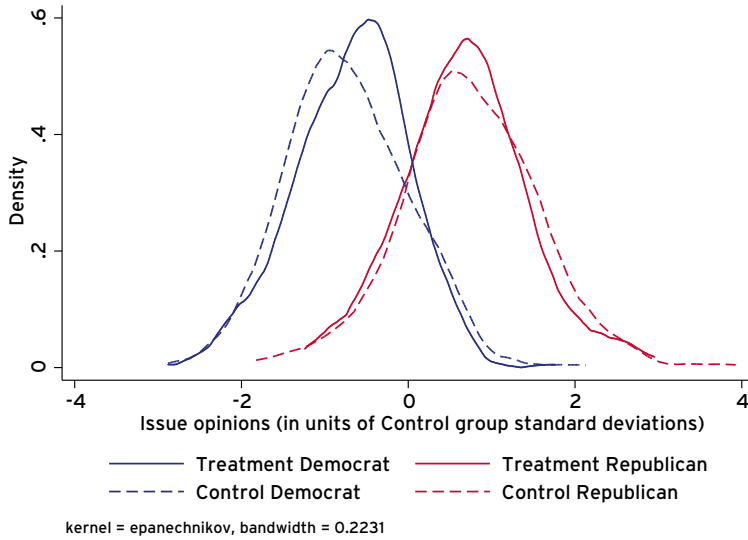
## KEY FINDINGS

Being off Facebook freed up an average of one hour to spend on other activities. How people use this extra time helps us understand which activities Facebook is crowding out, and this in turn tells us something about Facebook's effects. If Facebook time just replaces other social media or similar digital activities, the effects of deactivation might be small. If it replaces high-quality social interactions with family and friends, we might worry more about outcomes like (un)happiness, loneliness and depression. If it replaces consumption of high-quality news, we might worry more about impacts on political knowledge and polarisation.

Our surveys show that Facebook does not substitute for other digital activities – if anything, people reported spending less time on other social media and digital platforms while their Facebook accounts were deactivated. The deactivation group reported spending more time on offline activities, including face-to-face socialising and solitary activities like watching TV.

Our next set of findings focuses on news knowledge and political outcomes. Deactivating Facebook caused a significant reduction in total news consumption and news knowledge. Among other things, we find that those in the deactivation group were significantly worse at answering quiz questions about current issues in the news. At the same time, the deactivation group ended up significantly less polarised in a range of measures, including their views on policy issues such as immigration and policing (Figure 1). Our overall index of political polarisation fell by 0.16 standard deviations. As a point of comparison, prior work has found that a different index of political polarisation rose by 0.38 standard deviations between 1996 and 2018 (Boxell 2020). There is no detectable effect on political engagement, as measured by voter turnout in the midterm election.

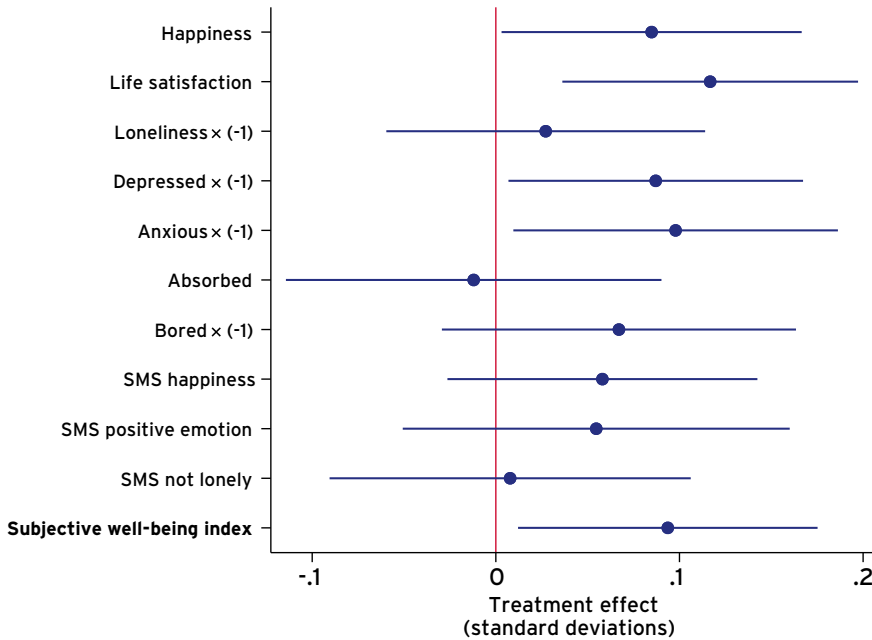
**FIGURE 1 EFFECTS ON POLARISATION**



Note: The dashed lines show the distribution of these views among control-group Democrats (blue) and Republicans (red). The solid lines represent the deactivation group. In both groups Democrats' views are well to the left of Republicans' views, but the inter-party differences are visibly smaller in the deactivation group, suggesting that deactivation moderated views in both parties.

In terms of wellbeing, we find that Facebook deactivation causes small but significant increases in self-reported individual life satisfaction and happiness, and significant decreases in self-reported levels of anxiety. We also elicited self-reported wellbeing using daily text messages, and find positive but statistically insignificant effects of Facebook deactivation on this outcome. As shown in Figure 2, an index of all measures together shows that deactivation caused significant improvements in overall wellbeing, with the overall index improving by 0.09 standard deviations. As a point of comparison, this is about 25-40% of the effect of psychological interventions including self-help therapy, group training and individual therapy, as reported in a meta-analysis by Bolier et al. (2013). These results are consistent with a recent quasi-experimental study finding that Facebook may have adverse effects on mental health (Braghieri et al. 2022). We find little evidence to support the hypothesis suggested by prior work that Facebook might be more beneficial for 'active' users – for example, users who regularly comment on pictures and posts from friends and family instead of just scrolling through their news feeds.

FIGURE 2 EFFECTS ON SUBJECTIVE WELLBEING



Note: Each point in the figure measures the effect of Facebook deactivation on wellbeing outcomes, measured in standard deviations. The lines to the left and right of each point indicate the 95% confidence interval. All outcomes are scaled so that the right of the figure indicates more positive outcomes. (Thus, measures of loneliness, depression, anxiety, and boredom are inverted.) The final row shows an index of all measures together, showing that deactivation caused significant improvements in overall well-being.

Finally, we measured whether deactivation affected people's demand for Facebook after the study was over, as well as their opinions about Facebook's role in society. As the experiment ended, participants assigned to the deactivation group reported planning to use Facebook much less in the future. Several weeks later, the deactivation group's reported usage of the Facebook mobile app was about 11 minutes (or 22%) lower than in control. In line with these self-reported measures, we found that 5% of the deactivation group still had their accounts deactivated nine weeks after the experiment ended. The deactivation group was also more likely to click on a post-experiment email providing information about tools to limit social media usage. Reduced post-experiment use aligns with our finding that deactivation improved subjective well-being, and it is also consistent with the hypotheses that Facebook is habit forming in the sense of Becker and Murphy (1988) or that people learned that they enjoy life without Facebook more than they had anticipated. A recent field experimental study on 'digital addiction' (Allcott et al. 2022) supports this notion, suggesting that self-control problems may cause 31% of social media use.

## BIG PICTURE

There is no doubt that many users perceive large benefits from Facebook. A majority of participants would require a payment of \$100 or more to deactivate Facebook for a month. Even after four weeks of deactivation, these valuations remained high and our participants continued to spend substantial time on Facebook every day. The results on news consumption and knowledge suggest that Facebook is an important source of news and information. Our participants' answers in free-response questions and follow-up interviews make clear the diverse ways in which Facebook can improve people's lives, whether as a source of entertainment, a way to organise a charity or an activist group, or a vital social lifeline for those who are otherwise isolated. Any discussion of social media's downsides should not obscure the basic fact that it fulfills deep and widespread needs.

At the same time, our results also make clear that the downsides are real. We find that four weeks without Facebook improves subjective wellbeing and substantially reduces post-experiment demand, suggesting that forces such as addiction may cause people to use Facebook more than they otherwise would. We find that while deactivation makes people less informed, it also makes them less polarised, consistent with the concern that social media have played some role in the recent rise of polarisation in the US.

The trajectory of views on social media – with early optimism about great benefits giving way to alarm about possible harms – is a familiar one. Innovations from novels to TV to nuclear energy have had similar trajectories. Along with the important existing work by other researchers, we hope that our analysis can help move the discussion from simplistic caricatures to hard evidence, and provide a sober assessment of the ways a new technology affects both individual people and larger social institutions.

## REFERENCES

Allcott, H, L Braghieri, S Eichmeyer and M Gentzkow (2020), "The Welfare Effects of Social Media", *American Economic Review* 110(3): 629-676.

Allcott, H, M Gentzkow and L Song (2022), "Digital Addiction", *American Economic Review* 112(7): 2424-2463.

Becker, G and K Murphy (1988), "A Theory of Rational Addiction", *Journal of Political Economy* 96(4): 675-700.

Bolier, L, M Haverman, G Westerhof, H Riper, F Smit and E Bohlmeijer (2013), "Positive Psychology Interventions: A Meta-Analysis of Randomized Controlled Studies", *BMC Public Health* 13(119).

Boxell, L (2020), "Demographic change and political polarization in the United States", *Economics Letters* 192.

Braghieri, L, R Levy and A Makarin (2022), “Social Media and Mental Health”, *American Economic Review* 112(11): 3660-3693.

Mosquera, R, M Odunowo, T McNamara, X Guo and R Petrie (2020), “The economic effects of Facebook”, *Experimental Economics* 23: 575-602.

Müller, K and C Schwarz (2021), “Fanning the Flames of Hate: Social Media and Hate Crime”, *Journal of the European Economic Association* 19(4): 2131–2167.

## ABOUT THE AUTHORS

**Hunt Allcott** is a Professor of Global Environmental Policy at Stanford University. He is the co-director of the Stanford Environmental and Energy Policy Analysis Center, a research associate at the National Bureau of Economic Research, an affiliate of ideas42 and Poverty Action Lab, and a member of the board of editors of the *American Economic Journal: Economic Policy*.

**Luca Braghieri** is an Assistant Professor in the Department of Decision Sciences at Bocconi University. He holds a PhD in economics from Stanford University and a bachelor’s degree from Harvard College. Most of his work is on topics related to applied microeconomics, behavioral economics, and political economy.

**Sarah Eichmeyer** is an Assistant Professor of Economics at Bocconi University. Her research spans the areas of public economics and political economy. She holds a PhD in economics from Stanford University, and undergraduate and master’s degrees from the University of Heidelberg and the University of Zurich, respectively.

**Matthew Gentzkow** is the Landau Professor of Technology and the Economy at Stanford University. He studies applied microeconomics with a focus on media industries. He received the 2014 John Bates Clark Medal, given by the American Economic Association to the American economist under the age of forty who has made the most significant contribution to economic thought and knowledge. He is a member of the National Academy of Sciences, a fellow of the American Academy of Arts and Sciences and the Econometric Society, and a senior fellow at the Stanford Institute for Economic Policy Research.

# CHAPTER 2

## Social media and mental health

**Luca Braghieri, Ro'ee Levy and Alexey Makarin**

Bocconi University; Tel Aviv University and CEPR; MIT Sloan School of Management

Over the last two decades, the mental health of adolescents and young adults in many countries has worsened considerably (Twenge et al. 2019). Data from the US shows that the fraction of individuals aged 18–23 who reported experiencing a major depressive episode in the past year almost doubled between 2008 and 2018 (NSDUH 2019). Similarly, over the same time period, suicides became more prevalent in the US and are now the second leading cause of death for individuals aged 15–24 years old (National Center for Health Statistics 2021).

Since the increased prevalence of mental illness among adolescents and young adults coincided with the diffusion of social media, researchers, journalists and policymakers alike began to wonder whether the two phenomena might be related (Twenge and Campbell 2019). In the autumn of 2021, a series of articles in the *Wall Street Journal* alleging that Meta (previously Facebook) was aware that Instagram had a negative effect on teenage girls' body image brought the relationship between social media and mental health to the forefront of public debate (Wells et al. 2021). Soon after, the US Congress held a committee hearing on the topic.

Despite the urgent need for studies on whether social media is detrimental to mental health, causal evidence remains scarce. Most existing papers estimate correlations between social media use and mental health (Bekalu et al. 2019, Berryman et al. 2018, Dienlin et al. 2017, Kelly et al. 2018; Lin et al. 2016, Twenge and Campbell 2019). A few experiments incentivise randomly selected participants to reduce their social media use and hence do estimate causal effects, but they do not concentrate primarily on mental health (Allcott et al. 2020, 2021, Mosquera et al. 2020).<sup>1</sup>

In a recent paper (Braghieri et al. 2022), we provide the most comprehensive causal evidence to date on the effects of social media on mental health by leveraging a unique natural experiment: the staggered roll-out of Facebook across US college campuses. Our empirical strategy allows us to estimate the short- to medium-run effects of Facebook on a rich set of mental health outcomes ranging from depression, to generalised anxiety

<sup>1</sup> A recent paper focuses on the causal effect of the internet - rather than social media - on mental health, and finds that access to high-speed internet increased incidence of mental disorders among young adults in Italy (Donati et al. 2022).



disorder, to anorexia.<sup>2</sup> Overall, we find that the introduction of Facebook at a college had a negative effect on student mental health, especially as far as depression and generalized anxiety disorder are concerned.

## ESTIMATING THE CAUSAL EFFECT OF FACEBOOK

Our research design leverages Facebook's gradual expansion across US colleges as a natural experiment. Facebook was created by Harvard undergraduate Mark Zuckerberg in February 2004. Initially, access to the platform was limited to Harvard students. Over the subsequent two and a half years, Facebook gradually expanded to other colleges in the US and abroad until eventually, in September 2006, it opened its doors to everyone in the world above the age of 13. The staggered nature of Facebook's roll-out allows us to compare changes in student mental health in colleges that just received Facebook access to changes in student mental health in colleges still without Facebook access in a difference-in-differences approach.

Although we study the expansion of a new technology, our study is not limited to a small subset of early adopters. When Facebook became available, colleges witnessed rapid and widespread adoption among students. Based on data provided by Facebook, we estimate that, in September 2005, approximately 85% of undergraduate students in colleges with access to Facebook had an account. Not only did Facebook spread rapidly and widely in the student population; usage was also intense. In early 2006, close to three-quarters of users logged into the site at least once a day, and the average user logged in six times a day (Hass 2006).

To estimate Facebook's effects on mental health, we rely on two datasets: one contains the dates in which Facebook was introduced at 775 US colleges; the other contains individual-level survey data about student mental health from the National College Health Assessment (NCHA). The database containing the Facebook introduction dates was constructed as follows. For the first 100 colleges that received access to Facebook, we relied on the introduction dates collected and made public in previous studies (Jacobs et al. 2015, Traud et al. 2012). For the remaining colleges, we collected introduction dates using the Wayback Machine, an online archive that contains snapshots of various websites at different points in time.<sup>3</sup> Our outcome variables come from the National College Health Assessment (NCHA), the largest and most comprehensive dataset on

2 The paper relates to an emerging literature, some of it featured in this CEPR eBook, exploiting the expansion of social media platforms to study the effects of social media on a variety of outcomes. The empirical strategy adopted in this paper is closely related to the one in Armona (2019), who leverages the staggered introduction of Facebook across U.S. colleges to study labor market outcomes more than a decade later. Enikolopov et al. (2020) and Fergusson and Molina (2020) exploit the expansion of the social media platform VK in Russia and of Facebook worldwide, respectively, to show that social media use increases protest participation. Bursztyn et al. (2019) and Müller and Schwarz (2020) exploit the expansion of VK and Twitter, respectively, and find that social media use increases the prevalence of hate crimes. Additional research on social media and political outcomes includes Enikolopov et al. (2018), Fujiwara et al. (2021), and Levy (2021). For a detailed overview, see Zhuravskaya et al. (2020). A unique feature of our setting is that it allows us to measure the effects of the sharp roll-out of the biggest social media platform in the world at a time in which very few close substitutes were available.

3 We thank Luis Armona for his collaboration in putting together the dataset containing Facebook expansion dates.

the mental health of US college students available at the time of Facebook's expansion (Leshner and Scherer 2021). We have access to the universe of responses to all NCHA survey waves administered between the spring of 2000 and the spring of 2008, the longest stretch of time around Facebook's early expansion in which the survey questionnaire did not vary. In order to allow us to carry out our analysis, the organisation administering the NCHA – the American College Health Association – generously provided us with a customised dataset that contained, together with the students' answers to the NCHA survey, a variable specifying the semester in which Facebook became available at the college attended by the survey respondent.<sup>4</sup>

Our main outcome variable is an index of poor mental health constructed by taking an equally weighted average of all the mental health questions in the NCHA survey inquiring about a respondent's recent past. We also analyse the effect of Facebook on three sub-indices: an index of questions about depression-related symptoms; an index of questions about other mental health conditions; and an index of questions about depression-related services, such as taking anti-depressants.

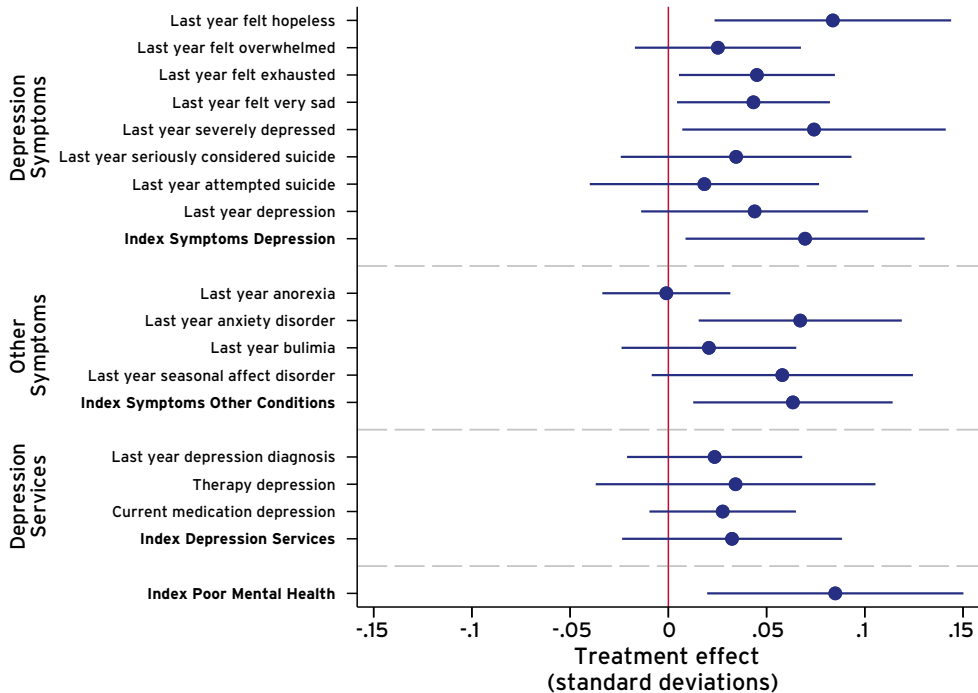
Mental health is a domain where self-reported outcomes are especially useful and self-reported symptoms are part of standard medical practice (Chan 2010). Still, the survey questions we analyse are not necessarily the questions medical professionals use in practice. Therefore, to validate that the NCHA survey questions measure mental health accurately, we conducted an original survey among more than 500 college students containing both the NCHA questions and the questions from canonical depression and generalised anxiety disorder screeners – the PHQ-9 and GAD-7, respectively – known to be highly predictive of actual medical mental illness diagnoses (Kroenke et al. 2001, Spitzer et al. 2006). We find that our index of poor mental health based on the NCHA questions is strongly correlated with the PHQ-9 and GAD-7 scores (correlation coefficients of 0.66 and 0.61 respectively), increasing our confidence that the NCHA survey is picking up the elements that feature into mental illness diagnoses.

4 For privacy reasons, our dataset does not contain college identifiers.

## THE EFFECT OF FACEBOOK ON MENTAL HEALTH

Figure 1 presents the causal difference-in-differences estimates of the impact of Facebook on mental health outcomes.<sup>5</sup>

**FIGURE 1 EFFECTS OF THE INTRODUCTION OF FACEBOOK ON STUDENT MENTAL HEALTH**



The results show that the introduction of Facebook at a college had a negative impact on student mental health. The effect size on the index of poor mental health is 0.085 standard deviation units. This corresponds to approximately 84% of the difference in the index of poor mental health between students in our sample with and without credit card debt. As an alternative point of comparison, the impact of introducing Facebook at a college is around 22% of the causal effect of a sudden unemployment spell (Paul and Moser 2009).

5 Specifically, we estimate the following two-way fixed effects (TWFE) model:

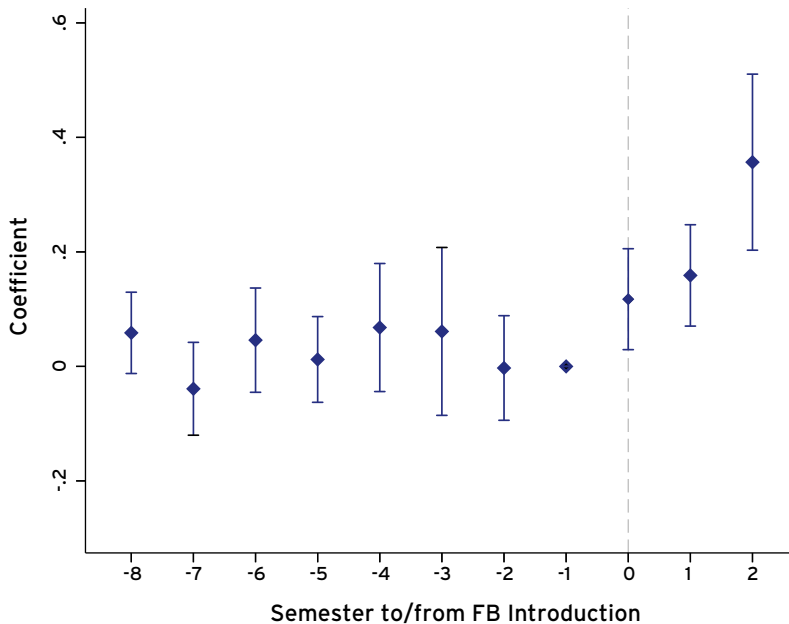
$$Y_{icgt} = \alpha_c + \delta_t + \beta \times \text{Facebook}_{gt} + \mathbf{X}_i \cdot \gamma + \mathbf{X}_c \cdot \psi + \varepsilon_{icgt}, \quad (1)$$

where  $Y_{icgt}$  represents an outcome for individual  $i$  who participated in survey wave  $t$  and attends college  $c$  that belongs to expansion group  $g$ ;  $\alpha_c$  indicates college fixed effects;  $\delta_t$  indicates survey-wave fixed effects;  $\text{Facebook}_{gt}$  is an indicator for whether, in survey wave  $t$ , Facebook was available at colleges in expansion group  $g$ ;  $\mathbf{X}_i$  and  $\mathbf{X}_c$  are vectors of individual-level and college-level controls, respectively. We estimate Equation (1) using OLS and cluster standard errors at the college level. We also address recent econometric concerns with staggered difference-in-differences research designs by showing robustness to the use of a variety of alternative estimators (Borusyak et al. 2021, Callaway and Sant'Anna 2021, De Chaisemartin and d'Haultfoeuille 2020, Sun and Abraham 2021).

The effects we find are strongest for depression and anxiety disorder. College-wide access to Facebook increased the number of students who reported experiencing severe depression or generalised anxiety disorder in the last year by 7% and 20%, respectively. Figure 1 shows that these estimates correspond to effects of around 0.07–0.08 in standard deviation units. This effect on severe depression is similar in magnitude to the effect observed in an experiment conducted by Allcott et al. (2020). Such similarity is striking, especially in light of the fact that the time period, survey question, target population, and empirical strategy in Allcott et al. (2020) are different from the ones in our paper. In contrast to depression and anxiety, we do not find significant effects on self-reports of anorexia and bulimia.

When estimating the effect of Facebook on mental health over time using an event-study regression, we find evidence that the effect increases as colleges are exposed to Facebook for more semesters (see Figure 2). Importantly for our empirical strategy, we do not find any significant effects or trends before Facebook is introduced at a college. The lack of pre-trends assuages potential concerns about our effect being driven by differential trends in mental health between colleges that received access to Facebook relatively early and colleges that received access relatively late.

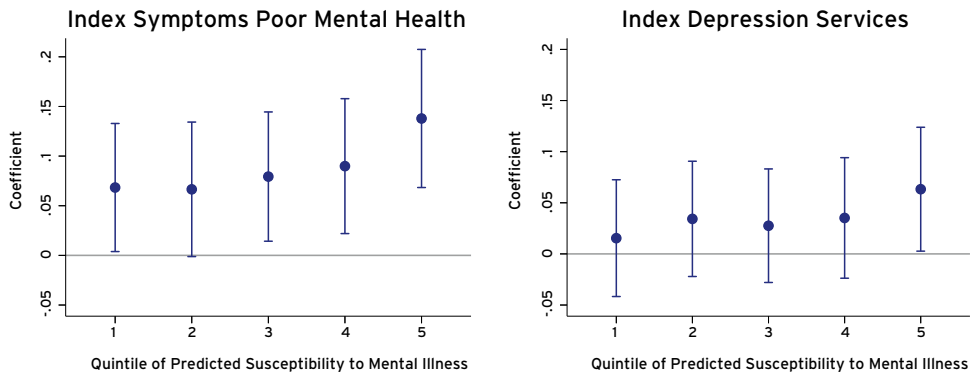
**FIGURE 2** EFFECTS OF FACEBOOK ON THE INDEX OF POOR MENTAL HEALTH BASED ON DISTANCE TO/FROM FACEBOOK INTRODUCTION



Note: Constructed using Sun and Abraham's (2021) estimator.

In order to study whether the effects of Facebook are concentrated among individuals who are particularly vulnerable to mental illness or whether they impact all students, we created a measure of predicted susceptibility to mental illness using a LASSO regression and studied heterogeneous treatment effects along that measure. The LASSO regression leverages a set of individual-level immutable characteristics such as gender and age to predict whether a student reported having ever received a mental illness diagnosis. As shown in Figure 3, the effects of Facebook on the index of poor mental health impact all students, but they are especially strong among students who are predicted to be most susceptible to mental illness. Furthermore, among those students, Facebook access significantly increased the take-up of psychotherapy and anti-depressants.

**FIGURE 3 HETEROGENEOUS EFFECTS BY PREDICTED SUSCEPTIBILITY TO MENTAL ILLNESS**

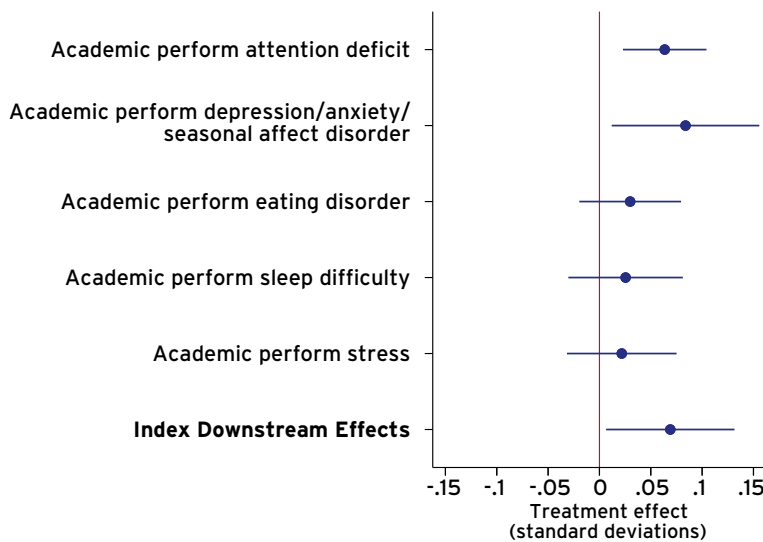


Does the effect of Facebook on mental health have negative downstream repercussions on academic performance? According to the students' reports, the answer is yes. The NCHA survey includes a host of questions asking students whether various conditions negatively affected their academic performance. We analyse all conditions related to mental health symptoms, along with an index summarising those symptoms. As shown in Figure 4, students were more likely to say that mental health issues negatively affected their academic performance after Facebook was introduced at their college. Consistent with our evidence suggesting that depression and anxiety-related disorders are the conditions most severely affected by the introduction of Facebook, we find the largest effect on a question asking about depression, anxiety and seasonal affect disorder. The number of students who reported that those conditions impaired their academic performance increased by three percentage points over a baseline of 13% as a result of the introduction of Facebook.

## Robustness

Our results pass numerous robustness checks. First, we observe null effects in placebo tests on variables that in principle should not be affected by the introduction of Facebook, such as our LASSO-predicted ‘susceptibility to mental illness’ variable that is based on baseline immutable characteristics. Second, the results remain similar in modified versions of our main specifications that take into account possible concerns related to (i) the construction of our index of poor mental health, (ii) the construction of our treatment variable, (iii) particular Facebook expansion groups driving the effects, (iv) other variables unrelated to Facebook driving the effects, and (v) possible violations of the parallel trends assumption (see the online appendix in Braghieri et al. 2022 for further details).

**FIGURE 4 DOWNSTREAM EFFECTS ON ACADEMIC PERFORMANCE**



## HOW DID FACEBOOK AFFECT MENTAL HEALTH?

So far, we have documented that Facebook access negatively affected student mental health. But what was the mechanism behind this effect? Recent scholarship identified two main channels whereby Facebook might directly affect mental health: unfavourable social comparisons (Appel et al. 2016) and disruptive internet use (Griffiths et al. 2014). Unfavourable social comparisons refers to the idea that users might use social media to compare themselves to others. To the extent that those comparisons are unfavourable, they might be detrimental to the users’ self-esteem and mental health. Disruptive internet use refers to the idea that social media might disrupt students’ ability to concentrate and to carry out their daily tasks, and lead to anxiety. Another possibility is that the

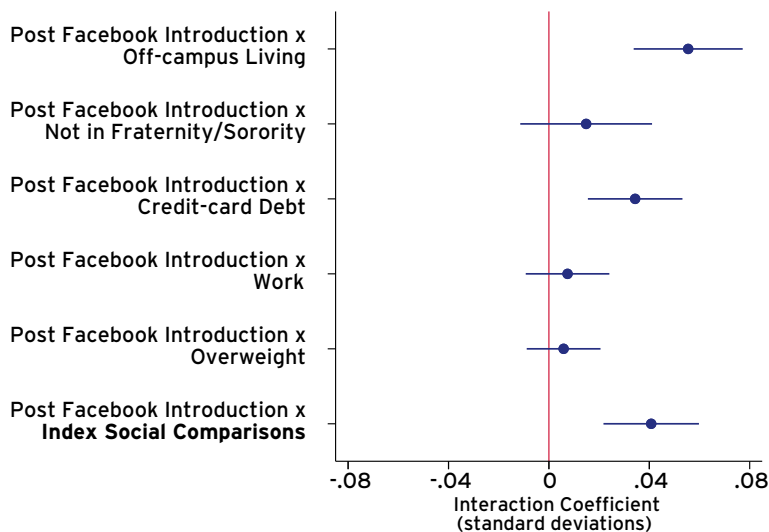
introduction of Facebook might lead to behavioural changes that, in turn, affect mental health. Overall, our evidence is most consistent with the unfavourable social comparisons channel.

### Unfavourable social comparisons

We find two pieces of evidence that suggest that Facebook's effect operated through social comparison.

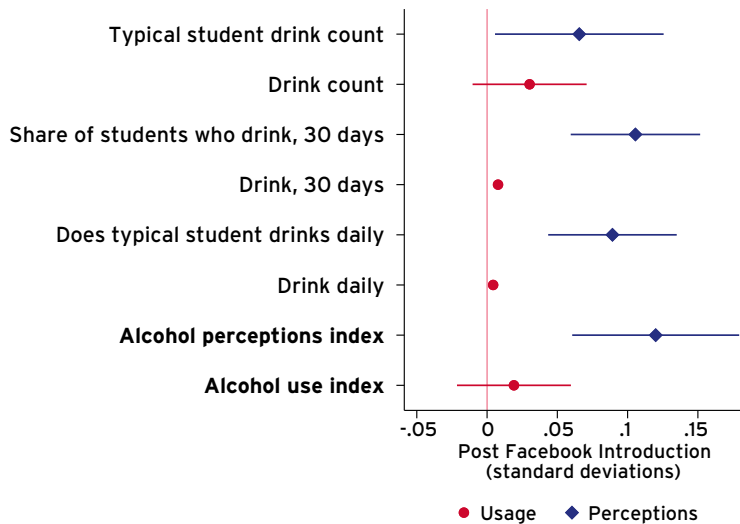
First, we focus on students who are more likely to be affected by unfavourable social comparisons: (i) students who live off-campus and are therefore less likely to participate in on-campus social life; (ii) students who have weaker offline social networks as measured by not belonging to a fraternity or sorority; (iii) students who have lower socioeconomic status as measured by carrying credit card debt or working part-time alongside studying; and (iv) students who are overweight. We aggregate these questions into an index of social comparison where respondents are considered to be at higher risk of unfavourable social comparisons if they have an above median number of the characteristics described above (e.g. they live off-campus, are overweight, and have credit card debt). Figure 5 shows that Facebook access had a more negative effect on students more likely to suffer from negative social comparisons. All the point estimates are positive and the estimates for off-campus living, credit card debt, and the index are statistically significant.

**FIGURE 5 HETEROGENEOUS EFFECTS AS EVIDENCE OF UNFAVOURABLE SOCIAL COMPARISONS**



Second, we test directly whether Facebook affected people's perceptions of their peers' social lives by estimating the impact of the roll-out of Facebook on students' perceptions of their peers' drinking behaviours.<sup>6</sup> Figure 6 shows that the introduction of Facebook increased the perceived prevalence of alcohol consumption among college students. Based on the questions asking students about their own alcohol consumption, the figure also shows that the increase in perceived alcohol consumption does not reflect an actual increase in consumption.

**FIGURE 6 EFFECTS ON ALCOHOL USE AND PERCEPTIONS AS EVIDENCE OF UNFAVORABLE SOCIAL COMPARISONS**



One explanation for a discrepancy between perceptions and reality in regards to alcohol is that students might have a hard time interpreting the content they observe on social media. In particular, they might forget that what they see on social media is a curated rather than representative version of their peers' lives. Indeed, we find an even stronger effect on perceptions among students living off-campus who have to rely more heavily on social media for information about their peers' behaviours. The changing perceptions could explain the negative effect on mental health, as inflated perceptions about others' social lives might make students feel worse about their own.

6 At the time, content related to alcohol featured prominently on Facebook.



### **Alternative channels**

We do not find evidence that the effect of Facebook on mental health operated through disruptive internet use. Facebook does not affect the share of students who report that the internet or video games affected their academic performance, as one would expect if Facebook were a distracting force.

We also do not find evidence that Facebook affected mental health indirectly by affecting other behaviours. Using a battery of questions in the NCHA survey, we find that Facebook did not affect drug use, assaults, sexual assaults or the answers to various questions related to relationships.

### **CONCLUSION**

In 2021, 4.3 billion individuals had a social media account, accounting for over half the world population and over 90% of internet users (We Are Social 2021). The repercussions of the rise of social media are thus likely to be far-reaching. We leverage the staggered introduction of Facebook across US colleges and find that the introduction of Facebook at a college had a negative effect on student mental health. Evidence points to unfavourable social comparisons as the leading mechanism.

Since our identification strategy delivers estimates of the effect of Facebook in the mid-2000s, one might wonder about the extent to which our estimates speak to the effects of social media today. Over the last fifteen years, Facebook introduced a host of new features, including the newsfeed algorithm, business pages, and videos. Although we cannot estimate the effect of these new features on mental health, we believe our estimates are still relevant because the main force driving our results – unfavourable social comparisons – is still a common feature of many social media platforms today. In fact, some of the new features introduced by Facebook might have exacerbated the effects of social comparisons: the information users receive on their peers is now richer (e.g. it includes videos), it is personalised by an algorithm, and content can be accessed at any time or place using a smartphone.

We emphasise that our analysis does not aim to estimate the overall welfare effects of social media; rather, it aims to shed light on a very important component of such a welfare calculation, namely, mental health. Clearly, social media might have positive effects on other outcomes affecting welfare. Indeed, the fact that individuals keep using social media despite the negative effects on subjective wellbeing and mental health suggests that social media platforms might have benefits that compensate for such costs. Ideally, future iterations of these platforms will be able to preserve the benefits while mitigating the mental health costs.

In terms of policy implications, our evidence on mechanisms suggests that regulators can consider interventions reminding the public that social media posts are not representative of people's real lives. Such intervention could include behaviour nudges on social media platforms or be part of digital education programs.

Overall, our results are consistent with the hypothesis that social media might be partly responsible for the recent deterioration in mental health among young adults. It is up to social media platforms, regulators and future research to determine whether and how these effects can be alleviated.

## REFERENCES

- Allcott, H, L Braghieri, S Eichmeyer, and M Gentzkow (2020), "The welfare effects of social media", *American Economic Review* 110(3): 629–676.
- Allcott, H, M Gentzkow and L Song (2021), "Digital addiction", NBER Working Paper No. 28936.
- Appel, H, A L Gerlach, and J Crusius (2016), "The interplay between Facebook use, social comparison, envy, and depression", *Current Opinion in Psychology* 9: 44–49.
- Armona, L (2019), "Online Social Network Effects in Labor Markets: Evidence From Facebook's Entry into College Campuses", available at SSRN 3381938.
- Bekalu, M A, R F McCloud, and K Viswanath (2019), "Association of social media use with social well-being, positive mental health, and self-rated health: disentangling routine use from emotional connection to use", *Health Education & Behavior* 46(2S): 69S–80S.
- Berryman, C, C J Ferguson, and C Negy (2018), "Social media use and mental health among young adults", *Psychiatric Quarterly* 89(2): 307–314.
- Borusyak, K, X Jaravel, and J Spiess (2021), "Revisiting event study designs: Robust and efficient estimation", arXiv preprint arXiv:2108.12419.
- Braghieri, L, R Levy, and A Makarin (2022), "Social media and mental health", *American Economic Review* 112(11): 3660–3693.
- Bursztyn, L, G Egorov, R Enikolopov, and M Petrova (2019), "Social media and xenophobia: evidence from Russia", NBER Working Paper No. 26567
- Callaway, B and P H Sant'Anna (2021), "Difference-in-differences with multiple time periods", *Journal of Econometrics* 225(2): 200–230.
- Chan, D (2010), "So why ask me? Are self-report data really that bad?", in *Statistical and methodological myths and urban legends*, Routledge, pp. 329–356.
- De Chaisemartin, C and X d'Haultfoeuille (2020), "Two-way fixed effects estimators with heterogeneous treatment effects", *American Economic Review* 110(9): 2964–96.

Dienlin, T, P K Masur, and S Trepte (2017), "Reinforcement or displacement? The reciprocity of FtF, IM, and SNS communication and their effects on loneliness and life satisfaction", *Journal of Computer-Mediated Communication* 22(2): 71–87.

Donati, D, R Durante, F Sobbrío, and D Zejcirović (2022), "Lost in the net? Broadband internet and youth mental health", IZA Discussion Paper No. 15202.

Enikolopov, R, A Makarin, and M Petrova (2020), "Social media and protest participation: Evidence from Russia", *Econometrica* 88(4): 1479–1514.

Enikolopov, R, M Petrova, and K Sonin (2018), "Social media and corruption", *American Economic Journal: Applied Economics* 10(1): 150–74.

Fergusson, L and C Molina (2020), "Facebook causes protests", Documentos CEDE 18002, Universidad de los Andes.

Fujiwara, T, K Müller, and C Schwarz (2021), "The effect of social media on elections: Evidence from the United States", NBER Working Paper No. 28849.

Griffiths, M D, D J Kuss, and Z Demetrovics (2014), "Social networking addiction: An overview of preliminary findings", Chapter 6 in K P Rosenberg and L C Feder (eds), *Behavioral Addictions*, Academic Press.

Hass, N (2006), "In your Facebook.com", *The New York Times*.

Jacobs, A Z, S F Way, J Ugander, and A Clauset (2015), "Assembling the Facebook: Using heterogeneity to understand online social network assembly", *WebSci '15: Proceedings of the ACM Web Science Conference*.

Kelly, Y, A Zilanawala, C Booker, and A Sacker (2018), "Social media use and adolescent mental health: Findings from the UK Millennium Cohort Study", *EclinicalMedicine* 6: 59–68.

Kroenke, K, R L Spitzer, and J B Williams (2001), "The PHQ-9: Validity of a brief depression severity measure", *Journal of General Internal Medicine* 16(9): 606–613.

Leshner, A I and L A Scherer (2021), *Mental health, substance use, and wellbeing in higher education: supporting the whole student*, National Academies Press.

Levy, R (2021), "Social media, news consumption, and polarization: Evidence from a field experiment", *American Economic Review* 111(3): 831–70.

Lin, L Y, J E Sidani, A Shensa, A Radovic, E Miller, J B Colditz, B L Hoffman, L M Giles, and B A Primack (2016), "Association between social media use and depression among us young adults", *Depression and Anxiety* 33(4): 323–331.

Mosquera, R, M Odunowo, T McNamara, X Guo, and R Petrie (2020), "The economic effects of Facebook", *Experimental Economics* 23(2): 575–602.

Mueller, K and C Schwarz (2020), “From hashtag to hate crime: Twitter and anti-minority sentiment”, available at SSRN 3149103.

National Center for Health Statistics (2021), “Health, United States, 2019: Table 007”.

NSDUH – National Survey on Drug Use and Health (2019), “2002–2019 National Survey on Drug Use and Health Final Analytic File (Codebook)”.

Paul, K I and K Moser (2009), “Unemployment Impairs Mental Health: Meta-Analyses”, *Journal of Vocational Behavior* 74(3): 264–282.

Spitzer, R L, K Kroenke, J B Williams, and B Löwe (2006), “A brief measure for assessing generalized anxiety disorder: The GAD-7”, *Archives of Internal Medicine* 166(10): 1092–1097.

Sun, L and S Abraham (2021), “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects”, *Journal of Econometrics* 225(2), 175–199.

Traud, A L, P J Mucha, and M A Porter (2012), “Social structure of Facebook networks”, *Physica A: Statistical Mechanics and its Applications* 391(16), 4165–4180.

Twenge, J M and W K Campbell (2019), “Media use is linked to lower psychological well-being: evidence from three datasets”, *Psychiatric Quarterly* 90(2): 311–331.

Twenge, J M, A B Cooper, T E Joiner, M E Duffy, and S G Binau (2019), “Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005–2017”, *Journal of Abnormal Psychology* 128(3): 185.

We Are Social (2021), *Digital 2021 April Global Statshot Report*.

Wells, G, J Horwitz, and D Seetharaman (2021), “Facebook knows Instagram is toxic for teen girls, company documents show”, *Wall Street Journal*.

Zhuravskaya, E, M Petrova, and R Enikolopov (2020), “Political effects of the internet and social media”, *Annual Review of Economics* 12: 415–438.

## ABOUT THE AUTHORS

**Luca Braghieri** is an Assistant Professor in the Department of Decision Sciences at Bocconi University. He holds a PhD in Economics from Stanford University and a bachelor’s degree from Harvard College. Most of his work is on topics related to applied microeconomics, behavioral economics, and political economy.

**Ro'ee Levy** is a Senior Lecturer (Assistant Professor) at the Tel Aviv University School of Economics and a CEPR Research Affiliate. His research focuses on political economy and social media. He studies the forces shaping social norms, political attitudes, and policy preferences, and their subsequent influence on individual behavior. Before joining Tel Aviv University, Ro'ee visited MIT as a post-doc and received his PhD from Yale University.

**Alexey Makarin** is an Assistant Professor of Applied Economics and a Mitsubishi Career Development Assistant Professor in International Management at the MIT Sloan School of Management. He earned his PhD in Economics from Northwestern University in June 2019. Alexey's research centres around political economy and the economics of digitization, with a special focus on the causal effects of social media platforms.

## **SECTION 2**

# **SOCIAL MEDIA AND VOTERS**



# CHAPTER 3

## Social media, news consumption and polarisation

**Ro'ee Levy**

Tel Aviv University and CEPR

The share of Americans getting news on social media has been steadily increasing over the recent past. In 2022, 70% of Americans consumed news on social media compared to fewer than one in eight Americans in 2008. When asked on which platforms they often get news, 18–29 year-olds report getting news on social media more than on any other medium, including television and news websites (Shearer 2018). As social media becomes a major source of news, there is growing apprehension over its effects on public opinion. A primary concern is that individuals are exposed to more news matching their ideology on social media. More pro-attitudinal news exposure could occur due to friends sharing like-minded content ('echo chambers') or as a result of algorithms prioritising content the user agrees with ('filter bubbles') (Pariser 2011). More exposure to like-minded news with a clear ideological slant may increase polarisation and threaten democracy (Sunstein 2017).

In a recent paper (Levy 2021), I collected novel data and conducted an experiment on Facebook to analyse news consumption on social media and its effect on polarisation. The paper has three main findings. First, individuals are exposed to more pro-attitudinal content and visit more pro-attitudinal websites through social media. Second, some of this behaviour can be explained by the algorithms governing social media. When individuals follow a Facebook page that matches their political opinions, they are more likely to see posts from that page, compared to a followed page that does not share the individual's opinions. Third, exposure to content on social media increases affective polarisation. Researchers tend to separate between affective polarisation, defined as the extent to which people of different parties dislike each other, and ideological polarisation, defined as the growing gap in opinions between the parties (Tucker et al. 2018). While I do not find evidence that exposure to ideological content increases ideological polarisation, I find that exposure to pro-attitudinal content increases affective polarisation, compared to counter-attitudinal content. Together, the results suggest that social media platforms may be increasing polarisation by exposing individuals to more pro-attitudinal content.



In this chapter, I discuss these three results. The first section analyses the news individuals are exposed to on social media. The second section explains the field experiment I conducted and presents results on how news exposure affects attitudes. The third section investigates why individuals are exposed to more pro-attitudinal content on social media.

### IS NEWS CONSUMED THROUGH SOCIAL MEDIA MORE SEGREGATED?

I analyse whether participants tend to consume like-minded news by merging the 2017 and 2018 Comscore WRDS Web Behavior Database Panels, from the Wharton Research Data Services at the University of Pennsylvania, with a dataset by Bakshy et al. (2015), defining the slant (ideological leaning) of 500 news domains. For each individual in the panel, I calculate the average slant of news sites visited through Facebook (i.e. by clicking a link in a Facebook post) and the slant of all other news sites visited. I then focus on the sample of participants who consumed news both through Facebook and through other websites and test whether news consumed through Facebook are more extreme and like-minded than other news.

Figure 1 shows that news consumed through Facebook are more extreme and pro-attitudinal. Figure 1a presents the distribution of the mean slant of news consumption and finds that news sites visited through Facebook are more extreme. For example, when visiting news sites through Facebook, 57% of individuals consume news that is on average more conservative than the *Wall Street Journal* or more liberal than the *Washington Post*, and when visiting news sites through other sources, only 39% of individuals consume such partisan news. However, this figure does not provide information on who consumes extreme news. Figure 1b shows that the most ideological individuals consume extreme news. The figure shows a clear correlation between the consumers' ideology, proxied based on their zip code, and the slant of their news consumption. The slope for news consumed through Facebook is steeper than the slope for news consumed through other sources, indicating that sites visited through Facebook tend to better match the consumers' ideology (for example, Republicans visit even more conservative sites through Facebook).

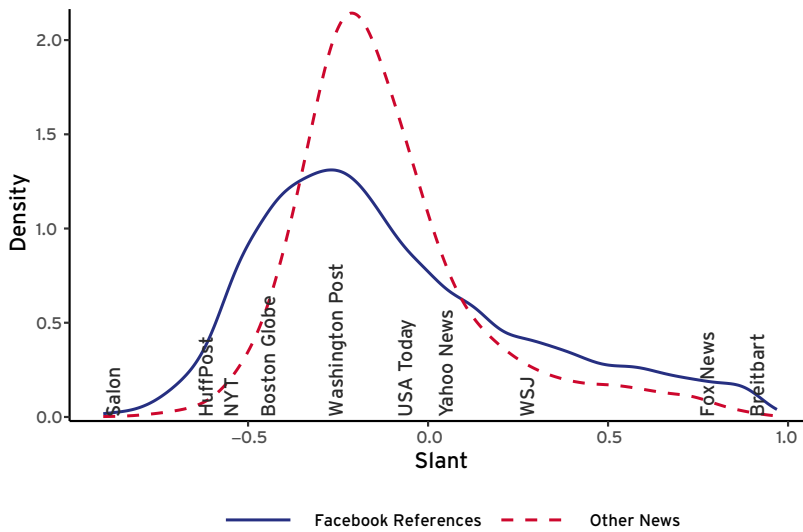
To quantify the difference between news consumed through social media and other news, I calculate segregation in news consumption, defined as the scaled standard deviation of the mean slant of news sites visited by individuals (Flaxman et al. 2016). I find that the segregation increases by over 50% when consuming news through Facebook compared to other news consumed. In other words, there is more variation in the average slant of news sites individuals visit when they click on links appearing on Facebook.

I complement this result with data from participants who installed a Google Chrome extension as part of the experiment, which is discussed in the next section. This additional dataset confirms that news consumed through social media is more segregated, indicating that this result does not stem from the characteristics of the Comscore

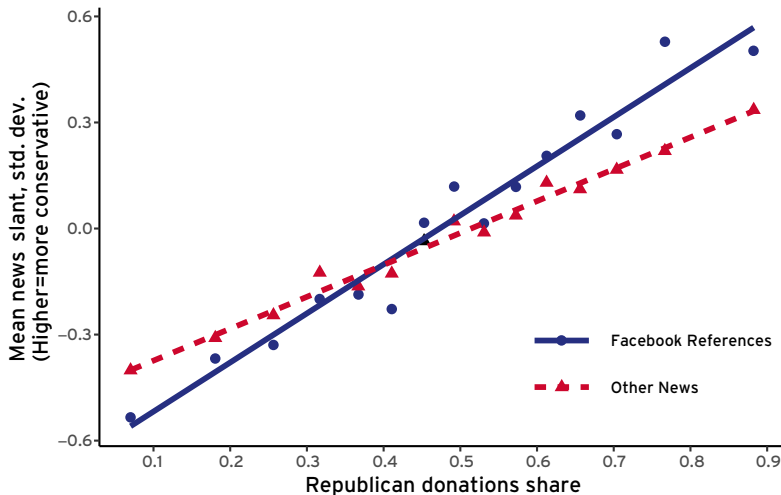
sample. I use the richer extension data to better understand the mechanisms leading to segregation. I find that the increase in segregation is mostly due to individuals clicking on posts posted by pages on Facebook and not posts by friends. This suggests that we should be more concerned with the personalisation of news outlets in social media feeds and less concerned with homophily. Finally, the extension data can be matched with self-reported data on party affiliation. This allows me to calculate isolation in online news consumption, defined as whether conservatives and liberals visit different websites (Gentzkow and Shapiro 2011). I find that isolation is much greater when visiting sites through social media.

**FIGURE 1 NEWS CONSUMPTION IN THE COMSCORE PANEL**

a) Distribution of mean news slant



b) Ideology and slant of news consumption



The results in this section are different from previous literature, which has often argued that concerns about echo chambers are overstated (Guess et al. 2018). There are at least three explanations for this discrepancy. First, there is disagreement about the definitions of echo chambers and filter bubbles. While I provide evidence that news consumption is more segregated on social media, individuals are not in complete echo chambers as they are still exposed to moderate and counter-attitudinal opinions in their feeds (Bakshy et al. 2015). Second, most news is still consumed through sources other than social media. Hence, even if news sites visited through social media are much more segregated, their effect on aggregate news consumption is limited (Flaxman et al. 2016). Third, some of the previous studies were conducted before social media became a popular source for news consumption. For example, Gentzkow and Shapiro (2011) find limited isolation in online news consumption, compared to offline sources, using data from 2009, while Peterson et al. (2019) find much greater isolation in 2016.

To conclude, the vast majority of Americans think that one-sided news is a very big, or at least a moderately big, problem on social media.<sup>1</sup> This section provides evidence that this concern is warranted. News accessed through Facebook is indeed more segregated and extreme than other online news, and although its current impact on total news consumption is limited, segregation may grow as more news is consumed through social media.

## THE EFFECT OF NEWS

### **A field experiment to estimate the effect of social media news exposure**

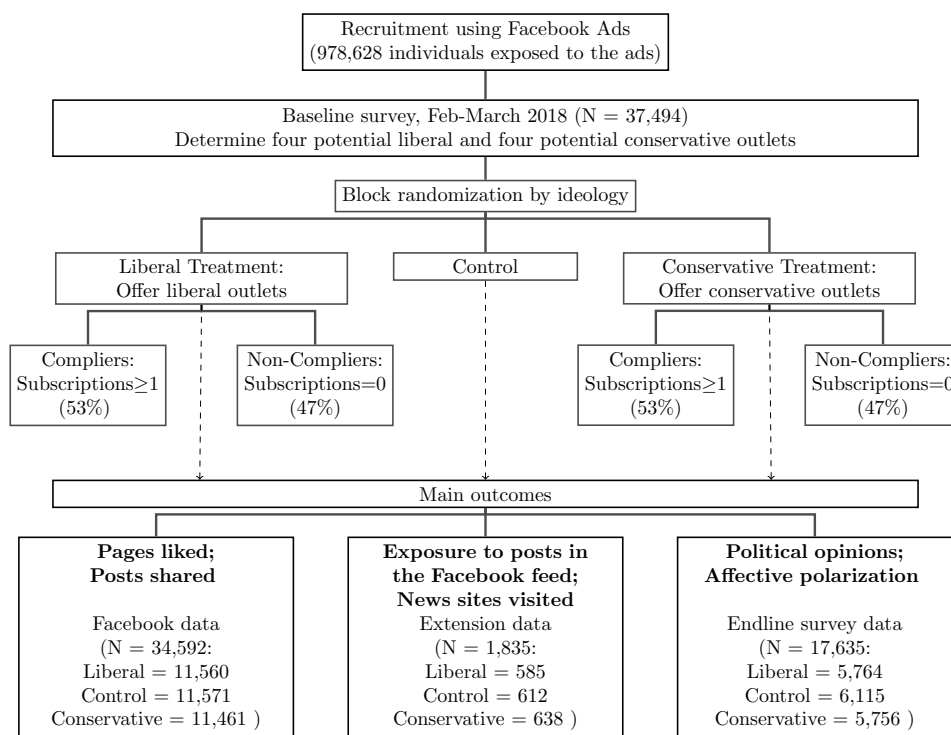
In February to March 2018, I conducted an experiment where participants were randomly assigned to three groups: a group offered the opportunity to like Facebook pages of four liberal news outlets (e.g., MSNBC), a group offered the opportunity to like pages of four conservative news outlets (e.g., Fox News) and a control group that was not offered any outlets. When Facebook users like an outlet's page on the platform, posts from the outlet may start appearing naturally in their feed, among many other posts they are exposed to (liking a post is similar to subscribing to specific content from an outlet, and I use the terms 'like' and 'subscribe' interchangeably in this chapter).

I designed the experiment to have high external validity. In contrast to lab studies, behaviour in the experiment occurs just as it does in the real world. Besides nudging individuals to like Facebook pages, the experiment did not directly intervene in any behaviour. The news supplied to participants was the actual news provided by leading media outlets. Facebook's algorithm determined which of the posts shared by the outlets appeared in the participants' Facebook feeds. Most importantly, participants decided whether to skip, read, click, or share posts. As a result, the effect of the intervention

1 Pew Research Center American Trends Panel Wave 51, July 2019.

is almost identical to the experience of millions of Americans who like popular news outlets on Facebook. I analyse three main datasets. First, attitudes were measured using a follow-up survey conducted two months after the experiment. Second, to measure compliance with the intervention and to test whether the treatment affected the posts people shared, I used Facebook's API to collect data, with participants' permission, on the pages people like on Facebook and the posts they share. Third, to analyse the effects of the experiment on participants' Facebook feeds and the news sites they visit, I asked a subset of participants who took the survey on Google Chrome to install an extension collecting this data. Figure 2 summarises the design of the experiment.

**FIGURE 2 EXPERIMENTAL DESIGN**

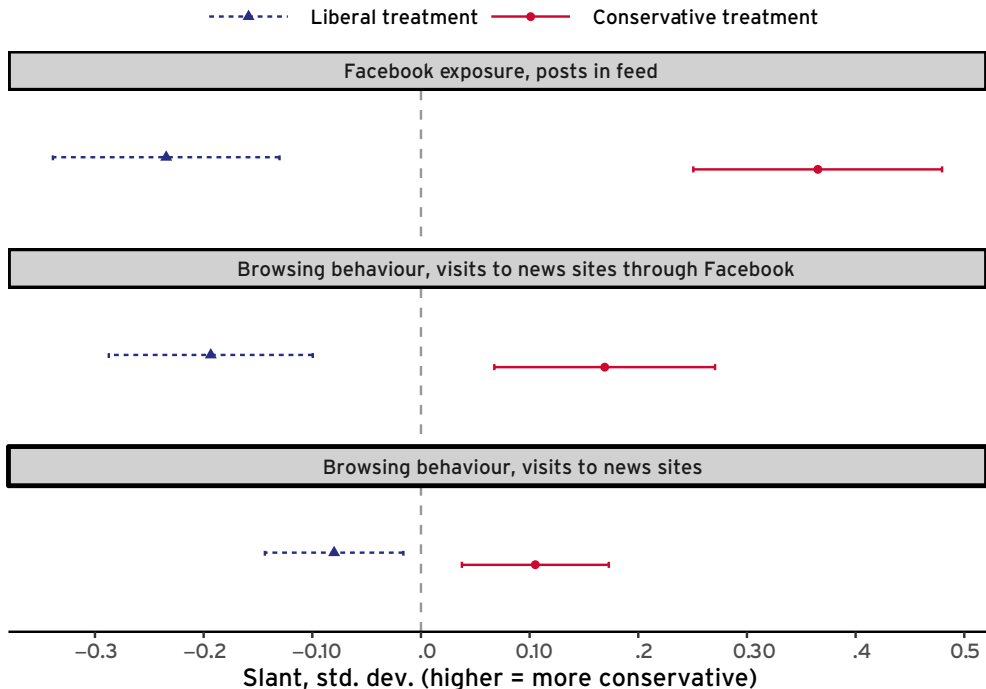


### Results: News consumption

Approximately half of the participants complied with the intervention by liking at least one of the pages offered. My first finding is that when posts from the offered outlets appeared in their feed, participants visited the websites of these outlets, even when the outlets were counter-attitudinal. This means that news is often consumed incidentally. Liberals may prefer liberal news and conservatives may prefer conservative news, but they are open to consuming news from other outlets when those outlets become more accessible.

The fact that individuals visited new websites as a result of the treatment does not mean that the average slant of the news they consumed changed. Participants could have adjusted the rest of their news consumption to keep their slant constant. For example, individuals randomly offered the *New York Times* could start consuming more articles from the outlet's website, but consequently consume less news from the *Boston Globe*, which offers a similar perspective. To test whether individuals re-optimize the sites they visit following an exogenous shock to their feed, I calculate the mean slant of all news exposure and consumption. The first panel of Figure 3 shows that when participants were randomly offered liberal or conservative outlets, their feed became substantially more liberal or conservative, respectively. The change in slant provides a strong first stage, which is useful when analysing the effect on political beliefs in the next subsection. It also allows me to test whether a change in the social media feed affects the slant of news sites visited or whether participants maintain a constant slant. The second and third panels of Figure 3 show that individuals do not fully re-optimize their news consumption as the treatments had a strong and significant effect on the slant of news sites visited by the participants. While many theories explain why individuals rationally consume news with a specific slant (Gentzkow et al. 2015), it seems that this news consumption could also stem from behavioural biases or search costs and not necessarily be a result of clear preferences.

**FIGURE 3** EFFECT OF THE TREATMENTS ON NEWS SLANT

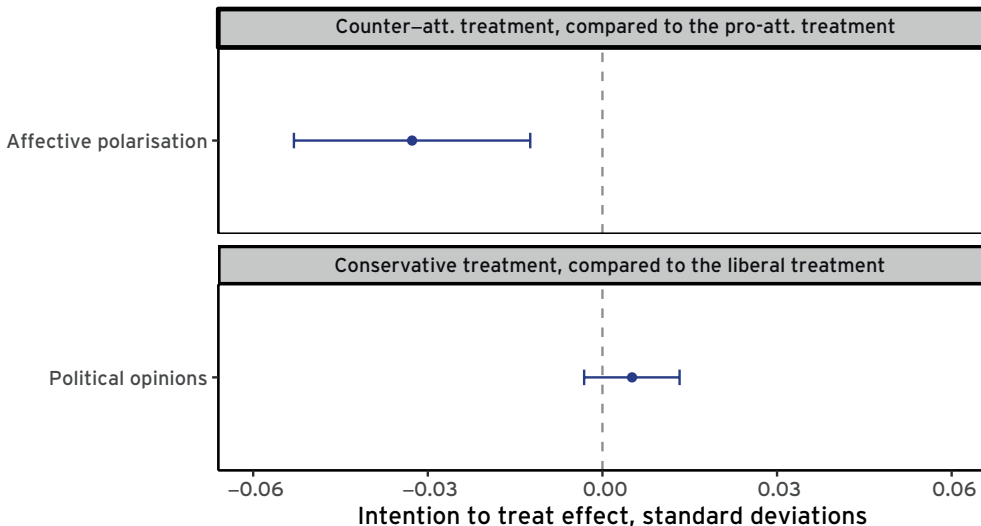


### Results: Attitudes

I estimate the effect on attitudes using two indices of questions from the follow-up survey. The political opinions index is composed of 20 survey questions focusing on domestic political issues covered in the news during the study period, such as the March for Our Lives movement. The affective polarisation index is composed of five outcomes summarising attitudes toward political parties. It includes questions such as how participants feel towards their own party and the opposing party (the feeling thermometer) and how upset would they feel if their son or daughter married someone from the opposing party.

The first panel of Figure 4 shows that counter-attitudinal news decreases affective polarisation compared to pro-attitudinal news. Focusing specifically on the feeling thermometer measures, I find that liking counter-attitudinal news outlets decreased the difference between participants' feelings towards their own party and the opposing party by 0.96 degrees on a 0-100 scale over two months, compared to liking pro-attitudinal news outlets.<sup>2</sup> For comparison, based on the American National Election Survey, this measure of polarisation increased by 3.83-10.52 degrees between 1996 and 2016. In other words, simply increasing exposure to content not aligned with one's view on Facebook (compared to news aligning with one's ideology) can mitigate a substantial portion of the increase in affective polarisation over the last two decades.

**FIGURE 4** EFFECT OF THE TREATMENTS ON POLITICAL OPINIONS AND AFFECTIVE POLARISATION



<sup>2</sup> This figure is the local average treatment effect based on instrumenting liking at least one pro- or counter-attitudinal outlet with whether those outlets were offered. The 'intention to treat' estimate is 0.58.

The takeaway from this result is not that Facebook explains all of the long-run trends in polarisation. In fact, descriptive evidence suggests that polarisation is not occurring only among Facebook users or only since Facebook became available (Boxell et al. 2018). However, the rise of polarisation may still be explained by a changing news environment that increases the consumption of like-minded news. Segregation in news consumption may have increased with talk radio, cable news and online blogs. The unique features of social media, which are discussed in more detail in the next section, are amplifying these trends and thus further increasing polarisation.

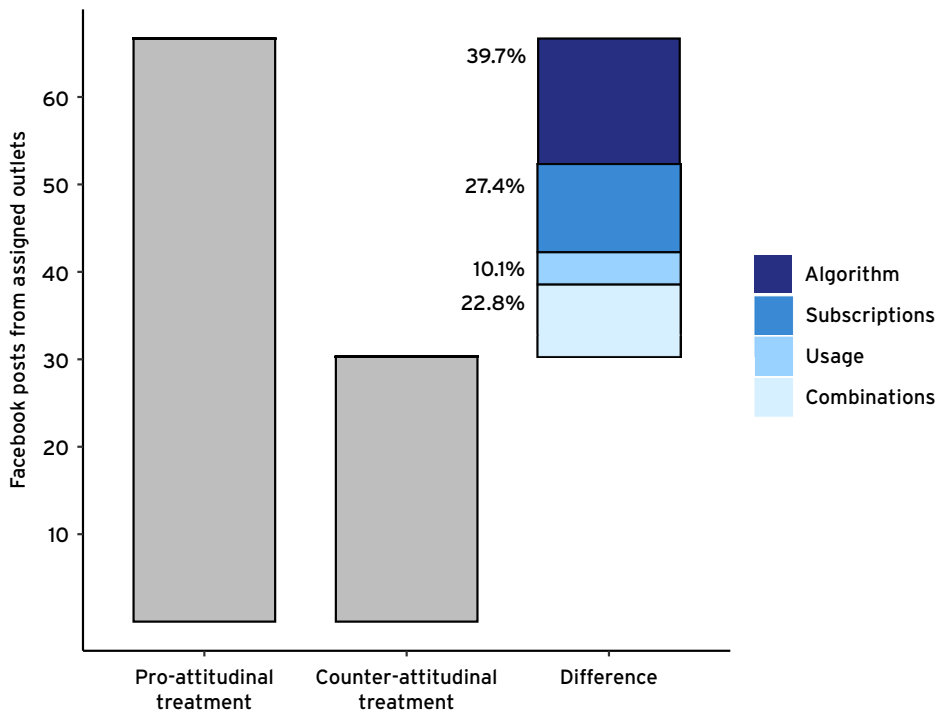
In contrast to the change in affective polarisation, the second panel of Figure 4 shows that posts from liberal outlets did not make individuals more liberal compared to posts from conservative outlets. Why did people's attitudes toward parties change while their opinions were not affected? One explanation consistent with the results is that participants exposed to counter-attitudinal news learned to rationalise the positions of the opposing party. Participants may have learned some of the opposing party's arguments and thus better understood why party supporters hold certain opinions, while still disagreeing with those opinions. More research is needed to understand why affective polarisation can increase without a corresponding increase in ideological polarisation, especially since there is clear evidence that affective polarisation has been growing in the US, while there is no consensus regarding the changes in ideological polarisation (Lelkes 2016, Iyengar et al. 2019).

### **EXPOSURE TO PRO-ATTITUDINAL CONTENT: SELECTION AND ALGORITHM**

In the experiment, participants who subscribed to pro-attitudinal outlets were more likely to see posts from these outlets in their feed, compared to individuals who subscribed to counter-attitudinal outlets. For example, a liberal who liked the Fox News page after a random offer observed far fewer posts from Fox News compared to a liberal who liked MSNBC following a random offer. I decompose the gap in exposure to posts shared by the pro- and counter-attitudinal outlets into three main forces: participants are less likely to subscribe to counter-attitudinal news outlets; Facebook's algorithm supplies fewer posts from counter-attitudinal outlets, conditional on participants subscribing to them; and participants use Facebook less often when offered counter-attitudinal outlets. This allows me to compare at least two channels often discussed in the literature: platforms decreasing exposure to specific content ('filter bubbles') and users choosing to consume content they agree with ('selective exposure').

Figure 5 shows that the strongest force associated with participants' increased exposure to pro-attitudinal news is the algorithm. Thus, even when individuals are willing to subscribe to outlets with a different point of view, Facebook's algorithm is less likely to show them content from those outlets. I also find evidence that participants prefer to subscribe to pro-attitudinal news outlets.<sup>3</sup>

**FIGURE 5** DECOMPOSING THE GAP BETWEEN EXPOSURE TO POSTS FROM THE OFFERED PRO-ATTITUDINAL AND COUNTER- ATTITUDINAL OUTLETS



The decomposition exercise provides novel evidence for the existence of a filter bubble, i.e. that Facebook's algorithm is more likely to expose individuals to news matching their ideology, conditional on following news from outlets. This does not imply that Facebook's algorithm intentionally increases segregation or that there is a causal effect of the match between a post's slant and a user's ideology on the ranking of a post in the feed. The ranking could be determined by many factors correlated with whether an outlet is pro- or counter-attitudinal. For example, it is possible that algorithmic ranking is mostly demand-driven and reflects users' past engagement with similar content. Still, the personalisation of news that individuals are exposed to on social media is an important departure from how news is supplied in other mediums. Until recently, the

<sup>3</sup> Changes in Facebook usage explain a smaller share of the increased exposure to pro-attitudinal content and this effect is only marginally significant.



engagement of an individual with news (e.g. the articles she read in the newspaper or the cable channels she chose to watch) did not affect her future supply of news. However, on social media, limited engagement with counter-attitudinal content may decrease future exposure to such content.

## CONCLUSIONS

The concern over the effects of social media is typically based on two assumptions: that social media exposes individuals to more like-minded news, and that such news affects their attitudes. I find evidence supporting both hypotheses. First, the difference in the news that Republicans and Democrats consume is greater when news sites are visited through social media, compared to other news sites visited. My results suggest that this is both due to individuals personalising their feed by following mostly like-minded outlets and due to the algorithm showing them more posts from these outlets. Second, I find that exposure to pro-attitudinal news increases affective polarisation. Together, the results of this study imply that social media platforms and their algorithms may be increasing polarisation.

However, the study also shows that news consumption is not fixed and that a simple, scalable nudge can substantially increase the consumption of counter-attitudinal news and decrease partisan animosity. Decision makers can promote similar interventions. While platforms already offer users suggested pages to follow or like, they can ensure that the pages they offer are more balanced and promote an ‘architecture of serendipity’ that exposes people to unexpected content (Sunstein, 2017). Regulators can require platforms to make their algorithms more transparent so users understand what news is being filtered out by companies. They can also require social media algorithms not to prioritise pro-attitudinal content. All of these interventions are likely to increase diversity in news exposure and thus potentially decrease polarisation, without forcing users to consume news they are not interested in.

The issues discussed in this chapter are likely to become more important over time. With improved artificial intelligence and bigger data, news consumption will become even more personalised. It is important to further understand the effects of these changes on attitudes and determine how digital news can prosper without creating insurmountable divisions in society.

## REFERENCES

Bakshy, E, S Messing and L A Adamic (2015), “Exposure to Ideologically Diverse News and Opinion on Facebook”, *Science* 348(6239): 1130–1132.

Boxell, L, M Gentzkow and J M Shapiro (2018), “Greater Internet Use is Not Associated with Faster Growth in Political Polarization among US Demographic Groups”, *Proceedings of the National Academy of Sciences of the United States of America* 115(3): 10612–10617.

Flaxman, S R, G Sharad and J M Rao (2016), “Filter Bubbles, Echo Chambers, and Online News Consumption”, *Public Opinion Quarterly* 80: 298–320.

Gentzkow, M and J M Shapiro (2011), “Ideological Segregation Online and Offline”, *Quarterly Journal of Economics* 126(4): 1799–1839.

Gentzkow, M, J M Shapiro and D F Stone (2015), “Media Bias in the Marketplace: Theory”, in *Handbook of Media Economics*, Vol. 1, Elsevier.

Guess, A, B Nyhan, B Lyons and J Reifler (2018), *Avoiding the Echo Chamber about Echo Chambers*, Knight Foundation.

Iyengar, S, Y Lelkes, M Levendusky, N Malhotra and S J Westwood (2019), “The Origins and Consequences of Affective Polarization in the United States”, *Annual Review of Political Science* 22(1): 129–146.

Lelkes, Y (2016), “The Polls-Review: Mass Polarization: Manifestations and Measurements”, *Public Opinion Quarterly* 80: 392–410.

Levy, R (2021), “Social media, news consumption, and polarization: Evidence from a field experiment”, *American Economic Review* 111(3): 831–70.

Pariser, E (2011), *The Filter Bubble*, The Penguin Press.

Peterson, E, G Shared and S Iyengar (2019), “Partisan Selective Exposure in Online News Consumption: Evidence from the 2016 Presidential Campaign”, *Political Science Research and Methods* 9(2): 1–17.

Shearer, E (2018), “Social media outpaces print newspapers in the US as a news source”, Pew Research Center, 10 December.

Sunstein, C (2017), *#Republic: Divided Democracy in the Age of Social Media*, Princeton University Press.

Tucker, J A, A Guess, P Barbera, C Vaccari, A Siegel, S Sanovich, D Stukal and B Nyhan (2018), *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature*, Hewlett Foundation.

## ABOUT THE AUTHOR

**Ro'ee Levy** is a Senior Lecturer (Assistant Professor) at the Tel Aviv University School of Economics and a CEPR Research Affiliate. His research focuses on political economy and social media. He studies the forces shaping social norms, political attitudes, and policy preferences, and their subsequent influence on individual behavior. Before joining Tel Aviv University, Ro'ee visited MIT as a post-doc and received his PhD from Yale University.

# CHAPTER 4

## Homophily, group size and the diffusion of political information in social networks

**Yosh Halberstam and Brian Knight**

Federal Deposit Insurance Corporation; Brown University

With the rise of social media as a key form of communication over recent decades, concern has arisen around the ‘echo chamber’ feature of key platforms, such as Twitter and Facebook. Unlike with traditional media outlets, such as radio, television and newspapers, each social media user is exposed to different information in their feed on a single platform. This is due to both algorithms, under which platforms selectively curate content for users, and the fact that users themselves actively choose which accounts to follow on social media (Levy 2021). On the latter mechanism, homophily – a tendency of individuals to form relationships with like-minded individuals – may further exacerbate this issue of echo chambers. Also, unlike traditional media outlets, which tend to be hierarchical in nature, social media users both consume information and produce content to which other users are exposed. This feature of social media might further contribute to echo chambers, in the sense that users only communicate with like-minded individuals via the two-way exchange of information. A final concern is that these issues can be exacerbated for minority groups, relative to majority groups, who, due to homophily, might have larger networks and hence be exposed to more information.

One important domain for this issue of homophily and echo chambers on social media is the political arena, given that voters often self-segregate into one of two groups (say, liberal or conservative). A tendency to create relationships with individuals of similar ideology has been previously documented (e.g. Iyengar et al. 2012, Malhotra and Huber 2017). In this chapter, we examine the implications of political homophily for the formation of links on social media platforms and the resulting flow of information.

Overall, social media has become an extremely important domain for politics. In terms of the consumption of political information, roughly one-half of adults report that they regularly get news from social media platforms.<sup>1</sup> In terms of the production of political information, according to a 2022 Pew Research Center survey of US adults, roughly one-third of all tweets are political in nature.<sup>2</sup>

1 [www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021](http://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021)

2 [www.pewresearch.org/politics/2022/06/16/politics-on-twitter-one-third-of-tweets-from-u-s-adults-are-political/](http://www.pewresearch.org/politics/2022/06/16/politics-on-twitter-one-third-of-tweets-from-u-s-adults-are-political/)

To investigate these issues around echo chambers on social media, in Halberstam and Knight (2016) we develop a simple theoretical model of social networks. The model incorporates political homophily, political communications and two groups: a majority political group and a minority political group. The model predicts that members of the majority group will tend to have more network connections and also tend to be exposed to more political information. Likewise, if users also tend to produce like-minded information, relative to information supporting the other group, then all users are exposed to more like-minded information, relative to information from the other group.

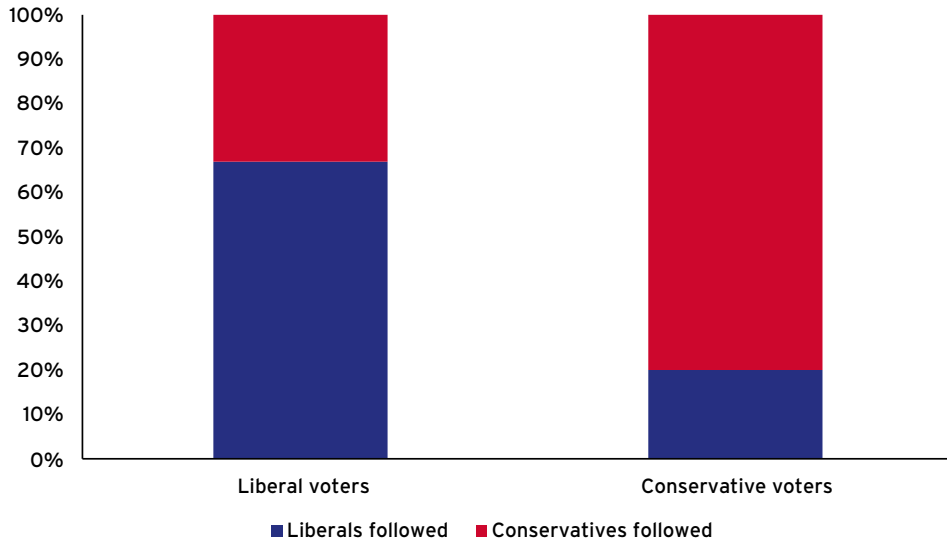
For several reasons, measuring the exposure of users to information on platforms in general, and like-minded information in particular, on social media is challenging. As noted above, every social media user is exposed to differing information on the same platform. Given this, researchers need to be able to access information on (1) the political affiliations of users in the political network, (2) the links between users in this network, (3) the content produced and transmitted by these users, and finally (4) the ideological nature of such content.

We attempt to overcome these challenges using data from Twitter. We begin by constructing a political network based upon users who follow accounts associated with major party candidates for the US Congress in the 2016 elections. Aggregating the followers across all candidates, this yields a political network of over 2 million Twitter users, who we term ‘voters’. To overcome the first challenge, we first measure the ideology of each voter based upon the number of Democrat and Republican candidates followed. So, a voter is coded as conservative if they follow more Republicans than Democrats and as liberal if they follow more Democrats than Republicans. To overcome the second challenge, we downloaded the list of followers of each of the 2 million accounts, allowing us to identify 90 million links between voters in this political network. To overcome the final two challenges, we focused on retweets by voters of candidates’ tweets along with mentions of candidates by voters. In both cases, we can use the partisan affiliation of the candidate – Democrat or Republican – as a proxy for the ideological content of this information. Finally, to test predictions around group size, we also created state-level subnetworks using information on user geography from Twitter. So, in more red states, such as Texas, Democratic voters will be in the minority, whereas in more blue states, such as California, Republican voters will be in the minority.

Using these data, we first investigate the degree of homophily in political connections. Despite the fact that our political network is only 36% liberal, 67% of followers of liberal voters are also liberal, with just 33% conservative, as shown in Figure 1. Likewise, 80% of followers of conservative voters are also conservative, with just 20% liberal, as shown in Figure 1. Thus, the Twitter political network exhibits a high degree of homophily and ideological segregation. The fact that conservative voters, who form a majority of our users, have a higher degree of like-minded followers is consistent with the predictions of

our theoretical model of social networks around group size. In particular, in the model, the majority ideological group, relative to the minority ideological group, tends to have more like-minded followers.

**FIGURE 1 HOMOPHILY IN THE POLITICAL NETWORK**

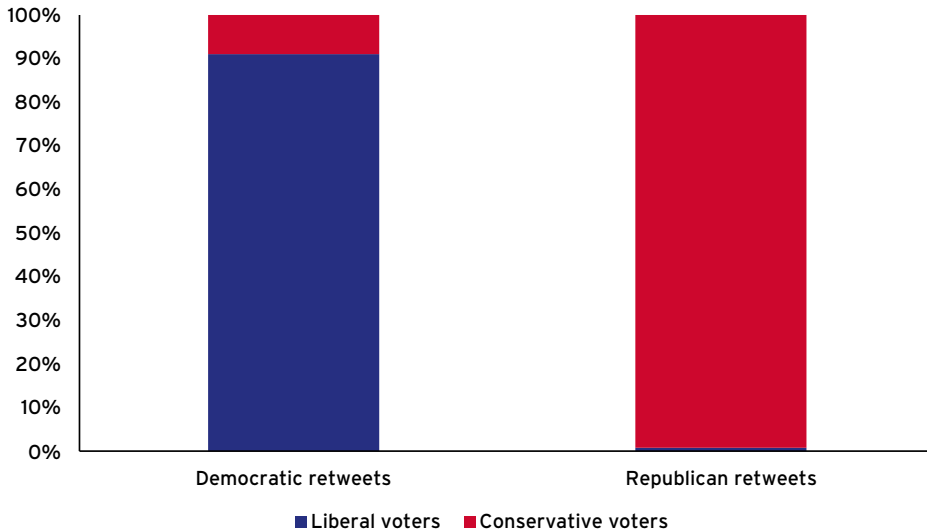


Of course, social media is not the first setting, either offline or online, to exhibit a high degree of homophily and ideological segregation. To compare our results to other settings, we create isolation indices that have been used in other academic studies. In short, we find that the degree of ideological segregation in our Twitter political network is similar to that in offline political social networks, defined as a network of individuals who sometimes discuss politics with each other, as reported in Gentzkow and Shapiro (2011). Moreover, the degree of political segregation in our Twitter political network is much higher than that in other type of interactions, suggesting that social media might be a force for increasing the degree of ideological segregation in society.

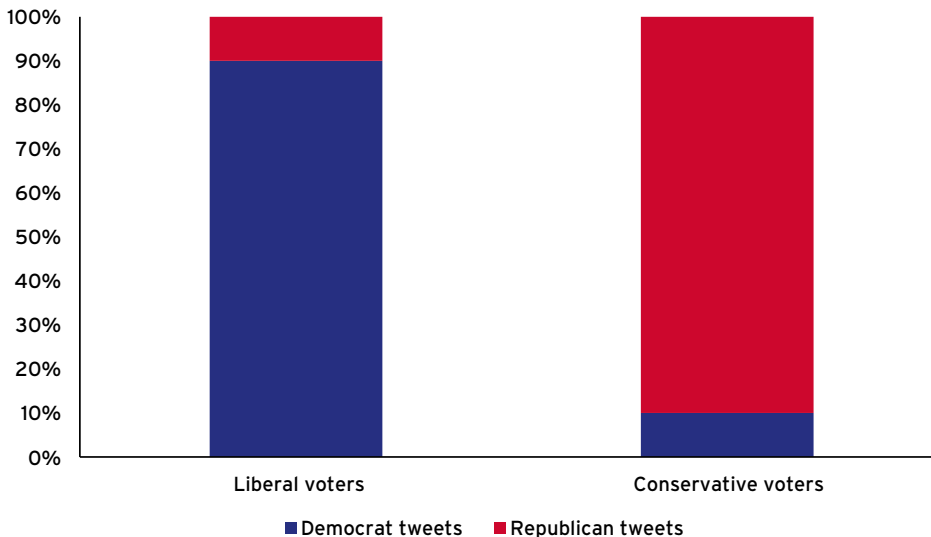
To further investigate the role of group size in homophily and echo chambers, we then turn to an analysis of the state-level networks. We first show that, consistent with the predictions of our theoretical model, in states that are more liberal, liberal voters tend to have more network connections. Likewise, in states that are more conservative, conservative voters tend to have more network connections. These networks also tend to exhibit a higher degree of homophily as the size of the group increases. So, for example, comparing a state with 70% conservatives to one with 60% conservatives, conservative voters will have both more network connections overall and a higher share of conservative followers in the more conservative state. This suggests that offline geographic segregation according to ideology may further exacerbate echo chambers on social media.

We next turn to an analysis of communication on the Twitter political network. Consistent with a key assumption of our model, we first document a high degree of the production of like-minded information, with liberal voters accounting for 91% of retweets of tweets from Democrat candidates and conservative voters accounting for nearly all (99%) of retweets of tweets from Republican candidates. We have similar findings for mentions of candidates, with 66% of mentions of Democrat candidates by liberal voters and 77% of mentions of Republican candidates by conservative voters.

**FIGURE 2 PRODUCTION OF LIKE-MINDED INFORMATION**



**FIGURE 3 EXPOSURE TO LIKE-MINDED INFORMATION**



Having established a high degree of homophily and the production of like-minded information, we next examine communications and the resulting exposure to ideological information in this political network. We again provide strong evidence of echo chambers. In particular, among liberal voters exposed to at least one candidate tweet in the political network (not including those from candidates that they directly follow), 90% of tweets to which they are exposed come from Democrat candidates. Likewise, among conservative voters exposed to at least one candidate tweets in the political network, an identical fraction of tweets (90%) to which they are exposed come from Republican candidates.

Turning to the role of group size in echo chambers and exposure to information, we next document that members of larger groups tend to be exposed to more information overall on Twitter. That is, a conservative in a conservative state receives more information on Twitter than does a conservative in a liberal state. Likewise, as the size of the majority group increases, the divergence in the total amount of information received by majority and minority voters increases. Correspondingly, these results extend to the political content of such information. That is, members of the majority group are, in a relative sense, exposed to more like-minded information and members of the minority group, also in a relative sense, exposed to more neutral information.

To the extent that newer information is more relevant for decision making than older information, homophily might have important implications for the speed of transmission and the resulting quality of such information. Indeed, in our theoretical model, we demonstrate that members of the majority group are exposed to information more quickly overall and that both groups receive like-minded information more quickly than they receive information from the other party. This pattern might potentially reinforce some of the findings described above with respect to the amount of information received.

To investigate these questions around the speed of transmission empirically, we use time and date stamps associated with tweets and re-tweets and measure the number of minutes that it takes for voters to receive a tweet, conditional on being exposed to the message. The unit of analysis in this case is the tweet user, allowing us to include tweet fixed effects. Using both linear regression methods and Cox survival analysis, we provide support for both predictions of the theoretical model. First, conditional on being exposed, larger groups are exposed to information more quickly overall. That is, going from the smallest groups (i.e. Democrat voters in very red states and Republican voters in very blue states) to the largest groups (i.e. Democrat voters in very blue states and Republican voters in very red states) is associated with a roughly 20-minute reduction in time to exposure. And, second, both groups tend to be exposed to like-minded information more quickly than they are exposed to information from the other party. That is, when the ideological nature of the content does not match that of the user, information takes an additional 80 minutes to arrive, again conditional on eventual exposure. To the extent that the quality of information decays over time and information becomes stale, these findings suggest a novel implication of echo chambers on social media.



Finally, based upon a content analysis of candidate tweets, which are classified as either political or non-political in nature, we investigate how the patterns described above differ between political and non-political information. Consistent with political homophily being more salient for political information, we show that our results are indeed weaker for non-political tweets. For example, liberals are exposed to only 6% conservative content for political information but 11% conservative content for non-political information. Likewise, we show that differences in the speed of transmission of information is higher for political information than for non-political information. That is, non-political tweets reach like-minded users 56 minutes faster and political tweets reach like-minded users 86 minutes faster, a difference of 30 minutes between political and non-political information.

Overall, these findings have several implications for the role of social media in society. They indicate that social media might further reinforce differences in the density of connections and the resulting exposure to information between larger political groups and smaller political groups. This can further disadvantage minority political movements. Our findings also suggest that social media has further contributed to increasing the echo chamber feature of interactions within broader society. And finally, to the extent that information on social media is influential in shaping beliefs and ideology (an issue not explicitly addressed in our study), then the rise of social media might have contributed to increased political polarisation, with liberals becoming more liberal and conservatives becoming more conservative.

## REFERENCES

- Gentzkow, M and J M Shapiro (2011), "Ideological segregation online and offline", *The Quarterly Journal of Economics* 126(4): 1799-1839.
- Halberstam, Y and B Knight (2016), "Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter", *Journal of Public Economics* 143: 73-88.
- Huber, G A and N Malhotra (2017), "Political homophily in social relationships: Evidence from online dating behavior", *The Journal of Politics* 79(1): 269-283.
- Iyengar, S, G Sood, and Y Lelkes (2012), "Affect, Not Ideology", *Public Opinion Quarterly* 76: 405-31.
- Levy, R (2021), "Social media, news consumption, and polarization: Evidence from a field experiment", *American Economic Review* 111: 831-70.

## ABOUT THE AUTHORS

**Yosh Halberstam** is a Senior Financial Economist at the FDIC Center for Financial Research in the Division of Insurance and Research. Before joining the FDIC in 2021, Yosh was Assistant Professor in the Economics Department at the University of Toronto. He received his PhD in economics from Northwestern University in 2011.

**Brian Knight** is a Professor of Economics at Brown University and a Research Associate at the National Bureau of Economic Research. He previously worked as an economist in the Division of Research and Statistics at the Federal Reserve Board and has held visiting faculty positions at Yale University and Harvard University. He received his PhD in 2000 from the University of Wisconsin-Madison. He is currently a co-editor at the American Economic Journal: Economic Policy and previously served as a co-editor at the Journal of Public Economics. He has also served on the Editorial Board at the American Economic Review.



# CHAPTER 5

## Political implications of the rise of mobile broadband internet

67

**Sergei Guriev,<sup>a,d</sup> Nikita Melnikov<sup>b</sup> and Ekaterina Zhuravskaya<sup>c,d</sup>**

<sup>a</sup>Sciences Po; <sup>b</sup>Nova School of Business & Economics; <sup>c</sup>Paris School of Economics; <sup>d</sup>CEPR

In the last 15 years, the world has experienced a major technological revolution – the rise of the mobile broadband internet powered by the so-called third and fourth generations of mobile telecommunications (3G and 4G, respectively). According to the International Telecommunications Union, in 2007 the world had only 0.04 active mobile broadband subscriptions per capita; in 2021, this number was 0.83. The growth has taken place in both developed and developing countries; in the latter, the respective numbers were 0.006 in 2007 and 0.73 in 2021. While 15 years ago broadband internet was accessed equally via mobile and cable-based technologies (there were 0.05 cable broadband subscriptions per capita in 2007), today, most users rely on mobile access: in 2021, there were only 0.17 cable broadband subscriptions per capita.

Unlike the previous generation of mobile technology (2G), mobile broadband internet (3G and 4G) allows seamless transfer of photos and videos. Not surprisingly, the spread of 3G contributed to a rapid growth of social media. Today, there are almost 3 billion Facebook users, 2.5 billion YouTube users, and 1.5 billion Instagram users. The vast majority of social media users access these platforms via mobile devices.

A prophetic book by Martin Gurri (2014, 2018) hypothesises that the arrival of mobile broadband has major political implications. Gurri argues that the self-immolation of Tunisian street trader Mohamed Bouazizi triggered the Arab Spring because it was recorded on a smartphone and went viral on social media. Indeed, a similar self-immolation of another street vendor, Abdesslem Trimech, that took place a few months earlier had gone unnoticed as it was not recorded and thus did not appear on social media.

Mobile broadband and social media not only contributed to triggering the Arab Spring, but also changed the way it was covered by mainstream media. Most of Al Jazeera's reporting of the Arab Spring came from mobile phone videos disseminated on social media, not from professional cameramen. This democratisation of coverage has dramatically impacted the way mainstream media work (Hatte et al. 2021, Cage et al. 2022).

The role of social media in the 2009-2010 Iranian ‘Twitter Revolution’ and the 2010-2011 Arab Spring generated a hope that the combination of 3G and social media would be a ‘liberation technology’ that keeps corrupt and autocratic governments accountable, informs citizens, and helps organise pro-democracy protests (Diamond and Plattner 2010). However, a concern has since emerged that mobile broadband and social media are also a ‘disinformation technology’ that helps to disseminate false news (Allcott and Gentzkow 2017, Melnikov 2023, Vosoughi et al. 2018) and populist messages (Tufekci 2018).

Which of these two views is correct? Is mobile broadband a liberation technology or a disinformation technology? On the one hand, recent studies have found that social media can indeed help expose corruption (Enikolopov et al. 2018) and organize protests (Enikolopov et al. 2020, Fergusson and Molina 2020, Manacorda and Tesei 2020). On the other hand, the internet and social media have also helped spread fake news (Allcott and Gentzkow 2017, Melnikov 2023) and contributed to increasing political polarisation (Halberstam and Knight 2016, Levy 2020, Melnikov 2023), xenophobia (Bursztyn et al. 2020) and hate crimes (Muller and Schwarz 2021). Certain governments have also successfully co-opted the internet, using it to advance their messaging (King et al. 2017, Qin et al. 2017, Chen and Yang 2020).

In our recent paper (Guriev et al. 2021), we carry out the first global analysis of the political implications of the rise of mobile broadband internet. We use data on 3G and 4G coverage in 2,232 subnational regions in 116 countries from 2008 to 2017 and survey data on almost a million individuals living in these regions.

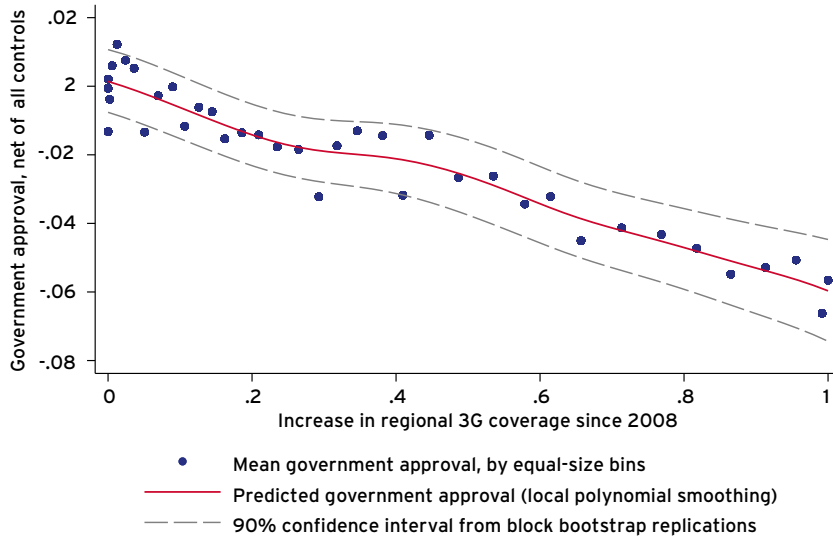
Our main result is that, on average, the increase in mobile broadband coverage reduces government approval (see Figure 1).

We show that this result is driven by negative information citizens receive about their governments on social media and other online resources. First, there is no effect of mobile internet on government approval when internet content is censored, and all of the effect comes from countries where internet content is not censored (see Figure 2).

Second, the effect of mobile broadband internet on government approval is especially strong in countries where traditional media are not free, and therefore citizens do not have access to alternative sources of independent political information (see Figure 3).

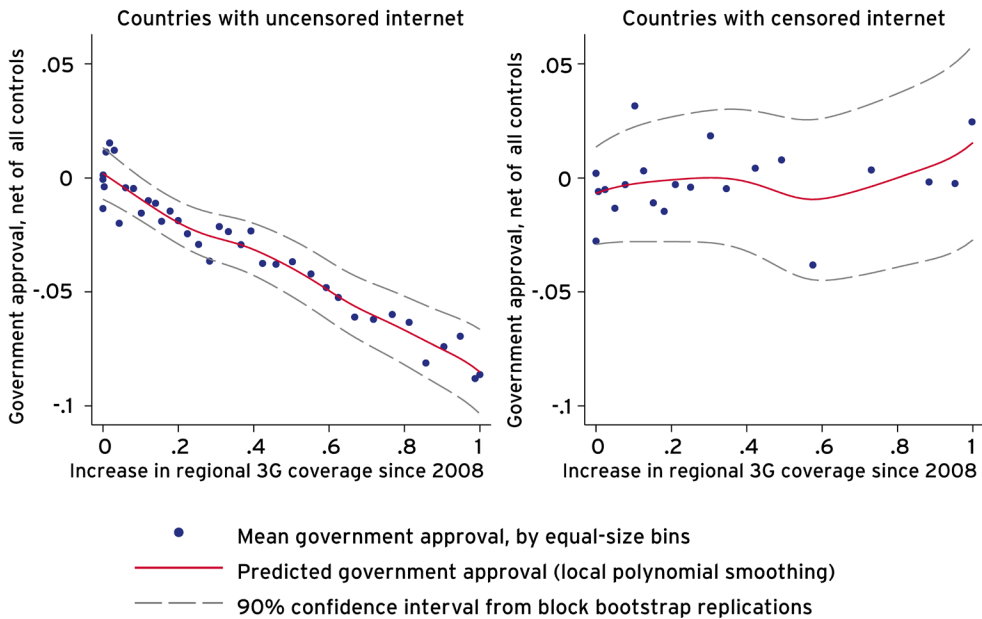
Third, mobile broadband internet indeed helps inform citizens about their government’s corruption scandals. In countries with actual incidences of corruption, citizens update their views of the level of corruption in the government upwards in places with access to mobile broadband internet, but not in places without mobile broadband internet (Figure 4).

**FIGURE 1 THE RELATIONSHIP BETWEEN REGIONAL 3G COVERAGE AND GOVERNMENT APPROVAL**

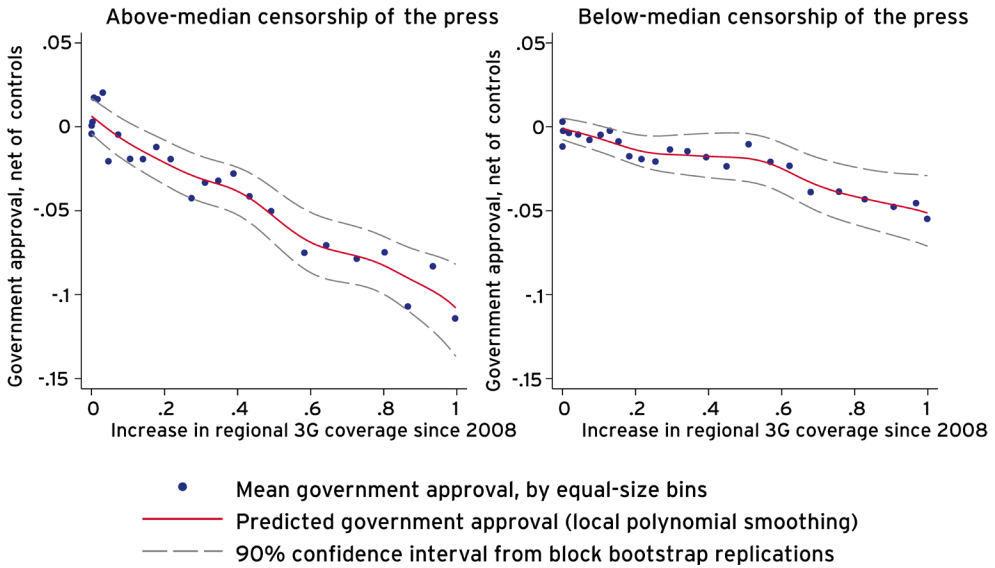


Note: The dots show the means of the respective outcome variables, net of all controls, by equal-size bins. The lines on the graphs show the predicted outcomes (Gaussian kernel, local polynomial smoothing). The confidence intervals are constructed by performing a block bootstrap at the level of the clusters.

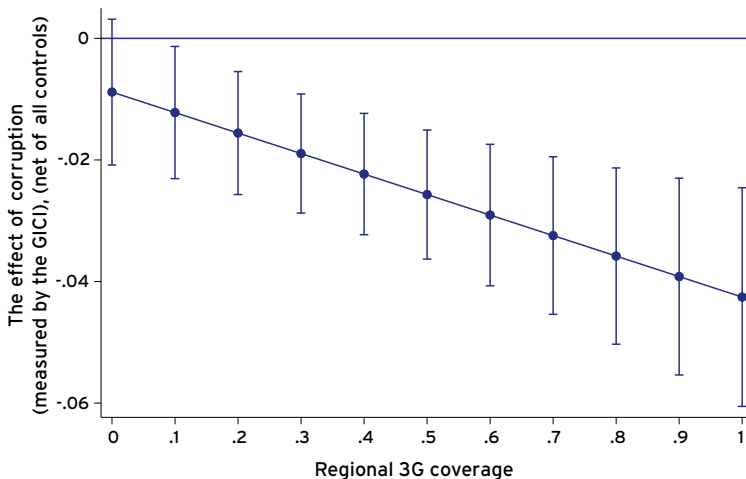
**FIGURE 2 THE RELATIONSHIP BETWEEN REGIONAL 3G COVERAGE AND GOVERNMENT APPROVAL SEPARATELY FOR COUNTRIES WHICH DO NOT AND WHICH DO CENSOR INTERNET CONTENT**



**FIGURE 3 THE RELATIONSHIP BETWEEN REGIONAL 3G COVERAGE AND GOVERNMENT APPROVAL SEPARATELY FOR COUNTRIES WITH ABOVE AND BELOW MEDIAN CENSORSHIP OF THE TRADITIONAL MEDIA AMONG THOSE COUNTRIES WHICH DO NOT CENSOR INTERNET**



**FIGURE 4 MOBILE BROADBAND COVERAGE AND ACTUAL VERSUS PERCEIVED CORRUPTION**



Note: The graph presents the marginal effect of an actual corruption incident on the perception of citizens that there is no corruption in government, depending on the level of regional mobile broadband coverage in the place of citizen's residence. The graphs also present 95% confidence intervals.

These results are consistent with the view that mobile broadband internet improves transparency and government accountability. Furthermore, we find that while the average global effect of 3G and 4G on government approval is negative, in the 'cleanest'

countries of the world (the top 10% in terms of low corruption), the effect is actually positive: when citizens learn that their government is actually clean – or at least cleaner than others – they are more likely to approve of it. In all other countries the effect is negative, but it is somewhat weaker in the most corrupt countries, in which citizens do not need to have access to mobile broadband internet to learn that their governments are corrupt.

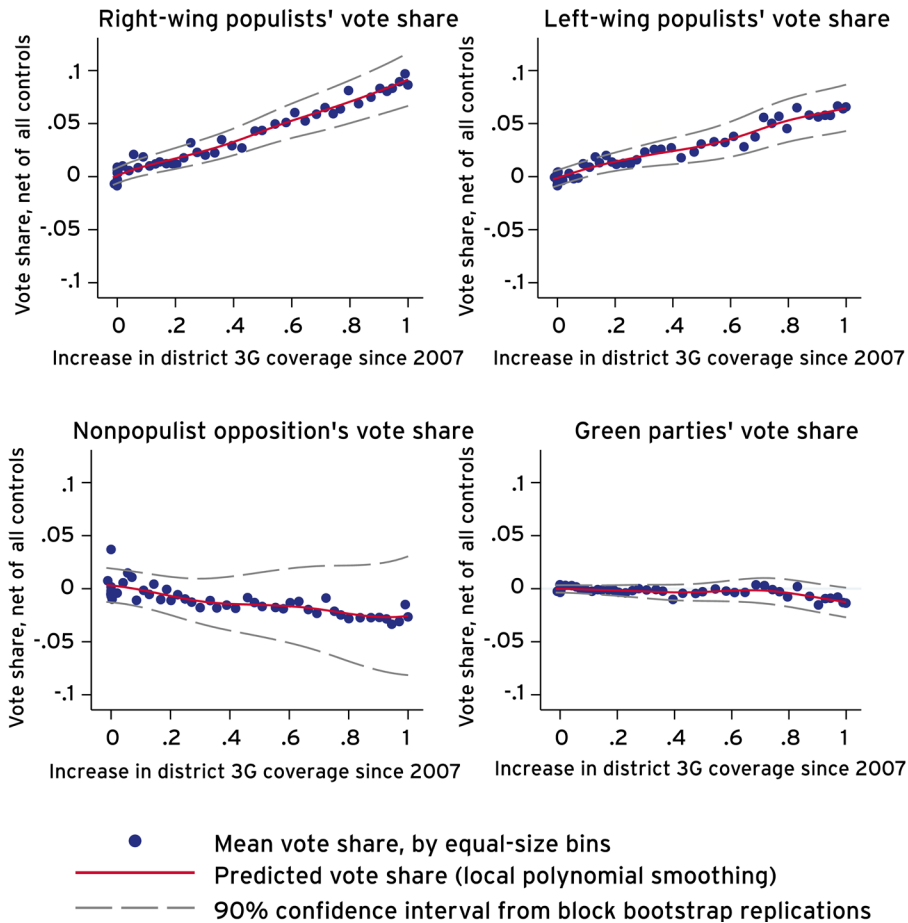
The negative average global effect implies that the negative updates of citizens' views in countries with intermediate corruption outweigh the positive ones in clean countries. This is puzzling, a priori. Indeed, if there is no systematic bias in the information received via new information technology and in the ex-ante beliefs of the public, then the negative and positive updates by the Bayesian public would balance each other out. There are two potential explanations for this puzzle. First, if social media is more conducive to disseminating negative information about the status quo no matter how good the government actually is, and the public is unaware of this asymmetry, one should expect the average effect of mobile broadband expansion on government approval to be negative. Second, if the public's ex-ante views are biased upwards – for example, because the mainstream elite-controlled media overstated the benefits of the status quo before the arrival of social media, as argued by Martin Gurri – an increase in transparency due to the mobile broadband expansion should also, on average, result in a downward shift in government approval.

Our results do indicate that mobile broadband is a 'liberation technology' in those places where the main grievances are related to corruption. However, we also find that 3G and 4G internet has helped populists in Europe. We focus on Europe for two reasons. First, European parliamentary elections are generally comparable in terms of political and electoral systems and integrity of elections. Second, European political parties are very well studied, and there is a conventional classification of European parties into populist and non-populist ones. We collect data on electoral outcomes in 398 subnational districts in 33 European countries (the EU28 plus Liechtenstein, Montenegro, Northern Macedonia, Norway and Switzerland) between 2007 and 2018, covering 102 parliamentary elections.

Between 2007 and 2018, the average subnational region in Europe experienced a major increase in mobile broadband coverage, with the share of the population with access to mobile broadband internet increasing from 37% to 90%. We show that this expansion of 3G and 4G internet in European regions led to a 4.7 percentage point decrease in the vote share of the incumbent party and increased the vote share for right-wing populists by 4.6 percentage points and of left-wing populists by 3.6 percentage points. This accounts for roughly half of the rise in populist vote share in Europe during this period (Guriev and Papaioannou 2022 and Figure 5). Interestingly, we find that the expansion of mobile broadband internet did not affect the vote share of the non-populist opposition, such as, for instance, green parties. (See Figure 5.)



**FIGURE 5 ELECTORAL IMPLICATIONS OF THE EXPANSION OF 3G AND 4G COVERAGE FOR THE OPPOSITION PARTIES**



Note: The dots represent the vote shares, net of all controls, by equal-size bins. The solid lines show the predicted vote shares (Gaussian kernel, local polynomial smoothing). The 90% confidence intervals are constructed by performing a block bootstrap at the level of the clusters.

Our results on European elections support the view that mobile broadband internet and social media are conducive to the spread of the populist narrative. This may be explained by several factors (Zhuravskaya et al. 2020). First, social media drastically reduces barriers to entry for politicians – which is especially beneficial for anti-establishment political movements. Second, social media provides non-intermediated and two-way communication with the voters. This is also particularly useful for populist politicians who claim to represent ‘ordinary people’ and who dislike the system of checks and balances imposed by the traditional media. Third, social media is more likely to disseminate simple narratives and critical messages. While our paper does not allow us to dig deeper into the mechanism relating the spread of mobile broadband and the rise of populism, an excellent related study by Manacorda et al. (2022) shows that at least part of this relationship is explained by the ‘political tribalism’ induced by social media

and mobile broadband. In particular, both self-sorting on social media into like-minded communities and social media algorithms that prioritise highly emotional content, due to it being more engaging, result in the spread of highly persuasive xenophobic and anti-immigrant messages, which mobilise right-wing-populist supporters.

A related study by Melnikov (2023) suggests that in the US, the mobile internet-driven increase in political polarisation was partly determined by online propaganda and misinformation.

Overall, both the optimistic view and the pessimistic view of the political implications of the spread of mobile broadband could be right, depending on what the real grievances in the society are. If they are related to government corruption (as in many developing countries), mobile broadband internet, which provides citizens with sources of political information independent of the government, is a crucial transparency tool that empowers the citizens. If, instead, the real grievances are related to immigration, globalisation or austerity, mobile broadband primarily becomes a platform for disseminating anti-establishment and populist views.

## REFERENCES

- Allcott, H and M Gentzkow (2017), “Social Media and Fake News in the 2016 Election”, *Journal of Economic Perspectives* 31: 211–236.
- Bursztyn, L, G Egorov, R Enikolopov and M Petrova (2020), “Social Media and Xenophobia: Evidence from Russia”, CEPR Discussion Paper No. 14877.
- Cage, J, N Hervé, and B Mazoyer (2022), “Social Media Influence Mainstream Media: Evidence from Two Billion Tweets”, CEPR Discussion Paper No. 17358.
- Chen, Y and D Yang (2020), “The Impact of Media Censorship: 1984 or Brave New World?”, *American Economic Review* 109(6): 2294–2332.
- Diamond, L and M F Plattner (2010), “Liberation Technology”, *Journal of Democracy* 21: 69–83.
- Enikolopov, R, M Petrova, and K Sonin (2018), “Social Media and Corruption”, *American Economic Journal: Applied Economics* 10(1): 150–174.
- Enikolopov, R, A Makarin, and M Petrova (2020), “Social media and protest participation: Evidence from Russia”, *Econometrica* 88(4): 1479–1514.
- Fergusson, L and C Molina (2020), “Facebook Causes Protests”, No. 018004, The Latin American and Caribbean Economic Association-LACEA.
- Gurri, M (2018), *The Revolt of the Public and the Crisis of Authority in the New Millennium*, Stripe Press (first edition in 2014).

Guriev, S and E Papaioannou (2022), “The Political Economy of Populism”, *Journal of Economic Literature*.

Guriev, S, N Melnikov and E Zhuravskaya (2021), “3G Internet and Confidence in Government”, *Quarterly Journal of Economics* 136(4): 2533-2613.

Halberstam, Y and B Knight (2016), “Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter”, *Journal of Public Economics* 143: 73–88.

Hatte, S., E. Madinier, and E. Zhuravskaya (2021). “Reading Twitter in the Newsroom: How Social Media Affects Traditional-Media Reporting of Conflicts.” CEPR Discussion Paper No. DP16167

King, G, J Pan and M E Roberts (2017), “How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument”, *American Political Science Review* 111(3): 484-501.

Levy, R (2021), “Social media, news consumption, and polarization: Evidence from a field experiment”, *American Economic Review* 111: 831–70.

Manacorda, M and A Tesei (2020), “Liberation Technology: Mobile Phones and Political Mobilization in Africa”, *Econometrica* 88(2): 533–567.

Manacorda, M, G Tabellini and A Tesei (2022). “Mobile Internet and the Rise of Political Tribalism in Europe”, mimeo, Queen Mary University of London.

Melnikov, N (2023), “Mobile Internet and Political Polarization”, mimeo, Nova SBE.

Muller, K and C Schwarz (2021), “Fanning the Flames of Hate: Social Media and Hate Crime”, *Journal of the European Economic Association* 19: 2131–2167.

Qin, B, D Strömberg and Y Wu (2017), “Why Does China Allow Freer Social Media? Protests versus Surveillance and Propaganda”, *Journal of Economic Perspectives* 31(1): 117–40.

Tufekci, Z (2018), “How Social Media Took us from Tahrir Square to Donald Trump”, *MIT Technology Review*.

Vosoughi, S, D Roy and S Aral (2018), “The Spread of True and False News Online”, *Science* 359: 1146–1151.

Zhuravskaya, E, M Petrova and R Enikolopov (2020), “Political Effects of the Internet and Social Media”, *Annual Review of Economics* 12: 415–438.

## ABOUT THE AUTHORS

**Sergei Guriev** joined Sciences Po as a tenured professor of economics in 2013 after running the New Economic School in Moscow in 2004-13. In 2016-19, he has been on leave from Sciences Po to serve as the Chief Economist of the European Bank for Reconstruction and Development (EBRD). In 2022, he was appointed Provost of Sciences Po. Since 2017, he has been a member of the Executive Committee of the International Economic Association. He is also a Research Fellow at the Centre for Economic Policy Research, London. He is a Global Member of the Trilateral Commission, Senior Member of the Institut Universitaire de France, Ordinary Member of the Academia Europaea, and an Honorary Foreign Member of the American Economic Association.

**Nikita Melnikov** is an Assistant Professor of Economics at Nova School of Business and Economics. He got his PhD from Princeton University in 2022. Nikita's research primarily focuses on political economy, media economics, and development economics.

**Ekaterina Zhuravskaya** is a Professor of Economics at the Paris School of Economics (EHESS), which she joined in 2010. She is a Research Fellow of CEPR in Political Economy and Development Economics programs. She got her PhD at Harvard University in 1999 and spent 10 subsequent years working as Professor at the New Economic School and as the Academic Director of the Center for Economic and Financial Research in Moscow. Her main academic interests are in political economy, with applications to media, economic history, and institutions. She is the recipient of the Birgit Grodahl Award and CNSR Silver Medal. She also was elected a Fellow of the Econometric Society.



# CHAPTER 6

## Mobile internet and the rise of communitarian politics

**Marco Manacorda,<sup>a,c</sup> Guido Tabellini<sup>b,c</sup> and Andrea Tesei<sup>a,c</sup>**

<sup>a</sup>Queen Mary University of London; <sup>b</sup>Bocconi University; <sup>c</sup>CEPR

Political conflict in modern democracies increasingly revolves around the cleavage between communitarian and universalist values. Communitarians emphasise loyalty to traditional communities and distrust of strangers, while universalists support generalised altruism and openness towards others, irrespective of their social or geographical distance. These different views of the world are systematically correlated with opposite political opinions – and ensuing differences in political demands – on issues like immigration, nationalism, civil rights and the composition of government spending (Enke 2020, Haidt 2012).

In Europe, support for communitarian parties has roughly doubled over the past 15 years. These parties prioritise the interests of insiders, such as the native-born population and those sharing prevalent cultural traits, at the expense of outsiders, such as foreigners, migrants and minorities. This same period has also been marked by the rapid diffusion of 3G and 4G mobile technologies, and the associated widespread use of mobile internet and social media. It is therefore natural to ask whether increased access to mobile internet is responsible for the success of communitarian parties and, if so, why.

In this chapter, which builds on our recent research work (Manacorda et al. 2023), we argue that this is the case. Our analysis draws on a rich literature in social psychology that suggests that social media promotes a heightened sense of, and make users particularly responsive to, messages that prime in-group identity and out-group derogation. This is in part the result of the highly competitive ‘attention economy’ in which digital products and services operate (Williams 2018, Wu 2017). Specifically, an influential body of work shows that, on the one hand, content characterised by moral outrage, fear and animosity is particularly effective in capturing users’ attention and driving online engagement (Berger and Milkman 2012, Brady et al. 2017, Vosoughi et al. 2018), thus creating strong incentives for both users and social media platforms to produce and disseminate such content. On the other hand, such animosity is better directed towards outgroup members, as targeting them offers greater returns in terms of online reach (Rathje et al. 2021) and poses a lower risk of offline retaliation (Crockett 2017).

Building on these insights, we argue that access to mobile internet and social media – and the associated exposure to an online environment where the in-group versus out-group distinction is hyper-salient – makes voters more communitarian in their policy views, increasing their distrust and intolerance of strangers and enhancing nationalist tendencies. By exacerbating communitarian attitudes in the population and by making political messages that capitalise on distrust of others particularly palatable to users, this technology in turn enhances the effectiveness of protectionist and nationalist propaganda and contributes to the electoral success of communitarian parties located on the right of the political spectrum.

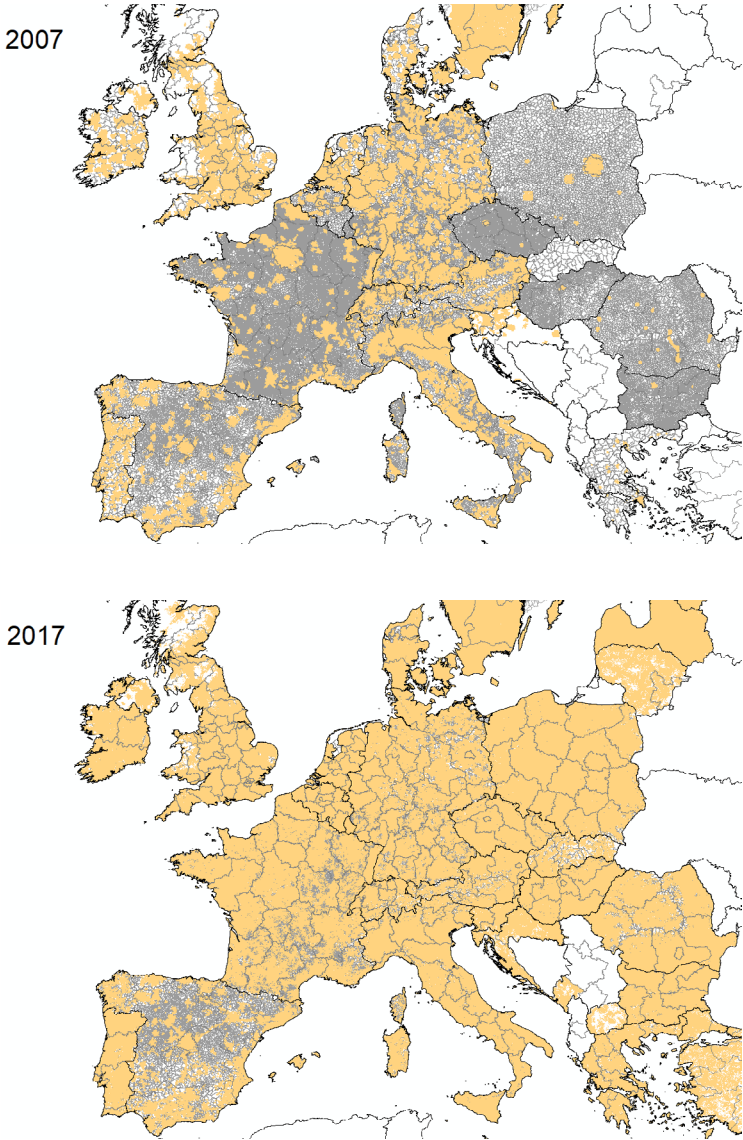
Our work contributes to the debate on the electoral consequences of mobile internet technology in Europe. The closest contribution is the influential work by Guriev et al. (2021), who show that, by exposing misgovernance and corruption, 3G and 4G mobile availability reduces voters' confidence in government and reduces the incumbent's chances of re-election, favouring the rise in support for populist parties in Europe. Our study advances the debate in two significant ways. First, we examine the communitarian platforms of political parties rather than their populist rhetoric or ideology. Specifically, we demonstrate that communitarianism and populism, while connected, are distinct phenomena. More significantly, we provide evidence that the effects of mobile internet technology are present for all communitarian parties, irrespective of whether they are classified as populist or not. Second, we analyse a novel mechanism of impact that is associated with shifts in voters' policy preferences and ideology, which stems from the tendency of this technology to promote tribalistic attitudes among users. In this regard, our work is closely related to that of Müller and Schwarz (2021), Bursztyn et al. (2019) and Bursztyn et al. (2020), who argue that exposure to social media can exacerbate intolerance towards minorities by changing individual attitudes and increasing their willingness to publicly express previously unacceptable social norm positions.

### **ESTIMATING THE POLITICAL CONSEQUENCES OF ACCESS TO MOBILE INTERNET AND SOCIAL MEDIA**

To perform our analysis, we draw on rich granular data on mobile signal availability and administrative data on votes by party in national parliamentary elections between 2007 and 2017, combined with data on party platforms from the Chapel Hill Expert Survey (CHES) that we use to identify communitarian parties. We focus in particular on six dimensions of parties' policy positions and ideologies that, broadly speaking, refer to support for a closed as opposed to an open society. These include positions on civil rights, immigration, multiculturalism, respect of minorities and attitudes towards supranational organisations. We rescale all variables in CHES so that higher values correspond to more communitarian positions. For each dimension, we define communitarian parties as those in the top decile of the respective continent-wide distribution.

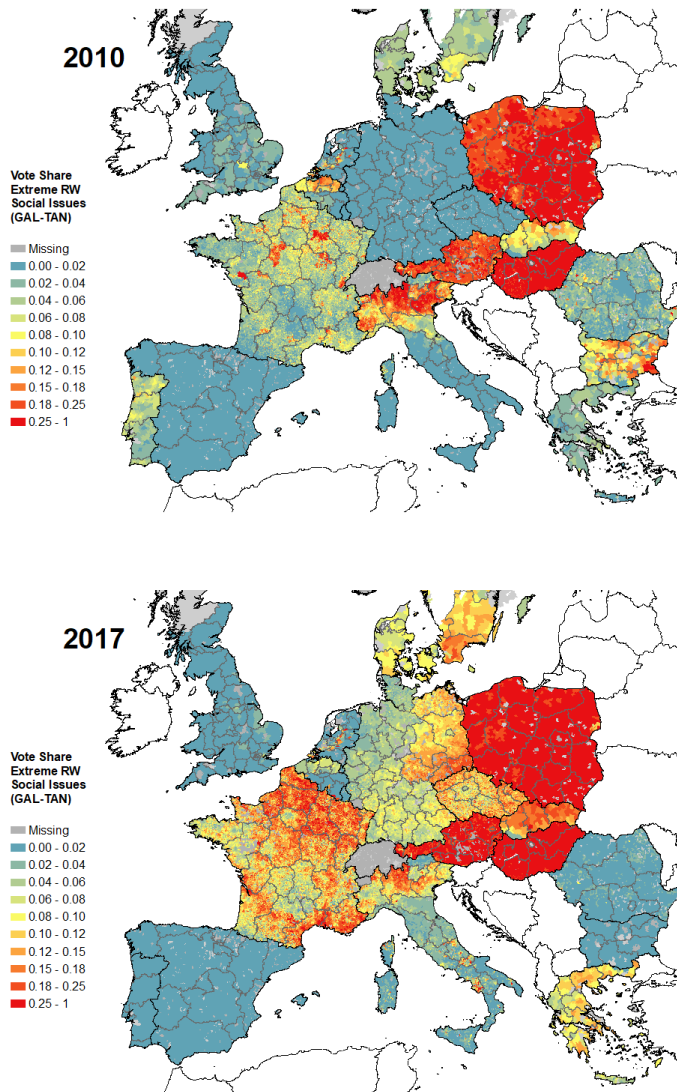
**FIGURE 1 TRENDS IN MOBILE INTERNET COVERAGE AND SUPPORT FOR COMMUNITARIAN PARTIES ACROSS EUROPEAN MUNICIPALITIES**

a) 3G and 4G mobile internet coverage





## b) Vote share for communitarian parties (GAL-TAN)

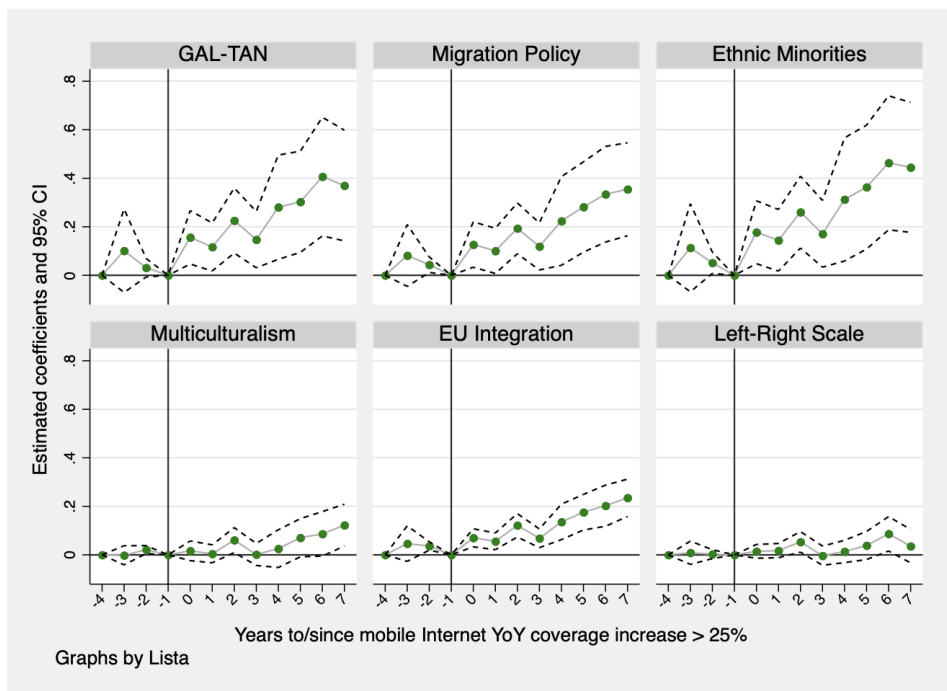


Our sample consists of 82,094 municipalities in 20 major European countries, accounting for almost half a billion people and 96% of the EU27 population. Figure 1 illustrates the level of granularity afforded by our data. The top and bottom panels show, respectively, the evolution of mobile internet coverage and vote share for communitarian parties between the beginning and the end of our sample period. The figure shows evidence of a pronounced increase in coverage both across and within countries, as well as a significant increase in support for communitarian parties throughout the continent and pronounced within-country variation. Notable trends include increased support for the Alternative for Germany (*Alternative für Deutschland*, or AfD) party in the former East Germany,

for the National Rally (*Rassemblement National*) party in the southern and northeastern peripheral areas of France, and for the Swedish Democrats (*Sverigedemokraterna*) in the southern regions of the country.

In order to identify the effect of mobile internet on voting outcomes, we exploit the differential variation in mobile internet coverage and electoral results across municipalities in the same country. Figure 2 presents graphical evidence based on an event-study design. We focus on the first year when a municipality experienced an increase in coverage of at least 25 percentage points over two consecutive years. We present coefficients from regressions of each outcome variable on indicators for different lags and leads (from -4 to +7) from the time when the change occurred. Independent of the outcome variable used, one can observe an increase in voting for communitarian parties occurring precisely in the year when the discrete increase in coverage occurs, with a positive gradient as time goes by. Importantly, the estimates also clearly indicate modest, and by and large statistically insignificant, pre-trend estimates up to lag -1. Ordinary least squares (OLS) regression estimates, which take into account a variety of area characteristics and cover all municipalities in the sample, support the previous findings. According to these estimates, the increase in mobile internet coverage across the continent during the sample period (from 68% to 97% of the population) is responsible for a rise in communitarian vote of between 0.4 and 1.3 percentage points.

**FIGURE 2** EVENT-STUDY ANALYSIS: CHANGES IN SUPPORT FOR COMMUNITARIAN PARTIES IN RESPONSE TO AN INCREASE IN MOBILE INTERNET COVERAGE



As one might be concerned about the possibly non-random allocation of coverage across municipalities, we also present two-stage least squares (2SLS) estimates based on a novel identification strategy that draws on insights from the literature in corporate finance. A number of studies in this area suggest that managers often pursue projects that benefit themselves, even if they are not in the best interests of the company. Décaire and Sosyura (2021), for example, provide evidence that CEOs in the US natural resources industry push for corporate investment projects in areas close to their residential properties, in order to gain direct returns in terms of property value appreciation. Closely related, an influential body of research shows that managers extract personal rents when the firm or the sector is performing well for reasons beyond the managers' control, due to shareholders' inattention that creates room for managers' slack (Bertrand and Mullainathan 2000, 2001). Based on these arguments, we thus instrument coverage in each municipality with its distance from the nearest birthplace of a telecommunications manager – where, in line with Décaire and Sosyura (2021), one would expect managers to benefit from over-investment – interacted with the country's growth rate in mobile phone coverage (a measure of sectoral demand growth affecting shareholders' attention). In line with insights from the corporate finance literature, we find strong first-stage evidence that municipalities closer to managers' birthplaces receive greater coverage, and that this effect is particularly pronounced during periods of higher sectoral growth.

A potential concern with this identification strategy is that the instrument may capture spurious trends in coverage and electoral outcomes along dimensions associated to a municipality distance from managers' birthplaces, as the latter tend to be larger and more affluent than the average country municipality. We take this concern to heart and provide a wealth of evidence corroborating our identification assumption. We show in particular that our findings remain unchanged after controlling for the log-distance to the largest municipality in the country, in the NUTS2 region and NUTS3 province, all interacted with year dummies. We also show that our results are similar when we increasingly restrict our sample to municipalities for which the nearest manager's birthplace is a small city (below 4,000 inhabitants). Taken together, these results suggest that our instrument is not simply capturing the distance to large urban centres, providing support for the exclusion restriction and underscoring the robustness of our empirical findings.

Our 2SLS results confirm the positive effect of mobile internet on the electoral success of communitarian parties in Europe. In particular, our estimates imply that the increase in mobile internet coverage in our sample period is responsible for an increase in the vote share of right-wing communitarian parties of between 2 and 3.4 percentage points – approximately one third to one half of their overall increase over the period – depending on the measure of communitarianism used. We also show that the effect is amplified by economic discontent, lower education and a large pool of users, as proxied by a greater share of younger individuals.

In sum, our analysis at the administrative level provides compelling evidence, through different identification strategies, that the spread of mobile internet has been a significant driver of the rise in communitarian parties vote shares across Europe. Our favoured interpretation for why this occurred, consistent with claims in the social psychology literature, is that access to mobile internet and social media shifted voters' opinions and policy preferences towards more communitarian positions, by over-exposing them to messages that strengthen in-group bias and animosity against out-groups, and by making political messages that capitalise on distrust of others particularly palatable to users.

### **CHANGES IN VOTERS' ATTITUDES AND ALTERNATIVE MECHANISMS**

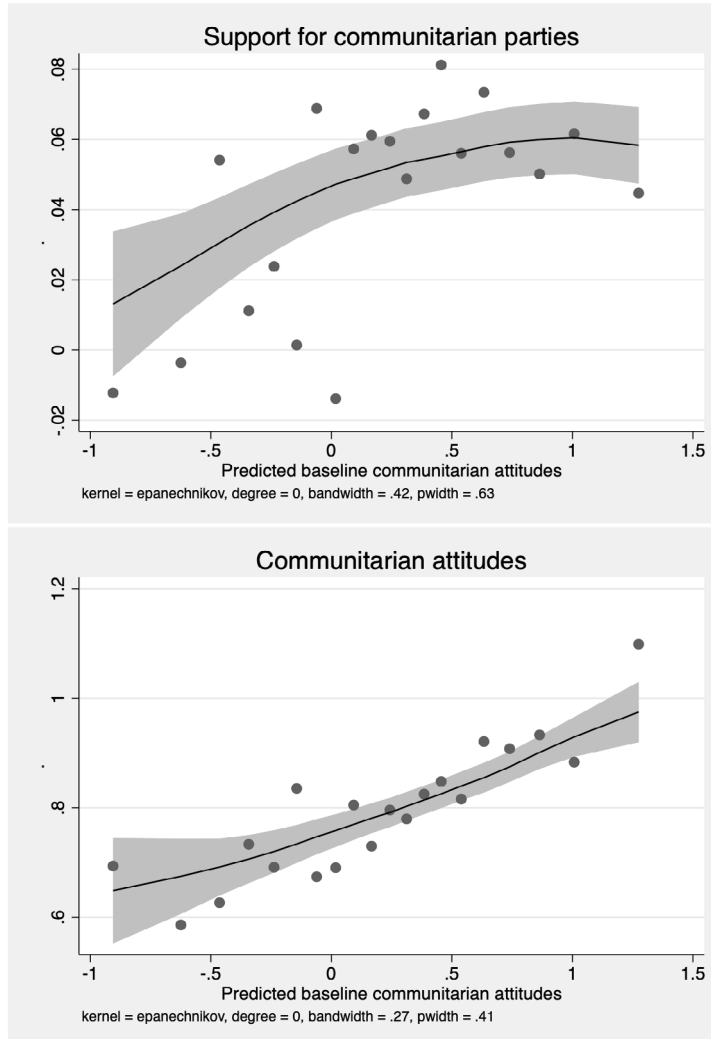
To corroborate this interpretation, we use individual-level data from the Integrated Value Surveys (IVS) to study the effects of mobile internet coverage on policy opinions and voting intentions. The IVS data report information on respondents' voting intentions as well as their ideological stance on a variety of policy issues, including measures of nationalism, attitudes towards individual freedoms and civil rights and identification with the local as opposed to the global community. Additionally, we consider variables capturing voters' attitudes towards migration, intolerance towards minorities, opposition to EU enlargement and ideological leaning on the left-right scale. We also compute a synthetic measure of communitarianism that is the principal component of all the above variables in IVS.

We start by showing that voters' preferences on these issues are predictive of their vote. To achieve this, we construct an indicator variable for each respondent in the IVS, which equals one if the individual's closest party is labelled communitarian according to CHES, and then regress this variable on voters' attitudes. We observe a strong congruence between voters' preferences and party platforms. Perhaps not surprisingly, communitarian voters are more conservative in their views, identifying themselves with the local community, showing distrust of strangers, resisting multiculturalism and opposing the extension of rights to ethnic and social minorities. This confirms that the communitarian dimensions of parties were salient to voters and explained their voting patterns.

Then, we show that mobile internet and social media had a strong effect in shifting voters' opinions towards more communitarian positions. Our OLS estimates display a significant and substantial effect on levels of communitarian attitudes, with increased mobile internet coverage over the period, leading to implied increases in communitarian positions ranging from 0.09 to 0.28 standard deviations, depending on the measure employed. Specifically, our findings indicate that mobile internet access made respondents more resistant to immigration, more nationalist, more intolerant of minority groups and less supportive of EU institutions. Unsurprisingly, we also find positive and precisely estimated effects of mobile internet on the synthetic measure of communitarianism.

Lastly, we delve into why extreme communitarian parties gained the most from these changes in voters' preferences. We illustrate this in Figure 3, where we explore which voters were most impacted by exposure to new technologies. The figure reports the estimated effect of access to mobile internet on both opinions and voting, separately by vingtiles of voters' predicted baseline communitarianism.

**FIGURE 3 THE EFFECT OF MOBILE INTERNET ON SUPPORT FOR COMMUNITARIAN PARTIES AND ATTITUDES: HETEROGENEOUS EFFECTS BY BASELINE COMMUNITARIANISM**



Two results emerge from the figure. First, differential changes in voting outcomes across groups closely resemble differential changes in opinions, reinforcing our hypothesis that the changes in voters' views brought about by new technology caused a shift in their

voting behaviour. Second, the rise in both communitarian attitudes and in support for communitarian parties is larger among those with higher baseline communitarianism levels. Since most of the rise in communitarian attitudes was concentrated at the top of the distribution, with marginally communitarian voters embracing extreme positions, extreme communitarian parties were better positioned than their more moderate counterparts to intercept these new extremist tendencies among the electorate.

We also investigate and discard numerous alternative mechanisms that could account for our findings. First, as previously mentioned, we show that the positive impact of mobile internet on voting is present for all communitarian parties, regardless of whether they are classified as populist or not. This suggests that the influence of mobile internet on voting that we unveil extends beyond particular characteristics of populism, such as its rhetoric or ideology. Second, we find no indications that mobile internet led to higher voter turnout or affected voting outcomes by altering the composition of the electorate.

Third, we also find no evidence that mobile internet favoured new political parties per se. This suggests that the effect that we estimate is not due simply to the circumstance that communitarian parties were newer than other parties, making them more capable of utilising new technologies or more appealing to voters when exposed to social media. Finally, due to our sample period preceding that of widespread circulation of fake news on social media, we also rule out that the effects we uncover are due to online misinformation.

## CONCLUSIONS

The widespread availability of mobile internet has brought about significant changes in social interactions, facilitating the diffusion of social media and transforming the mode and substance of political communication. Our recent work (Manacorda et al. 2023) reveals that one of the crucial political consequences of these technologies has been to boost support for parties with extreme right-wing and communitarian stances on social and cultural issues, a result that accords with a substantial body of evidence from social psychology on the impacts of social media. Whether or not this effect will persist in the future is a subject of debate.

On one hand, our analysis covers a period of extremely fast growth in mobile internet coverage and associated use of smartphones and social media. It is possible that the effects we detect are linked to the rapid transition to these technologies and that they will not persist over time. Moderation of content and fact-checking by online platforms could offset the tendency of these technologies to promote tribalistic attitudes. Furthermore, the time period we investigate is unique in many aspects, not only due to the emergence of social media. Other factors, such as the pressures of globalisation, immigration and labour-saving technologies, have also fuelled discontent and contributed to changes in voting patterns. Consistent with our findings that the effect of mobile internet and social media is larger in more economically and socially deprived areas, it is possible that our

results capture the effect of these new technologies at this specific economic juncture and that their impact would have been different in the absence of such major economic transformations.

On the other hand, it is also possible to speculate that some aspects of the economic model underlying these technologies, such as the high returns associated with capturing users' attention and the associated inclination to reward divisive content, will continue to encourage the dissemination of negative content aimed at minorities and out-group members. Specifically, emotionally charged topics such as immigration and the emergence of multicultural societies in a globalised world may continue to spread easily online and fuel the in-group versus out-group backlash and the resulting political demands that we have observed. Which of these different dimensions will prevail, as well as the potentially entirely new political effects associated with the emergence of AI technologies, are critically important questions that feature prominently in our future research agenda.

## REFERENCES

- Berger, J and K L Milkman (2012), "What Makes Online Content Viral?", *Journal of Marketing Research* 49 (2): 192–205.
- Bertrand, M and S Mullainathan (2000), "Agents with and without Principals", *American Economic Review* 90 (2): 203–208.
- Bertrand, M and S Mullainathan (2001), "Are CEOs Rewarded for Luck? The Ones without Principals are", *The Quarterly Journal of Economics* 116 (3): 901–932.
- Brady, W J, J A Wills, J T Jost, J A Tucker, and J J Van Bavel (2017), "Emotion shapes the Diffusion of Moralized Content in Social Networks", *Proceedings of the National Academy of Sciences* 114 (28): 7313–7318.
- Bursztyn, L, G Egorov, R Enikolopov, and M Petrova (2019), "Social Media and Xenophobia: Evidence from Russia", unpublished manuscript.
- Bursztyn, L, G Egorov, and S Fiorin (2020), "From Extreme to Mainstream: The Erosion of Social Norms", *American Economic Review* 110 (11): 3522–48.
- Crockett, M J (2017), "Moral Outrage in the Digital Age", *Nature Human Behaviour* 1(11): 769–771.
- Décaire, P and D Sosyura (2021), "CEO Pet Projects", unpublished manuscript.
- Enke, B (2020), "Moral Values and Voting", *Journal of Political Economy* 128(10): 3679–3729.
- Guriey, S, N Melnikov, and E Zhuravskaya (2021), "3G Internet and Confidence in Government", *The Quarterly Journal of Economics* 136(4): 2533–2613.

Haidt, J (2012), *The Righteous Mind: Why Good People are divided by Politics and Religion*, Vintage.

Manacorda, M, G Tabellini, and A Tesei (2023), “Mobile internet and the rise of communitarian politics”, CEPR Discussion Paper No. 18063.

Müller, K and C Schwarz (2021), “Fanning the Flames of Hate: Social Media and Hate Crime”, *Journal of the European Economic Association* 19 (4): 2131–2167.

Rathje, S, J J Van Bavel and S van der Linden (2021), “Out-group Animosity Drives Engagement on Social Media”, *Proceedings of the National Academy of Sciences* 118(26).

Vosoughi, S, D Roy, and S Aral (2018), “The Spread of True and False News Online”, *Science* 359 (6380): 1146–1151.

Williams, J (2018), *Stand out of our light: Freedom and resistance in the attention economy*, Cambridge University Press.

Wu, T (2017), *The attention merchants: The epic scramble to get inside our heads*, Vintage.

## ABOUT THE AUTHORS

**Marco Manacorda** is a Professor of Economics and Public Policy at Queen Mary University of London, a CEPR Research Fellow and a managing co-editor of *The Economic Journal*. His interests and expertise are at the intersection of political economy, labour and development economics.

**Guido Tabellini** is the Intesa SanPaolo Professor of Political Economics at Bocconi University, where he was Rector between 2008 and 2012. Before returning to Europe, he taught at Stanford and UCLA. He is a foreign honorary member of the American Academy of Arts and Sciences, a fellow of the Econometric Society, a Foreign Honorary Member of the American Economic Association, he received the Yrjo Jahnsson award by the European Economic Association, and he has been President of the European Economic Association.

**Andrea Tesei** is Associate Professor of Economics at Queen Mary University of London and a CEPR Research Affiliate. His research interests are in political economy and economic development. His research has been published in *The American Economic Review*, *Econometrica* and *The Review of Economics and Statistics*, among others.





# CHAPTER 7

## The effect of social media on elections: Evidence from the United States

**Thomas Fujiwara, Karsten Müller and Carlo Schwarz**

Princeton University; National University of Singapore; Bocconi University

Can social media affect election outcomes? A popular narrative holds that Twitter played a decisive role in both the 2016 presidential election in the US and the UK's Brexit referendum. Both Hillary Clinton and Donald Trump have argued that these factors were instrumental in the 2016 election outcome, as has Barack Obama (*The New Yorker* 2016). As Brad Parscale, Trump's digital media director in 2016, put it: "Facebook and Twitter were the reason we won this thing. Twitter for Mr. Trump. And Facebook for fundraising" (Wired 2016). In an interview with CBS News' *60 Minutes*, Trump himself declared: "I think I wouldn't be here if I didn't have social media."

Many see these statements as evidence of social media's broader influence on political polarisation and the emergence of right-wing populist politicians in many countries. US Federal Election Commissioner Ellen Weintraub, for example, has argued that Facebook "has no idea how seriously it is hurting democracy" (NPR 2020).

However, an alternative view holds that social media platforms are biased against conservative voices (e.g. *Wall Street Journal* 2020). In the 2016 presidential election, Trump notably received fewer votes from demographic groups with a higher propensity to use social media or the internet (Boxell et al. 2018). Indeed, Trump's broadest support came from older white voters without a college education in rural areas, who are among the least likely people to use social media actively (Hargittai 2015). Further, the content on social media platforms – particularly on Twitter – is disproportionately left-leaning. The Pew Research Center estimates that in 2018, 60% of American Twitter users identified as Democrat and only 35% as Republican, and 80% of Twitter users strongly disapproved of President Trump in 2019 (Pew Research Center 2019a, 2019b).

To evaluate such contrasting claims about the role of social media, we need estimates of its causal effect on election outcomes, something there is scarcely any evidence on.<sup>1</sup> The key challenge in estimating the causal effect of social media is that who uses them is far

1 On the other hand, there is a sizeable literature on the political effects of 'traditional' media, such as television, newspapers or radio, on election outcomes (see our paper, Fujiwara et al. 2021, for several references). There is also work on the effects of broadband and 3G internet access (e.g. Zhuravskaya et al. 2020). To the best of our knowledge, the only other evidence on the effects of social media on elections comes from Enikolopov et al. (2020), who mainly focus on protest participation. Experiments based on incentivising Facebook users to temporarily abstain from using it also find effects on self-reported political attitudes (Alcott et al. 2020, Mosquera et al. 2020).

from random. As such, an observed correlation between social media usage and voting patterns could also be the result of third factors, such as age or technology affinity, that are difficult to control for.

In recent work (Fujiwara et al. 2021), we tackle this challenge by studying the effect of Twitter on US federal elections. Specifically, we estimate the effect of Twitter's popularity at the county level. More precisely, we ask if increasing the number of platform users in a county affects its electoral results. To overcome concerns that counties with high Twitter usage also differ along other relevant characteristics, we exploit quasi-random variation in early Twitter adoption caused by the 2007 South by Southwest festival (SXSW), which left an imprint on the geography of Twitter usage until today.

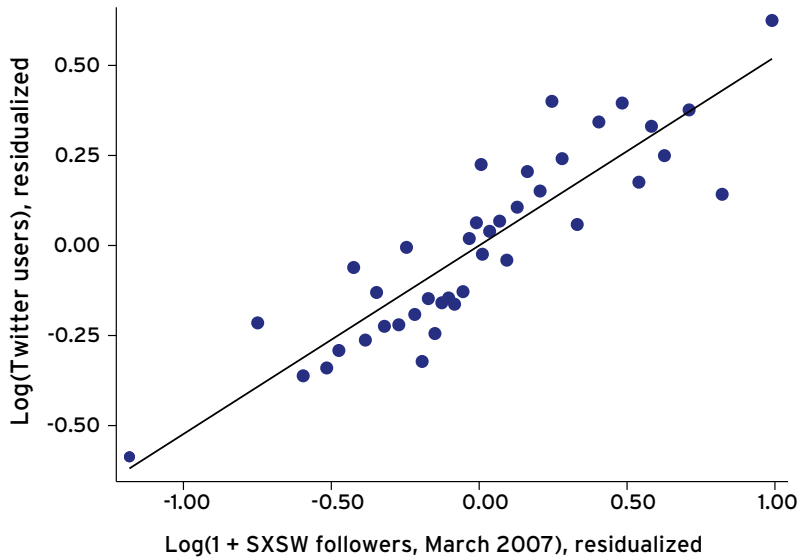
### **THE 2007 SOUTH BY SOUTHWEST FESTIVAL AND THE RISE OF TWITTER**

After its launch in March 2006, Twitter's popularity grew slowly. The turning point was an advertising campaign at the SXSW festival in March 2007 that allowed users to sign up to Twitter by sending a short text message with screens in the hallways showing tweets about the festival. This campaign jump-started the adoption of Twitter and quadrupled the platform's growth rate, spreading its popularity across the US.

Because network effects are key for social networks (that is, potential users prefer to join platforms that other people they know have joined), the campaign at SXSW influenced the geography of Twitter usage in the US. The early adopters at SXSW in 2007 spread the news about Twitter to their home counties, which led to an increase in sign-ups and Twitter activity there (Müller and Schwarz 2019). Even a decade later, Twitter still had more users in counties with relatively more early adopters at SXSW, highlighting the path dependency of social media adoption.

To measure the presence of these early Twitter adopters, we collected data from the Twitter page of the SXSW festival. For each county, we calculated the number of people following the SXSW Twitter account who joined the platform exactly in March 2007, the month of the festival. Figure 1 shows that the number of these SXSW followers has a strong positive association with the overall number of Twitter users in that county in 2014-2015. Importantly, only early adoption at SXSW in March 2007, but not previous interest in SXSW's account on Twitter, predicts Twitter usage as of 2014-2015. This suggests that the strong correlation between Twitter adopters at SXSW does not simply capture differences between counties depending on how likely they were to adopt Twitter initially or new technologies generally.

**FIGURE 1** EARLY ADOPTION AT SXSW PREDICTS FUTURE TWITTER USAGE



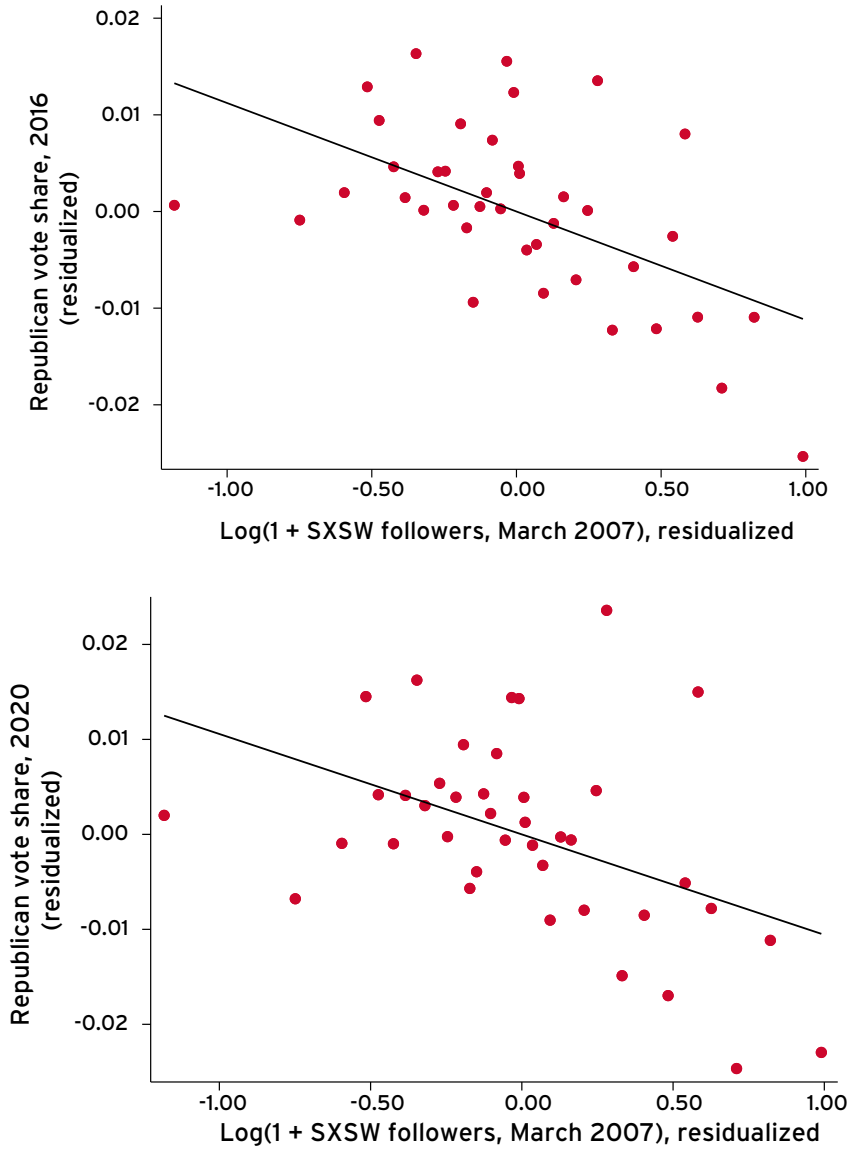
Note: This figure presents a binned scatter plot of the relationship between Twitter users in 2014-2015 and the number of SXSW followers who joined Twitter in March 2007. Variables are residualised by partialling out SXSW followers who joined before March 2007, population deciles, Census region fixed effects, as well as geographical, demographic, socioeconomic, China shock, and 1996 election control variables. The figure is constructed by dividing the x-axis variable into 40 equal-sized bins and plotting the average values of both variables in each bin. The fitted line is estimated using the unbinned data.

We use this initial shock to Twitter adoption to estimate the causal effect of Twitter usage on election outcomes. As we show in our paper, the shock to Twitter adoption induced by the SXSW festival is essentially uncorrelated with levels and trends in election outcomes before Twitter’s launch, as well as a host of county-level characteristics, conditional on geographic and socioeconomic controls and previous interest in SXSW on Twitter (measured via the number of SXSW followers that joined the platform before March 2007). In other words, Twitter adoption due to SXSW appears to be unrelated to any other characteristic that may influence election outcomes. This is what allows us to plausibly separate the effect of Twitter from other potential explanations.

## THE ELECTORAL EFFECTS OF TWITTER

Based on the differences in Twitter usage created by the SXSW festival, we investigate the impact of Twitter on the 2016 and 2020 US elections. We find that a 10% increase in a county’s number of Twitter users lowered the vote share of Republican presidential candidate Trump by 0.2 percentage points, both in the 2016 and 2020 elections. Figure 2 shows the negative relationship between the number of SXSW followers that joined in March 2007 and Trump’s vote share.

**FIGURE 2 TWITTER EARLY ADOPTERS AND THE REPUBLICAN VOTE SHARE**



Note: Panel A presents a binned scatter plot of the relationship between the Republican vote share in the 2016 presidential election and the number of SXSW followers who joined Twitter in March 2007. Variables are residualised with respect to SXSW followers who joined before March 2007, population deciles, Census region fixed effects, as well as geographical, demographic, socioeconomic, China shock, and 1996 election control variables.

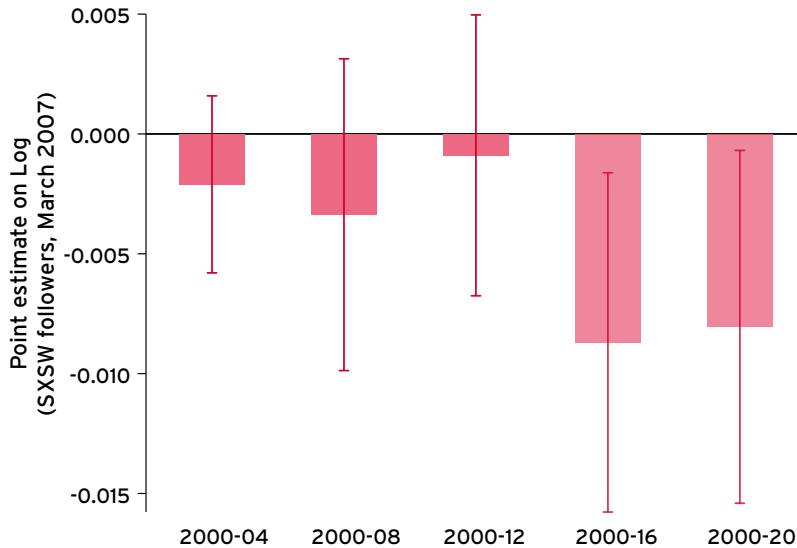
We benchmark our findings to other studies on the electoral effects of media by calculating what are known in this academic literature as ‘persuasion rates’. The implied persuasion rate of the Twitter effect we find is roughly 9%. This means that if 100 voters planning to vote Trump were exposed to Twitter, nine of them would change their mind and switch their vote. This persuasion rate is smaller than the estimated pro-Republican effect of Fox News (DellaVigna and Kaplan 2007, Martin and Yurukoglu 2017), the

pro-Democrat effect of the *Washington Post* (Gerber et al., 2009) or the effect of ‘get out the vote’ canvassing on turnout (Gerber and Green, 2000), but larger than the effect of an independent anti-Putin Russian TV channel on vote shares (Enikolopov et al. 2011) or the effect of TV rollout on turnout (Gentzkow 2006).

We confirm our findings using individual-level voting decisions measured in survey data from the 2016 Cooperative Congressional Election Study (CCES). The CCES moreover allows us to investigate how the Twitter effect differs by voters’ party affiliation. We find that the effects of Twitter on voting decisions are strongest for independents and moderates, who were more likely to change their voting decision in favour of the Democratic candidate (Clinton).

Importantly, the number of SXSW followers that joined in March 2007 is uncorrelated with the Republican vote share in earlier presidential elections. Figure 3 presents the correlations for 2004-2012 (in our paper, we show the results for all elections going back to 1924). This serves as a useful ‘placebo check’: if our results merely captured some unobserved differences across counties, these would likely also be correlated with vote shares before the widespread use of Twitter. Our findings are instead most consistent with the idea that Twitter had a causal effect on the 2016 and 2020 presidential elections.

**FIGURE 3 CORRELATION OF SXSW FOLLOWING WITH EARLIER PRESIDENTIAL ELECTIONS**



Note: This figure plots reduced form estimates from county-level regressions. They measure the effect of  $\text{Log}(1 + \text{SXSW followers, March 2007})$  on changes in the Republican vote share in presidential elections relative to the year 2000. All regressions control for SXSW followers who joined before March 2007, population deciles, and Census region fixed effects, as well as geographical, demographic, socioeconomic, China shock, and 1996 election results. Regressions are weighted by turnout in the 2000 presidential election. Whiskers represent 95% confidence intervals based on standard errors clustered by state.

Interestingly, the number of early adopters at SXSW 2007 is also uncorrelated with vote shares for House and Senate elections, even in the 2016 and 2020 elections (as well as the 2018 midterm races). Twitter adoption thus lowered Trump's vote share but did not hurt Republican candidates in congressional races on the same ballots on the same election days. This suggests that Twitter particularly mattered for the salient decision of which president to cast the vote for.

There are two takeaways from the results above. First, the negative effects of Twitter on the Republican vote share only appear in the 2016 and 2020 presidential elections. Second, the results bolster confidence in our empirical strategy and that we are likely capturing the causal effect of Twitter, not merely correlations, given that the number of SXSW followers that joined in March 2007 is uncorrelated with levels and trends of election results before 2016. Any alternative explanation would have to explain why an omitted variable is not influencing the outcomes of these other elections as well. The relationship depicted in Figure 2 is thus unlikely to be driven by a factor other than social media usage.

### **WHAT IS SPECIAL ABOUT THE 2016 AND 2020 ELECTIONS?**

We also investigate why we only find an effect of Twitter in the 2016 and 2020 presidential elections. A potential answer is that Trump was an atypical candidate, which may be why exposure to Twitter persuaded moderate voters to not vote for him without negatively affecting other Republican candidates in down-ballot races.

We provide two sets of results that corroborate this hypothesis. The first is that, when we estimate the effect of social media on the presidential primary races of 2016, we find that Twitter lowered support for Trump but had little effect on other Republican candidates (Cruz, Rubio, Kasich, Bush).

The second set of results is based on the analysis of over 460 million tweets mentioning the presidential candidates in the 2012, 2016, and 2020 elections. In particular, we document that political content on Twitter has a pro-Democratic slant. We classify the slant of tweets based on two complementary approaches: one based on the network users follow, and one using the text of tweets in a machine learning approach in the spirit of Gentzkow and Shapiro (2010). We find that the number and attention (proxied by 'likes') of tweets mentioning Trump was substantially larger than that of those mentioning Clinton and Biden. Moreover, tweets about Trump in 2016 and 2020 are 70% more likely to have Democratic rather than Republican slant.

Overall, a potential interpretation of our results is as follows. Users of Twitter (and other social media platforms) are more likely to be young, more educated, live in urban areas, and support the Democratic Party. This pro-Democrat slant did not manifest itself in a pronounced fashion before the 2016 presidential election because the Republican candidates were relatively moderate. In 2016 and 2020, however, Twitter became a

vehicle for spreading criticism of Trump. The sheer volume of this slanted content may have persuaded voters with weaker priors – independents and perhaps more moderate Republicans – to vote against Trump in the presidential election.

## CONCLUDING REMARKS

Overall, our results suggest that social media can affect election outcomes. This has potentially important implications for the ongoing debate on the regulation of social media platforms. Our findings suggest that, despite narratives about the impact of bots and fake news, Twitter was not instrumental for Trump's electoral triumph in 2016. If anything, Trump likely won despite, not because of Twitter. If Twitter indeed played an important role in the 2016 US election in Trump's favour, such an effect would have likely come from its indirect impact on the news media, widely used by journalists, through the extensive coverage Trump was able to create for himself using his inflammatory tweets. If social media platforms can affect the outcome of elections, an important question is which regulations (if any) should be imposed on them. This issue goes beyond elections: there is also evidence that social media can increase the likelihood of hate crimes and political polarisation (Bursztyjn et al. 2019, Müller and Schwarz 2019, Levy 2021).

It is important to highlight some caveats of our findings. First, they are silent on the effect of social media platforms other than Twitter, such as Facebook, because we use a source of variation based on a 'shock' that is specific to Twitter. Second, our research design cannot isolate the effect of particular types of social media content on Twitter, such as the potential role of foreign governments or misinformation ('fake news') and we instead speak to the aggregate effect of Twitter. We hope that future research can address these limitations and provide additional evidence on the effect of social media on political outcomes.

## REFERENCES

- Allcott, H, L Braghieri, S Eichmeyer and M Gentzkow (2020), "The Welfare Effects of Social Media", *American Economic Review* 110(3): 629–76.
- Allcott, H and M Gentzkow (2017), "Social Media and Fake News in the 2016 Election", *Journal of Economic Perspectives* 31(2): 211–36.
- Autor, D, D Dorn, G Hanson, K Majlesi et al. (2020), "Importing political polarization? the electoral consequences of rising trade exposure", *American Economic Review* 110(10): 3139–83.
- Barbera, P (2014), "How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the US", Paper prepared for the 2015 APSA Conference.
- Boxell, L, M Gentzkow and J M Shapiro (2018), "A Note on Internet Use and the 2016 U.S. Presidential Election Outcome", *PLOS One* 13(7): 1–7.



Bursztyn, L, G Egorov, R Enikolopo and M Petrova (2019), "Social media and xenophobia: evidence from Russia", NBER Working Paper No. w26567.

Enikolopov, R, M Petrova and E Zhuravskaya (2011), "Media and Political Persuasion: Evidence from Russia", *American Economic Review* 101(7): 3253–3285

Enikolopov, R, A Makarin, and M Petrova (2020), "Social Media and Protest Participation: Evidence from Russia", *Econometrica* 88(4): 1479–1514.

DellaVigna, S and M Gentzkow (2010), "Persuasion: Empirical Evidence", *Annual Review of Economics* 2(1): 643–669.

DellaVigna, S and E Kaplan (2007), "The Fox News Effect: Media Bias and Voting", *The Quarterly Journal of Economics* 122(3): 1187–1234

Fujiwara, T, K Müller, and C Schwarz (2021), "The Effect of Social Media on Elections: Evidence from the United States", working paper.

Gentzkow, M (2006), "Television and Voter Turnout", *Quarterly Journal of Economics* 121(3): 931–972

Gentzkow, M and J M Shapiro (2010), "What Drives Media Slant? Evidence from US Daily Newspapers", *Econometrica* 78(1): 35–71.

Gerber, A S and D P Green (2000), "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment", *American Political Science Review* 94(3): 653–663.

Gerber, A S, D Karlan, and D Bergan (2009), "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions", *American Economic Journal: Applied Economics* 1(2): 35–52.

Hargittai, E (2015), "Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites", *The Annals of the American Academy of Political and Social Science* 659: 63–76

Levy, R E (2021), "Social media, news consumption, and polarization: Evidence from a field experiment", *American Economic Review* 111(3): 831–70.

Martin, G J and A Yurukoglu (2017), "Bias in Cable News: Persuasion and Polarization", *American Economic Review* 107(9): 2565–2599.

Müller, K and C Schwarz (2019), "From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment", Working Paper.

Mosquera, R, M Odunowo, T McNamara, X Guo, and R Petrie (2020), "The Economic Effects of Facebook", *Experimental Economics* 23(2): 575–602.

NPR (2020), "Facebook Clamps Down On Posts, Ads That Could Undermine U.S. Presidential Election", 3 September.

Pew Research Center (2019a), *About One-In-Five Adult Twitter Users in the U.S. Follow Trump*.

Pew Research Center (2019b), *National Politics on Twitter: Small Share of U.S. Adults Produce Majority of Tweets*.

*The New Yorker* (2016), “Obama Reckons with a Trump Presidency”, 18 November.

*Wall Street Journal* (2020), “Twitter’s Partisan Censors”, 15 October.

*Wired* (2016), “Here’s How Facebook Actually Won Trump the Presidency”, 15 November.

Zhuravskaya, E, M Petrova, and R Enikolopov (2020), “Political Effects of the Internet and Social Media”, Working Paper.

## ABOUT THE AUTHORS

**Thomas Fujiwara** is an Associate Professor of Economics and International Affairs at Princeton University. His research is in political economy, with a substantive focus on developing countries. His work addresses questions related to representation both in elections and in the workplace. In particular, he has investigated how political participation shapes public policy, how electoral institutions affect representation, how voter behavior is shaped by different information and messages, and the role of gender norms in driving female participation in the workplace. He received his PhD from the University of British Columbia in 2011.

**Karsten Müller** is an Assistant Professor in Finance at the National University of Singapore Business School. His research focuses on finance, macroeconomics and political economy. One line of his work studies how politics and financial regulation determine which firms and households get access to credit in the economy and what the macroeconomic consequences may be. His second stream of research shows that social media can affect election outcomes and anti-minority sentiments offline. Before joining NUS, Karsten was a postdoctoral research associate at Princeton University. He earned his PhD in Business and Management from the University of Warwick in 2018.

**Carlo Schwarz** is an Assistant Professor in the Department of Economics at Bocconi University. He holds a PhD in Economics from the University of Warwick. His research interest lies in the field of applied microeconomics with a particular interest in policy-relevant questions. In his research, he combines causal inference strategies with techniques from text analysis, machine learning, and data science. His recent research investigates inequalities in the scientific process and their impact on the creation of knowledge. He also has written papers on the contributed of social media to the spread of hate crimes and anti-minority sentiments.



## **SECTION 3**

# **SOCIAL MEDIA AND POLITICIANS**



# CHAPTER 8

## Politics 2.0: The multifaceted effect of broadband internet on political participation

**Filipe Campante, Ruben Durante and Francesco Sobbrío**

Johns Hopkins University; National University of Singapore, ICREA-Universitat Pompeu Fabra and CEPR; Tor Vergata University of Rome

The internet and social media have dramatically changed the way individuals obtain, produce and exchange information. This transformation has had profound implications for various dimensions of social life, including political participation. The debate over the impact of the internet on political participation has largely been dominated by two alternative views. On the one hand, it has made it easier for citizens to acquire political information from more numerous and diverse sources (Gentzkow-Shapiro 2010). To the extent that better-informed citizens tend to be more politically involved (Feddersen and Pesendorfer 1996, 1999, Lassen 2005), the internet should favour political participation (e.g. Kaye and Johnson 2002, Tolbert and McNeal 2003). On the other hand, the internet offers unprecedented entertainment opportunities, which may crowd out users' consumption of political content, ultimately turning them into less informed and participative citizens (e.g. Prior 2005). These hypotheses are not new in the literature on the impact of media technologies, having previously been studied in relation to the diffusion of newspapers, radio and television (Stromberg 2004 Gentzkow 2006, Gentzkow et al. 2011).

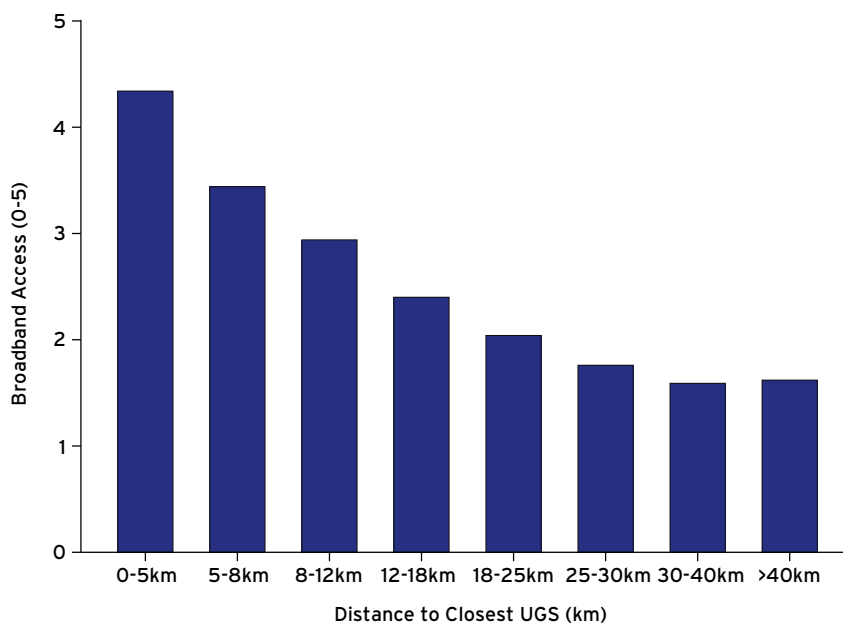
Most early studies on the impact of the internet on political participation (Czernich 2012, Falck et al. 2014, Jaber 2013, Larcinese and Miner 2012, Miner 2012) used the same approach, focusing on the tension between the aforementioned 'information' and 'crowding-out' channels, and looking primarily at the short-run effect on the most evident form of political participation: electoral turnout. However, the internet is fundamentally different from traditional media in ways that ought to be especially important from a political economy perspective. First, internet users not only consume content but provide it as well. In this respect, the internet offers citizens an unparalleled means of expressing their views, compared to TV or radio. Along the same lines, and again unlike traditional media, the internet also offers users a remarkably effective way to interact and coordinate with other people. Because of these features, and particularly with the advent of social media, the internet has brought about new opportunities for political discussion and mobilisation, made accessible to a wide range of political actors and with effects that take time to fully materialise.

For all these reasons, understanding how the internet ultimately affects political outcomes requires a different, broader perspective that considers the impact on different forms of political engagement – offline as well as online – both in the short and the longer run.

In our research (Campante et al. 2018), we employ such a broader perspective to investigate the causal impact of the introduction of high-speed internet on political participation in Italy over almost two decades (1996–2013). The Italian context is well suited for studying this question for several reasons. First, Italy is a country with solid democratic roots but where the mass media have traditionally been controlled by the government or by powerful private interests (Durante and Knight 2012). It is therefore an ideal setting to examine how the internet can help alleviating distortions in traditional media sectors. Second, during the period we analyse, the Italian political landscape, until then dominated by two mainstream coalitions, was shaken by the emergence of a new political actor, the Five Star Movement (M5S), which became the most voted for party in the 2013 national elections. Since the use of social media was one of the cornerstones of M5S' strategy, its rapid emergence provides a perfect opportunity to assess the impact of the internet on politics. Finally, the availability of data for seventeen years and five electoral cycles allows us to study both the short-run and long-run effect of the internet on the dynamics of political participation.

To identify the causal impact of high-speed internet, we exploit differences in the timing of the introduction of broadband technology across Italian municipalities between 2005 and 2011.<sup>1</sup> To deal with endogeneity in access to broadband, we exploit the fact that the diffusion of Asymmetric Digital Subscriber Line (ADSL) technology in a given municipality was affected by its relative position in the pre-existing voice telecommunications infrastructure. Specifically, because ADSL-based internet services could only be offered in municipalities connected to high-order telecommunication exchanges (Urban Group Stages, or UGSs) via fibre optic cable, we use the distance between a given municipality and the closest UGS – a good proxy for the investment required to connect the municipality – as a source of variation in the availability of high-speed internet. Figure 1 confirms that the quality of broadband access decreases steadily with the distance to the closest UGS. Because the pre-existing infrastructure was not randomly distributed, our identification strategy relies on interacting that distance with the time variation between the period before and after broadband became available, under the assumption that the correlation between distance and unobserved municipal characteristics did not change at that point in time, other than through the introduction of high-speed internet.

1 Data on ADSL coverage at the municipal level were provided by Osservatorio Banda Larga-Between, a joint venture between the main Italian telecommunications operators, the Italian Ministry for Telecommunications, and other private and public stakeholders. The data include information on the percentage of households with access to ADSL-based services in each Italian municipality for each year between 2005 and 2011 on an asymmetric six-point scale corresponding to the following brackets: 0%, 1%-50%, 51%-75%, 76%-85%, 86%-95%, and above 95%.

**FIGURE 1 ADSL COVERAGE IN 2006 AND DISTANCE TO CLOSEST UGS**

Note: The figure reports the average ADSL access score (from 0 to 5) in 2006 for municipalities located at various distances from the closest UGS (from 0-5 km to 40 km or more).

Source: Osservatorio Banda Larga-Between.

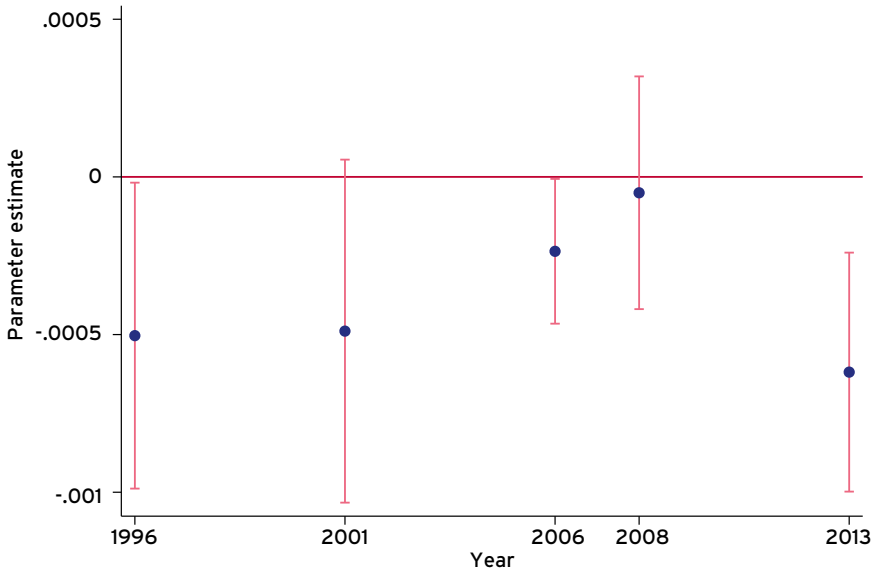
Our findings point to a complex chain of effects in which the impact of broadband internet availability changed over time, as new political actors emerged, in a ‘supply-side’ response to ‘demand-side’ changes in voter behaviour. First, we document that the diffusion of broadband internet led, initially, to a significant decline in electoral turnout in national parliamentary elections between 1996–2001 (pre-broadband) and 2006–2008 (post-broadband). This effect is sizeable – about 7 percentage points for going from no access to full broadband access – and very robust to the use of different measures of broadband access and various econometric specifications. However, this initial negative effect of the internet on turnout was largely reversed in the following elections, held in 2013. The evolution of the effect of broadband over time is depicted in Figure 2, which plots the results of a series of regressions of turnout on distance to the closest UGS, estimated separately for each election year.<sup>2</sup> It is evident that, prior to the advent of the internet, municipalities located farther away from UGSs (usually smaller and more remote) displayed lower rates of turnout than municipalities located closer to them.

<sup>2</sup> To account for the possibility of differential time trends, in all regressions we control flexibly for a range of municipal characteristics in baseline census year 2001 interacted with time variables (either a fourth order polynomial, a post-2001 dummy, or year dummies). Controls include population, age structure (i.e., share of the population aged 65 or more, and share aged 20-34 years), unemployment rate, and urbanisation. Furthermore, in all regressions we control for yearly population, and for province and local labour markets fixed effects.



However, in the first two post-internet elections, the gap between these two groups first decreased (in 2006) and then became indistinguishable from zero (in 2008), before reverting back to pre-internet levels in 2013.

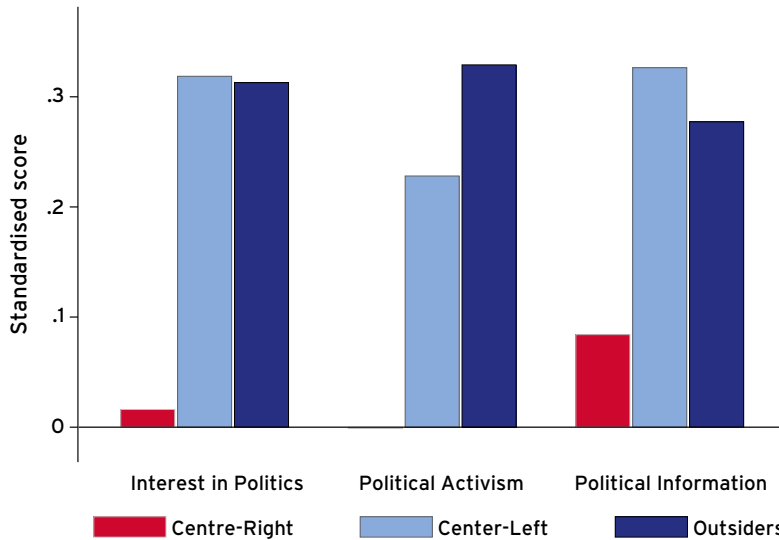
**FIGURE 2** COEFFICIENT FROM REGRESSION OF TURNOUT ON DISTANCE FROM CLOSEST UGS, 1996-2013



Note: The figure reports the estimated coefficient from a regression of electoral turnout on distance from the closest UGS separately for each election year. All regressions include the set of controls described in footnote 2.

To shed light on this nuanced pattern, we first document that the initial decline in turnout was especially detrimental to the electoral performance of ‘outsider’ parties, i.e. ideologically extreme forces outside of mainstream coalitions. Interestingly, the supporters of these parties are generally as well informed and politically active as supporters of mainstream parties, if not more so. This emerges clearly from Figure 3, which reports the average level of interest in politics, political activism and political information for voters for the centre-right coalition, the centre-left coalition and outsider parties based on data from the 2001 wave of the Italian National Election Study. This evidence seems hard to reconcile with the view that the effect of broadband internet operates through demobilisation induced by reduced information, because these voters are unlikely to be those disproportionately switching to online entertainment. Rather, it seems consistent with the diffusion of the internet having led to a particularly engaged but disenchanted segment of the electorate dropping out of the mainstream electoral process.

**FIGURE 3 INTEREST IN POLITICS, POLITICAL ACTIVISM AND POLITICAL INFORMATION BY VOTING CHOICE, 2001**



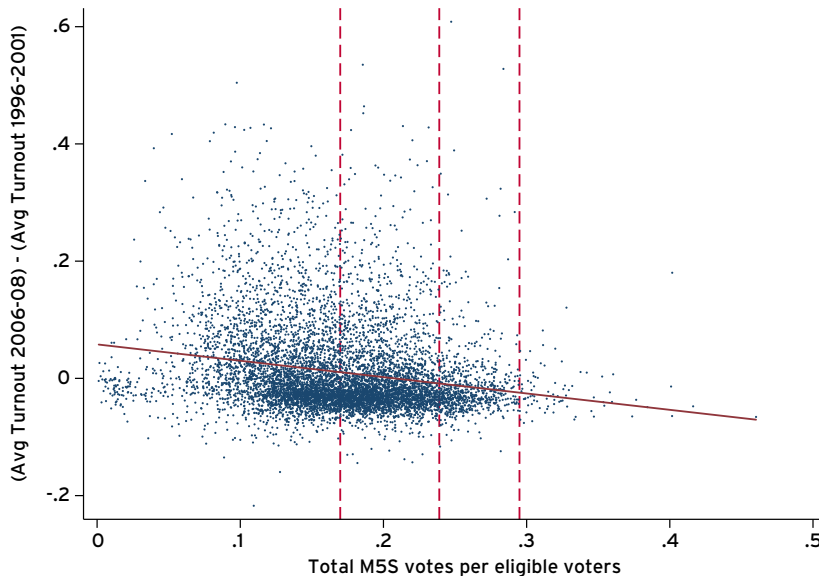
Note: The figure reports the average standardized score in the measures "Interest in Politics", "Political Activism" and "Political Information" separately for voters of the two mainstream coalitions (centre-right and centre-left) and of outsider parties. The measures are based on responses to questions included in the 2001 wave of the Italian National Election Study.

While access to broadband initially depressed electoral turnout, it appears to have fostered other and new forms of political engagement. In particular, using a unique dataset on the geographic distribution of local grassroots protest groups organised through the online platform Meetup.com, we show that the diffusion of broadband internet was associated with these groups forming earlier and growing faster.

More interestingly, the strength of these online local groups soon translated into greater support for M5S, at the time a largely web-based political movement that coalesced around those online groups and gradually evolved into a potent electoral force. Indeed, looking at local elections starting in 2008, we document that broadband internet access was positively associated with the presence of M5S on the ballots, and with electoral support for M5S local lists. These results suggest that political entrepreneurs eventually seized the opportunity to use the internet as a means of attracting a population of disenchanted (but engaged and 'connected') individuals that the internet itself had initially contributed to creating, favouring their return to mainstream electoral politics. This tendency was confirmed by the results for the 2013 parliamentary elections, when M5S ran for the first time at the national level, and which show a positive effect of broadband access on the electoral performance of M5S and other new, 'web-friendly' parties.

That the disenchantment with mainstream politics underpinned the rise of M5S is supported by the evidence in Figure 4. This plots the relationship between the electoral performance of the M5S in 2013 (on the horizontal axis) and the difference in average turnout between 2006–2008 and 1996–2001 (on the vertical axis) at the municipality

**FIGURE 4 M5S PERFORMANCE IN THE 2013 ELECTION AND CHANGE IN TURNOUT BETWEEN PRE-BROADBAND AND IMMEDIATELY POST-BROADBAND ELECTIONS**



Note: The figure plots M5S votes as a proportion of eligible voters against the difference in average turnout in 2006-2008 (post-broadband) and 1996-2001 (pre-broadband). Each dot represents one municipality. The solid line represents the fitted values (slope: -0.279, p-value: 0.000). The dashed vertical lines mark the 50th (0.1698), 90th (0.2390) and 99th (0.295) percentiles in the distribution of M5S votes as a proportion of eligible voters.

Taken together, our findings underscore what we might call the ‘general-equilibrium’ repercussions of change in media technology, exemplified by the onset of high-speed internet. Such a shock entails a shift on the ‘demand side’ of the political process, as voters react to the new medium. And yet, the latter is merely the first reaction in a more complex chain. Eventually, political entrepreneurs on the supply side take advantage of the opportunity presented by the initial demand-side movement – and of the possibilities and low barriers to entry that characterise the new medium itself – to enhance political mobilisation in ways that eventually feed back into and alter the initial landscape. Quite simply, using the classic framework of ‘exit, voice, and loyalty’ (Hirschman 1970), it seems that the new medium initially constituted an ‘exit’ option from the mainstream political process, but was eventually harnessed into a new ‘voice’ mechanism within that process. Within this framework, the transition from exit to voice is, in fact, quite natural once we consider that the effect of the mainstream political process on public-good provision ought to induce Hirschmanian ‘loyalty’, as it is essentially impossible to completely exit its reach. In addition, although we might expect a similar pattern to hold with different

instances of shifts in media technology, such a transition seems especially relevant in the case of the internet, with its low barriers to entry into the production and diffusion of content.

Ultimately, our study was among the first to document how the internet, and social media in particular, provided new political actors with a powerful tool to mobilise fringe voters disillusioned with mainstream politics. While in principle this process could foster political accountability, in the Italian context it ultimately paved the way for the rise of a new political party with marked populist traits. Far from being specific to Italy, the same process has since occurred in several other countries (Guriev et al. 2021, Manacorda et al. 2023).

## REFERENCES

- Campante, F, R Durante and F Sobbrío (2018), “Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation”, *Journal of the European Economic Association* 16(4): 1094-1136.
- Czernich, N (2012), “Broadband Internet and Political Participation: Evidence for Germany”, *Kyklos* 65: 31-52.
- Durante, R and B Knight (2012), “Partisan control, media bias, and viewer responses: Evidence from Berlusconi’s Italy”, *Journal of the European Economic Association* 10(3): 451-481.
- Enikolopov, R, A Makarin, and M Petrova (2016), “Social Media and Protest Participation: Evidence from Russia”, CEPR Discussion Paper No. 11254
- Falck, O, Robert Gold, and S Heblich (2014), “E-Lectons: Voting Behavior and the Internet”, *American Economic Review* 7: 2238-2265.
- Feddersen, T J and W Pesendorfer (1996), “The Swing Voter’s Curse”, *American Economic Review* 86(3): 408-424.
- Feddersen, T J and W Pesendorfer (1999), “Abstention in Elections with Asymmetric Information and diverse Preferences”, *American Political Science Review* 93(2): 381-398.
- Gentzkow, M (2006). “Television and Voter Turnout”, *Quarterly Journal of Economics* 121: 931-972.
- Gentzkow, M and J M Shapiro (2010). “What Drives Media Slant? Evidence from U.S. Daily Newspapers”, *Econometrica* 78(1): 35-71.
- Gentzkow, M, J M Shapiro, and M Sinkinson (2011), “The Effect of Newspaper Entry and Exit on Electoral Politics”, *American Economic Review* 101: 2980-3018.
- Guriev, S, N Melnikov, N and E Zhuravskaya (2021), “3G Internet and Confidence in Government”, *The Quarterly Journal of Economics* 136(4): 2533-2613.

Hirschman, A O (1970), *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*, Vol. 25, Harvard University Press.

Jaber, A (2013), “Individual Campaign Contributions and Voter Turnout: The Role of Broadband Network Availability”, Working Paper, Cornell University.

Kaye, B K and T J Johnson (2002), “Online and in the Know: Uses and Gratifications of the Web for Political Information”, *Journal of Broadcasting and Electronic Media* 46(1): 54-71.

Larcinese, V and L Miner (2012). “The Political Impact of the Internet on US Presidential Elections”, Working Paper, LSE and New Economic School.

Lassen, D D (2005), “The Effect of Information on Voter Turnout: Evidence from a Natural Experiment”, *American Journal of Political Science* 49(1): 103-118.

Tabellini, G, M Manacorda and A Tesei (2023), “Mobile Internet and the Rise of Communitarian Politics”, CEPR Discussion Paper No. 18063.

Miner, L (2012), “The Unintended Consequences of Internet Diffusion: Evidence from Malaysia”, Working Paper, New Economic School.

Prior, M (2005), “News v. Entertainment: How Increasing Media Choice Widens Gaps in Political Knowledge and Turnout”, *American Journal of Political Science* 49: 594-609.

Stromberg, D (2004), “Radio’s Impact on Public Spending”, *Quarterly Journal of Economics* 119: 189-221.

Tolbert, C J and R S McNeal (2003), “Unravelling the Effects of the Internet on Political Participation?”, *Political Research Quarterly* 56: 175-185.

## ABOUT THE AUTHORS

**Filipe Campante** is a Bloomberg Distinguished Professor of International Economics at Johns Hopkins University, and a Research Associate at the National Bureau of Economic Research (NBER). He is interested in political economy, development economics, and urban/regional issues. His research looks at what constrains politicians and policy makers beyond formal checks and balances: media, cultural norms, institutions, political protest. His work has appeared in leading academic journals, and received numerous mentions in media outlets all over the world. Born and raised in Rio de Janeiro, Brazil, he holds a PhD from Harvard University, an MA from the Pontifical Catholic University of Rio de Janeiro, and a BA from the Federal University of Rio de Janeiro, all in economics.

**Ruben Durante** is professor of economics at the National University of Singapore and ICREA researcher at Pompeu Fabra University. He is also affiliated faculty to the Barcelona School of Economics and IPEG, Research Advisor at Gallup US, and Research Fellow of CEPR, CESifo, and IZA. He has held visiting positions at Yale University, PSE,

and INSEAD. He studies political economy with a focus on the functioning and impact of traditional and new media in democratic societies. His work has been published in top journals in economics, political science, and management. His research has been supported by several funding agencies including the European Research Council through a Starting Grant, a Consolidator Grant, and a Proof of Concept Grant. He holds a BA in economics from the University of Messina, a Master in political economy from Sorbonne University, and a Master's and a PhD in Economics from Brown University.

**Francesco Sobbrío** is Professor of Economics at Tor Vergata University of Rome and CESifo Research Fellow. His research focuses on political economics and on the economics of media. Before joining Tor Vergata he was an Associate Professor at Luiss University, an Assistant Professor at Catholic University of Milan and a Jean Monnet Fellow at the European University Institute. He holds a BA in economics and social sciences from Bocconi University, a MSc in Economics from the University of Warwick, and a Ph.D. in Economics from the University of Southern California.



# CHAPTER 9

## New technologies and political competition: The impact of social media communication on political contributions

**Maria Petrova, Ananya Sen and Pinar Yildirim**

Universitat Pompeu Fabra and CEPR; Carnegie Mellon University; University of Pennsylvania

Campaign finance plays a significant role in democracies across the world (Bekkouche et al. 2022). Incumbent politicians hold a favourable position in raising resources for electoral races. It has been the case, especially in the US, that incumbents receive more media attention and endorsements compared to newer candidates (Prior 2006). Incumbents can also leverage their proximity to the office they hold to access resources that can lead to greater name recognition among the electorate. Resource constraints can prevent lesser-known and newer candidates from buying ad space in mainstream media, which is often expensive. This can create a vicious cycle giving an advantage to incumbents: barriers to enter into politics prevent newcomers from competing, and incumbency further erects barriers protecting the incumbents themselves. Given these advantages, incumbent politicians running for re-election in the US have about a 90% chance of winning (Levitt and Wolfram 1997).

The spread of the internet has led to the rise of digital platforms and new mass communication technologies, such as online social networks. Politicians have been increasingly embracing these new technologies and use them to communicate with their constituencies and hold political campaigns. There are a number of unique characteristics of social media that separate it from other communication technologies. Social media has a low cost of use and reduces barriers to entry for mass communication. It disseminates information in a real-time or fast fashion, working over networks of individuals. It can interact with other communication channels –substituting or amplifying the message sent through them (Enikolopov et al. 2022, Garcia-Jimeno et al. 2022). These unique characteristics of social media allow it to make an impact on public communication in general, and political communication in particular.

Motivated by these observations, we studied whether political communication on online social networks such as Twitter alter political outcomes, focusing on their effects on political competition. It is not uncommon for new technologies to alter competition among firms (Athey and Gans 2010). In politics, too, new technologies can alter the



dynamics and outcomes of electoral competition. The low cost of communication and the reduced barriers to reach out to the masses may allow social media to intensify political competition, compared to the periods and markets where traditional media are the only available communication technologies. Any individual is permitted, in theory, to set up a social media account and voice their opinions. This gives new politicians who lack resources an inexpensive alternative tool to compete against the incumbents and to raise funds for an electoral campaign.

In Petrova et al. (2021), we quantitatively estimate the impact of politicians' Twitter communication on the campaign contributions they receive. We analyse how the political contributions received by 1,834 candidates for the US Congress changed once these candidates opened a Twitter account and started communicating with it. The candidates cover those who ran over the three election cycles between 2009 and 2014. More specifically, we compare the weekly donations of politicians immediately before and after they started using Twitter across various US states, where Twitter has varying levels of success (i.e. penetration) relative to other online domains. We capture this variation by constructing a Twitter penetration measure calculating the relative (time or frequency of) use of Twitter to all other sites for each state. We use the Comscore online browsing panel to construct the measure. We also use a variety of other measures and an alternative data source – Simmons Oneview. If a politician is using Twitter in a state where the platform has already gained traction, the returns to using Twitter should be higher.

The outcome variable of interest is the amount of donations received by a politician in a particular week. Our data set includes every donation made to every candidate (above \$200), made publicly available by the Federal Election Commission (FEC). Since we are interested in the influence of social media on certain constituents (the broader public), we focus our analysis on donations below \$1,000 and those made by ordinary citizens. These small donations are more likely to be influenced by the direct communications of candidates (on Twitter and other media), while large donations from individuals and corporations or Political Action Committees (PACs) may have other motivations such as lobbying, and are less likely to respond to political communication (Ansolabehere and Snyder 2000).

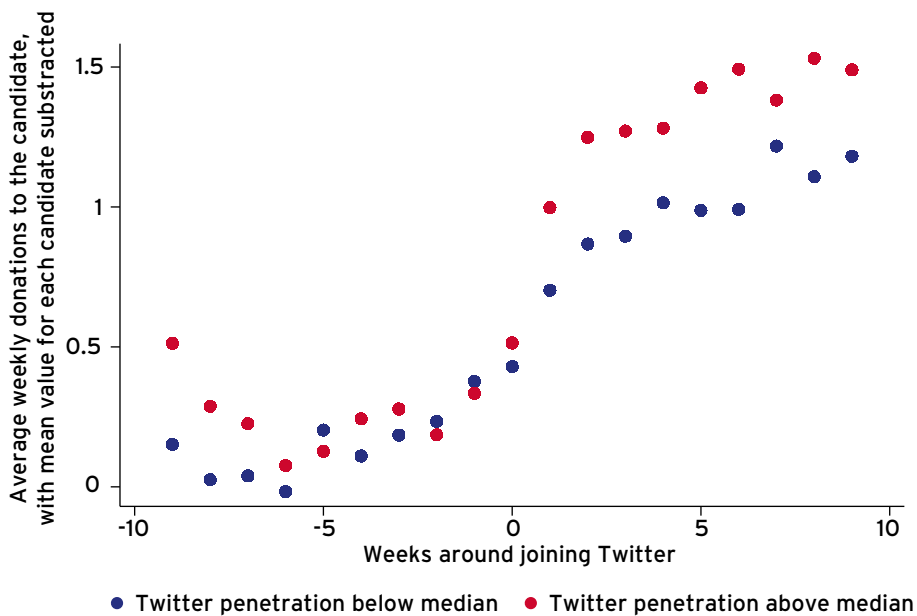
Our empirical strategy exploits the precise timing of opening a Twitter account by using the variation in donations before and after joining Twitter, and across areas of different Twitter penetration. We include politician-month fixed effects to control for politician-specific unobserved time-varying factors, such as being more progressive minded, more tech savvy, or being at a different stage of campaigning.

Before going into the details of the results, we turn to Figure 1, which demonstrates the key takeaway from our study. The figure suggests that there is a discontinuous increase in the political donations received right around the time when a politician joins Twitter, but only in the states with high Twitter usage. We build on this descriptive evidence by using

a more formal difference-in-differences analysis. In particular, we find that opening a Twitter account increases political donations by 2% on average. This effect is driven by new politicians who have not been elected to office before, who see up to a 3.1% increase in donations. Experienced politicians (those who have run for office previously) see no change in donations received. Importantly, if the use of Twitter leads to higher campaign contributions for new candidates, new communication technologies may intensify political competition, and even help reduce incumbency advantage.

We carry out a number of placebo or ‘balance’ checks to ensure that our estimates are not driven by unobservable factors that coincide with Twitter activity – such as political advertising and campaign events, or developing news stories about the politician. We collect data on campaign expenditures (made public by the FEC) to demonstrate that there is no similar discontinuous change on this dimension around the time politicians start using Twitter. Similarly, we use data on political ads on television (collected by Kantar Media) to find that there was no increase coinciding with the opening of the Twitter account. Finally, we also establish that there was no change in the news and blog coverage of the politicians around the time they joined Twitter. All these checks increase our confidence that the estimates are not driven by spurious correlations.

**FIGURE 1** CHANGE IN POLITICAL DONATIONS AROUND THE WEEK OF OPENING AN ACCOUNT ON TWITTER, IN AREAS OF HIGH VERSUS LOW TWITTER PENETRATION



Note: Political donations and joining Twitter, controlling for candidate and week fixed effects.

Source: Reproduced with permission from Petrova et al. (2021).

To interpret these magnitudes, we put them in the context of the existing literature. To do so, we compute persuasion rates (DellaVigna and Gentzkow 2010). Persuasion rates measure the percentage of individuals who were subjected to a persuasive message and changed their behaviour, among all who received the message and are not already persuaded. This is a scale-free measure that can be used to compare different channels of persuasion. The persuasion rates in our context range between 0.5% and 0.75%. These are at the lower end of persuasion rates reported in the literature, but comparable to the persuasion rates of 1% for direct mailing (Gerber and Green 2000) and between 0.1% and 1% for political advertising (Spenkuch and Toniatti 2018). One factor that may contribute to the low rates we measure is focusing on donations that are higher than \$200, and not observing the smaller donations. While Twitter's effectiveness compares to that of direct mailing and short political ads on TV, the cost of advertising and direct mailing can be significantly higher compared to that of operating a personal social media account.

After credibly establishing these baseline results, we move on to analysing the mechanism behind these results. We focus on two potential dimensions similar to advertising effects (Nelson 1974) through which information on Twitter might have impacted donors. The first is the information mechanism, whereby communicating through Twitter creates awareness about politicians who would be otherwise unknown to a hitherto untapped base of individuals. The second is the persuasion mechanism, whereby communicating on Twitter acts as a reminder to potential donors who are already aware of a politician, and helps persuade them to give.

We find that the information mechanism is more likely to be at play in explaining our findings. Evidence consistent with this mechanism comes from a set of different exercises. We first check if the donations a given politician receives come from repeat donors (i.e. individuals who have previously donated to the campaign of the politician) or new donors (i.e. those who have never donated before). We find that donations predominantly come from new donors. Put differently, Twitter communication expands the donor base of the politicians. When we look at the donation behaviour by repeat donors, we do not see a significant change in their donations, either for new or experienced politicians. Second, we find that the results are driven by an increase in donations received by candidates for the US House of Representatives, rather than those for the US Senate. Politicians serving in the House of Representatives are elected for a two-year period and tend to have lower name recognition compared to those elected to the US Senate, who serve for six years. Moreover, US states have many representatives, but only two Senate seats. As a result, creating awareness about one's candidacy and policy positions can be particularly beneficial for the House of Representatives candidates, compared to candidates for the Senate. Third, we find that politicians who had opened a Facebook account prior to starting to use Twitter did not see any increase in donations after joining Twitter. This finding is, again, consistent with social media communication rewarding politicians by disseminating new information about them.

What exactly is it in social media communication that informs or persuades donors? While we cannot pin down a single answer, we carried out a text analysis of the tweets posted by candidates to try to address this question. Since tweets from this period included only up to 140 characters (the limit was increased to 280 characters in November 2017), a meaningful analysis of tweet content is difficult. Nevertheless, using text and sentiment analysis, we report several observations. First, parsing the textual information of tweets, we find that politicians who send ‘more informative’ messages (e.g. by including a hyperlink/URL to additional information) see greater gains from opening an account. Second, use of a more inclusive language (e.g. words such as “we” instead of “I”) correlates with higher gains. Third, when we apply the linguistic inquiry and word count (LIWC) method (Pennebaker et al. 2015) to analyse the sentiment in tweets in terms of emotional, social and thinking styles, we find that a politician with a ‘plugged in’ social style is more likely to see a greater increase in donations. There is also a curious correlation between donations and using ‘anti-establishment’ language. Undoubtedly, there is more to communicating through social media – for instance, signalling one’s character, indicating being open or being accountable, caring about others, or a desire to listen to the public. Future research can study the social media language of politicians to examine such characteristics.

## TAKEAWAYS

Electoral races in the US depend on a variety of factors, with campaign funds being a central one. Resources for running an effective campaign became even more important with the Citizens United decision by the US Supreme Court in 2010 (Klump et al. 2016). With political campaigns becoming increasingly more expensive, a natural concern is whether challengers have enough opportunities to communicate with the electorate and to fundraise. The broad implication of our study is that the adoption and use of social media can alter electoral dynamics and make political elections more competitive, since this new technology offers politicians a relatively cost-effective alternative technology to communicate with the electorate, and reduces the gap in fundraising opportunities between new and experienced politicians.

Our research study opens avenues for further research. We look at the period between 2009-2014, when Twitter was still an emerging technology. There have been significant changes in the social media landscape since our study; Twitter, now rebranded as X, is now an established player in the social media market. A natural follow-up question is how a candidate should use these platforms upon joining them. Our analysis takes a first step by looking at correlational evidence between language, tweeting styles and donations as an outcome. Future studies can provide a more systematic examination of social media use. Moreover, politicians now also use platforms such as Instagram and TikTok, particularly to reach out to young voters. These platforms are different from Twitter and Facebook due to visual and video elements embedded as the primary means of communication. They are also geared towards entertainment, rather than being a source

of information. It may be informative to examine if and how political communication using visual, entertainment-oriented means can become a part of a campaign's strategy toolkit. Comparing the gains from these platforms relative to email or advertising campaigns can be telling regarding the course of future campaigns. Finally, with these tools in place, an open question is about the timing of these online communications and solicitations. Should politicians reach out to their potential donors and request donations right after a significant policy or legal decision (e.g. the US Supreme Court overturning *Roe vs. Wade*), or should they wait? We believe that all these topics are of consequence from the politician's as well as a societal perspective in this still rapidly evolving social media landscape.

## REFERENCES

- Ansolabehere, S and J M Snyder (2000), "Old voters, new voters, and the personal vote: Using redistricting to measure the incumbency advantage", *American Journal of Political Science* 44(1): 17-34.
- Athey, S and J S Gans (2010), "The impact of targeting technology on advertising markets and media competition", *American Economic Review* 100(2): 608-13.
- Bekkouche, Y, J Cage and E Dewitte (2022), "The heterogeneous price of a vote: Evidence from multiparty systems, 1993-2017", *Journal of Public Economics* 206, 104559.
- DellaVigna, S and M Gentzkow (2010), "Persuasion: Empirical evidence", *Annual Review of Economics* 2: 643-69.
- Enikolopov, R, M Rochlitz, K Schoors and N Zakharov (2022), "The Effect of Independent Online Media in an Autocracy", University of Pompeu Fabra Working Paper.
- García-Jimeno, C, A Iglesias and P Yildirim (2022), "Information Networks and Collective Action: Evidence from the Women's Temperance Crusade", *American Economic Review* 112(1): 41-80.
- Gerber, A S and D P Green (2000), "The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment", *American Political Science Review* 94(3): 653-663.
- Klumpp, T, H M Mialon, and M A. Williams (2016), "The business of American democracy: Citizens United, independent spending, and elections", *The Journal of Law and Economics* 59(1): 1-43.
- Levitt, S D and C D Wolfram (1997), "Decomposing the sources of incumbency advantage in the US House", *Legislative Studies Quarterly* 22(1): 45-60.
- Nelson, P (1974), "Advertising as information", *Journal of Political Economy* 82(4): 729-754.

Pennebaker, J W, R L Boyd, K Jordan and K Blackburn (2015), “The development and psychometric properties of LIWC2015”, UT Faculty/Researcher Works.

Petrova, M, A Sen and P Yildirim (2021), “Social media and political contributions: The impact of new technology on political competition”, *Management Science* 67(5): 2997-3021.

Prior, M (2006), “The incumbent in the living room: The rise of television and the incumbency advantage in US House elections”, *Journal of Politics* 68(3): 657-673.

Spenkuch, J L and D Toniatti (2018), “Political advertising and election results”, *The Quarterly Journal of Economics* 133(4): 1981-2036.

## ABOUT THE AUTHORS

**Maria Petrova** is ICREA Research Professor at the Universitat Pompeu Fabra and Barcelona Institute for Political Economy and Governance (IPEG), and an Affiliate Professor at Barcelona School of Economics. Maria is a Co-Editor of the *Journal of Public Economics* and a member of the Editorial Board of the *Review of Economic Studies* since 2016. She received PhD from Harvard University in 2008. She spent 2012 - 2013 as a Visiting Associate Research Scholar at the Center for the Study of Democratic Politics at Princeton University. In 2012-2013, she was also the Research Director at the Center for New Media and Society. Her research interests include mass media economics, political economy, and corporate governance. Her paper have been published in leading journals, including *American Economic Review*, *Quarterly Journal of Economics*, *Econometrica*, *American Political Science Review*, and *Management Science*.

**Ananya Sen** is an Assistant Professor at Carnegie Mellon University’s Heinz College. Ananya’s research interests centre around platforms with a special focus on the media, innovation and more broadly the digital economy. He uses a variety of empirical techniques to analyse data from field experiments as well as observational data to gain insight into broad research questions. Before moving to Carnegie Mellon, he was a Post-Doctoral Associate at MIT Sloan School of Management. He received a PhD in Economics from the Toulouse School of Economics.

**Pinar Yildirim** is Associate Professor of Marketing and Economics (by courtesy) at University of Pennsylvania. She is also a Faculty Research Fellow in the Productivity, Innovation, and Entrepreneurship Program at the National Bureau of Economic Research (NBER). In her research, Pinar analyzes such topics as social networks, advertising and competitive product positioning, the role of choice in two-sided platforms, and the incentive of platforms to moderate content. Her research has been published in leading peer-reviewed journals, including the *American Economic Review*, *Marketing Science*, *the Journal of Marketing Research*, *Quantitative Marketing and Economics*, and

*Management Science*. Pinar received a PhD from the University of Pittsburgh and has been on the editorial boards of several leading management journals, such as *Marketing Science* and *Journal of Marketing Research*.

**SECTION 4**  
**SOCIAL MEDIA AND**  
**MOBILISATION**





# CHAPTER 10

## Social media and protest participation: Evidence from Russia

121

**Ruben Enikolopov,<sup>a,c</sup> Alexey Makarin<sup>b</sup> and Maria Petrova<sup>a,c</sup>**

<sup>a</sup>Universitat Pompeu Fabra; <sup>b</sup>MIT Sloan School of Management; <sup>c</sup>CEPR

Social media is playing an increasingly important role in our lives. The political consequences of the advent of social media are now a hotly contested topic (Zhuravskaya et al. 2020). An optimistic view has been that social media would empower ordinary citizens and serve as a ‘liberation technology’ (Diamond and Plattner 2010) that would lead to faster democratisation in authoritarian countries (Shirky 2008) and make politicians more accountable (Besley and Prat 2006). Evidence of the effect of the spread of mobile phones in Africa seems to support the idea that new communication technologies can help mass political mobilisation (Manacorda and Tesei 2020). A more pessimistic view has been that autocratic governments would adjust and learn how to exploit social media to their own advantage, employing a combination of tools including online censorship, surveillance, and heavy use of online bots and trolls (Morozov 2011, Roberts 2018). In democracies, social media is now often blamed for the exacerbated political polarisation, the spread of xenophobic ideas, the proliferation of fake news and general negative effects on users’ wellbeing (Tufekci 2018, Allcott et al. 2020, Braghieri et al. 2022).

There is plentiful evidence that traditional media (newspapers, radio, TV) have had an important impact on political outcomes by providing political news and entertainment (DellaVigna and La Ferrara 2015, Enikolopov and Petrova 2015). In many respects, online media resemble traditional media and one should expect their persuasion effects to mirror those of traditional media. However, certain features of new media – of social media, in particular – are distinct. The two most important distinguishing features of new social media are low barriers to entry and reliance on user-generated content. Low entry barriers make gatekeeping the spread of political information much less effective, allowing new entrants who were previously sidelined by the political establishment. By providing an outlet for the opposition and for whistleblowers, social media makes it harder for political and business actors to hide potentially harmful information, increasing their accountability (Besley and Prat 2006). Low entry barriers can also have social costs. For example, social media can be used to spread extremist ideas, increasing their reach and potentially their influence. Low barriers to entry could also undermine the reputation mechanisms that serve to guard the quality of information of traditional media outlets (Gentzkow and Shapiro 2006, Cagé 2020) and lead to the spread of misinformation and fake news. Low barriers to entry also vastly increase the choice of news sources and,

arguably, allow users to tailor their news sources to their pre-existing preferences more finely than traditional media allow; this potentially could give rise to ‘echo chambers’ and lead to increased political polarisation.

By allowing horizontal flows of information through two-way communication between users, social media facilitates coordination between people, thus potentially making it easier to organise collective actions such as street protests. At the same time, online protest activity in social media could crowd out offline actions necessary for real political change in autocracies (Gladwell 2010). User-generated content and two-way communication in social media could also change the way politicians and citizens interact (Bessone et al. 2022). Social media allows politicians to receive immediate feedback on policy actions, discuss policy proposals and measure political discontent. Such feedback could be used for policy improvements, but it could also be used for oppression and political surveillance. In addition, the low cost of creating automated accounts and the ability to post content using anonymous or impersonated accounts enable the manipulation of online content seen by real users, potentially leading to political persuasion. Also, the data that online platforms collect about their users could be (and have been) used to target specific groups of users to make such manipulations more effective.

The combination of low barriers to entry and horizontal flows of information could make social media especially important in facilitating street protests. Low barriers to entry in social media make it easier to spread information critical of the government, which is especially important in autocratic regimes where traditional media is under a tight control by the government. This increases the number of informed citizens who are unhappy with their government and thus potentially ready to take part in political protests. Furthermore, horizontal flows of information between users of social media allow them to exchange logistical information about the upcoming events and coordinate their tactics on the spot. This helps solve collective-action problems and increases the probability that protests actually take place by increasing the probability that people who are potentially ready to participate in political protests actually do participate.

In our paper (Enikolopov et al. 2020),<sup>1</sup> we provide causal estimates of the political effects of social media in a non-democratic environment, focusing on the effect on participation in political protests. Testing empirically the effect of social media on political protests is methodologically challenging because social media usage is endogenous to individual and community characteristics. In addition, protests are typically concentrated in one or a few primary locations, as was the case for Tahrir Square in Egypt or Maidan in Ukraine. Hence, geographic variation in protests is often very limited. Temporal variation in protest intensity can provide evidence on the association between the activity and the content on social media and subsequent protests (Acemoglu et al. 2017), but not on the causal impact of social media availability.

1 The results of which hold after correcting for the mistakes in two control variables (Enikolopov et al. 2023).

To understand whether social media can indeed promote protest participation, we study an unexpected wave of political protests in Russia in December 2011 triggered by electoral fraud in parliamentary elections, coupled with an analysis of the effect of social media on support for the government. Our empirical setting allows us to overcome the limitations of previous studies for two reasons. First, there was substantial geographic and temporal variation in both protest activities and the penetration of the major online social networks across Russian cities. For example, among the 625 cities in our sample, 133 witnessed at least one protest demonstration on 10–11 December 2011, the first weekend after the elections. Second, particularities of the development of VKontakte (VK), the most popular social network in Russia, allow us to exploit quasi-random variation in the penetration of this platform across cities and ultimately identify the causal effect of social media penetration on political protests.

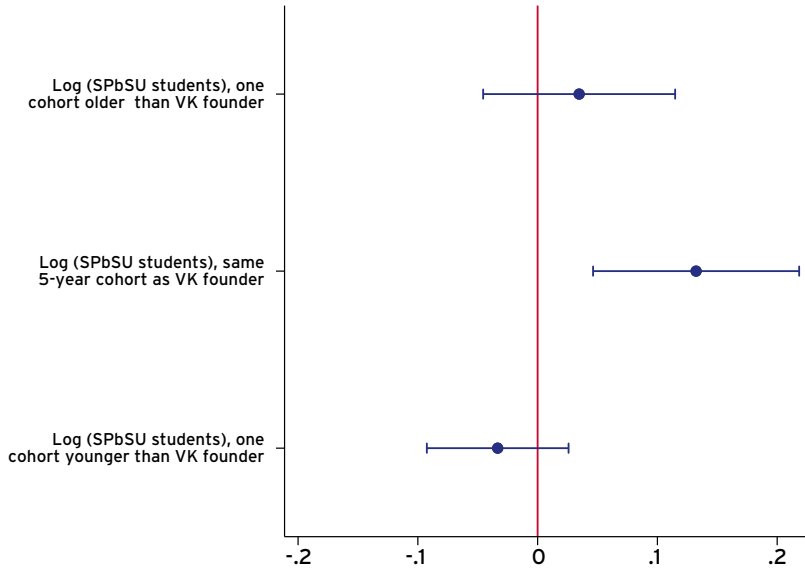
Our identification is based on the information about the early stages of VK's development. VK was created in 2006 by Pavel Durov, a student at Saint Petersburg State University (SPbSU). This online social network, analogous to Facebook in functionality and design, was the first mover in the Russian market and secured its dominant position with a user share of over 90% by 2011. Initially, users could only join the platform by invitation through a student forum of the university, which was also created by Durov. As a result, the vast majority of early users of VK were Durov's fellow students at SPbSU. This, in turn, made friends and relatives of these students more likely to open an account early on. Since SPbSU attracted students from across the country, this sped up the propagation of VK in the cities these students had come from. Network externalities magnified these effects and, as a result, the distribution of the home cities of Durov's classmates had a long-lasting effect on VK penetration.

We exploit this feature of VK development in our empirical analysis by using the origin of students who studied at SPbSU in the same five-year cohort as the VK founder as an instrument for VK penetration in the summer of 2011, controlling for the origin of the students who studied at SPbSU several years earlier and later. Thus, our identification is based on the assumption that temporal fluctuations in the number of students coming to SPbSU from different Russian cities were not related to unobserved city characteristics correlated with political outcomes. In the first-stage regression, we find that the distribution of the home cities of the students who studied at SPbSU at the same time as Durov predicts the penetration of VK across cities in 2011, whereas the distribution of the home cities of the students who studied at SPbSU several years earlier or later does not (see Figure 1).

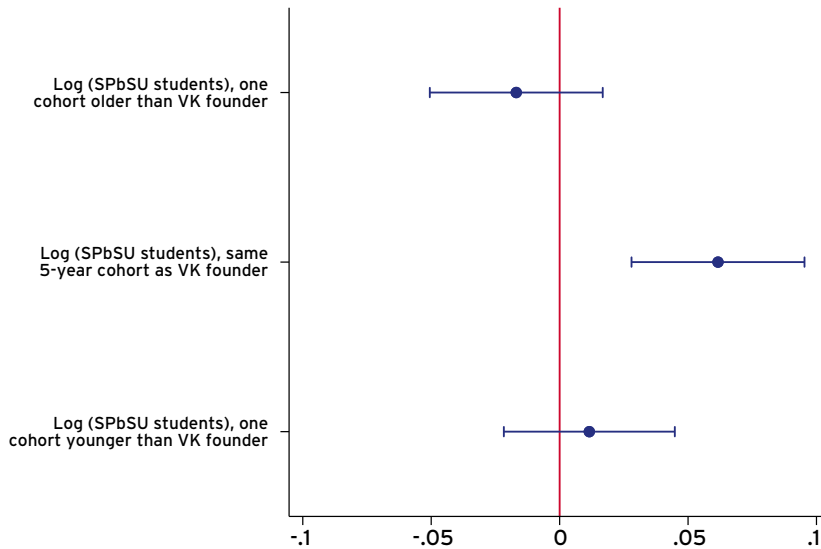
Using this instrument, we estimate the causal impact of VK penetration on the incidence of protests and protest participation. In the reduced-form analysis, we show that fluctuations in the student flow from Russian cities to SPbSU over time predict the incidence of protests in December 2011 (Figure 2) and the number of protest participants. The corresponding IV estimates indicate that the magnitude of the effect is sizable: a 10% increase in the number of VK users in a city led to both a 4.6 percentage point increase

in the probability of there being a protest and a 19% increase in the number of protest participants the first weekend after the elections. Non-parametrically, we document that there exists a threshold of VK penetration below which there is no relation between VK penetration and protests (see Figure 3).

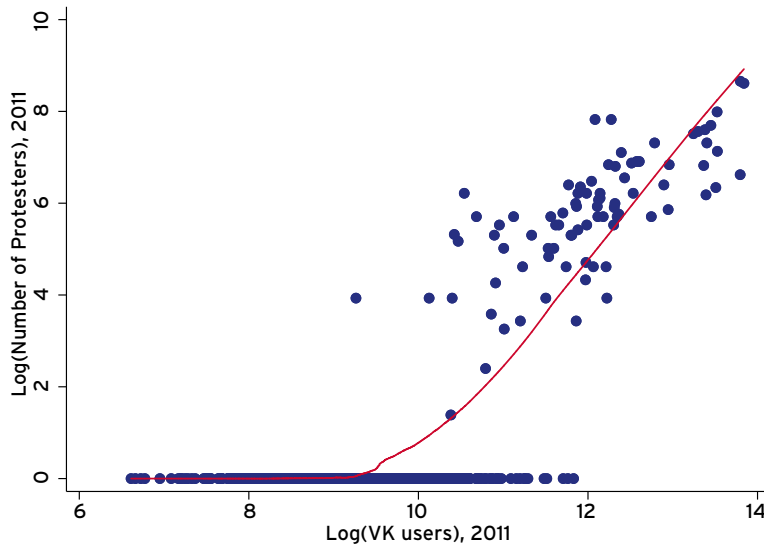
**FIGURE 1 VK PENETRATION IN 2011 AND SPBSU STUDENT COHORTS**



**FIGURE 2 INCIDENCE OF PROTESTS IN 2011 AND SPBSU STUDENT COHORTS**



**FIGURE 3 NONPARAMETRIC RELATIONSHIP BETWEEN VK PENETRATION AND NUMBER OF PROTESTERS**



As a falsification test, we show that VK penetration in 2011 does not predict protest participation in the same cities before the creation of VK using three different protest instances: anti-government protests at the end of the Soviet Union (1987–1992), labour protests in 1997–2002, and social protests in 2005. We also show that VK penetration in 2011 was not related to voting outcomes before the creation of VK. We also replicate our first-stage regressions using the information on the cities of origin of the students who studied in more than 60 other major Russian universities. We find that the coefficient for our instrument – VK founder’s cohort at SPbSU – lies at the top end of the distribution of the corresponding coefficients in other universities, while the coefficients for younger and older cohorts lie close to the medians of the corresponding distributions, consistent with our identifying assumptions.

We highlight two channels through which social media could lead to protest participation in a non-democracy. On the one hand, low barriers to entry make it much more difficult for the regime to limit the spread of potentially harmful information that would lead to more anti-government sentiments in the population. We call this the *information channel*. On the other hand, the fact that social media relies on user-generated content facilitates horizontal information flows, which could lower the costs of coordination and thus alleviate the collective action problem (Ostrom 1990). We call this the *coordination channel*. If this channel is at work, the chances that people take out to the streets could go up even if the number of individuals in opposition does not increase. In our context, VK was used heavily for tactical coordination of protests in Russia in 2011–2012. For instance, for almost every city with a protest, activists created VK protest communities, where people could exchange the logistical details.

We find no evidence that supports the information channel being at work in our context. In particular, we study VK's impact on the pro-government vote and attitudes toward the regime. We show that, consistently across all elections after the creation of the social network, VK led, if anything, to a higher, not lower, pro-government vote. We also do not find any evidence for increased political polarisation since there was no jump in negative attitudes toward the regime or in the opposition vote. Finally, we analyse the political content on VK and find that, on average, it was neutral or positive towards the government.

However, we do find evidence in favour of coordination. We find that the number of VK users in online protest communities was positively associated with the incidence and the size of the protests. We also find that the impact of social media on protests was stronger in larger cities, where logistical coordination tends to be more critical. Finally, we show that protests tend to be smaller in cities where, conditional on the total number of social media users, the user base was more fractionalized between Facebook and VK. A more divided user base matters because it may lead to less horizontal information flows between users of different social networks and, as a result, more difficult logistical coordination. These findings may also be consistent with the importance of peer pressure and social image, and we explore this hypothesis further in a companion project (Enikolopov et al. 2022).

Overall, our paper provides evidence that social media penetration had a causal effect on both the incidence and the size of the protest demonstrations in Russia in December 2011. Additional evidence suggests that social media affects protest activity by reducing the costs of collective action, rather than by spreading information critical of the government or by increasing political polarisation. Thus, our results imply that social media induces coordination and alleviates the collective action problem.

While our results confirm the earlier claims of the digital optimists, we note that these results may not generalise to other settings. The Russian protests of 2011 were unexpected, and the government did not have time to prepare for them. If a government is aware of this effect of social media on political protests, it may counteract it by censoring online content related to collective action (King et al. 2014, Ananyev et al. 2019) or manipulating the information in social media (King et al. 2017). Furthermore, the reduction in coordination costs can also have its dark sides, for example by possibly leading to more extremism and hate crimes (Bursztyn et al. 2019). More research is needed to understand whether similar results hold for other outcomes and in different contexts.

## REFERENCES

Acemoglu, D, T A Hassan and A Tahoun (2017), "The Power of the Street: Evidence From Egypt's Arab Spring", *Review of Financial Studies* 31(1): 1-42.

- Allcott, H, L Braghieri, S Eichmeyer, and M Gentzkow (2020), “The Welfare Effects of Social Media”, *American Economic Review*, 110(3): 629-676.
- Ananyev, M, D Xefteris, G Zudenkova, and M Petrova (2019), “Information and Communication Technologies, Protests, and Censorship”, working paper.
- Besley, T, and A Prat (2006), “Handcuffs for the Grabbing Hand? Media Capture and Government Accountability”, *American Economic Review* 96(3): 720-736.
- Bessone, P, F R Campante, C Ferraz and P Souza (2022), “Social Media and the Behavior of Politicians: Evidence from Facebook in Brazil”, NBER Working Paper No. 30306.
- Braghieri, L, R E Levy and A Makarin (2022), “Social media and mental health”, *American Economic Review* 112(11): 3660-3693.
- Bursztyn, L, G Egorov, R Enikolopov, and M Petrova (2019), “Social Media and Xenophobia: Evidence from Russia”, Working paper.
- Cagé, J (2020), “Media Competition, Information Provision and Political Participation: Evidence from French Local Newspapers and Elections, 1944-2014”, *Journal of Public Economics* 185: 104077.
- DellaVigna, S and E La Ferrara (2015), “Economic and Social Impacts of the Media”, in *Handbook of Media Economics*, Vol. 1, Elsevier.
- Diamond, L and M Plattner (2010), “Liberation Technology”, *Journal of Democracy* 21: 69-83.
- Gentzkow, M and J M Shapiro (2006), “Media Bias and Reputation”, *Journal of Political Economy* 114(2): 280-316.
- Enikolopov, R, and M Petrova (2015), “Media Capture”, in *Handbook of Media Economics*, Vol. 1, Elsevier.
- Enikolopov, R, A Makarin, and M Petrova (2020), “Social Media and Protest Participation: Evidence from Russia”, *Econometrica* 88(4): 1479-1514.
- Enikolopov, R, A Makarin, M Petrova, and L Polishchuk (2022), “Social Image, Networks, and Protest Participation”, working Paper.
- Enikolopov, R, A Makarin, and M Petrova (2023), “Online Corrigendum to “Social Media and Protest Participation: Evidence From Russia”, *Econometrica* 91(3): 1-24.
- Gladwell, M (2010), “Small Change”, *The New Yorker*, 27 September.
- King, G, J Pan, and M E Roberts (2014), “Reverse-Engineering Censorship in China: Randomized Experimentation and Participant Observation”, *Science* 345(6199): 1-10.



King, G, J Pan, and M E Roberts (2017), “How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument”, *American Political Science Review* 111(3): 484-501.

Manacorda, M and A Tesei (2020), “Liberation Technology: Mobile Phones and Political Mobilization in Africa”, *Econometrica* 88(2): 533-567.

Morozov, E (2011), *The Net Delusion: The Dark Side of Internet Freedom*, Perseus Books.

Roberts, M (2018), *Censored: Distraction and Diversion Inside China's Great Firewall*, Princeton University Press.

Shirky, C (2008), *Here Comes Everybody: The Power of Organizing Without Organizations*, AllenLane.

Tufekci, Z (2018), “How Social Media Took Us from Tahrir Square to Donald Trump”, *MIT Technology Review*.

Zhuravskaya, E, R Enikolopov, and M Petrova (2020), “Political Effects of the Internet and Social Media”, *Annual Review of Economics* 12: 415-438.

## ABOUT THE AUTHORS

**Maria Petrova** is ICREA Research Professor at the Universitat Pompeu Fabra and Barcelona Institute for Political Economy and Governance (IPEG), and an Affiliate Professor at Barcelona School of Economics. Maria is a Co-Editor of the *Journal of Public Economics* and a member of the Editorial Board of the *Review of Economic Studies* since 2016. She received PhD from Harvard University in 2008. She spent 2012 - 2013 as a Visiting Associate Research Scholar at the Center for the Study of Democratic Politics at Princeton University. In 2012-2013, she was also the Research Director at the Center for New Media and Society. Her research interests include mass media economics, political economy, and corporate governance. Her paper have been published in leading journals, including *American Economic Review*, *Quarterly Journal of Economics*, *Econometrica*, *American Political Science Review*, and *Management Science*.

**Alexey Makarin** is an Assistant Professor of Applied Economics and a Mitsubishi Career Development Assistant Professor in International Management at the MIT Sloan School of Management. He earned his PhD in Economics from Northwestern University in June 2019. Alexey's research centres around political economy and the economics of digitization, with a special focus on the causal effects of social media platforms.

**Maria Petrova** is ICREA Research Professor at the Universitat Pompeu Fabra, Researcher at the Barcelona Institute for Political Economy and Governance (IPEG), Affiliate Professor at Barcelona Graduate School of Economics, and the Alfa-Bank Associate Professor of Economics at the New Economic School. Maria is a Co-Editor of the *Journal of Public Economics* and a member of the Editorial Board of the *Review of Economic*

*Studies* since 2016. She received PhD from Harvard University in 2008. She spent 2012 - 2013 as a Visiting Associate Research Scholar at the Center for the Study of Democratic Politics at Princeton University. In 2012-2013, she was also the Research Director at the Center for New Media and Society. Her research interests include mass media economics, political economy, and corporate governance.



# CHAPTER 11

## Social media and mobilisation

131

**Leopoldo Fergusson and Carlos Molina**

Universidad de los Andes; Massachusetts Institute of Technology

On 25 January 2011, thousands of Egyptians took to the streets to demand change. A few weeks later, Wael Ghonim, an internet activist who helped coordinate the protests and was incarcerated during the events, summarised the emotions: “if you want to liberate society, all you need is the internet”.

Many shared the enthusiasm during the Arab Spring, especially with regards to one of the internet’s most disruptive innovations: online social media. One Egyptian went as far as naming his daughter Facebook, honouring the platform’s role in freeing the country from autocracy.<sup>1</sup>

Social media continues to feature in the news as a major contributing factor to recent waves of citizen mobilisation. Still, there is now greater recognition that some protests may be ineffective. Egypt and almost every country involved in the Arab Spring failed to deliver the democratic promises. Moreover, social media may be used to attack democracy, not just to demand or protect it. The Capitol Riots against the 2021 US democratic presidential transition provide just one example.

On the whole, how important has social media been for citizen mobilisation? What mechanisms explain its influence? Moreover, what have been the broader political implications? The answers to these questions now seem more complex than one might have imagined in 2011. When invited five years later to talk again at TED, Ghonim himself stated: “Remember I said in 2011 that if you want to liberate society, all you need is the internet? Well, I was wrong”.

### THE SOCIAL EFFECTS OF SOCIAL MEDIA

Even if much remains to be understood, abundant research has shed light on these questions. Like traditional media, social media facilitates a one-way transmission of information from the platform to the users watching on their computers and cell phones. This might change their beliefs and opinions, enticing them to take action.

1 See <https://www.huffpost.com/entry/baby-named-facebook-egypt-n-825934>

However, perhaps the main reason social media attracts attention as a catalyst for citizen mobilisations is that it also involves information exchange in multiple directions, where people communicate with each other and learn what others think and do.

Scholars have noted that the multi-directional exchange of information through social media may facilitate collective action for at least three reasons. First, strengthening horizontal communication among users weakens the obstacles to tactical coordination, like sharing the location of a protest and other logistical issues (e.g. Little 2016, Enikolopov et al. 2020). Second, altering beliefs about how many others are also willing to act (e.g. Edmond 2013, Barbera and Jackson 2020) might convey information about how safe or relevant it is to protest. Third, the desire to project one's social image may motivate (or deter) participation (e.g. Enikolopov et al. 2017, Cantoni et al. 2019).

Studies of the content of and activity on online social networks that evaluate the role that platforms like Twitter and Facebook play during protests also hint at these 'social effects'. Participants learn about protests and are encouraged to participate by information gathered through social media, directly or indirectly via friends. Evidence from Turkey, Ukraine, Occupy Wall Street, Chile and Tahir Square (e.g. Jost et al. 2018, Tufekci and Wilson 2012 Tucker et al. 2015, Valenzuela et al. 2012, Valenzuela 2013) reveals that Twitter and Facebook are used to share information on key logistical issues, to disseminate motivational appeals emphasised in social psychological theories of protest participation, and to publicise visuals from the demonstrations. Across several contexts, studies have also found that social media activity coordinates and predicts citizen mobilisation.

### **BUT DOES SOCIAL MEDIA MOBILISE PEOPLE?**

While shedding light on *how* social media influences collective action, several of the studies mentioned above are not designed to determine how much additional protest activity can be attributed to these tools. Finding out whether this is the case is not as simple as it may seem. Indeed, if online social networks had not been available, protesters might have used traditional ways to coordinate and communicate. Global Positioning System (GPS) devices and applications provide a helpful analogy. Do people drive more since the appearance of apps like Waze, which track their location and suggest a route? Probably. But many journeys would likely have occurred without the technology. The same might have happened with social media and protests.

So how can we answer this deceptively simple question? Ideally, we would need an experiment. Since an experiment is not often feasible, a few studies have relied on methodologies that approximate the ideal controlled experiment.

Manacorda and Tesei (2020) and Christensen and Garfias (2018) rely on expansions of mobile phone coverage to evaluate their impact on protests in Africa and a panel of countries, respectively. Mobile phones are not social media, so their impact can reflect

broader influences. But they do provide information, connect individuals and, on the case of smartphones, connect to the internet and online social networks. In fact, both studies emphasise inherently 'social' mechanisms.

Focusing on social media specifically, Enikolopov et al. (2020) examine variation in the expansion of VKontakte (VK), Russia's leading social network, to identify the impact of network penetration on political protests. They find that if VK penetration increases by 10%, the probability of a protest goes up by 4.6% and the number of protesters by 19%. Also, they attribute these effects to lower coordination costs rather than the spread of information against the government. Qin et al. (2021) look at a period of rapid expansion of social media in China to show how retweets, by users in one city, of blog posts from other cities affect the geographic spread of protests and strikes in China. They attribute the effect to tacit coordination and emotional reactions.

In short, we now know that social media, at least in some contexts, increases citizen mobilisation. We also know that this effect differs from traditional media precisely because of social media's *social* communication. Nevertheless, the best evidence focusing on social media is limited to specific contexts, political junctures, and non-democratic regimes or weak democracies. Can these findings for specific contexts be generalised? When and where do protest movements respond to social media? Can we directly explore how regime type and other circumstances shape the reaction to increased social media?

## **SOCIAL MEDIA AND PROTESTS: A GLOBAL EXPERIMENT**

In a recent study (Fergusson and Molina 2021), we tackle the challenge of exploring, at a global scale and for over one decade, the effect of social media on citizen mobilisation, the mechanisms at stake, and the broader political implications.

To do so, we use an 'experiment' that history naturally gave us: the fact that Facebook was initially launched in English around 2006, and then expanded to be available in several additional languages.

This 'experiment' is helpful because as Facebook became available in, say, French, countries and people who speak that language could access the platform more than others who did not yet have Facebook in the language they speak. This increased access reflects that internet users interpret and use the platform more efficiently in their main language. Also, even if some people can understand the platform well enough in their second language, they will likely use it more when their 'peers' (friends, politicians, businesses) enjoy greater access with the language barrier gone.

We, therefore, explore whether people with access to Facebook in their language, whom we call 'Facebook Speakers', use Facebook and protest more than those without access. Moreover, we explore whether countries with more Facebook Speakers experience more protests on aggregate and experience any positive political change.

Figure 1 summarises these findings. To illustrate the variation in our dataset, Panel A documents Facebook's expansion, showing the number of language-specific platforms launched since the initial English version in 2006. The share of speakers increases as new versions arrive, and the languages launched earlier have, on average, a stronger impact on the number of speakers than those launched later. Nevertheless, even later languages matter because in some regions a significant share of the population speaks those languages.

Panel B shows that the expansion of these language-specific platforms increases Facebook use globally and in countries with a very diverse set of observable characteristics (including diverging population levels, age distributions, urbanisation and urban growth, sex ratios, fertility rates, income per capita, education, linguistic fragmentation or polarisation, the share of English-speaking population, and measures of religiosity). Consistently, when exploiting the language spoken by survey respondents using individual-level data, the arrival of these language-specific platforms increases use among people with very distinct individual-level traits (e.g. age, sex, education, and wealth).

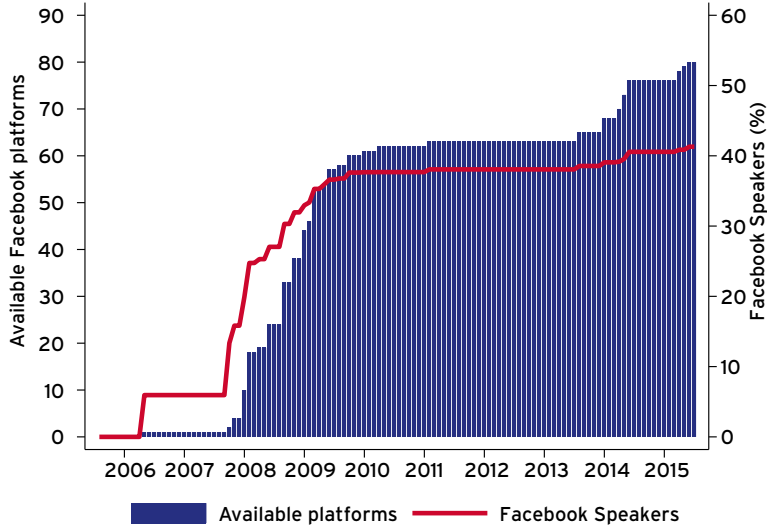
A related relevant question to assess how generalisable our effects are is by how much individuals whose Facebook engagement increases when a platform becomes available in their native language (technically, the 'compliers') differ from the 'average' person. As we show in Panel C, the answer is not much. The set of compliers is not only very diverse but also very similar to the average person.

Finally, in Panel D, we track protest activity before and after Facebook launches a new language. We observe the resulting increase in protests where there are now more Facebook speakers.<sup>2</sup> The figure shows that the expansions into new Facebook platforms increase protests by approximately 19%.

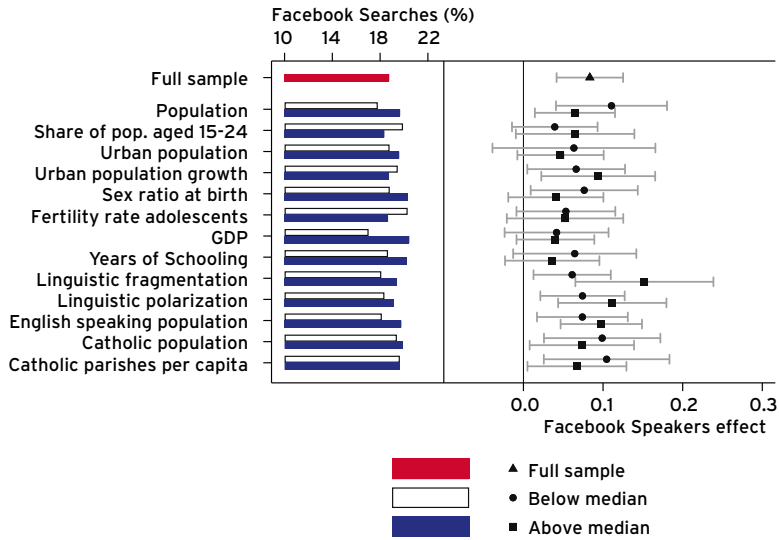
2 In the figure we report estimates following Sun and Abraham (2020) that allow for heterogeneous treatment effects across groups and time.

**FIGURE 1 FACEBOOK EXPANSION TO NEW LANGUAGES AND ITS EFFECTS ON FACEBOOK USE AND PROTESTS**

a) Facebook expansion over time

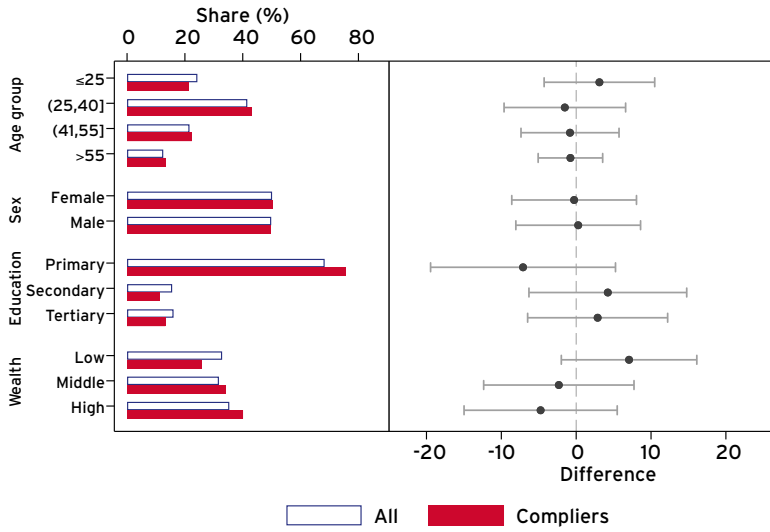


b) Effect of Facebook Speakers on Facebook use breaking the sample by countries' characteristics

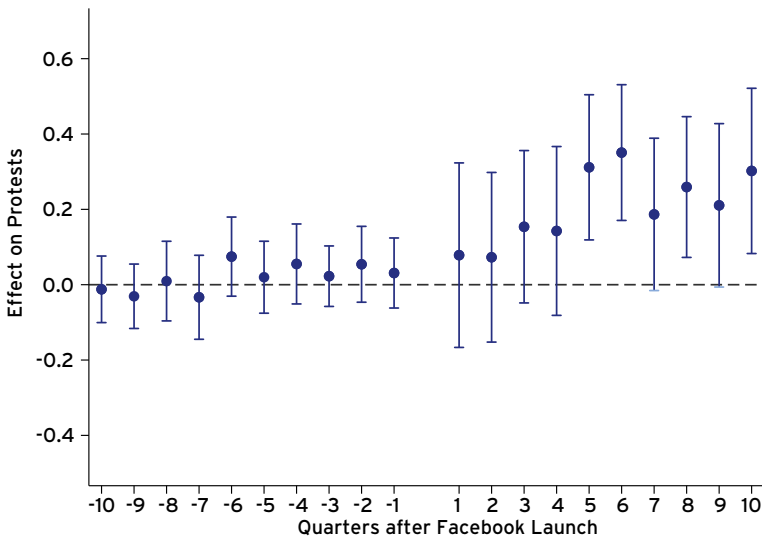




c) Distribution of characteristics of all respondents and compliers



d) Facebook Speakers' effect on protests



Note: A Facebook Speaker is a person who can access the platform in their native language.

## THE BROADER IMPLICATIONS

When examining why Facebook caused protests, it appears that it is not merely because of the one-way information it provides. Indeed, we find no impact of Facebook on a wide range of political views or the use of traditional information sources.

We also examine an essential social effect beyond the three more common categories (tactical coordination, beliefs about others, and social image) mentioned before: having a platform to talk and share opinions inspires me to think that I have freedom of expression and that I can then take that freedom (first manifested online) as protests in the offline world. Indeed, we find that Facebook speakers declare more freedom of expression than non-speakers, as reported in Panel A of Figure 2. Facebook has been a 'liberation technology' (Diamond 2010) in the sense that access to the social network increases, by an appreciable average of 10%, the chances that people report freedom to say what they think, join political organisations, vote, and state their political opinions.

Despite these effects on protests and perceived freedom of expression, many have grown increasingly disillusioned with social media after seeing how these tools not only empower people against oppressive regimes but are also used to spread fake news or empower regimes against the opposition and even against other countries.

So, is Facebook at all effective in bringing about positive political change? According to our study, largely not. We do not find any improvement in measures of governance or democracy. Facebook expansions did not noticeably increase the chances of regime change, did not improve indices of the quality of democracy, and did not improve various measures of the quality of governance. We find that the effects of Facebook Speakers on these outcomes are not just statistically not significant, but also quantitatively small.

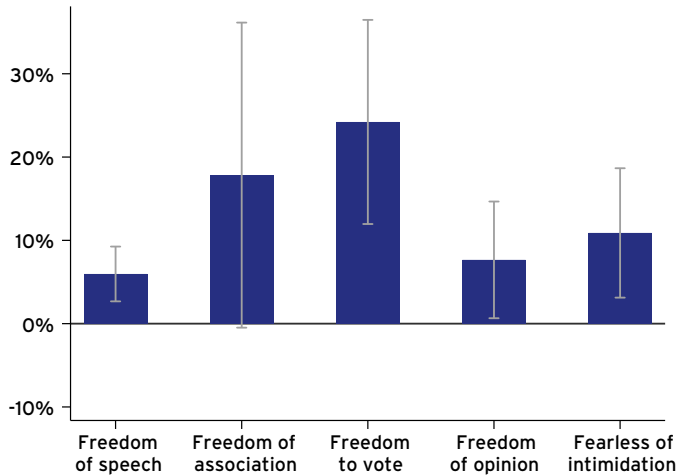
When we explore why Facebook failed to influence political regimes more broadly, we find evidence for at least three reasons. First, Facebook was actively used not just by people to protest but by governments and other powerful groups to offset the effects of citizen protests. Protests against the opposition are most important in the least democratic areas. Thus, this counteracting force limiting possible broader effects of protests appears to be relevant precisely where it matters most.

Second, although Facebook produces protests, other traditional power structures like political parties or traditional media trump its importance during critical junctures. Evidence along these lines comes from noticing that Facebook has a more limited impact on protests during electoral campaigns, as shown by Panel B of Figure 2. We interpret these results as suggestive that, during elections, traditional media presence and party organisation might substitute for Facebook's otherwise significant influence on protests.

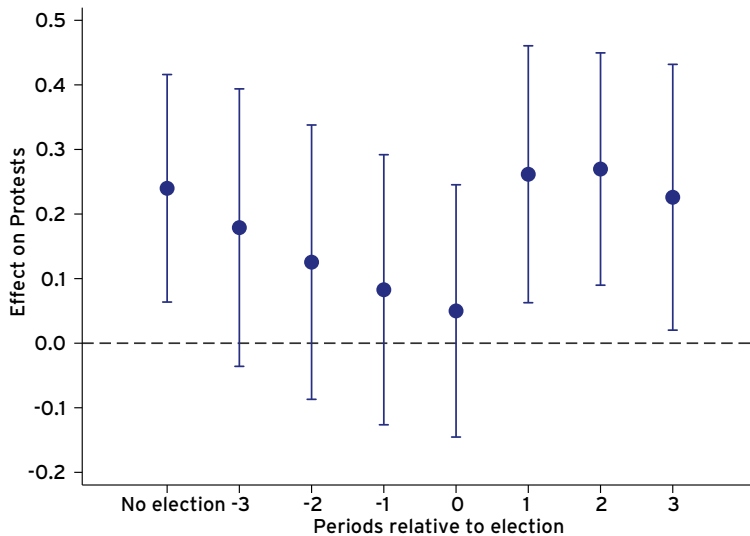
Finally, Facebook was ineffective in changing other forms of political participation. It did not increase voting, interest in and discussion of politics, participation in organisations, the signing of petitions, ability to reach politicians, participation in partisan activities, or identification with parties.

**FIGURE 2** ADDITIONAL EFFECTS OF FACEBOOK EXPANSION TO NEW LANGUAGES

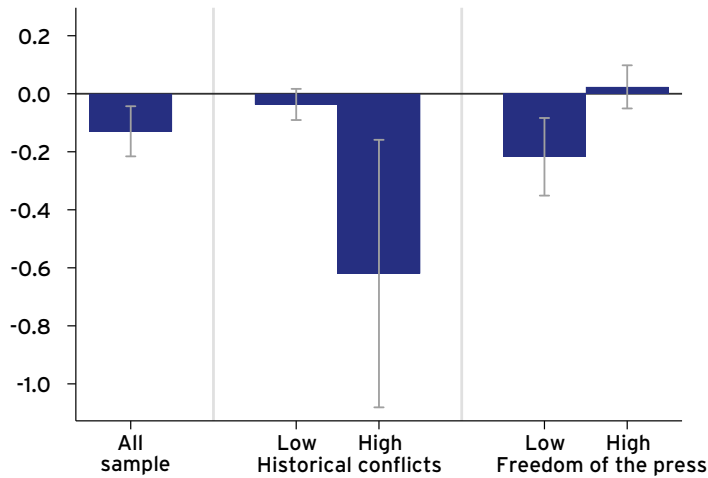
a) Facebook Speakers' effect on freedom of expression



b) Facebook Speakers' effect on protests by election timing



c) Facebook Speakers' effect on violence



While our evidence suggests that Facebook is ineffective in bringing out positive political change, we do find that Facebook is decisive in decreasing violent conflict, as shown in Panel C of Figure 2. This effect is quantitatively meaningful: a one standard deviation increase in Facebook Speakers reduces the number of violent internal conflicts by 6% of a one standard deviation. One likely reason is that Facebook itself, and the protests it generated, helped express dissent that would otherwise have turned more violent. That Facebook Speakers report an increase in their perceived freedom of expression is consistent with this explanation. Also, protests tend to increase more in countries with a history of protests. In turn, and as the figure shows, the decrease in internal conflict is driven by countries with a history of persistent conflict. This suggests that, in areas with more underlying reasons for conflict, Facebook's protests help voice discontent that would otherwise turn violent.

Second, social media helps make violence more visible to the world and to relevant third parties like journalists, human rights organisations, and others who may help deter violence (Durante and Zhuravskaya 2018). Indeed, the decrease in conflict occurs in countries without good access to public information or free media, precisely where Facebook should be more important to increase exposure.

## CONCLUSIONS AND IMPLICATIONS

The deceptively simple questions of whether social media causes social mobilisation and whether it produces any positive political change have relatively nuanced answers.

Yes, it causes protests, largely because of the social nature of social media beyond the one-way information transmission. Our study examining the impacts of Facebook globally also implies an empowering liberation effect: Facebook increased perceived political freedom (freedom to vote, to associate, and to express views without intimidation).

However, on average, Facebook has been largely ineffective in producing positive political change. One reason is that these are tools that can be used to repress change as much as they can produce it and because protests have not always transformed into other forms of political participation at crucial times. Instead, Facebook's expansion reduced conflict, possibly because it helped voice discontent that would otherwise have turned violent and deterred violence by increasing visibility.

Some of our findings are relevant in practice. For instance, some countries have attempted to include regulations to control information on social platforms. The finding that these platforms matter mostly for their coordination effects rather than the pure one-way information effect suggests some of these policies may be ineffective.

More broadly, policymakers should recognise that, as with any general-purpose technology, social media has many effects that change over time as different players evolve and adapt (Tucker et al. 2017). Our study confirms that the effects of social media on mobilisation are different for countries with different political institutions as well as for different actors within similar political institutions.

## REFERENCES

- Barbera, S and M O Jackson (2020), "A model of protests, revolution, and information", *Quarterly Journal of Political Science* 15 (1): 297–335.
- Cantoni, D, D Y Yang, N Yuchtman and Y J Zhang (2019), "Protests as strategic games: Experimental evidence from Hong Kong's antiauthoritarian movement", *The Quarterly Journal of Economics* 134 (2): 1021–1077.
- Christensen, D and F Garfias (2018), "Can you hear me now? How communication technology affects protest and repression", *Quarterly Journal of Political Science* 13(1): 89–117.
- Diamond, L (2010), "Liberation technology", *Journal of Democracy* 21: 69.
- Durante, R and E Zhuravskaya (2018), "Attack when the world is not watching? US news and the Israeli-Palestinian conflict", *Journal of Political Economy* 126 (3): 1085–1133.
- Edmond, C (2013), "Information manipulation, coordination, and regime change", *Review of Economic Studies* 80 (4): 1422–1458.
- Enikolopov, R, A Makarin and M Petrova (2020), "Social media and protest participation: Evidence from Russia", *Econometrica* 88 (4): 1479–1514.
- Enikolopov, R, A Makarin, M Petrova and L Polishchuk (2017), "Social image, networks, and protest participation", available at SSRN.
- Fergusson, L and C Molina (2021), "Facebook causes protests", Documento CEDE 018002, Universidad de Los Andes.

Jost, J T, P Barberá, R Bonneau et al. (2018), “How social media facilitates political protest: Information, motivation, and social networks”, *Political Psychology* 39 : 85–118.

Little, A T (2016), “Communication technology and protest”, *The Journal of Politics* 78 (1): 152–166.

Manacorda, M and A Tesei (2020), “Liberation technology: Mobile phones and political mobilization in Africa”, *Econometrica* 88 (2): 533–567.

Qin, B, D Strömberg and Y Wu (2021), “Social media and collective action in China”, CEPR Discussion Paper 16731.

Sun, L and S Abraham (2020), “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects”, *Journal of Econometrics* 225(2): 175–199.

Tucker, J, M Metzger, D Penfold-Brown, R Bonneau, J Jost and J Nagler (2015), “Protest in the age of social media”, *Medium*, 8 January.

Tucker, J A, Y Theodoris, M E Roberts and P Barberá (2017), “From liberation to turmoil: social media and democracy”, *Journal of Democracy* 28 (4): 46–59.

Tufekci, Z and C Wilson (2012), “Social media and the decision to participate in political protest: Observations from Tahrir Square”, *Journal of Communication* 62(2): 363–379.

Valenzuela, S (2013), “Unpacking the use of social media for protest behavior: The roles of information, opinion expression, and activism”, *American Behavioral Scientist* 57(7): 920–942.

Valenzuela, S, A Arriagada and A Scherman (2012), “The social media basis of youth protest behavior: The case of Chile”, *Journal of Communication* 62 (2): 299–314.

## ABOUT THE AUTHORS

**Leopoldo Fergusson** (PhD, MIT) is a Professor of Economics at Universidad de los Andes in Bogotá, Colombia, where he also directs the Center of Studies on Economic Development (CEDE). He has held visiting positions at MIT, the University of Chicago, Harvard, and Universitat de Barcelona. His academic work focuses on political economy and development, covering media, state capacity, conflict, and inequality. He co-leads the initiative Teaching and Researching Equitable Economics from the South (TREES) and is the scientific director of the Evidence in Governance and Politics (EGAP) Latin American Hub.

**Carlos Molina** is a PhD candidate in Economics at the Massachusetts Institute of Technology. He received a BA in Economics at the Universidad de Antioquia in 2014 and an MA in Economics at the Universidad de Los Andes in 2017. His research lies at the intersection of political economy and development economics.



# CHAPTER 12

## Liberation technology: Mobile phones and political mobilisation in Africa

143

**Marco Manacorda and Andrea Tesei**

Queen Mary University of London and CEPR

Can digital information and communication technology foster mass political mobilisation? In our 2020 piece (Manacorda and Tesei 2020), we use a variety of georeferenced data for the whole of Africa covering a 15-year span to investigate this question and explore channels of impact.

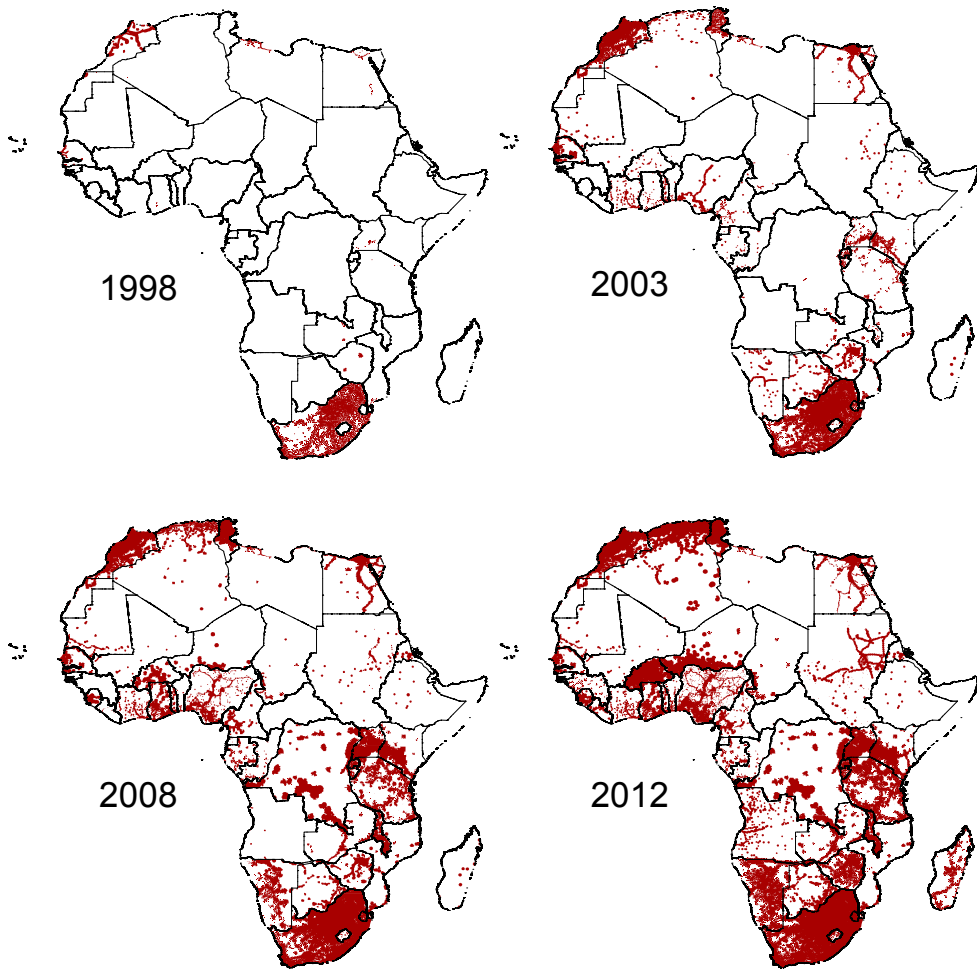
The spread of digital information and communication technology (ICT) has fed a wave of optimism about its use as a ‘liberation technology’ capable of helping the oppressed and disenfranchised around the world. According to this argument, popularised by political sociologists and media scholars alike (Diamond 2010, Shirky 2011), mobile phones and the internet, thanks to the opportunity they offer for two-way, multi-way and mass communication, in addition to their low-cost, decentralised, open-access nature, have the potential to foster citizens’ political activism and even lead to mass political mobilisation, especially when civic forms of political participation are de facto or lawfully prevented.

Despite the popularity of this argument, credible empirical evidence of the effect of ICT, and in particular of mobile phones, on political mobilisation is scant. The channels of impact are also not well understood. With the exception of a few studies that focus on the role of the internet and social media in protest participation (Acemoglu et al. 2018 for Egypt, Enikolopov et al. 2020 for Russia), a large body of research has focused on the effect of traditional media and the internet on civic forms of participation such as voting (Gentzkow 2006, Falck et al. 2014).

In this chapter, we study the role played by mobile phones on political mobilisation across the whole of Africa, one of the continents with the fastest rates of adoption in mobile phone technology (Figure 1) and theatre to some of the most spectacular episodes of mobilisation in recent years (food riots swept the continent between 2007 and 2008 and mass civil unrest – the Arab Spring – erupted in the northern countries between 2010 and 2012). Importantly, mobile phone technology adoption in many countries in the continent happened against the backdrop of a practically non-existent fixed line infrastructure and, because of this, this is claimed to have had unprecedented consequences on the life of African citizens, especially the poor (Aker and Mbiti 2010).



FIGURE 1 2G MOBILE NETWORK ROLLOUT: AFRICA 1998-2012



Notes: The figure reports georeferenced data on 2G mobile phone coverage for all of Africa at five-year intervals between 1998 and 2012.

Source: GSMA.

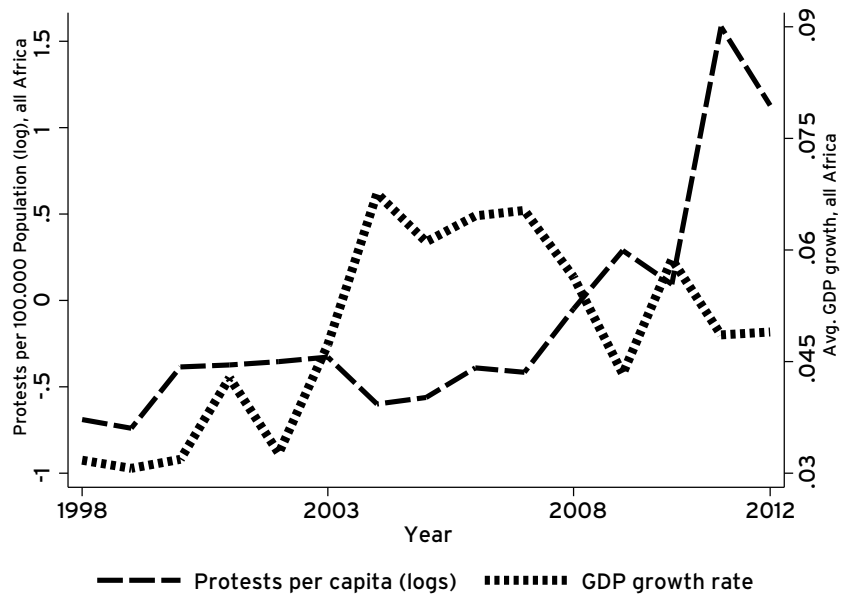
*Increased information* and communication enabled by mobile phones have the potential to trigger collective action through information spreading and acquisition. By granting access to unadulterated information, digital ICT also has the potential to offset government propaganda, which can curb discontent via misinformation and persuasion, especially when media are under the control of the government or other interest groups (Edmond 2013).

This effect is reinforced when strategic complementarities are at play. In particular, when the returns to political activism increase or the costs of participation decrease with the number of others participating (Barbera and Jackson 2017, Passarelli and Tabellini 2017), mobile phone technology can also foster mass mobilisation through *increased coordination*.

Despite there being good reasons to speculate that mobile phones can foster political activism, there are also good reasons to believe the opposite. Governments can use this technology as a surveillance or propaganda tool, hence making protests less – rather than more – likely (Morozov 2012) and digital ICT can discourage social capital accumulation and the establishment of ‘strong ties’ that are instrumental to mass mobilisation (Gladwell 2010), ultimately leading to political apathy. Hence, whether mobile phone availability fosters or discourages protest activity remains an open empirical question.

An additional consideration is that while the ‘liberation technology’ argument suggests that protests should arise in response to the availability of mobile phones, an established body of evidence shows that protests do not happen in a void and tend to be particularly likely during economic downturns, something that is also clearly evident in our data (Figure 2). This is because worse economic conditions both reduce the opportunity costs of protest participation and provide reasons for grievances. This suggests that while mobile phones may play a role in fostering protest provision, this effect would be expected to be stronger during recessions, when an independent trigger for protests exists.

**FIGURE 2 GDP GROWTH AND PROTESTS PER CAPITA**



Notes: The figure reports continent-wide log protests per 100,000 individuals (dashed line) and the rate of GDP growth (dotted line) as a function of time. Continent-wide GDP growth is obtained as a population-weighted average of GDP growth in each country.

In order to address these questions, we combine data on local 2G mobile phone coverage from the Global System for Mobile Communications Association (GSMA) with data on the occurrence of protest events, all based on compilations of news wires, for – effectively – the entire continent between 1998 and 2012. The advantage of these data is their level of geographical detail. GSMA data provide information on mobile phone signal availability at a level of geographical precision of between approximately 1 km<sup>2</sup> and 23 km<sup>2</sup> on the ground, depending on the country. Protest data from the Global Database on Events, Location and Tone (GDELT) (Leetaru and Schrodtt 2013) – a large, open-source data set relying on an automated textual analysis of news sources – provide precise coordinates on the location of protest events. Additional manually compiled data on protest occurrence that we use to corroborate data from GDELT – the Armed Conflict Location and Event Data Project, or ACLED (Raleigh et al. 2010) and the Climate Change and African Political Stability (CCPAS) Social Conflict Analysis Database, or SCAD (Salehyan et al. 2012) – also provide precise coordinates of such events.

The very detailed level of geographical disaggregation of the data allows us to compare changes in the incidence of protests across very small areas within the same country. This alleviates the obvious concern that our estimates of impact capture a spurious correlation between mobile phone coverage and protests across countries and time. Moreover, the high level of geographical details and the relatively long time span allow us to estimate differential effects across areas with different characteristics as well as at different points along the business cycle.

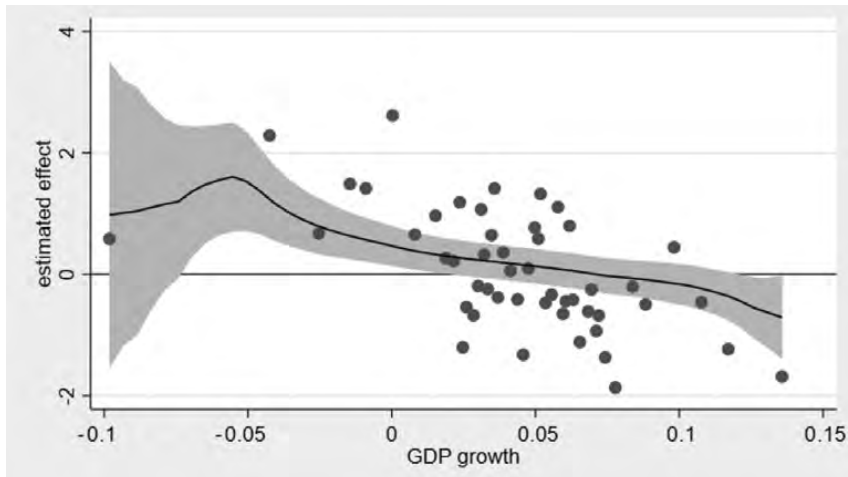
In our analysis we include a very large set of area-specific controls, yet one might still be concerned that protests and mobile phone adoption are spuriously correlated. To circumvent this problem, we propose and implement a novel instrumental variable strategy that exploits the fact that frequent electrostatic discharges during storms damage mobile phone infrastructures and negatively affect connectivity (Andersen et al. 2012),<sup>1</sup> hence discouraging adoption. Using NASA satellite-generated data, we show indeed that areas with higher than average incidence of lightning strikes display slower adoption of mobile phone technology over the period. We hence use the variation on coverage adoption induced by differential levels of lightning activity across areas to derive causal estimate of the impact of mobile phones on protest activity.

In line with the nuanced version of the ‘liberation technology’ argument, we find strong and robust evidence that mobile phones are instrumental to mass political mobilisation, although this only occurs during economic downturns. In order to get a sense of magnitude, our estimates imply that a one standard deviation fall in GDP growth (approximately 4 percentage points) leads to a differential increase in protests per capita between an area with full mobile phone coverage compared to an area with no coverage of between 8% and 23%, depending on the measure of protest used. Effects manifest

1 International Telecommunication Union, “Emission levels and test methods for wireline telecommunication networks to minimize electromagnetic disturbance of radio services” (<https://www.itu.int/rec/T-REC-K.60/en>).

exclusively during recessions; we find no effects during good economic times. This is clearly evident in Figure 3, which shows the relationship between protests and predicted mobile phone coverage (as induced by our instrument) at different levels of GDP growth.

**FIGURE 3 NON-PARAMETRIC ESTIMATES OF THE RELATIONSHIP BETWEEN GDP GROWTH AND PROTESTS**



Notes: The figure reports 2SLS estimates of the effect of coverage on protests by pairs of percentiles of GDP growth distribution, estimated non-parametrically. Point estimates are reported in the figures as dots. We superimpose a kernel-weighted local polynomial regression where each observation is weighted by the inverse of the square of the standard error of the associated estimate. We use a polynomial of degree 0 and an Epanechnikov kernel function, with a 'rule-of-thumb' bandwidth. The graph reports this estimated regression fit as well as the 95 percent confidence interval around the prediction.

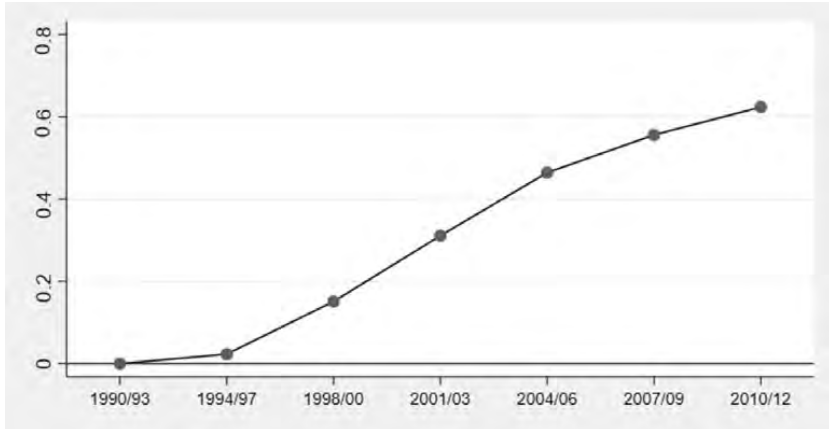
We also uncover important dimensions of heterogeneity, with the effects being particularly pronounced in urban areas, in areas with a legacy of conflict, in non-democratic countries, and when traditional media are captured by the state.

In order to corroborate a casual interpretation for our estimates, we show in Figure 4 that our instrument is uncorrelated with protests in periods when mobile phone technology was unavailable, suggesting that the effect of lightning strikes on protest is credibly attributable to the mediating effect of mobile phone coverage.

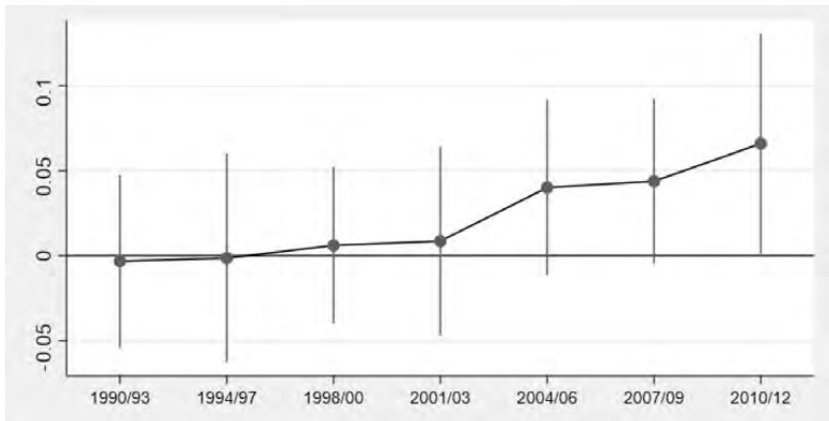
We complement the analysis with microdata from the Afrobarometer, which provide information on individual protest participation as well as precise area of residence for a subset of countries/years. Remarkably, results based on micro data closely mimic those obtained using information on protest occurrence from newswires. These data also allow us to shed some light on the mechanisms of impact and quantify their role. In line with the two mechanisms of impact highlighted above – increased information and increased coordination – we show that both are at a play, each accounting for about half of the reduced-form effect we uncover.

**FIGURE 4 RELATIONSHIP BETWEEN LIGHTNING STRIKES AND PROTESTS AS 2G MOBILE NETWORK ROLLS OUT**

a) Share of African population covered by 2G mobile signal



b) Reduced-form relationship between lightning strikes and protests



Notes: Panel A reports the continental trend in mobile phone coverage by three-year sub-periods. Coverage is set at 0 for all cells before 1995 (the year in which mobile phone technology was first introduced in Africa) and is linearly interpolated at the cell level between 1995 and 1998 (the first year in our data). Panel B reports the estimated coefficients from the reduced-form regression of log protests per 100,000 people (winsorized at the 99th percentile) on the instrument by three-year sub-periods and the corresponding 90 percent confidence intervals.

In sum, our analysis suggests that ICT, and in particular the ubiquitous use of mobile phones for calling and messaging, indeed helped promote mass mobilisation in Africa, especially when reasons for grievance existed and citizens had reasons to blame the government for the poor state of the economy.

Of course, our analysis refers to the period when 2G technology spread fast around the world and it largely predates the period of advent of 3G to 5G technologies, which have increasingly allowed for access to email, mobile internet and social media.

The question then is whether the effects we uncover during the transition to full mobile phone coverage are still at a play, and whether the availability of new technologies has had a different effect on citizens' political participation as well as government responses.

A growing body of research has recently turned to studying the effect of mobile internet on voting outcomes. In particular, Guriev et al. (2021) show that 3G mobile signal has reduced citizens' confidence in governments worldwide and reduced chances of the incumbent's re-election, while Manacorda et al. (2022) show that these technologies have increased political tribalism and communitarianism in Europe, with an ensuing support for right-wing populist parties.

## REFERENCES

Acemoglu, D, T A Hassan and A Tahoun (2018), "The Power of the Street: Evidence From Egypt's Arab Spring", *The Review of Financial Studies* 31(1): 1-42.

Aker, J C, and I M Mbiti (2010), "Mobile Phones and Economic Development in Africa", *Journal of Economic Perspectives* 24(3): 207-232.

Andersen, T B, J Bentzen, C-J Dalgaard, and P Selaya (2012), "Lightning, IT Diffusion, and Economic Growth Across US States", *Review of Economics and Statistics* 94(4): 903-924.

Barbera, S and M Jackson (2017), "A Model of Protests, Revolution, and Information", HiCN Working Paper No. 243.

Diamond, L (2010), "Liberation Technology", *Journal of Democracy* 21(3): 69-83

Edmond, C (2013), "Information Manipulation, Coordination, and Regime Change", *Review of Economic Studies* 80(4): 1422-1458.

Enikolopov, R, A Makarin, and M Petrova (2020), "Social Media and Protest Participation: Evidence from Russia", *Econometrica* 88(4): 1479-1514.

Falck, O, R Gold, and S Heblich (2014), "E-Lectons: Voting Behavior and the Internet", *American Economic Review* 104(7): 2238-2265.

Gentzkow, M (2006), "Television and Voter Turnout", *Quarterly Journal of Economics* 121(3): 931-972.

Gladwell, M (2010), "Small Change", *The New Yorker*, 27 September.

Guriev, S, N Melnikov, and E Zhuravskaya (2021), "3G Internet and Confidence in Government", *The Quarterly Journal of Economics* 136(4): 2533-2613.

Leetaru, K, and P A Schrodtt (2013), "GDELT: Global Data on Events, Location, and Tone, 1979-2012", International Studies Association Annual Conference.

Manacorda, M and A Tesei (2020), “Liberation Technology: Mobile Phones and Political Mobilization in Africa”, *Econometrica* 88(2): 533–567.

Manacorda, M Tabellini, G and A Tesei (2022) “Mobile Internet and the Rise of Political Tribalism in Europe”, CEP Discussion Paper No. 1877.

Morozov, E (2012), *The Net Delusion: The Dark Side of Internet Freedom*, Public Affairs.

Passarelli, F and G Tabellini (2017), “Emotions and Political Unrest”, *Journal of Political Economy* 125(3): 903–946.

Raleigh, C, A Linke, H Hegre, and J Karlsen (2010), “Introducing ACLED—Armed Conflict Location and Event Dataset”, *Journal of Peace Research* 47(5): 651–660.

Salehyan, I, C S Hendrix, J Hamner, C Case, C Linebarger, E Stull, and J Williams (2012), “Social Conflict in Africa: A New Database”, *International Interactions* 38(4): 503–511.

Shirky, C (2011), “The Political Power of Social Media”, *Foreign Affairs* 90(1): 28–41.

## ABOUT THE AUTHORS

**Marco Manacorda** is a Professor of Economics and Public Policy at Queen Mary University of London, a CEPR Research Fellow and a managing co-editor of *The Economic Journal*. His interests and expertise are at the intersection of political economy, labour and development economics.

**Andrea Tesei** is Associate Professor of Economics at Queen Mary University of London and a CEPR Research Affiliate. His research interests are in political economy and economic development. His research has been published in *The American Economic Review*, *Econometrica* and *The Review of Economics and Statistics*, among others.

## **SECTION 5**

# **SOCIAL MEDIA AND HATRED**





# CHAPTER 13

## Social media and Xenophobia: Evidence from Russia

153

**Leonardo Bursztyn,<sup>a</sup> Georgy Egorov,<sup>b</sup> Ruben Enikolopov<sup>cd</sup> and Maria Petrova<sup>cd</sup>**

<sup>a</sup>University of Chicago; <sup>b</sup>Northwestern University; <sup>c</sup>Universitat Pompeu Fabra; <sup>d</sup>CEPR

One of the important roles of social media is to help people find individuals ‘like themselves’ and become part of a group or community. Indeed, existing evidence indicates that the internet can make it easier to meet like-minded people (van Alstyne and Brynjolfsson 2005, Putnam 2000, Sunstein 2017), and that social media tends to reinforce this effect (see Barberá 2020 for a recent overview). One of the main distinguishing features of social media is the existence of very low barriers to entry (Zhuravskaya et al. 2020), which dramatically increases the number of users of all different kinds. This allows individuals to find groups and communities based on their interests, even if these are fringe and unusual. The existing literature has examined how social media communities can be a positive force – for example, they help with the coordination of activities, ranging from leisure to pro-democratic political rallies (Enikolopov et al. 2020). However, helping find like-minded people can also have negative consequences. Already in 2001, Sunstein (2001) argued that these types of discussions between individuals with similar opinions could become a “breeding ground for extremism” (p. 71). Imagine those who hold an extreme opinion, such as xenophobia. For such individuals, finding like-minded people in the real world might be a difficult task, so social media would be particularly helpful in connecting people holding similar fringe views.

There is a small, but growing strand of literature that examines the effect of traditional and social media on the promotion of hate crimes and genocides. In this line of research, most papers document an immediate effect of posts on social media on hateful actions. For example, Müller and Schwarz (2021) find that anti-refugee sentiment on Facebook predicts daily changes in crimes against refugees in Germany. Similarly, Müller and Schwarz (forthcoming) find that anti-Muslim hate crimes in the United States have gone up in counties with a high penetration of Twitter users, but only since the start of Donald Trump’s presidential campaign.<sup>1</sup>

<sup>1</sup> Other studies have instead focused on the effect of other, traditional mass media on violence, including Yanagizawa-Drott (2014) on the Rwandan genocide, DellaVigna et al. (2014) on the war in the former Yugoslavia, and Adena et al. (2015) on antisemitic violence in Germany in the 1930s.

Our prior paper (Bursztyn et al. 2020) complements the literature by focusing on the longer-term, causal effects of the proliferation of social media on xenophobic attitudes and ethnic hate crimes. Identifying a causal effect of social media is challenging, since both access to and consumption of social media are not randomly assigned. In our paper, we thus follow the approach of Enikolopov et al. (2020), which exploits the roll-out of the main Russian social media platform, VKontakte (VK). This online social network was the first mover in the Russian market and kept a dominant position for many years, with a user share of over 90% by 2011. VK was launched in October 2006 by Pavel Durov, an undergraduate student at Saint Petersburg State University (SPbSU) at the time. Initially, an invitation through a university forum curated by Durov was required to join the platform. As a result, a very large share of early users of VK were Durov's classmates. Friends and relatives of these students were therefore more likely to join the platform earlier on, once open registration started. The idiosyncratic variation in the distribution of the cities of origin of Durov's classmates therefore seeded the initial conditions for city-level penetration, which in turn could have led to long-term effects. Following this logic, Enikolopov et al. (2020) use fluctuations in the distribution of SPbSU students across cities as an instrument for the city-level penetration of VK. Using this approach, we evaluate the effect of higher VK penetration on attitudes and hate crimes towards other ethnicities, combining existing survey data, a newly collected dataset on hate attitudes from a survey experiment, and data on hate crimes, between 2007 and 2015.

We show that exposure of individuals in a city to social media increases the share of individuals with extreme positions, such as xenophobia. This increased prevalence of extremists can itself increase hate crimes, but this is especially true in cities with a high pre-existing level of xenophobia (because social media's ability to connect like-minded people falls on a fertile ground) and for crimes with multiple perpetrators (because coordination is particularly important for such crimes, although online social groups could provide valuable information or nudging even for crimes committed by single individuals). At the same time, perhaps counterintuitively, the increase in the share of extremists does not necessarily increase the share of people who openly agree with these extreme opinions; on the contrary, the share of people who hide these extreme opinions is likely to increase. The reason for this is that social media increases the share of extreme views on both sides of the spectrum (so there are more very tolerant people as well), thus increasing the importance of social image and stigma from expressing fringe opinions.

Guided by a theoretical framework, we designed and conducted an online survey experiment in the summer of 2018, with over 4,000 respondents from 124 cities. Given the potential for social desirability bias in the elicitation of sensitive opinions, a 'list experiment' was used. Under this approach, respondents are asked to indicate the number of statements with which they agree from a list, but not *which* ones. By adding the statement of interest to a random subgroup of respondents, one can estimate the share of respondents agreeing with this statement without being able to identify who

exactly agrees with it (except for those who agree with all the statements). The statement of interest, borrowed from existing surveys, was “I feel annoyance or dislike toward some ethnicities.”

Using this approach, we find a positive effect of social media penetration on elicited ethnic hostility, i.e. the share of respondents that hold xenophobic attitudes, both using the ordinary least squares (OLS) and the instrumental variables (IV) approach. Under the IV approach, for example, a 10% increase in social media penetration raises the share of people agreeing with the xenophobic statement by 9.5 percentage points. Moreover, the magnitude of this effect is larger for those with lower education (17.3 percentage points) compared with those with higher education (6.9 percentage points), and for younger people (21.3 percentage points) compared with older people (4.3 percentage points, not statistically significant at conventional levels).

To understand expressions of xenophobia, the survey asked the subjects in the control group (i.e. those randomly assigned to the list not containing the statement about xenophobia) a direct question on whether they agreed with the statement. What is the effect of social media on the expression of ethnic hostility without the cover provided by the list experiment? Any difference across the two elicitation methods would give a measure of stigma associated with the expression of xenophobic opinions – at least in a survey context. The two methods indeed led to different rates of agreement with the statement: approximately 38% for the list experiment and 33% for direct elicitation. This indicates the potential presence of stigma associated with the expression of xenophobic opinions in the survey.

What is the effect of social media penetration on such stigma? The findings indicate a positive effect: a 10% increase in social media penetration on average increases the percentage of people unwilling to admit to xenophobia by 11.6%.

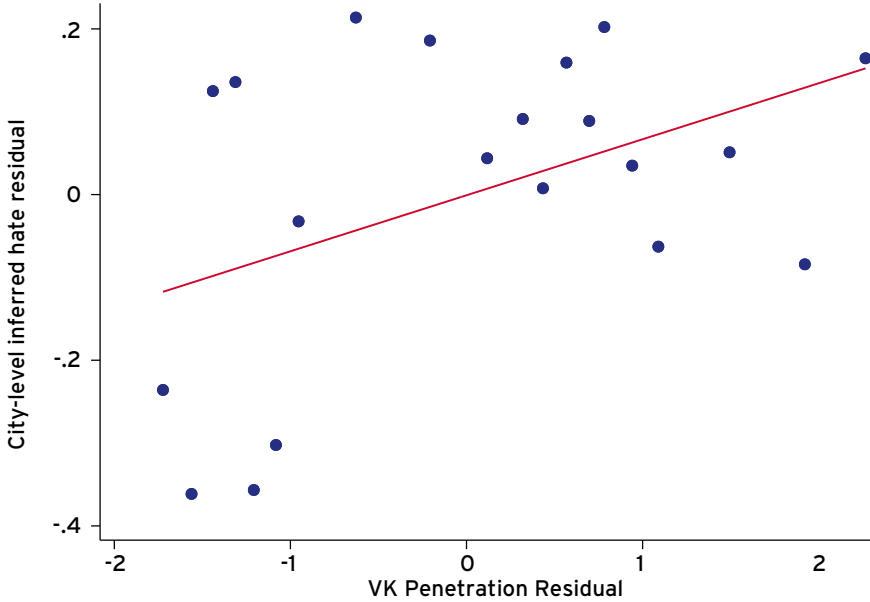
The intuition behind this effect is developed in the theoretical framework in our paper. There, proliferation of social media increases citizens’ propensity to interact with like-minded individuals, which is shown to increase polarisation. This increase in polarisation implies that more people hold fringe beliefs, and in particular more are xenophobic. At the same time, a more polarised society makes individuals more likely to be judged by extreme individuals on both sides of the spectrum, which increases social image concerns. The model shows that this effect may be strong enough to decrease expression of xenophobia despite an increase in the number of people holding such views, and in fact dominates under some simple distributional assumptions.

TABLE 1 SOCIAL MEDIA AND ETHNIC HOSTILITY, ELICITED FROM LIST EXPERIMENT

Subsample	List experiment elicited hostility						
	Full sample (1)	Male (2)	Female (3)	Low education (4)	High education (5)	Younger (6)	Older (7)
Log (number of VK users 2011)	0.950***	0.758**	0.997***	1.734***	0.693***	2.103***	0.427
Weak instrument robust confidence 95% sets	(-0.26, 2.037) [0.277]	(-.585, 2.101) [0.343]	(-.205, 2.199) [0.307]	(.008, 3.699) [0.501]	(-.354, 1.739) [0.267]	(.676, 4.280) [0.548]	(.685, 1.539) [0.248]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Younger/older SPbSU student cohorts	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	124	122	116	124	111	121	116

Notes: Results of the two-sample two-stage least squares estimation. Unit of observation is a city. Robust standard errors in square brackets. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic controls include fifth polynomial of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummies for regional and county centres, distances to Moscow and St.Petersburg, log (average wage in 2011), dummy for the existence of a university in a city, internet penetration in 2011, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census, and nationalistic party vote share in 2003 (pre-social media).

**FIGURE 1 VK PENETRATION AND INFERRED HATE**



Notes: This figure presents binned scatter plot for the OLS version of the relationship between VK penetration and city-level inferred hate, as reported in column (1) of Table 2. City-level inferred hate is computed as the difference in means in the number of options chosen by the respondents in treatment and control groups.

What is the effect of social penetration on hate crimes? Data on hate crimes come from a database maintained by SOVA Center, a Moscow-based Russian independent nonprofit organisation. The dataset covers incidents of violent hate crime, which include murders, assaults, batteries, and death threats, and has been collected consistently since 2007. We focus on the period 2007 to 2015.

We find an effect of social media penetration on hate crime, driven by cities with higher pre-existing level of nationalism. Pre-existing nationalist sentiment at the city level is proxied by the vote share of the Rodina ('Motherland') party in the parliamentary election of December 2003, the only election this party participated in and also the last parliamentary election before the creation of VK. The heterogeneity by pre-existing nationalism is especially strong for hate crimes conducted by multiple perpetrators. Numerically, the results imply that the effect of a one standard deviation increase in social media penetration ranges from being close to zero at the minimum level of nationalist party support to a 21.7% increase in the total number of hate crime victims at the maximum level of nationalist support.

TABLE 2 SOCIAL MEDIA, HATE CRIME AND PRE-EXISTING NATIONALISM, 2007-2015

	Log (# of hate crimes)			Log (# of ethnic hate crimes)			Log (# of non-ethnic hate crimes)		
	Total (1)	Single perpetrator (2)	Multiple perpetrators (3)	Total (4)	Single perpetrator (5)	Multiple perpetrators (6)	Total (7)	Single perpetrator (8)	Multiple perpetrators (9)
Log (number of VK users), 2011 x nationalist party support in 2003	10.034** (2.955, 20.653) [4.334]	5.334 (2.391, 17.0106) [3.604]	9.324** (2.709, 19.246) [4.050]	8.889** (2.524, 18.438) [3.897]	4.575* (.141, 13.443) [2.715]	8.066** (1.520, 17.885) [4.008]	6.356 (2.852, 20.371) [4.291]	1.373 (-2.346, 6.952) [2.277]	5.313 (2.565, 16.304) [3.365]
Log (number of VK users), 2011	0.052 (-.430, .774) [0.295]	0.195 (-.209, 1.004) [0.247]	0.004 (-.720, .729) [0.296]	-0.012 (-.455, .653) [0.271]	0.250 (-.098, .947) [0.213]	-0.149 (-.059, .324) [0.290]	0.298 (-.232, 1.331) [0.319]	0.039 (-.250, .471) [0.176]	0.363 (-.078, 1.243) [1.269]
Weak instrument robust confidence 95% sets	4.578*** [2.259]	1.685 [1.227]	4.322* [2.218]	4.297** [1.968]	0.942 [0.878]	4.418** [2.124]	1.475 [1.531]	0.516 [0.949]	0.676 [1.214]
Nationalist party support in 2003	-7.107 [5.568]	-4.764 [4.102]	-6.613 [5.128]	-6.289 [4.835]	-4.191 [3.221]	-4.634 [4.864]	-6.164 [5.2223]	-1.608 [2.572]	-5.903 [4.462]
Log (SPbSU students), one cohort younger than VK founder x nationalist party support in 2003	-2.178 [2.823]	0.203 [2.282]	-3.370 [2.630]	-1.625 [2.358]	0.216 [1.628]	-2.887 [2.282]	-1.709 [2.444]	0.428 [1.408]	-1.790 [1.895]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	625	625	625	625	625	625	625	625	625
Kleibergen-Paap F-statistics	6.449	6.449	6.449	6.449	6.449	6.449	6.449	6.449	6.449
Full effect at min. level of nationalist party	-0.429 (-1.160, .058)	-0.060 (-.576, .455)	-0.443 (-1.165, 0.39)	-0.438* (-1.035, -.039)	0.031 (-.243, .441)	-0.536** (-1.194, -.098)	-0.016 (-.501, .711)	-0.027 (-.468, .414)	0.108 (-.298, .718)
Weak instrument robust confidence 95% sets	2.168** (.452, 4.742)	1.320 (-.181, 4.323)	1.971** (.349, 4.402)	1.863* (.259, 4.269)	1.215* (.032, 3.580)	1.552 (-.093, 4.018)	1.629 (.747, 5.158)	0.328 (-.579, 1.689)	1.483* (.782, 4.289)

Notes: Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls include logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional centre, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census. Robust standard errors clustered by region in brackets. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Taken together, the findings in our paper and other recent papers indicate the potential for social media to increase hate and extremism – from attitudes to crimes. These results add to a growing body of evidence suggesting that social media is a complex phenomenon that has both positive and negative effects on the welfare of people (Allcott et al. 2020, Braghieri et al. 2022).

## REFERENCES

Allcott, H, L Braghieri, S Eichmeyer, and M Gentzkow (2020), “The Welfare Effects of Social Media”, *American Economic Review* 110(3): 629–76.

Adena, M, R Enikolopov, M Petrova, V Santarosa, and E Zhuravskaya (2015), “Radio and the Rise of the Nazis in Prewar Germany”, *Quarterly Journal of Economics* 130(4): 1885–1939.

Barberá, P (2020), “Social Media, Echo Chambers, and Political Polarization,” in N Persily and J A Tucker (eds), *Social Media and Democracy: The State of the Field, Prospects for Reform*, Cambridge University Press.

Braghieri, L, R Levy, and A Makarin (2022), “Social media and mental health”, *American Economic Review* 112(11): 3660–3693.

Bursztyn, L, G Egorov, R Enikolopov and M Petrova (2020), “Social Media and Xenophobia: Evidence from Russia”, CEPR Discussion Paper No. 14877.

DellaVigna, S, R Enikolopov, V Mironova, M Petrova, and E Zhuravskaya (2014), “Cross-border media and nationalism: Evidence from Serbian radio in Croatia”, *American Economic Journal: Applied Economics* 6(3): 103–32.

Enikolopov, R, A Makarin, and M Petrova (2020), “Social Media and Protest Participation: Evidence from Russia”, *Econometrica* 88: 1479–1514.

Müller, K and C R Schwarz (2021), “Fanning the Flames of Hate: Social Media and Hate Crime”, *Journal of the European Economic Association* 19(4): 2131–2167.

Müller, K and C R Schwarz (forthcoming), “From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment”, *American Economic Journal: Applied Economics*.

Putnam, R D (2000), *Bowling Alone: The Collapse and Revival of American Community*, Simon & Schuster.

Satyanath, S, N Voigtländer, and H-J Voth (2017), “Bowling for fascism: Social capital and the rise of the Nazi Party”, *Journal of Political Economy* 125(2): 478–526.

Sunstein, C (2001), *Republic.Com*, Princeton University Press.

Sunstein, C (2017), *#Republic: Divided Democracy in the Age of Social Media*, Princeton University Press.



van Alstyne, M and E Brynjolfsson (2005), “Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities”, *Management Science* 51(6): 851–868.

Yanagizawa-Drott, D (2014), “Propaganda and Conflict: Evidence from the Rwandan Genocide”, *The Quarterly Journal of Economics* 129(4): 1947–1994.

Zhuravskaya, E, M Petrova, and R Enikolopov (2020), “Political Effects of the Internet and Social Media”, *Annual Review of Economics* 12(1): 415–438.

## ABOUT THE AUTHORS

**Leonardo Bursztyn** is the Saieh Family Professor of Economics at the University of Chicago. He is also an Editor of the *Journal of Political Economy*, the co-director of the Becker Friedman Institute Political Economics Initiative, and the founder and director of the Normal Lab. His research uses experiments, often combined with observational data and economic theory, to better understand how individuals’ main economic decisions are shaped by their social environments. His work has examined educational, labor market, financial, consumption, and political decisions, both in developing and developed countries, and has been published in all leading economics journals and featured extensively in major media outlets. Leonardo is a Research Associate at the National Bureau of Economic Research (NBER), a fellow at the Bureau for Research and Economic Analysis of Development (BREAD), and an affiliate at the Abdul Latif Jameel Poverty Action Lab (J-PAL) and at the Pearson Institute. He is also the recipient of a 2016 Sloan Research Fellowship. He received his PhD in Economics at Harvard University in 2010.

**Georgy Egorov** is James Farley/Booz, Allen & Hamilton Research Professor at the Kellogg School of Management, Northwestern University. He joined the Kellogg faculty in 2009 after receiving his PhD in Economics from Harvard University. His research interests include political economy of democracies and non-democracies, dynamics of institutions, culture, and topics in economic theory. His papers have been published in leading journals, including *Econometrica*, *American Economic Review*, *Quarterly Journal of Economics*, *Journal of Political Economy*, *Review of Economic Studies*, and *American Political Science Review*.

**Ruben Enikolopov** is an ICREA Research Professor at UPF and Barcelona Institute of Political Economy and Governance and a visiting professor at the New Economic School. His research interests include political economy, economics of mass media, and development economics. He has published in *American Economic Review*, *Quarterly Journal of Economics*, *Review of Economic Studies*, *Econometrica*, *Proceedings of National Academy of Science*, *American Political Science Review* and *American Economic Journal: Applied Economics*. Enikolopov is a Research Fellow of the Center

for Economic and Policy Research (CEPR) and Research Affiliate of the International Economic Growth Center (IGC). He received a PhD in Economics at Harvard University in 2008.

**Maria Petrova** is ICREA Research Professor at the Universitat Pompeu Fabra and Barcelona Institute for Political Economy and Governance (IPEG), and an Affiliate Professor at Barcelona School of Economics. Maria is a Co-Editor of the *Journal of Public Economics* and a member of the Editorial Board of the *Review of Economic Studies* since 2016. She received PhD from Harvard University in 2008. She spent 2012 - 2013 as a Visiting Associate Research Scholar at the Center for the Study of Democratic Politics at Princeton University. In 2012-2013, she was also the Research Director at the Center for New Media and Society. Her research interests include mass media economics, political economy, and corporate governance. Her paper have been published in leading journals, including *American Economic Review*, *Quarterly Journal of Economics*, *Econometrica*, *American Political Science Review*, and *Management Science*.



# CHAPTER 14

## Can social media spur offline hatred?

163

**Karsten Müller and Carlo Schwarz**

National University of Singapore; Bocconi University

Many have blamed social media for a dizzying array of worrisome societal trends, from beliefs in fake news and depression rates among teenagers to the results of the 2016 US presidential election. A recent literature has begun to study the offline effects of social media. Zhuravskaya et al. (2020) review the literature on the political effects of the internet and social media. In an early study, Acemoglu et al. (2018) found a link between Twitter and protest turnout during Egypt's Arab Spring, while Giavazzi et al. (2020) show that, after terrorist attacks, Twitter messages in Germany become more similar to those of the right-wing party Alternative für Deutschland (AfD), and that such shifts in discourse are correlated with voting patterns. Enikolopov et al. (2020) exploit variation in social media usage in Russia explained by connections to the founder of the country's most popular social media platform, VKontakte, and relate this variation to the incidence of protests.

However, perhaps one of the most pressing concerns has been the spread of incendiary online content, which often turns into outright hateful speech when targeted at minorities. Commentators have argued that such online rhetoric may also have real-life consequences. The idea is that social media 'echo chambers' could reinforce people's extreme beliefs and, in turn, push potential perpetrators over the edge to carry out hate crimes.

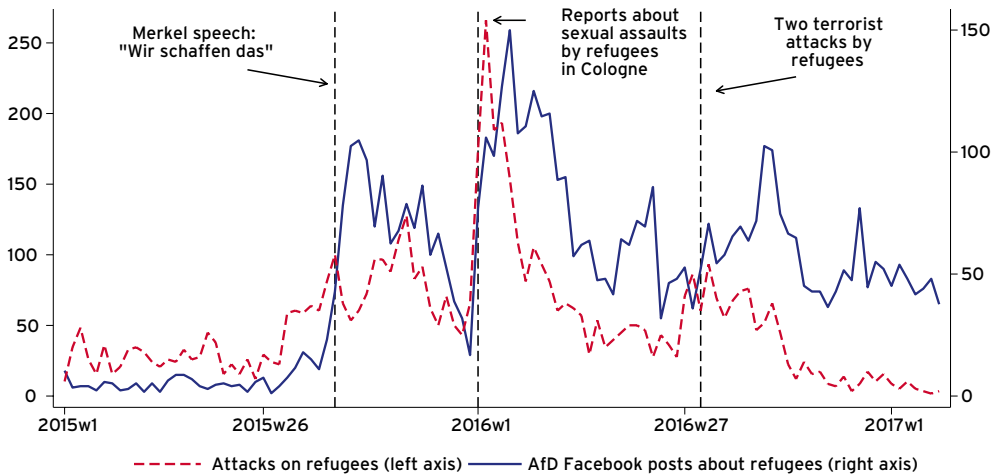
In two recent papers, we investigated if exposure to hateful content online can motivate violent offline action targeting minorities. Each of these papers, which we will describe in turn, focuses on a notable case study in which many commentators alleged a connection between social media and offline hatred. The first paper investigates whether right-wing social media content played a role in the spike in violence against refugees in Germany during the height of the refugee crisis between 2015 and 2017. The second paper examines former President Donald Trump's Twitter activity during his political rise and hate crimes against minorities, focusing on Trump's anti-Muslim rhetoric.

## FACEBOOK AND ANTI-REFUGEE VIOLENCE IN GERMANY

The 2015 refugee crisis shook Europe as millions of refugees fleeing the Syrian Civil War and conflict-ridden countries such as Afghanistan, Iraq, and Eritrea arrived at EU borders. Germany alone took in more than a million refugees following a speech by former Chancellor Angela Merkel, which included the noteworthy statement “*Wir schaffen das*” (“We will manage this”).

The large increase in refugees in Germany also caused an unfortunate rise in the number of violent attacks on refugees and refugee homes, which we study in Müller and Schwarz (2021). To relate these offline events to the online discourse, we propose a simple measure of anti-refugee sentiment on German social media based on the Facebook page of the largest far-right party in Germany, Alternative for Germany (*Alternative für Deutschland*, or AfD). In the 2017 federal elections, AfD would become the third-largest faction in the German parliament, overshadowing the smaller traditional coalition partners of the two main parties (SPD and CDU). What stood out about the AfD was its social media presence: in 2017, its Facebook page was the most ‘liked’ page out of all the German parties, giving the party and its followers substantial reach. Because the discourse about refugees on AfD’s Facebook page was noticeably more hostile and often outright hateful, it provides a good starting point for measuring anti-refugee sentiment.

**FIGURE 1 REFUGEE POSTS ON SOCIAL MEDIA AND ANTI-REFUGEE INCIDENTS OVER TIME**



Note: This figure plots the number of posts about refugees on the Facebook page of AfD and the number of anti-refugee incidents in Germany over time.

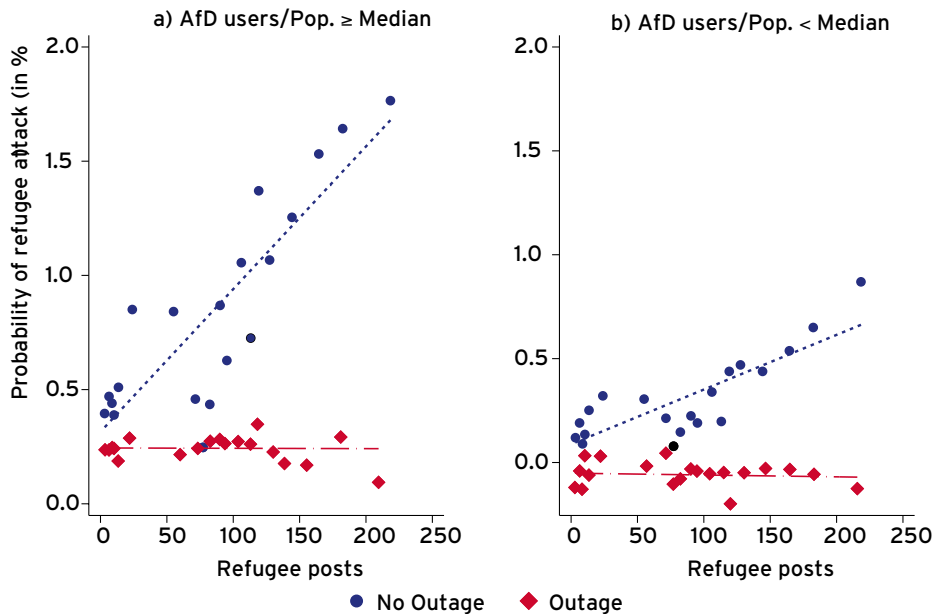
Figure 1 visualises the number of anti-refugee incidents and the measure of online anti-refugee sentiment in Germany over time. These time series are highly correlated. Both real-life attacks and anti-refugee sentiments spiked after Chancellor Merkel’s “*Wir schaffen das*” speech outlining her commitment to accepting a sizeable number of

refugees into Germany. They also spiked after several widely publicised sexual assaults on New Year's Eve 2016 in Cologne by immigrants, some of which were refugees, and further saw a rise after two terrorist attacks carried out by refugees in mid-2016. However, this relationship does not necessarily imply that social media has the power to affect real-life events. As just one example, the co-movement of these time series could also reflect the offline and online reactions to underlying political and news events.

To overcome this challenge, our paper uses two types of quasi-experiments to show that social media likely has a causal effect: (1) weekly variation in the (frequent) incidence of local internet outages across German towns, and (2) the (rare) incidence of Germany-wide Facebook outages. The intuition is straightforward. If there is a causal link between hateful online content and offline hate crimes, we would expect the number of attacks against refugees to go down following internet and Facebook outages. We would also expect the correlation between Germany-wide online anti-refugee sentiment and offline incidents to disappear during such outages.

Figure 2 visualises the main finding of our study. Each panel plots the relationship between the amount of anti-refugee sentiment on social media (x-axis) and the probability of anti-refugee incidents (y-axis). Panel (a) shows the relationship for municipalities where right-wing social media usage is above the median, while panel (b) shows the relationship for municipalities below the median. It is immediately apparent that the relationship between anti-refugee sentiment and anti-refugee attacks is far stronger for municipalities with many AfD Facebook users.

**FIGURE 2 THE CORRELATION BETWEEN ANTI-REFUGEE CONTENT ON FACEBOOK AND THE NUMBER OF HATE CRIMES IN A MUNICIPALITY GOES TO ZERO DURING LOCAL INTERNET OUTAGES**



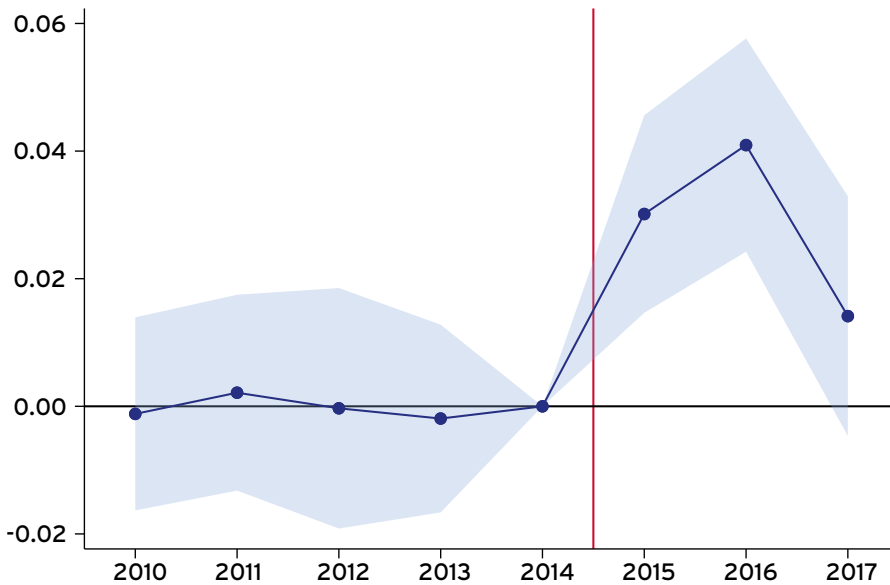
Even more striking is the effect of internet outages, shown by the grey diamonds. The relationship between anti-refugee sentiment on social media and anti-refugee incidents completely disappears for municipalities suffering an internet outage. This strongly suggests that exposure to hateful online content does indeed matter for the spread of offline violence.

### TWITTER AND ANTI-MUSLIM HATE CRIME IN THE UNITED STATES

In the second study, we turn our attention to Donald Trump, who has often been cited as a prime example of how inflammatory comments on social media can increase anti-minority sentiment. Critics have claimed that Trump's rhetoric on social media may have serious consequences. Minnesota congresswoman Ilhan Omar, for example, has linked tweets by Trump targeting her Muslim faith to "an increase in direct threats on my life – many directly referring or replying to the president's video". In a particularly tragic example of the potential influence of social media, the perpetrator of a terrorist attack on a mosque in New Zealand in 2019 used Facebook to livestream the shooting, which more than 200 people followed.

Can anti-minority messaging by elites have real-life effects? We study this question in Müller and Schwarz (2020). We start by documenting that the number of anti-Muslim hate crimes doubled following the 2016 presidential primaries, and this increase was particularly strong in counties with high Twitter usage.

**FIGURE 3 ANTI-MUSLIM HATE CRIMES SPIKED IN COUNTIES WITH HIGH TWITTER USAGE FOLLOWING TRUMP'S PRESIDENTIAL CAMPAIGN**



In the US, the incidence of anti-Muslim hate crimes – one metric of Islamophobic sentiment – has increased since 2015, starting around the beginning of Donald Trump’s presidential campaign. As seen in Figure 3, this rise in hate crimes also has been concentrated in areas where many people use Twitter, which was not the case before. In our paper, we also show that this pattern is considerably stronger in counties that already had active hate groups before the platform became popular, while it matters little where hate groups were less prevalent. This result indicates that social media can reinforce existing hateful attitudes towards minorities.

## FROM CORRELATION TO CAUSALITY

Although there has been a rise in both anti-Muslim hate crimes and President Trump’s anti-Muslim rhetoric on Twitter, how can we know whether the social media channel causes real-life actions?

To get at the issue of causality, we investigate the impact of the 2007 South by Southwest conference and festival (henceforth SXSW), which marked an early tipping point for the adoption of Twitter in the United States. We show that SXSW 2007, which only had around 7,000 registered attendees at the “Interactive” part of the event, left its imprint on the geography of Twitter usage across US counties until today.

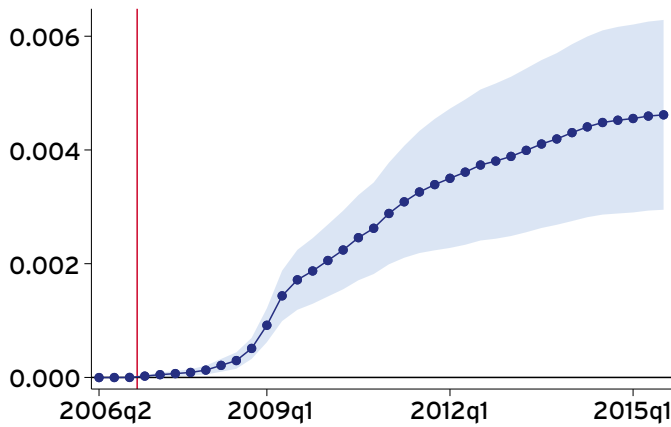
At the 2007 SXSW festival, Twitter held a launch event with a special option that allowed users to join Twitter by simply sending a text message, and screens in the main hallways showed tweets about the event. These measures proved to be highly effective in spurring Twitter adoption. The daily volume of tweets increased from around 20,000 to 60,000.

Our analysis in Müller and Schwarz (2020) exploits the fact that the home counties of SXSW attendees received a boost in the number of early-stage Twitter users around the time of the 2007 event, when excited festival-goers spread the news about the platform to their friends and families. These inflows of additional early adopters put these counties on a differential growth path, which ultimately resulted in a higher level of Twitter usage.

The paper provides several pieces of evidence to support the idea that these early adopters were vital to the rise of Twitter. First, we compare counties with and without new SXSW followers in March 2007 and observe an uptick in Twitter adoption with the beginning of SXSW that persists until the end of 2015. Figure 4 traces the impact of early adopters on Twitter usage per capita in their home counties, where the line marks the beginning of the SXSW festival. The data exhibit an S-shaped adoption curve typical for the diffusion of innovations, starting from the day of the festival, as more and more user signed up over time. The estimates imply that a one standard deviation increase in SXSW followers who signed up in March 2007 increased Twitter adoption by around 22% by the end of 2015.



**FIGURE 4** EARLY ADOPTION AT SXSW 2007 PREDICTS HIGHER TWITTER USAGE IN THE HOME COUNTIES OF THE FESTIVAL'S ATTENDEES



Second, we find that early Twitter adopters were indeed largely connected to the SXSW festival. In March 2007, as many as 60% of Twitter users followed the SXSW festival or followed somebody that did. With the diffusion of Twitter over time, this share decreased to around 5% today, as the platform's usage spread to broader subsections of the population.

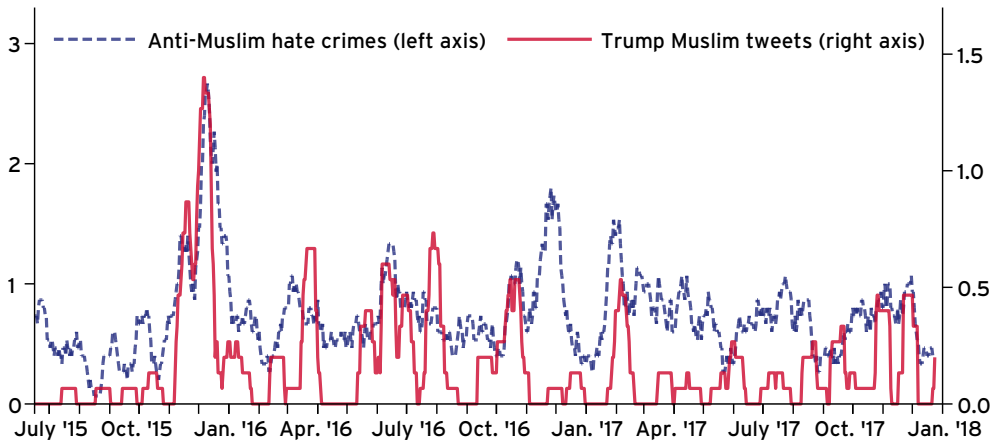
With this source of quasi-random variation in Twitter usage, we can confirm the link between social media and hate crimes. Around the start of Trump's presidential campaign, the home counties of the early Twitter adopters saw an apparent upward shift in anti-Muslim sentiments. Our estimates suggest that one standard deviation higher Twitter use was associated with a 38% increase in hate crimes against Muslims.

### TRUMP'S TWEETS AND ANTI-MUSLIM SENTIMENT

To make sense of these broad patterns, we analyse one potential driver of anti-Muslim sentiments: Donald Trump's Twitter feed. Similar to our analysis in Germany, we found a clear pattern in the data. Trump's tweets about Muslims tend to be followed by a higher frequency of anti-Muslim hashtags appearing on Twitter, increased reporting about Muslims by cable news stations, and a bump in hate crimes targeting Muslims.

Figure 5 shows the number of Trump's anti-Muslim tweets and the number of anti-Muslim hate crimes over time. A strong temporal correlation is immediately apparent. The causal connection between these two-time series is again far harder to establish. In our paper, we provide some suggestive evidence based on the former President's golf habit, which appears to affect his behaviour on Twitter.

**FIGURE 5 CORRELATION BETWEEN TRUMP TWEETS ABOUT MUSLIMS AND ANTI-MUSLIM HATE CRIMES**



In essence, on days that President Trump played golf, his tweets also contained fewer references to daily politics and more about minorities, especially Muslims. A potential reason is that his social media manager, Dan Scavino (Trump's former caddie), is the likely source of many of the former president's inflammatory tweets. In our analysis, we show that Trump's tweets about Muslims induced by his golf habit are still highly predictive of anti-Muslim hate crimes in the following days. This suggests that Trump's social media activity might spur offline behaviour.

By analysing more than 100 million tweets by Trump's followers, we also show that his hateful messaging begets more hate online. Not only are Trump's anti-Muslim tweets widely shared, his followers also produce more xenophobic content in response, causing a spike in the hashtags #StopIslam and #BanIslam.

### **BROADER PERSPECTIVES**

Even though these two projects investigate different countries, different social media platforms and different minority groups, many parallels emerge. In both contexts, the empirical evidence suggests that what politicians say on social media matters. While social media may have many positive effects, it also matters for propagating anti-minority sentiments.

Other authors have also shown that social media affects anti-ethnic hate crime in Russia (Bursztyn et al. 2022). Moreover, during the Covid-19 pandemic, there has been a well-documented wave of anti-Chinese sentiments on social media, particularly after incendiary tweets by President Trump (Hswen et al. 2021), and these online sentiments coincided with increases in anti-Chinese hate crime. Our evidence tells us that this correlation is unlikely to be a coincidence.

Moreover, the broad reach of social media means that it can influence the general public's perceptions of hate crimes, which might ultimately make them more likely to tolerate them. The worry is that non-stop exposure to hateful messages that some social media users are subjected to could desensitise them to hate speech and encourage them to see hate crimes as more acceptable. This does not mean that social media is solely responsible for hate crimes. What matters is how politicians and other influential figures use it.

It is also important to highlight that the connection between media and violence is not exclusive to social media. Existing work (e.g. Yanagizawa-Drott 2014, Adena et al. 2015) finds that traditional media can also contribute to ethnic hatred and violence. Other research has linked television (Card and Dahl 2011) and movies (Dahl and DellaVigna 2009) to short-lived spikes (or decreases) in violence. Bhuller et al. (2013) document increases in sex crime associated with the roll-out of broadband internet in Norway; Chan et al. (2016) find a correlation between broadband availability and hate crimes in the US. In this regard, social media may have amplified existing problems with traditional media platforms through its interactive character and low entry barriers.

### **POLICY IMPLICATIONS**

While limited in scope, the existing evidence on social media provides some important insights into its potential negative real-life consequences. Unfortunately, this work does not yet give us a good sense of the effective solutions that could tackle online hate speech and its potential offline consequences. This makes it challenging to evaluate initiatives by social media platforms and policymakers to moderate social media content, an issue that has proven to be extraordinarily controversial.

One piece of experimental evidence on the effectiveness of content moderation comes from recent work by Jiménez Durán (2022). He provides some of the first experimental evidence for the efficacy of content moderation and documents that randomly reporting hateful posts on Twitter increases the likelihood they are removed by the platform, but does not appear to significantly affect the number of tweets these hateful users send; nor does it stop them from posting more hateful content. However, reporting hateful posts does cause a reaction by other users, who reply in greater numbers to attack the original hateful content and increase their total time spent on Twitter. He concludes that while content moderation may not discipline users to stop posting hateful content, it does mobilise opposition from other users. Importantly, it also does not appear to be revenue-destroying for platforms, given that it increases the overall time users spend on Twitter.

What seems clear is that, especially when used by powerful groups or individuals, social media can cause harm to minorities. How the trade-off between preserving free speech and preventing harm can be managed should be a key priority for policymakers.

Our findings suggest that platforms like Facebook and Twitter are responsible for taking action against the normalisation of hate online. The major social media companies are already under pressure to do more to protect minority groups from hate speech. But we believe such initiatives have a long way to go. What will be absolutely key is the availability of data to support independent research that can help determine which policies work and which do not. On this front, forcing social media companies to provide easy and free access to much more data that are currently available should be an obvious policy that we hope can gather widespread political support.

## REFERENCES

Acemoglu, D, T A Hassan and A Tahoun (2018), “The power of the street: Evidence from Egypt’s Arab Spring”, *The Review of Financial Studies* 31(1): 1-42.

Adena, M, R Enikolopov, M Petrova, V Santarosa and E Zhuravskaya (2015), “Radio and the Rise of the Nazis in Prewar Germany”, *The Quarterly Journal of Economics* 130(4): 1885-1939.

Bhuller, M, T Havnes, E Leuven and M Mogstad (2013), “Broadband internet: An information superhighway to sex crime?”, *Review of Economic Studies* 80(4): 1237-1266.

Bursztyn, L, G Egorov, R Enikolopov and M Petrova (2019), “Social media and xenophobia: evidence from Russia”, NBER Working Paper No. w26567.

Card, D and G B Dahl (2011), “Family violence and football: The effect of unexpected emotional cues on violent behavior”, *The Quarterly Journal of Economics* 126(1): 103-143.

Chan, J, A Ghose, and R Seamans (2016), “The Internet and Racial Hate Crime”, *MIS Quarterly* 40(2): 381-404.

Dahl, G, and S DellaVigna (2009), “Does movie violence increase violent crime?”, *The Quarterly Journal of Economics* 124(2): 677-734.

Enikolopov, R, A Makarin and M Petrova (2020), “Social media and protest participation: Evidence from Russia”, *Econometrica* 88(4): 1479-1514.

Giavazzi, F, F Ighhaut, G Lemoli and G Rubera (2020), “Terrorist attacks, cultural incidents and the vote for radical parties: Analyzing text from Twitter”, NBER Working Paper No. w26825.

Hswen, Y, X Xu, A Hing, J B Hawkins, J S Brownstein, and G C Gee (2021), “Association of “#covid19” Versus “#chinesevirus” With Anti-Asian Sentiments on Twitter: March 9–23, 2020”, *American Journal of Public Health* 111: 956-964.

Jiménez Durán, R (2022), “The Economics of Content Moderation: Theory and Experimental Evidence from Hate Speech on Twitter”, available at SSRN.

Müller, K and C Schwarz (2021), “Fanning the flames of hate: Social media and hate crime”, *Journal of the European Economic Association* 19(4): 2131-2167.

Müller, K and C Schwarz (2020) “From hashtag to hate crime: Twitter and anti-minority sentiment”, conditionally accepted at *AEJ: Applied Economics*.

Yanagizawa-Drott, D (2014), “Propaganda and conflict: Evidence from the Rwandan genocide”, *The Quarterly Journal of Economics* 129(4): 1947-1994.

Zhuravskaya, E, M Petrova and R Enikolopov (2020), “Political Effects of the Internet and Social Media”, *Annual Review of Economics* 12(1): 415-438.

## ABOUT THE AUTHORS

**Karsten Müller** is an Assistant Professor in Finance at the National University of Singapore Business School. His research focuses on finance, macroeconomics and political economy. One line of his work studies how politics and financial regulation determine which firms and households get access to credit in the economy and what the macroeconomic consequences may be. His second stream of research shows that social media can affect election outcomes and anti-minority sentiments offline. Before joining NUS, Karsten was a postdoctoral research associate at Princeton University. He earned his Ph.D. in Business and Management from the University of Warwick in 2018.

**Carlo Schwarz** is an Assistant Professor in the Department of Economics at Bocconi University. He holds a PhD in Economics from the University of Warwick. His research interest lies in the field of applied microeconomics with a particular interest in policy-relevant questions. In his research, he combines causal inference strategies with techniques from text analysis, machine learning, and data science. His recent research investigates inequalities in the scientific process and their impact on the creation of knowledge. He also has written papers on the contributed of social media to the spread of hate crimes and anti-minority sentiments.

## **SECTION 6**

# **SOCIAL MEDIA IN AUTOCRACIES**



# CHAPTER 15

## The political economy of social media in China

175

**Bei Qin, David Strömberg and Yanhui Wu**

Hong Kong Baptist University; University of Stockholm; University of Hong Kong and CEPR

The arrival of social media has generated a huge information shock to democracies and autocracies alike. One of the most striking examples is China, the world's largest autocracy. On one hand, China has a vast population of 'netizens' on social media platforms, such as Weibo, WeChat, and TikTok, which have become part and parcel of everyday life. On the other hand, government censorship of the media is pervasive. Freedom House has consistently ranked China among the countries with the lowest level of internet freedom in the last decade. Yet, eruptions of information on social problems and political issues are allowed on social media from time to time. Apparently, the Chinese government has been trying to turn social media to its advantage through strategic management of online information flows.

Whether the Chinese government's information control strategy will succeed is a critical question not only for China but also for the rest of the world. Given the level of economic development and social media penetration, events in China may be indicative of the way forward for other developing countries. Moreover, as a leader in the use of artificial intelligence (AI) technology for political control, China is currently exporting its surveillance technology (perhaps its model of information control as well) to other authoritarian countries.

In this chapter, we discuss the political information landscape on Chinese social media and the political impact of the circulation of such information. Our discussion is based primarily on our recent research (Qin et al. 2017, 2021) examining the content of 13.2 billion posts published between 2009 and 2013 on Sina Weibo (the Chinese equivalent of Twitter). After a brief description of the background, we will discuss three topics: propaganda, monitoring of officials, and surveillance of collective action (with a special emphasis on surveillance).

### BACKGROUND

Among the numerous social media in China, Sina Weibo (Weibo hereafter) is the most prominent platform for breaking news and political commentary. It was introduced in August 2009, in a market that was vacant due to the forced exit of Twitter and Facebook.



After a period of rapid growth, Weibo reached a peak of 500 million reported users in 2012. Since 2013, it has lost some ground to WeChat, a mobile phone-based social networking service, but has remained influential for public discussion of social issues.

Prior to the advent of social media, information on surveillance and monitoring was much more limited. It is very costly to collect information directly from citizens in a country as large and diverse as China. The many layers of self-interested governments in the Chinese political system create substantial barriers to information transmission. Traditional media are controlled mostly by local governments, which lack sufficient incentive to supply information for monitoring and propaganda (Qin et al. 2018).

Unlike traditional media, Chinese social media is directly controlled by the National Office of Information Control under the close supervision of the Chinese Communist Party's Central Propaganda Department. The centralised control of social media prevents local governments and politicians from censoring material that is detrimental to them. This in turn opens a channel of communication between social media users and the central government on issues where their interests are aligned, such as fighting corruption at local levels. Meanwhile, Chinese governments at all levels have opened their own Weibo accounts in an effort to sway public opinion; some even hire internet trolls or commentators to create posts to their advantage. In 2012, Weibo reported approximately 50,000 accounts operated by government offices or individual officials.

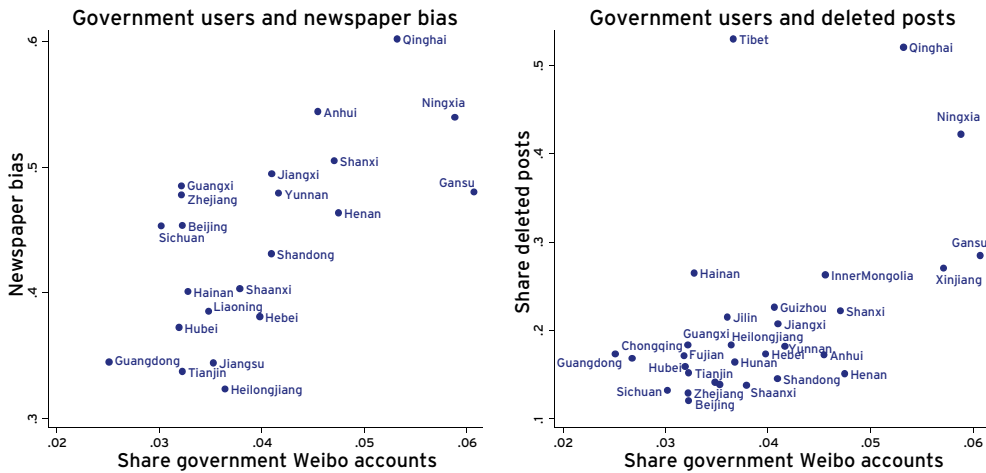
## PROPAGANDA

Propaganda is indispensable to authoritarian rulers. As the channels expand from traditional media to social media, the Chinese government has followed suit to disseminate pro-government information and propaganda.

In Qin et al. (2017), we provide the first external estimate of the Chinese government's presence on social media. We identify government accounts from user profiles as well as from post content. We combine this with a machine learning algorithm (support vector machine) to ascertain the likelihood of each user in our data being affiliated with the government.

According to our estimation, there are 600,000 government-affiliated Weibo accounts, ten times more than what was reported by the company. These accounts contribute 4% of all posts on political and economic issues. As shown in Figure 1, the share of government users is higher in areas with more extensive censorship of social media posts as measured by Bamman et al. (2012), and in areas where newspapers are more compliant with the Party line as measured in Qin et al. (2018). This is consistent with political influence being the main motive of the government-affiliated accounts.

**FIGURE 1 SHARE OF GOVERNMENT USERS ON SINA WEIBO ACROSS PROVINCES VERSUS NEWSPAPER BIAS AND CENSORSHIP**



These posts can be used to identify local officials charged with corruption before any government action. We examine 200 corruption cases involving high-ranking government or Party leaders. For comparison, we construct a matched sample of 480 politicians who were not charged with corruption. The comparison politicians hold similar political positions to, and are located in areas geographically near to, the charged politicians. We count the number of posts mentioning the name of each of these 680 politicians and the number of posts mentioning both the politician and words referencing corruption.

Such posts, published one year before the first government action, predict which politicians will eventually be charged with corruption. The reason may be that these individuals are indeed more corrupt than the comparison officials. It could be also the case that the central government lifts censorship of corruption posts or even plants its own posts about leaders who have lost political support and who will later be charged with corruption. To investigate whether the post-planting story is true, we examined a well-reported scandal involving Bo Xilai, a high-ranking official. We find that there was blanket censoring of posts mentioning Bo during his investigation, but no evidence that censorship focused on posts that were supportive of Bo or that there was a trend in corruption posts prior to his downfall.

## COLLECTIVE ACTION

This section gives an account of the effectiveness of Weibo posts in surveillance of collective action events, such as conflicts, protests, and strikes. Our discussion addresses two central questions. First, given the extensive censorship and noisy information on social media, how effective are the existing posts in detecting collective action events? Second, to what extent do these posts help spread events across regions?

Although information that explicitly calls for anti-regime action hardly exists on Weibo, we are able to obtain millions of posts about collective action in our data. They are categorised by topic modelling, as shown in Figure 2.

FIGURE 2 HOT TOPICS IN POSTS ABOUT COLLECTIVE ACTION



Further examination of post content reveals that many posts talk about the causes of these events, such as corruption or wage arrears, and criticise the government for bad policies and misconduct. Posts expressing anger and sympathy for protesters are among the most forwarded (retweeted) posts. In contrast, posts that contain logistical information (e.g. specifying where and when to meet) or that mention protest tactics are extremely rare, most likely due to censorship. These findings support a significant notion of the Chinese government's censorship strategy: content useful for surveillance is allowed while posts with information only useful for protesters are censored.

## EFFICACY OF SURVEILLANCE

To evaluate whether the posts about collective action can be used for surveillance, in Qin et al. (2017) we analyse 545 large collective action events (primarily protests and strikes) that took place in mainland China between 2009 and 2012. Using a method based on keyword counts, we are able to predict these events one day in advance. Table 1 demonstrates a simplified version of our approach. One day before and on the day of these events, there are abnormally high numbers of Weibo posts containing keywords related to each event type in the city where the event occurred. In principle, a technology-aided authoritarian leader can note these abnormal bursts of posts and sense the underlying grievances by analysing the posts on the preceding day of the event.

We evaluate our prediction using AUROC, a popular measure of the accuracy of a model's predictive ability.<sup>1</sup> The surveillance tool we develop has an AUROC of 0.87 for predicting strikes and 0.96 for predicting anti-Japan protests, close to or above 0.9, which is the threshold for a prediction being viewed as excellent. This is probably a lower bound on the accuracy of the actual surveillance systems used by Chinese government agencies, which have invested heavily in building AI systems that exploit information on social media.

**TABLE 1** EVENT PREDICTION AND DETECTION

VARIABLES	Conflict	Protest	Strike	Anti-Japan	Coalmine accident
# Weibo posts: day of event	6.1	62.6	167.3	2036.6	3.0
# Weibo posts day before event	3.4	54.3	48.1	924.6	0.7
# Weibo posts: no event	0.7	4.4	2.5	4.5	1.2

<sup>1</sup> AUROC measures the area under the ROC curve. A ROC curve shows the trade-off between type 1 and type 2 errors in prediction and was first employed in World War II to evaluate methods that analyzed radio signals to identify Japanese aircrafts.

## SPREAD OF COLLECTIVE ACTION

In Qin et al. (2021), we investigate whether information diffusion on Weibo facilitates the spread of protests and strikes, despite the absence of logistics and tactics information. This is generally regarded as the main driver of the social media effects on protests in a freer informational environment.

We first demonstrate that information about collective action travels fast and wide on Weibo. We take advantage of our unique data on post forwarding. The fact that a user forwards a message indicates that the user has viewed it. Thus, forwards are an effective measure of information diffusion (e.g. Kwak et al. 2010). From a subset of 3 million first forwards on which we have their precise timing and locations, it is clear that information about protests and strikes disseminates rapidly and extensively. Approximately 30% of the forwards occur within one hour of the posting of the original messages, and 80% within one day. After one hour, the mean distance between the user who posts a message and the user who forwards it is over 800 km.

We then measure information flows on Weibo between city pairs. Before Weibo was introduced, there were no such flows. By counting the number of posts originated from each city that are forwarded by users in another city at a certain point of time, we measure how information flows across a pair of cities increase over time.

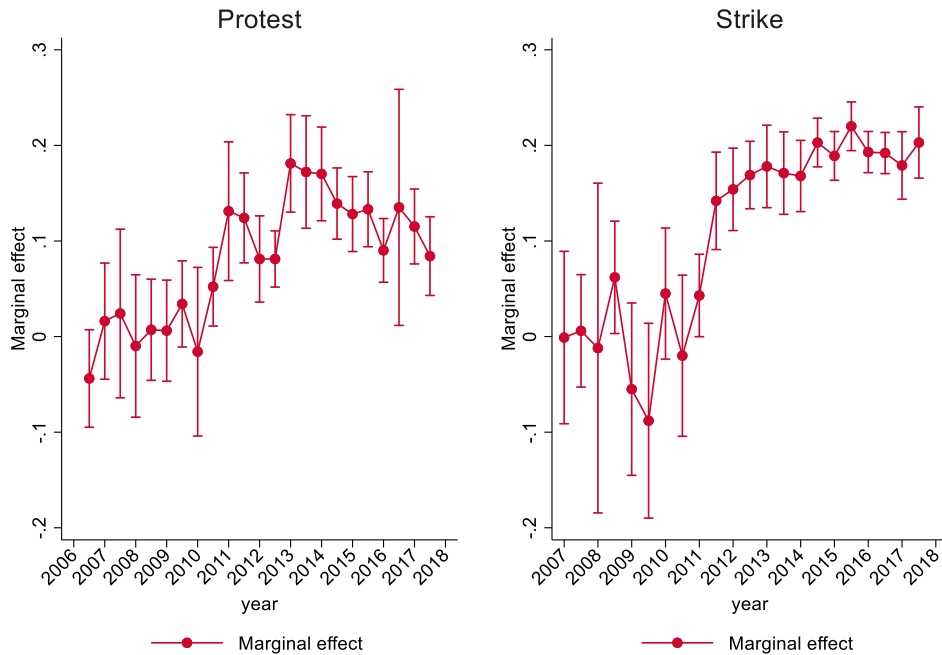
Finally, we investigate whether events spread more easily across cities which become more connected via social media. We find that protests and strikes spread as rapidly as within two days through Weibo connectedness. As shown in Figure 3, both protests and strikes start to roll out in 2010–2011 across cities that eventually become highly connected through Weibo. For strikes, which are politically less sensitive, the spread continues at high levels until 2017. In contrast, the spread of protests across cities connected through Weibo covaries with changes in political sensitivity and the strictness of censorship. The spread effect reduces in 2012, when Xi Jinping replaced Hu Jintao as General Secretary of the Chinese Communist Party, and falls significantly after 2014, when censorship became stricter as measured by the Press Freedom Index published by Reporters Without Borders.

We also explore the mechanisms that drive the spread of events via social media. Previous studies focus on how protesters use social media to organise events or disseminate protest tactics. For instance, Enikolopov et al. (2020) report that two thirds of the Russian cities in their sample have social media communities created to organise protest demonstrations. However, our inspection shows that such content is absent on Weibo, likely due to censorship.

While we rule out explicit organisation and learning tactics as important mechanisms, our empirical findings are consistent with other potential mechanisms. In a random sample of 1,000 posts and the 100 most forwarded posts about protests, the majority of them describe ongoing events, discuss their causes, or express anger and sympathy. Such

content can spread the protests through emotional reaction if angry people spontaneously protest against social injustices, as suggested in Pasarelli and Tabellini (2017). It is also consistent with tacit coordination. These posts may help protesters implicitly coordinate their actions across cities when people anticipate that simultaneous protests may increase the chance of pressuring the involved parties to deal with the problem and reduce the risk of punishment (Edmond 2013, Little 2016, Barberà and Jackson 2020).

**FIGURE 3 SPREAD OF PROTESTS AND STRIKE THROUGH SOCIAL MEDIA**



**CONCLUSION**

Surveillance, monitoring, and propaganda are three potentially prominent political uses of social media by authoritarian regimes. Nevertheless, information that is crucial to these uses may also be destructive to the regime. For example, information that is valuable for surveillance may spread collective resistance, information that is useful for monitoring officials may erode public trust of the government, and propaganda may undermine the popularity of social media. These trade-offs put authoritarian leaders in a difficult position in their management of information flows. Our research sheds light on some of these trade-offs.

Our analysis leads to the conclusion that social media is an efficient tool for surveillance of local collective action events in China. Protests and strikes can be predicted one day in advance with excellent accuracy. However, this comes at a cost for the regime. The published posts and forwards have a sizeable effect on the spread of both protests and

strikes across cities. The lesson is that, to limit the spread of protests, an authoritarian government must shut down the discussion of causes and emotional reactions instrumental in surveilling collective action.

We also find that social media posts help predict which government officials will be charged with corruption one year before government action. Regarding propaganda, we estimate that 600,000 government-affiliated accounts operating on Weibo generated approximately 4% of all posts referencing political and economic issues. The geographic distribution of these accounts indicates that political influence is the main motive behind these government-affiliated accounts.

## REFERENCES

- Bamman, D, B O'Connor, and N A Smith (2012), "Censorship and Deletion Practices in Chinese Social Media", *First Monday* 17(3).
- Barberà, S and M O Jackson (2020), "A model of protests, revolution, and information", *Quarterly Journal of Political Science* 15(3): 297-335.
- Edmond, C (2013), "Information Manipulation, Coordination, and Regime Change", *Review of Economic Studies* 80(4): 1422-1458.
- Egorov, G, S Guriev, and K Sonin (2009), "Why resource-poor dictators allow freer media: A theory and evidence from panel data", *American Political Science Review* 103(4): 645-668.
- Enikolopov, R, A Makarin, and M Petrova (2020), "Social media and protest participation: Evidence from Russia", *Econometrica* 88(4): 1479-1514.
- Kwak, H, C Lee, H Park, and S Moon (2010), "What is Twitter, a social network or a news media?", *Proceedings of the 19th International Conference on World Wide Web*, pp. 591-600.
- Little, A T (2016), "Communication technology and protest", *The Journal of Politics* 78(1): 152-166.
- Lorentzen, P (2014), "China's Strategic Censorship", *American Journal of Political Science* 58(2): 402-414.
- Passarelli, F and G Tabellini (2017), "Emotions and political unrest", *Journal of Political Economy* 125(3): 903-946.
- Qin, B, D Strömberg, and Y Wu (2017), "Why does China allow freer social media? Protests versus surveillance and propaganda", *Journal of Economic Perspectives* 31(1): 117-40.
- Qin, B, D Strömberg, and Y Wu (2018), "Media bias in China", *American Economic Review* 108(9): 2442-76.

Qin, B, D Strömberg, and Y Wu (2021), “Social Media and Collective Action in China”, CEPR Discussion Paper No. 16731.

## ABOUT THE AUTHORS

**Bei Qin** is Associate Professor at the Department of Accountancy, Economics and Finance, Hong Kong Baptist University. Her research interest is media economics, political economics, and development economics. She received her Ph.D. in Economics from Stockholm University.

**David Strömberg** is Professor of Economics at the Economics Department of Stockholm University. His research interests include media economics and political economics. He received his PhD in Economics from Princeton University.

**Yanhui Wu** is Associate Professor of Economics and of Management and Strategy at the University of Hong Kong and a research fellow of the Centre for Economic Policy Research (CEPR). Prior to HKU, he was Assistant Professor of Finance and Business Economics at the University of Southern California. His research interests include media economics, organizational economics, digital economy, and the Chinese economy. He received his PhD in Economics from the London School of Economics.





# CHAPTER 16

## Social media in autocracies

185

**David Y. Yang**

Harvard University

Freedom of expression is the foundation of any free, democratic and civic society. As a fundamental human right, it underpins most other human rights and allows them to flourish. Instrumentally, access to accurate information is crucial to the well-functioning of democratic systems and market economics (e.g. DellaVigna and Gentzkow 2010, Strömberg 2015). However, as of 2022, only about 18% of the world's internet user population enjoys internet free from censorship.<sup>1</sup>

Media censorship is concentrated among autocratic regimes. In fact, media censorship is a hallmark of autocracies. Countries such as China spend a tremendous amount of resources to block foreign websites so that uncensored, regime-threatening information is out of citizens' reach (Roberts 2018). Scholars have long suggested that censorship is key to the popular support and stability of these regimes (e.g. Ford 1935). More recently, scholars document that the expansion of 3G mobile networks around the world has reduced government approval, although existing media censorship counteracts the new media technology (Guriev et al. 2020). Nonetheless, direct empirical evidence about the effect of removing censorship is limited.

Does providing access to an uncensored internet lead citizens to acquire politically sensitive information? Does the acquisition of politically sensitive information change citizens' beliefs, attitudes and behaviours? The answers to these questions are far from clear. Citizens with access to uncensored internet may not seek out politically sensitive information, due to lack of interest in politics, fear of government reprisal, and unawareness or distrust of foreign news outlets. Even if they do acquire such information and become fully informed, their attitudes and beliefs may not change.

### EXPERIMENTAL SETUP

In order to answer these two questions, my co-author and I conducted a field experiment in China between 2015 and 2017 (Chen and Yang 2019). At the time (and to a large extent this remains true today), access blockage introduced by the Chinese regime could be bypassed through a range of censorship circumvention tools using proxy servers or traffic

<sup>1</sup> Freedom House's *Freedom on the Net* report assesses 89% of the world's internet user population (<https://freedomhouse.org/report/freedom-net/2022/countering-authoritarian-overhaul-internet>).

data encryption (e.g. a virtual proxy network, or VPN). More than a dozen such tools were available to Chinese internet users, and as of the beginning of the experiment, there was no law that explicitly regulated their usage. These tools are in general relatively inexpensive, but only approximately 3% of internet users in China regularly use them to bypass censorship (Roberts et al. 2010), which is *prima facie* evidence that citizens may not demand access to an uncensored internet.

We randomly assigned 1,800 university students in Beijing to either a control condition in which their internet use was subject to status quo censorship, or to a treatment condition in which they were given tools to bypass internet censorship for free for 18 months.<sup>2</sup> A subset of the treated students also received temporary encouragement for four months to visit the Chinese edition of the *New York Times*, which is blocked by China's censorship apparatus.

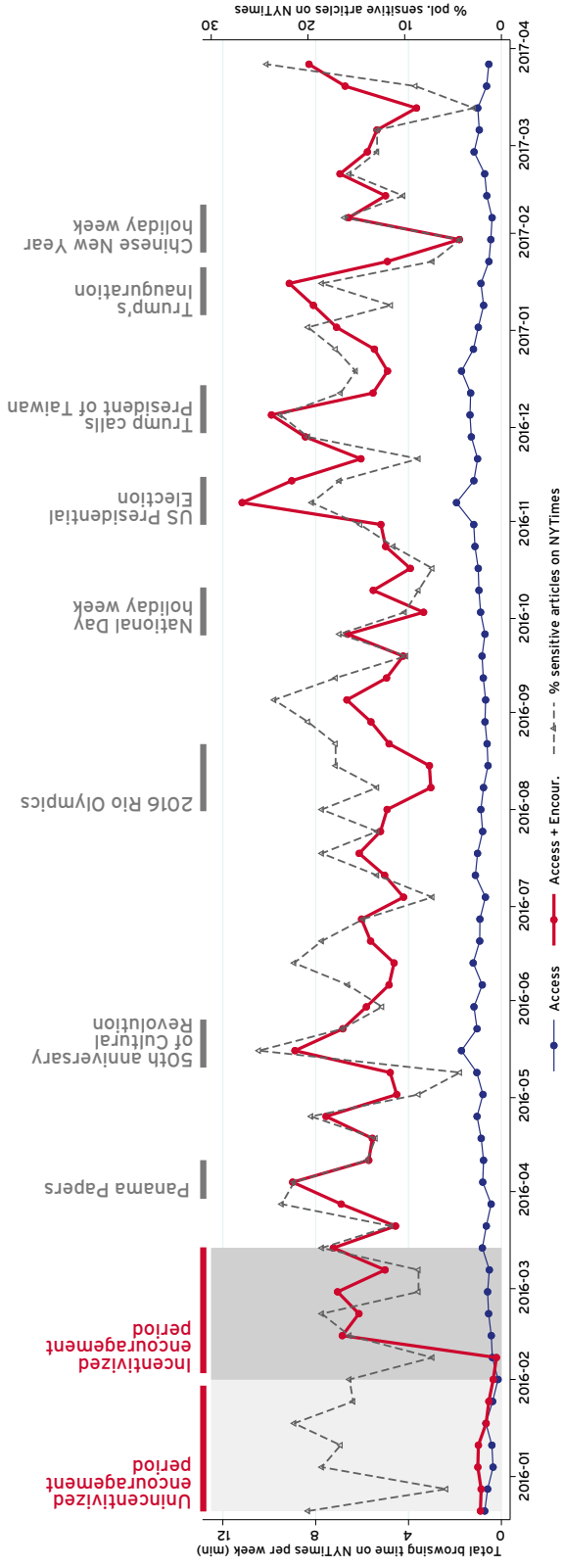
## FINDINGS

We find three main results. First, access to uncensored internet alone has little impact on students' acquisition of politically sensitive information. Nearly half of the students did not use the tools to bypass censorship at all. Among those who did, almost none spent time browsing foreign news websites that are blocked (see the blue line in Figure 1). These numbers indicate that students' low demand for uncensored, politically sensitive information is an important reason why they do not consume such information, in spite of the low cost.

Second, modest and temporary incentives to visit Western news outlets lead to large and persistent increases in students' acquisition of politically sensitive information. Students spent on average 435% more time on the *New York Times* even after the incentivised encouragement ended (see red line in Figure 1). This persistent increase suggests that demand is not inherently low, and in particular, fear of government reprisal is unlikely to be the reason students do not demand sensitive information. Rather, an important factor shaping students' low demand appears to be their underestimation of the value of uncensored information. A period of exposure to foreign news outlets persistently increases students' reported trust of these outlets and makes them willing to pay a higher price for the access (see Figure 2).

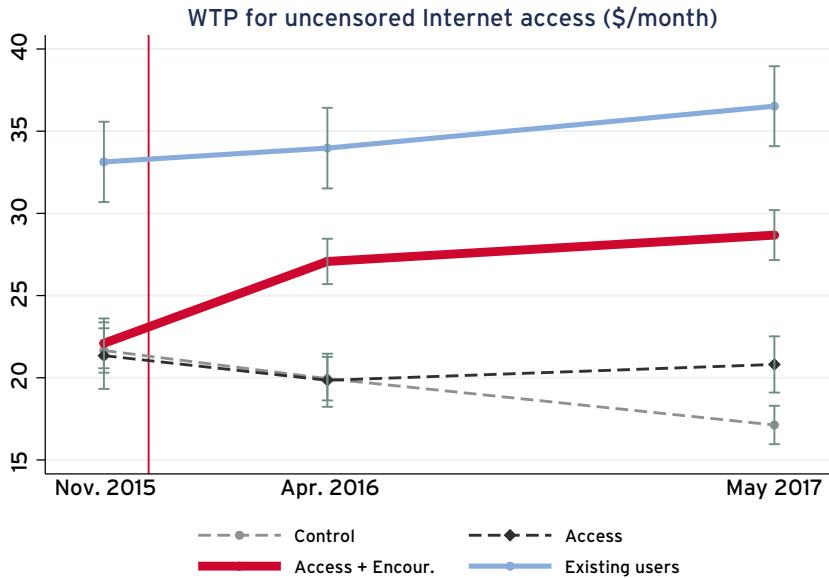
<sup>2</sup> Approximately 21% of the students already purchase tools bypass censorship before our experimental intervention. They are much richer and liberal minded than the average student. We assign the access treatment only among those students are not yet existing users.

**FIGURE 1 AVERAGE TOTAL BROWSING TIME ON THE NEW YORK TIMES PER WEEK (MINUTES)**



Note: Figure shows the average total browsing time on the *New York Times* per week, among students received only the access treatment and those who received both access and encouragement treatment. Dotted line (y-axis on the right-hand side) indicates the proportion of articles published on the *New York Times* Chinese edition that are politically sensitive.

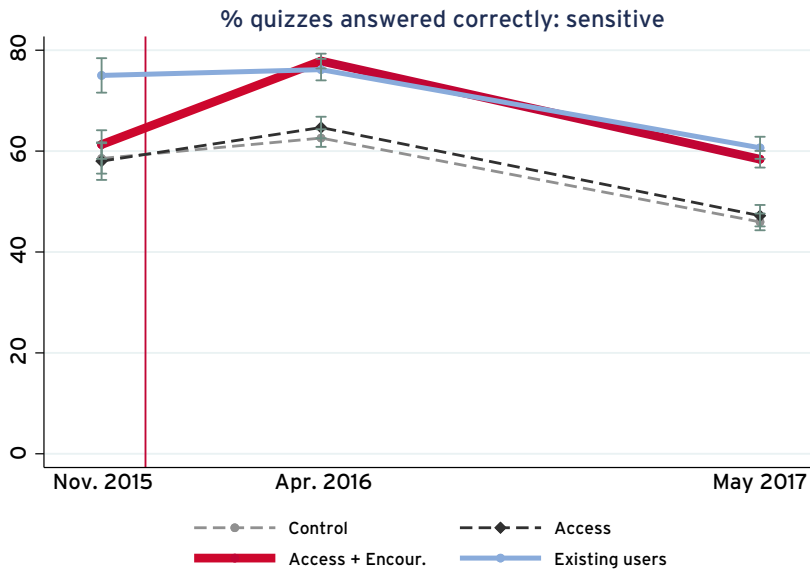
**FIGURE 2 WILLINGNESS TO PAY FOR ACCESSING CENSORSHIP CIRCUMVENTION TOOLS (US\$ PER MONTH)**



Note: Figure shows the average level of willingness to pay for accessing censorship circumvention tools among students in control group, those who received only the access treatment, those who received both access and encouragement treatments, and the existing users, across the baseline survey (November 2015), midline survey (May 2016), and endline survey (April 2017).

Third, acquisition of politically sensitive information brings broad, substantial, and persistent changes to students' knowledge, beliefs, attitudes, and intended behaviours. Acquisition, as a result of free access and temporary incentives, makes students (1) more knowledgeable of current events and notable figures censored on domestic media, as well as politically sensitive events in the past; (2) more pessimistic about Chinese economic growth and stock market performance in the near future, revealed in an incentive-compatible manner; (3) more sceptical of the Chinese government, less satisfied with its performance, and more likely to demand changes in Chinese institutions; and (4) more willing to take actions to incite changes, more likely to plan on leaving China through foreign graduate schools, and more likely to report having pulled out investments in the Chinese stock market (among the small number of students who invest). The effect is the largest among students who have limited access to alternative sources of uncensored information (e.g. those from disadvantaged backgrounds).

**FIGURE 3 AVERAGE PERCENTAGE OF QUIZZES ON RECENT POLITICALLY SENSITIVE NEWS EVENTS CORRECTLY ANSWERED**



Note: Quizzes cover a set of events occurred during the two months prior to each survey wave; they do not cover events that are directly mentioned in the encouragement treatment.

## IMPLICATIONS

Taken together, our findings suggest that demand-side factors are important for understanding how internet censorship works in China today. Censorship in China is effective not only because the regime makes it difficult to access sensitive information, but also because it fosters an environment in which citizens do not demand such information in the first place. Depending on citizens' demand for uncensored information, the censorship apparatus in China can be either robust or fragile. After years of censorship and propaganda campaigns, the current level of demand is low. As a result, initiatives such as the Lantern Project that passively supply access to uncensored internet to citizens in authoritarian regimes are unlikely to be as effective as some might imagine. In fact, the Chinese government may not need to bear the extremely high costs of fully 'sealing' its internet, as it can afford to leave some holes open. The masses may not begin to respond to negative news shocks, information-demanding elites may not be irritated, and business interests relying on global Internet connections may not be sacrificed.

This demand-driven censorship is not unique to contemporary China. The current Russian regime enforces repressive censorship over TV, while leaving the internet, and in particular the social media landscape, largely uncensored. Similarly, during the Cold War, the East German government employed heavy propaganda and censorship campaigns, while simultaneously allowing its citizens to purchase, *de facto*, antennae to access West German TV if they were sufficiently interested.

Nevertheless, our findings do not imply that the Chinese regime can safely eliminate the Great Firewall. The current cost of circumventing censorship imposes a huge campaign cost on foreign news outlets. Without such costs, outlets such as the New York Times might begin to campaign and effectively raise demand among Chinese readers. Removing the Great Firewall could also raise the demand for uncensored information by signalling to citizens the quality of foreign media, making foreign media consumption more socially acceptable or inducing shifts in domestic outlets' news reporting. The demand for uncensored information, once raised, is likely to persist and can generate substantial pressure on the censorship apparatus.

## FUTURE RESEARCH DIRECTIONS

There are many interesting questions about social media (and media in general) in autocratic regimes. I see three promising directions where more progress can be made to help us understand the role media plays in autocratic regimes.

First, media censorship is an important component of autocrats' toolkit, which also contains tools such as propaganda, repression, and elite co-optation (Guriev and Triesman 2021). How do these tools interact with one another? Does deploying censorship and propaganda lead to a decrease in repression? How do online and offline control interact – are they strategic complements or substitutes? How do these tools responsive to broad social, economic and (geo)political conditions? We know very little empirically about these issues.

Second, understanding the supply of information under a censored environment is key to our knowledge of how media operates in autocratic regimes. Regarding domestic media, Zhuang (2022) shows that intergovernmental competition in China affects the degree of censorship imposed on local media, thus affecting the landscape of information available on censored, domestic media. Regarding foreign media, Chen and Han (2022) document that foreign media outlets become more negative in their coverage on China once they are blocked by the Great Firewall. These studies point to exciting new avenues as we seek to understand the relationship between information supply on (censored) domestic media and (uncensored, but perhaps self-censored) foreign media, and ultimately how they shape and react to consumer demand.

Finally, as social media platforms around the world are battling misinformation and engaging in content moderation, there is a growing literature that seeks to understand the effects of content moderation. For example, Beknazar-Yuzbashev et al. (2023) study how blocking of toxic content affects user engagement on Facebook, Twitter and YouTube, while Jiménez Rurán et al. (2022) analyse the effect of content moderation on hate speech in Germany. The knowledge – both theoretical and empirical – that we have gained from media censorship in autocratic regimes can become relevant in the context of content moderation, despite the very difference in the objective functions featured in each situation.

## REFERENCES

- Beknazar-Yuzbashev, G, R Jiménez Durán, J McCrosky and M Stalinski (2022), “Toxic Content and User Engagement on Social Media: Evidence from a Field Experiment”, available at SSRN.
- Chen, H and L Han (2022), “Do the Media Bow to Foreign Economic Powers? Evidence from a News Website Crackdown”, HKUST CEP Working Paper No. 202201.
- Chen, Y and D Y Yang (2019), “The Impact of Media Censorship: 1984 or Brave New World?”, *American Economic Review* 109(6): 2294–2332.
- DellaVigna, S and M Gentzkow (2010), “Persuasion: Empirical Evidence”, *Annual Review of Economics* 2(1): 643–669.
- Ford, G S (1935), *Dictatorship in the Modern World*, The University of Minnesota Press.
- Guriey, S and D Treisman (2019), “Informational Autocrats”, *Journal of Economic Perspectives* 33(4): 100–127.
- Guriey, S, N Melnikov and E Zhuravskaya (2020), “3G Internet and Confidence in Government”, *The Quarterly Journal of Economics* 136(4).
- Jiménez Durán, R, K Müller and C Schwarz (2022), “The Effect of Content Moderation on Online and Offline Hate: Evidence from Germany’s NetzDG”, CEPR Discussion Paper 17554.
- Roberts, M E (2018), *Censored. Distraction and Diversion Inside China’s Great Firewall*, Princeton University Press.
- Roberts, H, E Zuckerman, J York, R Faris and J Palfrey (2010), *2010 Circumvention Tool Usage Report*, Berkman Klein Center for Internet & Society.
- Strömberg, D (2015), “Media and politics”, *Annual Review of Economics* 7(1): 173–205.
- Zhuang, M (2022), “Intergovernmental Conflict and Censorship: Evidence from China’s Anti-Corruption Campaign”, *Journal of the European Economic Association* 20(6).

## ABOUT THE AUTHOR

**David Y. Yang** is a Professor in the Department of Economics at Harvard University and Director of the Center for History and Economics at Harvard. David is a Faculty Research Fellow at NBER, a Global Scholar at CIFAR, and a fellow at BREAD. David’s research focuses on political economy. In particular, David studies the forces of stability and forces of changes in authoritarian regimes, drawing lessons from historical and contemporary China. David received a B.A. in Statistics and B.S. in Business Administration from University of California at Berkeley, and PhD in Economics from Stanford.





**SECTION 7**  
**SOCIAL MEDIA AND LEGACY**  
**MEDIA**



# CHAPTER 17

## Social media and legacy media

195

**Sophie Hatte and Ekaterina Zhuravskaya**

Ecole Normale Supérieure de Lyon; Paris School of Economics and CEPR

In the Web 2.0 world, social media has changed the way journalists report the news. Online platforms help traditional media to promote their content and get feedback through comments, and serve as real-time indicators of demand for news on different topics. Furthermore, user-generated content on social media endows journalists with opportunities for reporting unmet in pre-Web 2.0 world. Participants in and witnesses of newsworthy events record and post their recordings online. This allows journalists to learn details of breaking news they did not witness, find new sources, and gather more accurate information.

The use of social media by legacy media is not new. It exploded over the past decade and is now a staple of contemporary journalism – in the US, for example, more than nine out of ten journalists use social media (Pew Research Center 2022). While there is a consensus among media scholars that social media has become a first-order determinant of journalists' work (e.g. Adornato 2016, Lysak et al. 2012, von Nordheim et al. 2018), prior literature provides no systematic analysis of the effect of social media on the news produced by legacy media. This question is important as, even in the age of social media, the general public consumes a lot of news that is originally produced and reported by legacy media. Traditional news organisations are still seen by many as trusted gatekeepers of information (Pew Research Center 2021).

Two recent papers have taken a more systematic approach to studying the effects of social media on legacy media. Cagé et al. (2022) show that traditional media cover stories that are trending on Twitter. This evidence is based on the content analysis of online editions of French mainstream media and French-language Twitter. In a paper with Etienne Madinier (Hatte et al. 2022), we go a step further and ask *how* the inflow of user-generated content on social media shapes the production of news by legacy media, particularly when reporting is hard and costly and journalists cannot witness the newsworthy events themselves. Such situations arise when reporting is dangerous, when newsworthy events are unpredictable or when they take place in sites inaccessible to journalists. Wars and other violent conflicts often share all of these criteria. We focus on one such highly socially mediated war – the Israeli-Palestinian conflict – and document substantial changes brought about by user-generated content released from the conflict zone on US TV reporting about this conflict. Importantly, we show not only that the amount of coverage of the Israeli-Palestinian conflict by US television networks is

affected by social media from the conflict zone, but also that on-site social media changes the way the legacy media chronicles the conflict. Both the paper by Cagé et al. (2022) and our paper establish a causal effect of social media on legacy media. Focusing on war reporting allows us to open the black box of how exactly the content of legacy-media war reporting is changed by social media.

### **SOCIAL MEDIA IN THE NEWSROOM**

Back in 2010, Alan Rusbridger, then the editor-in-chief of *The Guardian*, explained during the Andrew Olle Media Lecture that: “[n]ews organizations still break lots of news. But, increasingly, news happens first on Twitter. [...] There are millions of human monitors out there who will pick up on the smallest things and who have the same instincts as the agencies—to be the first with the news. As more people join, the better it will get.”<sup>1</sup> Rusbridger’s forecast has been confirmed a decade later: journalists from well-established traditional media outlets extensively source information on social media, and Twitter clearly ranks at the top of the list (e.g. Moon and Hadley 2014).

### **THE CITIZEN JOURNALISM IN THE WEB 2.0 WORLD**

Journalists and open-source investigators working on conflicts (e.g. Patrikarakos 2017, Higgins 2021) stress that user-generated content on social media transforms eyewitnesses of events into ‘citizen journalists’. The Syrian civil war was one of the first socially mediated conflicts; participants of every conflict since have posted a lot of online information. The invasion of Ukraine by Russia is a vivid example of how social media is changing the way violent conflicts are reported on traditional media. Nowadays, major television networks such as the BBC, CNN, FOX and Al Jazeera, irrespective of their political leanings, dedicate teams of people to processing, geo-referencing and verifying videos, photos and text posted by ordinary citizens as well as war actors online. This information proves to be an invaluable source used in the production of news on conflicts.

The story of Farah Baker, a 16-year-old Palestinian girl who tweeted in English from Gaza during the 2014 Gaza War, illustrates the citizen journalist–legacy media nexus in the context of conflict (e.g. Patrikarakos 2017: 21-37). In the summer of 2014, Farah became known across the whole world (Reading 2016, Patrikarakos 2017). Legacy media and news agencies across the globe covered the content she posted on her Twitter account, in which she described the Israeli bombing raids on her town during the war (Figure 2).

David Patrikarakos, in his wonderful book *War In 140 Characters: How Social Media Is Reshaping Conflict in the Twenty-first Century*, writes:

1 [www.theguardian.com/media/2010/nov/19/alan-rusbridger-twitter](http://www.theguardian.com/media/2010/nov/19/alan-rusbridger-twitter) (accessed 2 August 2022).

*“The majority of articles [by legacy-media outlets] were based on her tweets and the narrative around them. In effect, they treated her Twitter feed like a newswire service; a tweet became comparable to an associate press bulletin” (Patrikarakos 2017: 34).*

In her tweets, Farah described what she saw and heard. She also posted videos of what she filmed through her window.

*“This is the car which was bombed at my house door #Gaza #GazaUnderAttack,” she tweeted on July 26 with accompanying photo of the destroyed vehicle... But it was the detailing of her emotions—her fear for her safety and for that of her family, especially her little sister, Lamar—that was by far the most powerful and popular element of her output” (Patrikarakos 2017: 27).*

Some of the Farah’s tweets went viral. Often this happened when journalists and opinion makers with many followers re-tweeted her first.

*“Tweets begat retweets, which begat greater audiences, which begat news coverage, which begat demonstrations, which begat yet more news coverage, most of it pro-Gaza” (Patrikarakos 2017: 35).*

In an interview with Patrikarakos, Farah said:

*“[with Twitter] more people ... can see what you write, and crucially, journalists use it as a source. People on the ground tweeting photos and descriptions of events during wartime have become invaluable—especially as they often tweet or post from areas too dangerous for journalists to go... It allows the victims of war to gain a voice and the world to view—with greater detail than ever before—just what exactly is happening inside zones of conflict” (Patrikarakos 2017: 25).*

## **USER-GENERATED CONTENT POSTED ONLINE CHANGES THE REPORTING OF CONFLICT BY LEGACY MEDIA**

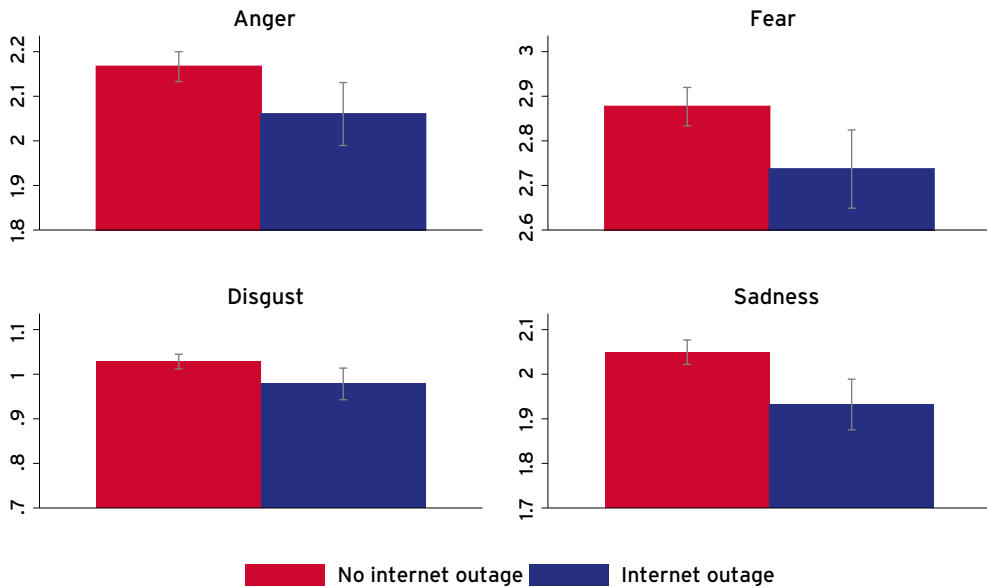
How does user-generated content posted on social media affect the reporting of violent conflict by legacy media? In our paper, we gather rich data on the Israeli-Palestinian conflict to address this question. Specifically, we explore how the content posted on Twitter from the conflict zone in Israel and Palestine affects the coverage of the conflict by eight major television networks in the United States (ABC, CBS, NBC, CNN, FOX, MSNBC, PBS, and Bloomberg).

To identify the effect of user-generated content posted online on the reporting of the conflict by US television, we exploit exogenous variation in access to the internet for ordinary people located in the conflict zone, stemming from internet outages in Israel and Palestine. Two sources of internet outages are combined in our study: lightning strikes and technical failures. Lightning strikes cause electrostatic discharges and power surges that cause outages. They can thus significantly reduce connectivity to both fixed

and mobile internet, especially when providers and users do not have access to power surge protection tools. Another important source of internet outages are software failures, which can lead to a fall in connectivity between a certain geographical area and the rest of the World Wide Web. Using a frontline methodology developed by computer scientists, we detect days when internet users in Israel and Palestine could not connect to the internet. The number of tweets about the conflict from the conflict zone drops significantly on the day of the internet outage, and tweeting rebounds the day after the outage episode ends. This is largely driven by the inability of the most vulnerable Twitter users to connect to the platform: tweets from ordinary people (or ‘citizen journalists’) are most affected compared to tweets posted by business and organisations, news outlets and local leaders because ordinary people do not have satellite phones or power surge protection tools. In our paper, we discuss the validity of using outages stemming from lightning strikes and technical failures as a source of exogenous variation. We show that the outages are unrelated to the conflict events and are not caused by the conflict actors.

What does US television report on when the internet and social media are available in the conflict zone compared to when they are unavailable? We find that the increased inflow of user-generated content posted from Israel and Palestine translates into more conflict coverage on US television, both at the extensive and intensive margins. This result is obtained after controlling for import drivers of news coverage, notably the current and past intensity of the conflict and non-conflict-related news pressure in US television. The extent of conflict coverage, for instance, is larger when there are intensive attacks in the conflict area and it is smaller when there are important newsworthy events in the United States, which journalists are compelled to cover and which crowd out coverage of the Israeli-Palestinian conflict. We find that, on average, an internet outage causes a 2.5 percentage point decrease in prime-time conflict coverage on US TV and a 6.5 minute decrease in the length of conflict-related broadcast per day.

More importantly, online content also substantially changes the content and the tonality of conflict coverage. Online content about the conflict in Israel and Palestine makes US television news about the conflict more emotional. Figure 1 shows that words with higher negative emotional intensity (i.e. words associated with some intensity to anger, fear, disgust and sadness) are used more frequently during US television news stories about the conflict zone when the internet is not muted by outages in Israel and Palestine. The effects we estimate are sizeable. One way to interpret their magnitude is to compare them to the emotional intensity of different words in the emotion lexicon. The estimated effects imply that internet outages in the conflict zone make the US TV news about the conflict less emotional, with a difference in the average emotional intensity of the broadcasts with and without an internet outage in the conflict zone comparable to the difference between the words “disagreement” and “catastrophe”.

**FIGURE 1 INTERNET OUTAGES AND EMOTIONAL INTENSITY OF CONFLICT COVERAGE**

Note: The figure presents an illustration of the reduced-form relationship between emotional intensity of coverage of the conflict by US TV news and internet outages in the conflict zone. The graphs summarise the emotions scores of conflict zone-related broadcasts across days and TV networks separately for the days with and without internet outage. The unit of observation is a day x TV network. Sample: all days and TV networks with at least one story about the conflict zone. We use measures based on the use of emotional words divided by the total number of words.

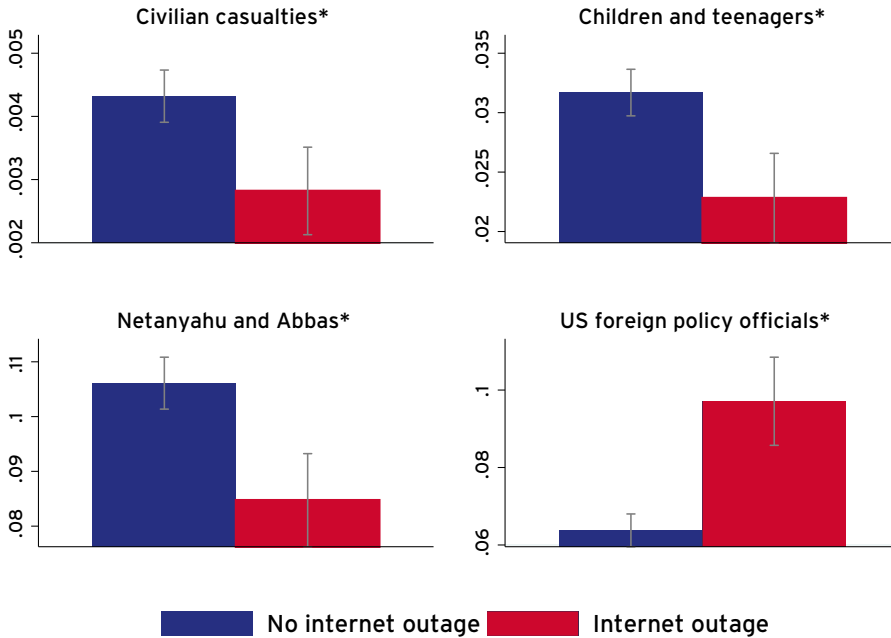
Source: Hatte et al. (2022)

User-generated content posted from the conflict zone also has a significant impact on the topics of conflict-related broadcasts on US television. These topics are identified first by simple keyword counts, identifying relevant lists of keywords to capture specific topics. When users in the conflict zone have access to the internet and social media, TV news segments in the United States about the conflict focus more on the stories of civilians caught in the conflict and less on the role of the US Secretary of State or elections, as illustrated in Figure 2. These television stories are more likely to explicitly mention Twitter and Facebook. This analysis is cross-validated by an alternative approach to identifying the concrete topics of the conflict-related US TV broadcast, exploiting a frontline machine-learning (LDA) algorithm.

Importantly, television stories about the conflict contain more details about the events on the ground when social media is not muted by outages in the conflict zone: the US television stories produced on the Israeli-Palestinian conflict mention the names of heavy armaments more often and refer more to specific small geographic locations, especially in Palestine.



**FIGURE 2 INTERNET OUTAGES AND KEYWORDS MENTIONED BY STORIES ON THE CONFLICT**



Note: The figure presents an illustration of the reduced-form relationship between emotional intensity of coverage of the conflict by US TV news and internet outages in the conflict zone. The graphs summarise the emotions scores of conflict zone-related broadcasts across days and TV networks separately for the days with and without internet outage. The unit of observation is a day x TV network. Sample: all days and TV networks with at least one story about the conflict zone. we use measures based on the use of emotional words divided by the total number of words.

Source: Hatte et al. (2022)

In our paper, we study US television networks with very different ideological leanings, which allows us to explore variation in the social media effect across networks. No clear cross-network heterogeneity arises from our analysis. However, social media in the conflict zone appears to make news stories about the conflict on the different US TV networks more similar to each other. To understand whether the results are heterogeneous with respect to the attitudes of the TV network's editors towards the two sides of the conflict, we explore data on the coverage of the conflict by Al Jazeera America – a Qatari network available in the United States. The content of news stories on the eight US television networks becomes more similar to that on Al Jazeera America when social media from the conflict zone is not muted by internet outages. This is an important finding which suggests that ideological differences are erased when the legacy media cover the suffering of civilians. There are not two sides to a story when civilians get caught in a violent conflict.

We also assess how social media changes which side of this highly asymmetric conflict gets more coverage as a result of the social media presence in the conflict zone. While we find that civilian suffering on both sides gets much more coverage, due to the asymmetry of the conflict and given that Palestinians experience much higher civilian casualties, the

increase in mentions of Palestinian suffering by legacy media due to social media from the conflict zone is seven times higher than the increase in mentions of Israeli suffering. These results suggest that social media helps the narrative of the side that suffers a higher civilian death toll and serves as a tool for levelling the playing field in the information sphere in conflicts in which military capabilities are highly asymmetric.

## REFERENCES

Adornato, A C (2016), “Forces at the Gate: Social Media’s Influence on Editorial and Production Decisions in Local Television Newsrooms,” *Electronic News* 10(2): 87–104.

Cagé, J, N Hervé, and B Mazoyer (2022), “Social Media Influence Mainstream Media: Evidence from Two Billion Tweets”, CEPR Discussion Paper DP17358.

Hatte, S, E Madinier, and E Zhuravskaya (2022), “Reading Twitter in the Newsroom: Web 2.0 and Traditional-Media Reporting of Conflicts”, CEPR Discussion Paper DP16167.

Higgins, E (2021), *We Are Bellingcat: An Intelligence Agency for the People*, Bloomsbury Publishing.

Lysak, S, M Cremedas, and J Wolf (2012), “Facebook and Twitter in the newsroom”, *Electronic News* 6(4): 187–207.

Moon, S J and P Hadley (2014), “Routinizing a new technology in the newsroom: twitter as a news source in mainstream media,” *Journal of Broadcasting and Electronic Media* 58(2): 289–305.

Patrikarakos, D (2017), *War In 140 Characters: How Social Media Is Reshaping Conflict In The Twenty-first Century* (1st edition), Basic Books.

Pew Research Center (2021), *At 100 Day Mark: Coverage of Biden Has Been Slightly More Negative Than Positive, Varied Greatly by Outlet Type*, Technical Report.

Pew Research Center (2022), *Journalists Sense Turmoil in Their Industry Amid Continued Passion for Their Work*, Technical Report.

Reading, A (2016), *Gender and Memory in the Global Age*, Palgrave Macmillan UK.

von Nordheim, G, K Boczek, and L Koppers (2018), “Sourcing the Sources,” *Digital Journalism* 6(7): 807–828.

## ABOUT THE AUTHORS

**Sophie Hatte** is an Assistant Professor at Ecole Normale Supérieure de Lyon and a research affiliate at the Center for Economic Research on Governance, Inequality and Conflict. Before joining ENS de Lyon, she was a postdoctoral researcher at the University of Lausanne. Her research focuses on issues related to political economy, development, and public economics.

**Ekaterina Zhuravskaya** is a Professor of Economics at the Paris School of Economics (EHESS), which she joined in 2010. She is a Research Fellow of CEPR in Political Economy and Development Economics programs. She got her PhD at Harvard University in 1999 and spent 10 subsequent years working as Professor at the New Economic School and as the Academic Director of the Center for Economic and Financial Research in Moscow. Her main academic interests are in political economy, with applications to media, economic history, and institutions. She is the recipient of the Birgit Grodahl Award and CNSR Silver Medal. She also was elected a Fellow of the Econometric Society.

# CHAPTER 18

## Contagion from social media to mainstream media

203

**Julia Cagé,<sup>ac</sup> Nicolas Hervé<sup>b</sup> and Béatrice Mazoyer<sup>ab</sup>**

<sup>a</sup>Sciences Po; <sup>b</sup>Institut National de l'Audiovisuel; <sup>c</sup>CEPR

Many recent papers have shown that social media has changed society (Fujiwara et al. 2021, Levy 2021), yet television remains by far the most popular source of news. So how can we explain the outsized influence of social media? In this chapter, we rely on our recent paper (Cagé et al. 2022) to quantify what many long suspected: that Twitter affects publishers' production and editorial decisions.<sup>1</sup>

We also add a new set of perspectives to the literature studying the welfare effects of social media (Allcott et al. 2020). While there are widespread fears that new technologies are worsening editorial quality, we investigate whether the contagion from social to mainstream media varies depending on the characteristics of the outlets – in particular, on whether they offer digital news for free. The influence of social media may indeed increase information inequality (Kennedy and Prat 2019), which would in turn affect voting outcomes.

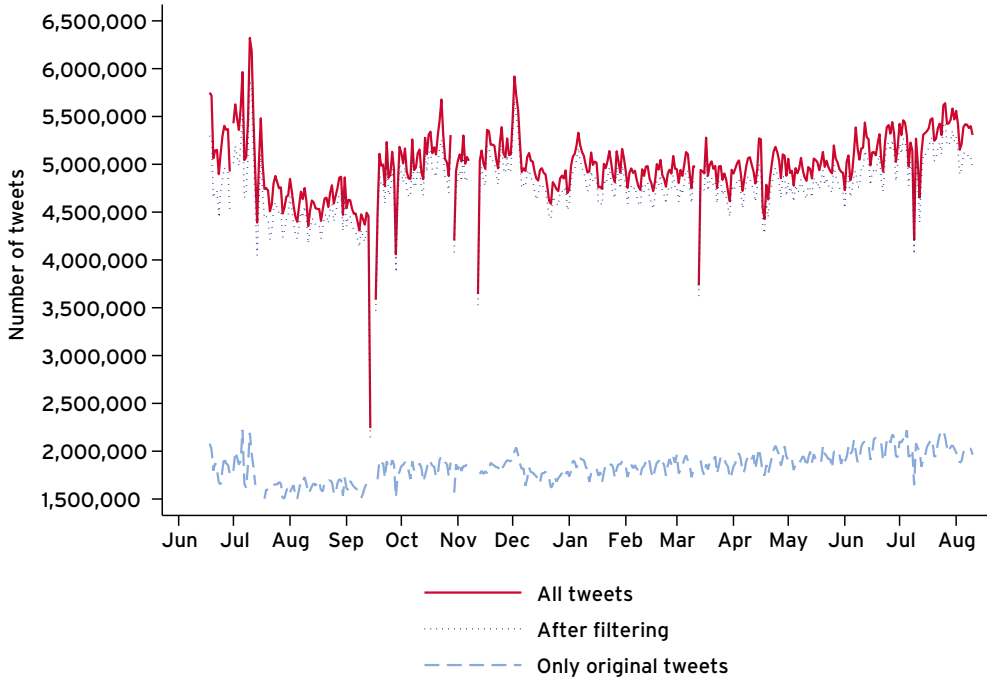
To do so, we proceed in three steps. First, we collect a representative sample of all the tweets produced in French between August 2018 and July 2019 and combine it with the content published online by all the mainstream media outlets (encompassing newspapers, television channels, radio stations, pure online media, and news agencies' dispatches). Our dataset, which contains around 1.8 billion tweets, covers around 70% of all the tweets in French (including retweets) during this time period. Figure 1 plots the daily distribution of the number of tweets.

For each of these tweets, we collect information on their 'success' on Twitter (number of likes, of comments, etc.), as well as information on the characteristics of the user at the time of the tweet (e.g. its number of followers). To construct this unique dataset, we have

1 Interesting empirical evidence is provided by Hatte et al. (2021), who study the effect of Twitter on the US TV coverage of the Israeli-Palestinian conflict. Compared to this work, our contribution is threefold. First, we focus on the overall activity on Twitter and collect a large representative sample of about 70% of all tweets (about 1.8 billion tweets) rather than the tweets associated with a small number of keywords. Second, we develop an instrument for measuring popularity shocks on Twitter, based on the structure of the network, which could be of use in different contexts. Finally, we investigate whether there are heterogeneous effects depending on the media characteristics, in particular their business model and their reliance on advertising revenues.

combined the sample and the Filter Twitter application programming interfaces (APIs), and selected keywords. Figure 2 summarises our data collection setup, and Mazoyer et al. (2018, 2020) provide details.

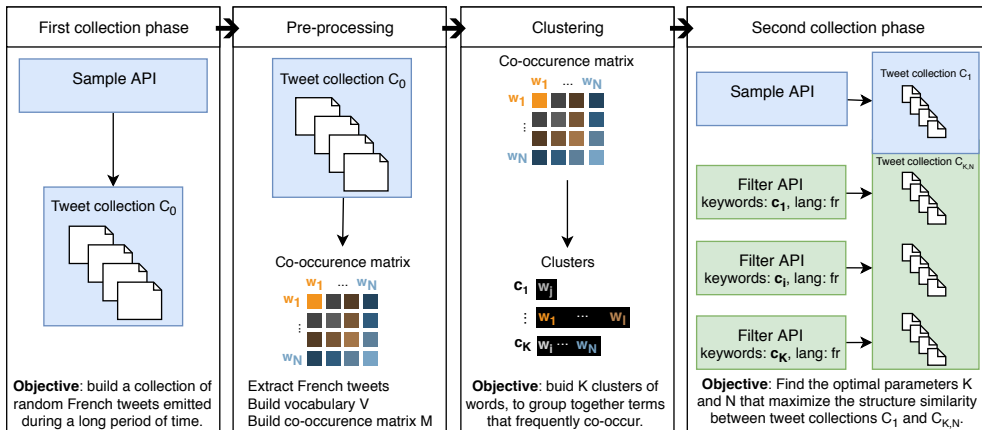
**FIGURE 1 DAILY DISTRIBUTION OF THE NUMBER OF TWEETS IN THE SAMPLE**



Note: The figure plots the daily number of tweets included in our dataset. The red line plots all the tweets, the dotted blue line plots these tweets once we apply the filter, and the green dashed line plots only the original tweets. Time period is 18 June 2018 to 10 August 2019. The few number of days without information comes from exceptional days when the server collapsed and we were thus unable to capture tweets in real time.

Source: Cagé et al. (2022).

**FIGURE 2 DIAGRAM OF OUR EXPERIMENTAL SETUP TO SELECT THE BEST TWEET COLLECTION METHOD**



Source: Cagé et al. (2022).

Second, we develop novel algorithms to identify all the ‘news stories’ covered both on social and traditional media. An ‘event’ here is a cluster of documents (tweets and media articles) that discuss the same news story. For example, all the documents (tweets and media articles) discussing the Hokkaido Eastern Iburi earthquake on 6 September 2018 will be classified as part of the same event. Events are detected by our algorithms using the fact that the documents share sufficient semantic similarity. In a nutshell, regarding Twitter, our approach consists in modelling the event detection problem as a dynamic clustering problem, using a ‘first story detection’ (FSD) algorithm (see Mazoyer et al. 2022 for more details). To detect the news events among the stories published online by traditional media outlets, we follow Cagé et al. (2020) and describe each news article by a semantic vector (using TF-IDF) and use the cosine distance to measure their semantic similarity. Used jointly with temporal constraints, we can cluster the articles to form the events. Finally, to generate the intersection between social media events and mainstream media events, we rely on the Louvain community detection algorithm (Blondel et al. 2008), as illustrated in Figure 3.

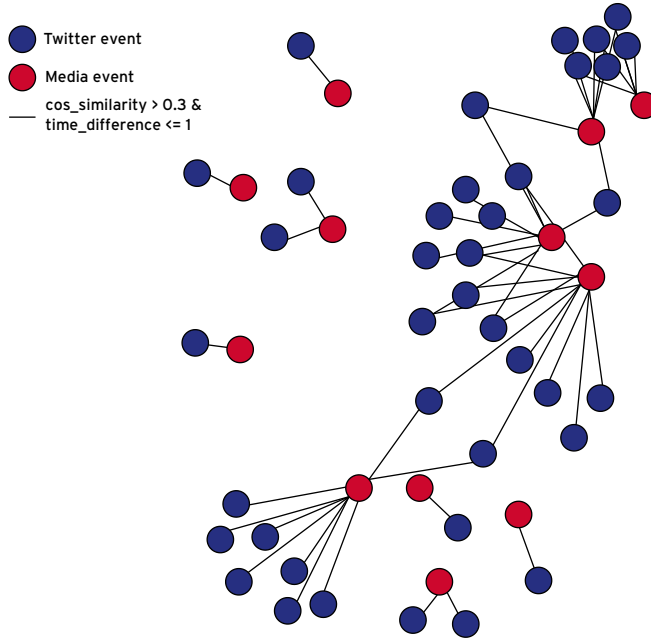
We identify 3,992 joint events – i.e. events that are covered both on social and on traditional media – out of which 3,904 originate first on Twitter.

Third, we rely on the structure of the social media network – and in particular, on the centrality of its users – to isolate ‘exogenous’ shocks to the popularity of the stories on Twitter (measured by the number of tweets, retweets, likes, etc.). In other words, we isolate variations in the popularity of stories on Twitter independently of the intrinsic interest of these stories. To do so, we leverage the enormity of our dataset to propose a novel instrumental variable strategy: our instrument is the interaction between the first Twitter users’ centrality in the network – measured computing PageRank centrality (Page et al. 1999) just before the event – and the news pressure in the social media at the time of the first tweets on the event. News pressure is measured by the number of interactions generated by all the tweets published in the hour preceding the first tweet in the event.

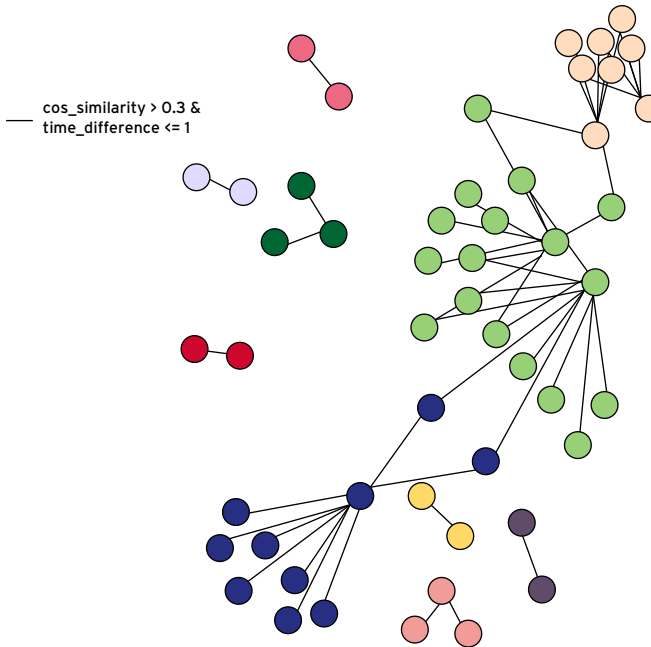
Our identification assumption is that, once we control for the direct effect of centrality and news pressure, the interaction between users’ centrality and news pressure should only affect traditional news production through its effect on the visibility of the tweets on Twitter. While we cannot rigorously test this assumption, we show that our instrument is not correlated with a number of observable event characteristics, such as its topic (economy, sport, environment, etc.) or the number of named entities (mentions of places such as Paris, or of individuals such as Boris Johnson).

FIGURE 3 GRAPHICAL REPRESENTATION: BUILDING THE JOINT EVENTS

a) Building the similarity network



b) Applying Louvain algorithm

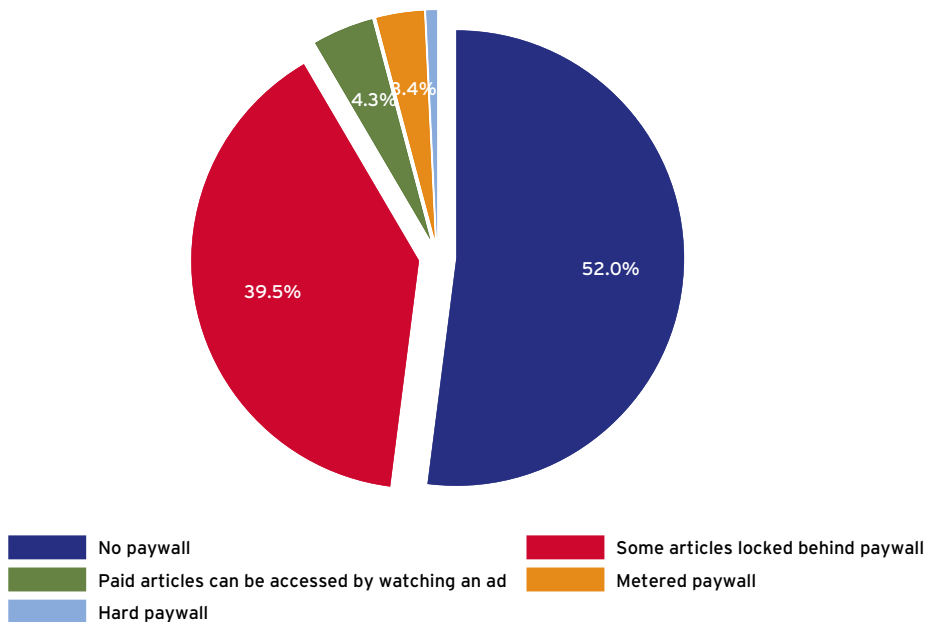


Our findings are enlightening: everything else equal – and in particular, independently of the newsworthiness of a story – a 50% increase in the number of tweets posted before the first media article on a story leads to an increase in the number of news articles covering the story corresponding to 17% of the mean (see Table 7 in Cagé et al. 2022 for detailed estimates). That is to say, Twitter sets the agenda of media coverage in a quantitatively meaningful way.

Why is this so? A growing literature in journalism studies highlights the fact that social media plays an important role as a news source. Consistent with this idea, we show that the magnitude of the effect is higher for the media outlets that have a high number of journalists with a Twitter account, pointing toward the role of the monitoring of Twitter by journalists.

But the use of the platforms as journalistic sources is not the only factor at play here; in particular, we investigate whether the magnitude of the contagion between social and mainstream media depends on the outlet's business model. For each of the outlets in our dataset, we collect information on whether it uses a paywall (at the time of the data collection), the characteristics of this paywall (e.g. soft versus hard), and the date of introduction of the paywall. This information is summarised in Figure 4.

**FIGURE 4 NEWS EDITORS' BUSINESS MODELS**

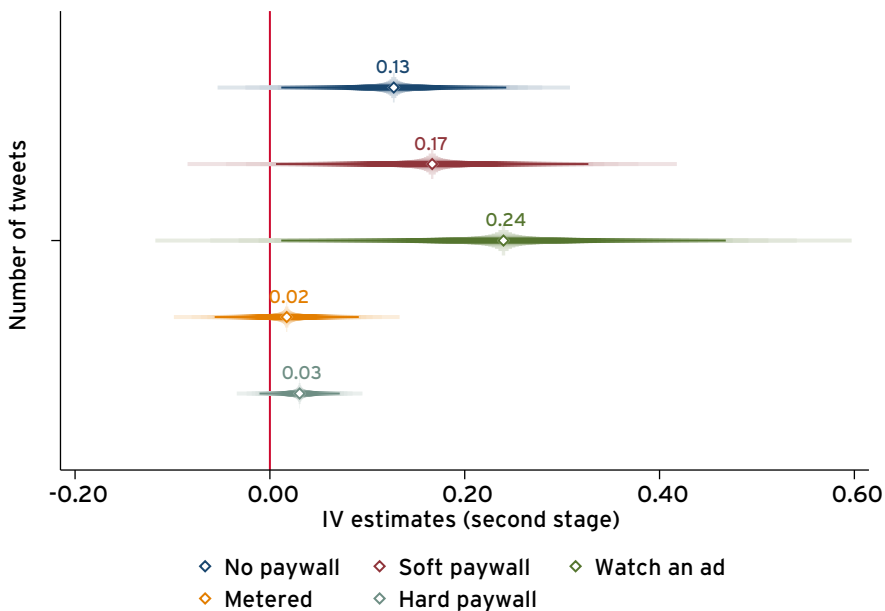


Note: The figure breaks down the media outlets in our sample by their online business model. Fifty-two percent of the outlets in our site do not have a paywall ("no paywall"), and 4.3% condition the reading of the paid articles on the fact of watching an advert ("paid articles can be accessed by watching an ad"). Of the outlets that do have a paywall, we distinguish between three models: hard paywall, metered paywall and soft paywall ("some articles locked behind paywall").



In Figure 5, we show that the magnitude of our effects is much larger for the media outlets that fully or strongly rely on advertising revenues than for those whose online content is behind a paywall (and thus mainly depend on subscriptions). For the former, a 50% increase in popularity leads to an increase in news coverage corresponding to 22.0% (no paywall), 20.3% (soft paywall) and 21.1% ('watch an ad' paywall) of the mean, compared to 6.2% of the mean for the outlets using a metered paywall, a coefficient that is furthermore not statistically significant. In other words, Twitter influences mainstream media because of short-term considerations generated by advertising revenue-bearing clicks. This result is in line with findings in Sen and Yildirim (2015), who show, using data from a leading Indian national daily newspaper, that editors' coverage decisions regarding online news stories are influenced by the observed popularity of the story, as measured by the number of clicks received. According to our findings, media outlets whose revenues mostly come for subscriptions, in contrast, seem to be much less influenced by social media.

**FIGURE 5 THE IMPACT OF THE POPULARITY OF A STORY ON TWITTER, DEPENDING ON THE PAYWALL**



Note: The figure plots the results of the IV estimation of the popularity of a story on Twitter (as measured by the logarithm of the number of tweets and instrumented by the interaction between users' centrality and news pressure) on the number of articles. The time period is 1 August 2018 to 30 November 2018. We consider separately (1) the media outlets that do not have a paywall, (2) those that use a soft paywall, (3) those that require consumers to watch an ad before being allowed to read the articles, (4) the media outlets that use a metered paywall, and finally (5) the outlets that rely on a hard paywall.

Source: Cagé et al. (2022).

While there are widespread fears that new technologies are worsening editorial quality – in particular, because they have led to savings in the newsroom, which in turn have reduced the quality of news provision and the production of original content – our findings imply that they are disproportionately worsening it for people who cannot afford

or are unwilling to pay for news. Put another way, because media outlets whose content is available online for free tend to be more influenced by the popularity of stories on Twitter than those using a paywall, the platform generates an increase in information inequality, making disadvantaged citizens further vulnerable to manipulation (Kennedy and Prat 2019). This may in turn affect voting outcomes and increase political inequality.

Our findings, which capture the effects of a variation in popularity that is uncorrelated with a story's underlying newsworthiness, also suggest that social media may provide a biased signal of what readers want, which may in turn explain why, as highlighted by survey data, a significant share of the population is not interested in the news produced by the media (and might thus decide not to consume news). Twitter users are indeed not representative of the general news-reading population. This points to a negative effect of social media driven by the production side, consistent with recent changes in both *The Guardian's* and *The New York Times's* social media guidelines, which highlight the fact that journalists tend to rely too much on Twitter as both a reporting and feedback tool, and that it may distort their view of who their audience is.

Turning to the demand for news and using audience data, we finally show that the news articles covering events that are more popular on Twitter do not get more views compared to the other articles, further suggesting that the journalists' reliance on Twitter might distort the information they produce compared to the preferences of the citizens.

The findings of our research paper, summarised in this chapter, shed new light on our understanding of how editors decide on the coverage for stories, and should to be taken into account when discussing policy implications of the recent changes in media technologies. Social media matters for democracy beyond what we could have expected. Not only does it impact the users who spend time on the platforms, but there is contagion from social to mainstream media. This contagion questions the business model of the legacy media, as well as the welfare effects of the platforms. In particular, from our results, one may call into question whether citizens would be better informed in the absence of Twitter, and wonder whether social media may be harmful to both journalism and democracy.

## REFERENCES

Allcott, H, L Braghieri, S Eichmeyer, and M Gentzkow (2020), "The Welfare Effects of Social Media", *American Economic Review* 110 (3): 629–676.

Blondel, V D, J-L Guillaume, R Lambiotte, and E Lefebvre (2008), "Fast unfolding of communities in large networks", *Journal of Statistical Mechanics: Theory and Experiment* 2008: P10008.

Cagé, J, N Hervé, and M-L Viaud (2020), "The Production of Information in an Online World", *The Review of Economic Studies* 87(5): 2126–2164.

Cagé, J, N Hervé, and B B Mazoyer (2022), “Social Media Influence Mainstream Media: Evidence from Two Billion Tweets”, CEPR Discussion Paper 17358.

Fujiwara, T, K Muller, and C Schwarz (2021), “The Effect of Social Media on Elections: Evidence from the United States”, NBER Working Paper 28849.

Hatte, S, E Madinier, and E Zhuravskaya (2021), “Reading Twitter in the Newsroom: How Social Media Affects Traditional-Media Reporting of Conflicts”, CEPR Discussion Paper 16167.

Kennedy, P J and A Prat (2019), “Where Do People Get Their News?”, *Economic Policy* 34(97): 5-47.

Levy, R (2021), “Social Media, News Consumption, and Polarization: Evidence from a Field Experiment”, *American Economic Review* 111(3): 831-870.

Mazoyer, B, J Cagé, C Hudelot, and M-L Viaud (2018), “Real-time collection of reliable and representative tweets datasets related to news events”, in *CEUR Workshop Proceedings*, Vol. 2078.

Mazoyer, B, J Cagé, N Hervé, and C Hudelot (2020), “A French Corpus for Event Detection on Twitter”, *Proceedings of the Twelfth Language Resources and Evaluation Conference*.

Mazoyer, B, N Hervé, C Hudelot, and J Cagé (2022), “Short-Text Embeddings for Unsupervised Event Detection in a Stream of Tweets”, *Advances in Knowledge Discovery and Management* 10.

Page, L, S Brin, R Motwani, and T Winograd (1999), *The PageRank Citation Ranking: Bringing Order to the Web*, Technical Report 1999-66, Stanford InfoLab.

Sen, A and P Yildirim (2015), “Clicks and Editorial Decisions: How Does Popularity Shape Online News Coverage?”, available at SSRN.

## ABOUT THE AUTHORS

**Julia Cagé** is an Associate Professor of Economics in the Department of Economics at Sciences Po Paris, and a CEPR Research Fellow in Public Economics, Industrial Organization and Economic History. She is also co-director of the Laboratory for Interdisciplinary Evaluation of Public Policies (LIEPP)’s “Evaluation of Democracy” research group. She completed her PhD at Harvard University in 2014. Her work has been published in leading journals in economics and she has authored books on the topic of media economics, political participation and political attitudes.

**Nicolas Hervé** is the Head of the Research Department at INA (French National Audiovisual Institute). His current research activities focus primarily on the digital social sciences and humanities and, more specifically, on the information ecosystem using machine learning and data mining approaches. He coordinates the OTMedia research platform that enables transdisciplinary studies on the media world and the propagation of online information. He obtained his Engineer degree in 2005 from the Conservatoire National des Arts et Métiers (CNAM, Paris) and his PhD in 2009 from Paris-Sud University (Orsay).

**Béatrice Mazoyer** works as a research engineer at the médialab of Sciences Po Paris, an interdisciplinary research laboratory that investigates the role of digital technology in our societies. Her research focuses on the collection and analysis of social media data, using natural language processing and data mining methods. She completed her PhD at Paris-Saclay University in 2020.

The emergence of social media has reshaped the way humans communicate, interact and coordinate with each other. Assessing the impact of that transformation on politics has been one of the great social science questions of the last or decade or so, and will continue to occupy researchers for a long time to come. This book provides a snapshot of how economists in particular have been trying to answer this question. It contains 18 chapters, written by some of the leading scholars working on the topic, summarising empirical evidence on different dimensions of the political impact of social media.

The book starts by considering how social media platforms have affected the overall wellbeing of their users. It then goes over how they have changed the behaviour of voters, particularly through news consumption, and whether it can be linked to phenomena such as increased polarisation or the rise of populism. The next section looks at how politicians have responded to the new environment, and how that in turn has affected elections. The following two sections address the coordination role of social media, asking how it has affected political mobilisation and, on the negative side, the spread of political hatred. Another section focuses on how social media has changed politics in the autocratic context of China. Lastly, the final chapters shed light on how the political role of other, so-called legacy media has been impacted by the new technologies.

Put together, the contributions described in this book showcase how the ubiquity of social media, the nature of the networks that emerge through it, and the absence of barriers to entry in producing and broadcasting content all converge to make this technology a uniquely consequential transformation in the media environment.

ISBN 978-1-912179-80-0



187 boulevard Saint-Germain | 75007 Paris | France  
33 Great Sutton Street | LONDON EC1V ODX | UK  
TEL: +44 (0)20 7183 8801 | EMAIL: [CEPR@CEPR.ORG](mailto:CEPR@CEPR.ORG)  
[WWW.CEPR.ORG](http://WWW.CEPR.ORG)