Protecting political discourse from online manipulation: the international human rights law framework

Kate Jones[*]

Journal Article

European Human Rights Law Review

E.H.R.L.R. 2021, 1, 68-79

Subject
Constitutional law

Other related subjects
Human rights; Information technology; International law

Keywords
Freedom of expression; Human rights; International law; Online intermediaries; Political activities; Politics; Right to vote

Legislation cited
International Covenant on Civil and Political Rights 1966 art.17, art.18, art.19, art.25

E.H.R.L.R. 68 **Abstract**
Diverse forms of online interference in political debate can be grouped together as manipulation, which poses a significant threat to our democracies, whether conducted by foreign or domestic actors. This article argues that both states and internet platforms should respond to this threat, and their responses should be guided by international human rights law. It draws on the UN Guiding Principles on Business and Human Rights to explain how international human rights law places obligations on states and responsibilities on corporate actors. It explains how both foreign and domestic state actors have obligations not to engage in manipulation themselves and to protect their populations from human rights abuses that manipulation may cause. It discusses internet platforms' responsibilities to respect human rights. It then discusses briefly the substance of the principal human rights that may be infringed or engaged by online manipulation of political debate: the right to participate in public affairs and to vote, the freedoms of thought and of opinion, the right to privacy and freedom of expression. It concludes that both states and platforms should tackle online manipulation by taking steps to protect human rights, while avoiding blunt responses that themselves infringe human rights.

1. Introduction

The banning of US President Trump from Facebook and Twitter in January 2021 highlights both how widespread abusive behaviour is in the virtual political arena, and the apparently unfettered discretion of digital platforms to permit or curb it. This incident may be the apex of a period in which both domestic political actors and overseas powers have taken advantage of unregulated online environments to distort political discourse. The manipulation of political debate online is a crisis that can no longer be ignored nor left to its hosts, the online platforms, to curb without public sector involvement.

The manipulation of political debate presents a large and expanding threat to democracy. Since first hitting the headlines in the aftermath of the US presidential election in 2016, it has increasingly become a hazard of election campaigns and political discussion all over the world. Its techniques are being adopted both by overseas, often state-sponsored actors and by domestic political campaigners. Fact-checking, media literacy and support for robust independent media, while all important, are insufficient defences against technologies which are constantly developing, often covert, and have generally been implemented without regard to public or democratic interest.

Until now, different aspects of manipulation of political debate have been discussed separately and attempts to apply human rights law to them have been piecemeal. For example, we see increasing references ***E.H.R.L.R. 69*** to freedom of expression in discussions of content moderation by internet platforms, but little attention to human rights in discussion of technology that

may deter people from voting or polarise their views. Moreover, until January 2021, there has often been an assumption that attempts to tackle manipulation should be left to the internet platforms, an assumption made without regard to the state's duty to protect its population against abuses of human rights by non-state actors.

This article argues that international human rights law requires both states and internet platforms to take steps to quell manipulation online, and that international law ought to form the conceptual framework by which to frame responses to the challenge of online manipulation.

2. Online manipulation of political debate

We need a common terminology to describe interference in elections and other political debate. "Manipulation" is an appropriate term, defined as:

"[T]he action or an act of managing or directing a person, etc., esp. in a skilful manner; the exercise of subtle, underhand, or devious influence or control over a person, organization, etc.; interference, tampering".[1]

"Manipulation" captures the three essential elements of interference in political debate: that it is performed with the intention to manipulate; that it is hidden from those it seeks to influence; and that it is performed in order to exert influence or control over its audience in an underhand or devious manner. It does not restrict the wide array of forms that interference may take, nor the motivations of those engaging in it. The possibilities for manipulation in creation, distribution and audience targeting of online content allow for myriad forms, many of them not easily visible to researchers. The boundaries of manipulation have not yet been defined, such that the distinction between legitimate campaigning and manipulation is currently blurry. This area requires further research: the aftermath of the 2020 US Presidential election has raised unexplored questions about, for example, the deliberate generation or reinforcement of cult-like organisations and the deliberate adoption of online harassment strategies.

In practice, online manipulation of political debate, including election campaigns and referenda, is conducted both by international actors, often orchestrated with state support ("foreign interference operations") and by domestic actors, whether government agency, political party, private actor or individual ("domestic online manipulation campaigns"). The two sources of manipulation have often been discussed separately, as foreign interference operations have received considerable attention from cyber security experts and public international lawyers. From a human rights perspective, the same norms apply to both forms of manipulation, save as affected by the jurisdictional limits of human rights law, as discussed below.

Disinformation is one egregious example of manipulation: the intentional sharing of false, distorted or manipulated information in order to cause harm.[2] Its most significant characteristic is not the falsity of the information but the intention to cause harm. True information taken out of context or subjective opinion shared with similar intent can be equally problematic. One prominent monitor of disinformation defines it as comprising "adversarial narratives", narratives often developed from seeds of truth that are manipulated with the aim of creating opposition between societal groups.[3] Disinformation should be distinguished from misinformation, information that is false but is not created or shared with the intention of causing harm. ***E.H.R.L.R. 70***

Another patent example of manipulation is the use of networks of inauthentic accounts to increase the reach of material. Such online manipulation services are openly available on the internet and can easily be purchased.[4] To give a crude example, the more "likes" and comments a post has (for example, because a political party has bought them or a foreign actor deploys bots or employees to post them), the more social media algorithms will promote it to the heart of newsfeeds and online debate.

Other forms of manipulation are manifold. They may range from manipulative message content (for example words and pictures that are emotive, that deliberately divide and polarise their audience or undermine trust in government, including deliberate sowing of conspiracy theories, rumours, confusion or social discord, discrediting institutions or prominent individuals); to deliberately exaggerating the reach and impact of messages (for example by opaque use of advertisements, harnessing of groups, deploying techniques to distort platform algorithms); to use of synthetic techniques (such as deepfakes).[5] Facebook's reporting on "coordinated inauthentic behaviour",[6] Twitter's on "platform manipulation"[7] and Google's on "coordinated influence operation campaigns"[8] each track some of the online manipulation identified by them.

Manipulation is widespread and growing, and increasingly originates at the behest of major governments and political parties themselves.[9] In 2019—prior to 2020's COVID-19 "infodemic"—foreign interference operations in political discourse were found to originate from seven countries. The Oxford Internet Institute assessed that domestic organised online manipulation

campaigns, run by political parties or government agencies, took place in 70 countries (as compared to 48 in 2018); and that in 26 countries, authoritarian regimes were using computational propaganda to stymie political debate.[10] Inauthentic behaviour was recently assessed to account for over 25% of Twitter traffic between 2 and 20 May 2020, including 50% of traffic related to US conspiracy theories.[11] The types and scale of manipulation are likely to develop further with emerging data analytics and sentiment analysis technology.[12]

3. International human rights law: structure of obligations

International human rights law ("IHRL") should be at the heart of efforts to tackle manipulation. The UN Human Rights Council has repeatedly affirmed that "the same rights that people have offline must also be protected online".[13] IHRL places obligations on governments to comply with it, and responsibilities on political campaigners and online intermediaries to respect it in their activities. The rights engaged are discussed below. ***E.H.R.L.R. 71***

IHRL is an atypical body of public international law in that the obligations it creates for a state are not only owed to other states, but simultaneously to individuals both within its jurisdiction, and to some extent extraterritorially outside its jurisdiction. While international human rights treaties place legal obligations only on states, since the UN Human Rights Council adopted the non-binding UN Guiding Principles on Business and Human Rights in 2011[14] it has been widely accepted that business enterprises have responsibilities to respect the norms of IHRL, without jurisdictional limitation.[15]

IHRL is therefore relevant to regulation of and accountability for foreign interference operations and domestic online manipulation campaigns in three ways. Regarding foreign interference operations conducted or sponsored by State A in respect of an election in State B, IHRL is relevant in considering:

> 3.1 State A's obligations to individuals in State B ("international accountability"), to the extent State A has "extraterritorial jurisdiction" in respect of individuals in State B;
>
> 3.2 State B's obligations to individuals (and occasionally groups) within its jurisdiction ("domestic regulation and accountability"); and
>
> 3.3 the responsibilities of platforms, wherever located, to respect the human rights of individuals in State B ("corporate responsibilities").[16]

Points 3.2 and 3.3 apply equally in respect of domestic online manipulation campaigns. All three points are considered in turn below.

3.1 International human rights law: state accountability for foreign interference operations

Existing IHRL may offer some scope for state accountability for foreign interference operations, but only to the extent that its norms are ones of extraterritorial jurisdiction (i.e. to the extent that State A must respect them as regards the impact of its activities on individuals in State B).

In other contexts, the prevalent view of states, courts, expert bodies and academics is that, subject to the specific wording of each treaty, the state's obligations under international and regional human rights treaties, as well as customary international law, are primarily owed to individuals within its territory. Over the last 20 years, there has been growing acceptance that the state's civil and political rights treaty obligations are owed to an individual outside its territory if the state has physical power or control over the individual, either personally (for example because the state is detaining the individual overseas) or because the individual is in an area subject to the state's effective control (for example Turkey in respect of Northern Cyprus[17]).[18] A minority of states, notably the United States, considers that human rights obligations are always territorial, without extraterritorial exception.[19]

While this is by no means an established position, some expert bodies and courts may be moving towards the view that the state has "jurisdiction" to the extent that it exercises power or effective control over enjoyment of the right, rather than over an individual or territory. For example, the UN Human Rights ***E.H.R.L.R. 72*** Committee considers that the state's jurisdiction in respect of the right to life extends to "all persons over whose enjoyment of the right to life [the state] exercises power or

effective control".[20] Alternatively it has been argued that the state's duty to respect, or not to interfere with, an individual's human rights has no jurisdictional limits.[21]

In the field of surveillance, the UN Human Rights Committee has called on the US to take "all necessary measures to ensure that its surveillance activities, *both within and outside the United States*, conform to its obligations under the Covenant"[22] and the UN Human Rights Council has adopted preambular language emphasising "that unlawful or arbitrary surveillance and/or interception of communications… violate or abuse the right to privacy… *including when undertaken extraterritorially* …" [emphasis added].[23] In contrast, the UK Investigatory Powers Tribunal concluded that the UK has no obligation to respect the private life of individuals outside the UK, even if it receives surveillance information on them.[24] A majority of the experts drafting Tallinn Manual 2.0 took the view that "*physical* control over territory or the individual is required before human rights law obligations are triggered" in respect of cyber-related activity, whereas a minority considered that the state's power or effective control over the individual's exercise or enjoyment of the right is sufficient.[25] The Office of the UN High Commissioner for Human Rights has taken a hybrid approach, proposing as regards surveillance that a state's human rights obligations apply wherever a state exercises "power or effective control in relation to digital communications infrastructure", for example "through direct tapping or penetration of that infrastructure".[26]

Given these divergences of view, and the lack of expert, judicial and inter-state consideration of foreign interference operations to date, it is not yet established whether foreign interference operations conducted by State A may violate the human rights of individuals in State B. There may be a trend amongst some states and human rights bodies towards the position that they do, but this trend would doubtless be contested by other states.

If the state does owe such extraterritorial human rights obligations, it does not follow that IHRL is well-tailored to providing accountability for what is essentially a state attack on another state's electoral or political system. Moreover, even if jurisdiction is established, the inherently covert nature of interference entails that both establishing attribution and securing a remedy are likely to be problematic.[27] It is submitted that if State A runs a foreign interference operation in State B, public international law should set a threshold for condemning that interference as wrong rather than requiring assessment of its impact on individuals' rights of freedom of expression, privacy, thought etc in the circumstances of the case.[28]

Arguably the collective right of self-determination (common Article 1 of the International Covenant on Economic, Social and Cultural Rights ("ICESCR") and International Covenant on Civil and Political Rights ("ICCPR"), and generally recognised as customary international law) could develop to play a role ***E.H.R.L.R. 73*** here. There is not yet jurisprudence on whether the right of self-determination entails that the people of a state have the right to conduct their domestic election process and political debate free of foreign interference. Arguably this might be part of the ordinary meaning of the right of peoples to "freely determine their political status" (common Art.1(1) ICCPR and ICESCR). The possibility has attracted a little academic discussion.[29] Such a right may entail a number of practical difficulties, including identification of the will of the people of the state and establishing what degree of interference, or what impact on an election or political dialogue, would amount to violation of that will.[30] If the right were to develop in this way, it would provide a legal basis for condemnation of foreign interference operations but otherwise would be unlikely to guide states or platforms in assessing and handling online manipulation campaigns.

3.2 International human rights law: domestic regulation and accountability for online manipulation

IHRL entails several obligations for governments in tackling online manipulation within their own states. First, governments must not engage in online manipulation that breaches IHRL. This is an important point, as at present a significant proportion of inauthentic behaviour reported by platforms consists of manipulation campaigns originating from domestic militaries, officials and ruling political parties.

Second, governments must take appropriate measures to protect their populations from manipulation conducted by others.[31] There are four dimensions to this obligation. First, governments must take proactive measures to defend society against the potential impact of attempts at manipulation, for example by supporting a robust independent and diverse media, educating the public to guard against manipulation, sponsoring fact-checkers and educating political parties and campaigners as to the rules and accepted practice.

Second, governments should clearly identify and condemn manipulation campaigns that interfere with human rights. Unless they do so, we risk a race to the bottom as political parties all over the world adopt online manipulation techniques that are not clearly unacceptable, with platforms fighting a rear-guard action to control them. To be able to identify manipulation, states should require much more transparency from platforms and should conduct a multi-stakeholder conversation to distinguish legitimate political debate from manipulation.[32]

Third, governments should guide platforms as to how they should strike a balance between fostering open and pluralistic speech and avoiding online manipulation.[33] Some governments are working on initiatives in this respect, such as the UK's Defending Democracy Programme.[34] The European Commission's European Democracy Action Plan, complementing the proposed Digital Services Act, is adopting this approach, proposing to tackle "manipulative techniques" including "distortion of information, misleading the audience and manipulative tactics" as well as disinformation, while "fully safeguard[ing] ***E.H.R.L.R. 74*** freedom of expression".[35] But to date most governments have not provided adequate guidance to platforms. Some have largely left platforms to manage online manipulation, while others have responded with blunt measures that violate the right to freedom of expression, such as banning falsehoods (for example Singapore's Protection from Online Falsehoods and Manipulation Act 2019).[36]

Governments should not leave platforms to manage manipulation without guidance, for at least five reasons. First, doing so leaves platforms' efforts to meet their human rights responsibilities vulnerable to political controversy, with platforms at risk from political or commercial pressure to yield to political demands. The 2020 US election campaign provides a vivid illustration. Second, doing so relies on an assumption of political neutrality and human rights compliance on the part of platforms, which (whether or not justified to date) may not be the case in future. Manipulation can easily jump to platforms that pay less attention to human rights and democracy.[37] Third, doing so assumes that platforms are immutable, but governments should not shy away from requiring changes to platform structures and operating practices if necessary to avoid human rights violations. Fourth, as online manipulation raises novel issues of wide-ranging significance, it is the state which has the democratic mandate and guardianship of the public interest that both inform and justify performance of the sensitive balancing of competing rights and interests required by IHRL. Although platforms make decisions in individual cases, they should be guided by governments on overarching principles. If there is concern that government involvement in guiding platforms may lead to risks of state control over speech and information, governments should ensure independent input or oversight. Fifth, as the implementation of IHRL varies between national and regional contexts within the parameters of international obligation, so each state should guide platforms on the requirements of human rights law in its own jurisdiction.[38] For example, just as the regulation of political advertising offline is standard in some European states but not in the US, so states' approaches to a human rights assessment of restriction of political advertising online are likely to vary. Each state should also work with platforms to ensure that they respect IHRL in their own jurisdiction, significant given reports that the extent of platforms' activity to tackle manipulation varies significantly around the world.

Fourth, Governments' IHRL obligations entail that they should provide appropriate accountability not only for those who engage in online manipulation, but also for platforms.[39] Platforms' total or limited immunity in many jurisdictions in respect of content hosted on their platforms[40] has generated a welcome environment of pluralistic expression. If retained, it should be compensated for by other measures of accountability, including platform transparency, independent oversight, some form of redress for individuals and regular multi-stakeholder dialogue. Such accountability must be carefully calibrated to encourage good platform behaviour but avoid a chilling effect on speech by incentivising content removal. Domestic and regional litigation testing the application of human rights law to online manipulation of political debate could provide further guidance on appropriate accountability.

There are jurisdiction and conflict of laws issues at domestic level. In the view of the Human Rights Committee, states have a duty to ensure that corporate activities within their jurisdiction, but having a "direct and reasonably foreseeable impact on the [rights] of individuals outside their territory" are consistent ***E.H.R.L.R. 75*** with human rights standards.[41] Some states, including the US and the Netherlands, dispute this view.[42] In practice, a platform may face conflicting obligations from the state where it is headquartered to censor content on its platform and from a state where it is operating to respect freedom of expression, or vice versa. While a headquarters state may require a platform to respect human rights globally, the reality is that the platform's only options may be to meet the demands of the state where it is operating or to leave that country.

3.3 Human rights law: the corporate responsibility to respect: platforms and online manipulation

While IHRL does not impose obligations directly on private companies, all business enterprises have a non-binding "responsibility to respect" human rights, an element of the UN Guiding Principles on Business and Human Rights.[43] This responsibility concerns all the rights in the International Bill of Rights,[44] regardless of where the activity takes place or where those affected are located. It entails that businesses should adopt a clear policy commitment to human rights; a human rights due diligence process to identify, prevent, mitigate and account for how they address their impacts on human rights; and remediation processes in respect of any adverse human rights impacts.[45] The responsibility applies in respect not only of the enterprise's own activities, but also in respect of "human rights impacts that are directly linked to [its] operations, products or services by [its] business relationships".[46] Companies should report regularly and adequately to external stakeholders on how they address human rights impacts identified.[47]

The responsibility of platforms to respect human rights has been urged by UN and regional Special Rapporteurs,[48] recognised by the Committee of Ministers of the Council of Europe[49] and, as regards the processing of personal data, by the UN Human Rights Council.[50] Some platforms have assumed voluntary commitments in respect of freedom of expression and privacy through membership of the Global Network Initiative.[51] A growing number of platforms have decided to structure their developing content moderation and privacy practices around human rights norms, albeit that their implementation of human rights responsibilities lacks transparency and is nascent, patchy and underdeveloped. Even as regards those platforms which profess commitment to human rights, evidence from the non-Western world is that there is little implementation and that online manipulation is costing fragile political systems dearly.[52] This account was attested to in 2020 by Facebook whistleblower Sophie Zhang. ***E.H.R.L.R. 76*** [53]

To enable guidance from and accountability to states, as well as academic, civil society and media review, platforms should be much more transparent about the manipulation they see, how their operating systems interact with it and the action they take and are planning to develop to counter it, building on the periodic reporting some are already providing. Platforms should ensure that their own operating systems do not inadvertently support and further manipulation.

Leaving aside content moderation and privacy, platforms have not expressly structured their efforts to curb manipulation around human rights, particularly freedom of thought and opinion and the right of political participation. Doing so would assist them by providing common standards by which to structure and communicate their efforts against online manipulation, while also helping them reconcile their responses to manipulation with protection of freedom of expression. This may entail that platforms look closely at their own structures and algorithm designs to ensure that they are not themselves manipulative or facilitating manipulation by others. The human rights principally engaged by online manipulation are discussed below.

4 Which human rights are engaged by online manipulation?

Both governments and platforms should pay close regard to the following rights in tackling online manipulation of political debate: the right to participate in public affairs and to vote; the rights to freedom of thought and freedom of opinion; the right to privacy; and the right to freedom of expression. As discussed above, it is possible that the collective right of self-determination may also play a role in respect of foreign interference operations, if interpreted to entail that the people of a state have the right to conduct their domestic election and political process free of foreign interference. For the purposes of this section, the rights in the ICCPR are discussed and are assumed to reflect customary international law, except where otherwise indicated. Variations between ICCPR and regional human rights treaties are not discussed.[54]

4.1 Right to participate in public affairs and to vote (Article 25 ICCPR)

This right includes the right not only to vote, but to engage in public debate and discussion. Online manipulation that reduces citizens' ability to engage in democratic debate and voters' ability to glean information and make up their mind freely is incompatible with the right to vote. Such manipulation might include: interruptions to internet access, such that citizens are unable to access information or participate in debate; interferences with election infrastructure so that voters are unable to exercise their right to vote or to have their vote counted; deliberately misleading voters over how to vote, for example as to the date of the election or where or how they may vote; deliberately distorting voters' ability to form their will freely (see discussion of rights to freedom of thought and opinion, below); and deliberately deterring voters from voting.[55] Deliberate dissuasion of candidates from standing or speaking their views, for example through trolling or incitement to hatred, also interferes with the right. ***E.H.R.L.R. 77***

4.2 Rights to freedom of thought and freedom of opinion (Articles 18 and 19 ICCPR)

As former Google strategist James Williams has described,[56] and as given popular expression in *The Social Dilemma*,[57] manipulative techniques online can affect individual thoughts and opinions on a scale never seen in the analogue world. The rights to freedom of thought and freedom of opinion have a key role to play in tackling online manipulation. They are absolute rights, as the Human Rights Committee has stressed in its General Comments on both articles.[58] They are rights that have traditionally been largely taken for granted, and there is little expert comment or jurisprudence on them. Their parameters are not yet clear: we now need an interdisciplinary discussion to establish the boundary between legitimate political campaigning techniques and unacceptable online manipulation.[59] In 2005, Nowak suggested that infringements of freedom of thought may be limited to involuntary influence over opinions.[60] Assessing the right to freedom of opinion in detail, Aswad proposes that "deliberate efforts to influence through non-consensual means violate this right when they rise to the level of either overwhelming mental autonomy or manipulating one's reasoning".[61]

Individuals of course are constantly receiving influences over their thoughts and opinions, including deliberate attempts to persuade through advertising and political argument. But the digital world amplifies exponentially the opportunities for governments and business enterprises to know the thoughts of individuals, to shape them without individuals being aware that this is happening, and for individuals to be penalised or discriminated against on the basis of their views—all of which may violate the rights to freedom of thought and opinion. As Facebook commented in July 2020, "it's critical that we, as a society, have a broader discussion about what is acceptable political advocacy and take steps to deter people from crossing the line."[62] To inform that conversation, governments, researchers, the media and civil society need much more transparency from platforms, political campaigners and industry regarding influence techniques being developed and deployed.

To be clear, restricting manipulative techniques does not restrict freedom of expression: the restriction is not over what is said but over techniques used in its generation or amplification. Restricting manipulative techniques is akin to banning subliminal advertising or hypnosis, which some states have done in the analogue world.

4.3 Right to privacy (Article 17 ICCPR)

The right to privacy includes a right for the individual to choose not to divulge their personal data, a right to opt out of trading in and profiling on the basis of their personal data, and a right to have their data processed only with their consent or for a legitimate purpose, and with their knowledge.[63] These rights were carefully protected in the analogue era, with individuals having genuine choice in ticking boxes for consent to data retention and sharing. In the digital era, these protections have been swept aside, as online ***E.H.R.L.R. 78*** behaviour is widely monitored (often through notional "consent to cookies" or "legitimate interest") and combined with offline profiles to create profiles of voters. European and British data protection laws are inadequate to stem this tide, as their bases for processing ("consent", "democratic engagement", "legitimate interests") are interpreted so broadly as not to impose meaningful limits. Those running online manipulation campaigns may be able to gather and use an extensive compilation of personal data without legitimate basis in order to micro-target messages without recipient awareness, consent or choice, inconsistently with the right to privacy.[64]

4.4 Right to freedom of expression (Article 19 ICCPR)

In the battle against manipulation, the right to freedom of expression, which includes the right "to seek, receive and impart information and ideas through any media and regardless of frontiers",[65] is a vital bulwark of an open and uncensored Internet. It is generally a breach of this right to interrupt internet connectivity, and generally a breach of this right to censor communication.

Importantly, freedom of expression constrains responses to manipulation. It entails that generally, states must not encourage or require platforms to take down disinformation or other manipulative material on the basis of its content (for example, on the basis that it's false); but only to remove, label or restrict its circulation on the basis that it forms part of an online manipulation campaign. Material can exceptionally be removed because of its content if to do so would be lawful, meet a legitimate aim and be necessary for one of the purposes in art.19(3) ICCPR, such as protection of public health or national security, or if it is hate speech that incites discrimination, hostility or violence. Particularly in the political context, any such restriction must be tightly constrained to avoid its application for authoritarian ends.

The French Government's 2018 Law Against the Manipulation of Information, which withstood a freedom of expression challenge to the Constitutional Court,[66] may be an example of a content-based restriction that meets the requirements of art.19(3).[67] During election campaign periods it permits the most egregious disinformation, of a nature that risks disturbance of the peace or compromise to the outcome of an election, to be suppressed on the order of a judge.

Some academics argue that freedom of expression imposes obligations on states in respect of disinformation. Milanovic and Schmitt argue that disinformation *systematically* disseminated by the state to its own inhabitants may breach the individual's right to seek and receive information (an element of the right to freedom of expression), especially when accompanied by suppression of true information.[68] Bayer et al. posit that the state breaches the right to freedom of expression if it fails to ensure an information environment in which individuals can access true information through diversity in public discourse and media environment.[69] Both formulations are novel, and both base breach on the state's distortion of the information environment as a whole. It is important to be clear that these propositions do *not* assert that all disinformation breaches (or undermines) the right to freedom of expression, nor that governments are obliged to suppress all falsehoods or assure that all political speech online is factually true. As discussed, a general ban on false or unverified information would breach freedom of expression.

5 Conclusion

Much of the world's political debate now takes place online and free of editorial control. The openness and reach of the internet have brought great benefits for political debate: tremendous increases in the number and diversity of voices, opportunities to participate, and access to information in politics. But they have also brought online manipulation campaigns: distortions to electoral and political discussion, both foreign and domestic, that threaten to undermine democracy.

Manipulation interferes with the right to freedom of thought and opinion, and the right to participate in public affairs and to vote. When it exploits personal data, it interferes with the right to privacy. If it systematically distorts the information environment, arguably it may interfere with the right to freedom of expression. It may be arguable that foreign interference operations breach the right of self-determination.

Both states and platforms should, tackle online manipulation. Given the potential power wielded by those who respond to manipulation - ultimately with the potential to remove a public platform from a state's leaders or opposition - those responses should be governed by transparent rules and subject to accountability. International human rights law provides an appropriate framework for those rules. It requires that states and state-controlled bodies should refrain from engaging in foreign or domestic manipulation themselves, and should also refrain from blunt responses that stymie open political debate, such as shutting down the Internet or banning categories of speech. States have a duty to protect their populations, not only by supporting a robust and independent media and digital literacy but also by taking active steps to condemn manipulation, guiding platforms on how they should manage manipulation, and providing accountability. Platforms have a responsibility to respect international human rights law in tackling manipulation, not only in content moderation but more widely in their efforts to combat inauthentic behaviour. But in order to tackle the manipulation crisis adequately by reference to IHRL, the starting point needs to be a multi-stakeholder conversation to establish and advertise the parameters of acceptable political advocacy.

Kate Jones

Footnotes

1 *Oxford English Dictionary, 3rd edn, (Oxford: Oxford University Press, 2000)*.

2 C. Wardle and H. Derakshan, "Information Disorder: Toward an interdisciplinary framework for research and policymaking" (Council

of Europe, 2017) https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c [Accessed 14 December 2020] p.20.

3 Global Disinformation Index, "Adversarial Narratives: A New Model for Disinformation" (2019) https://disinformationindex.org/wp-content/uploads/2019/08/GDI_Adverserial-Narratives_Report_V6.pdf [Accessed 12 October 2020].

4 NATO StratCom COE Singularex, "The Black Market for Social Media Manipulation" (2018) https://www.stratcomcoe.org/download/file/fid/79968 [Accessed 12 October 2020].

5 J. Pamment, "The EU's Role in Fighting Disinformation: Taking Back the Initiative" (Carnegie Endowment for International Peace, 2020) https://carnegieendowment.org/2020/07/15/eu-s-role-in-fighting-disinformation-taking-back-initiative-pub-82286 [Accessed 14 December 2020] p.4-5.

6 Defined as "groups of pages or people work[ing] together to mislead others about who they are or what they are doing" by Facebook's Head of Cybersecurity Policy: *N. Gleicher, "Coordinated Inauthentic Behaviour Explained" (6 December 2018), Facebook Newsroom* https://about.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/ [Accessed 12 October 2020].

7 Twitter, "Platform manipulation and spam policy" (September 2020) https://help.twitter.com/en/rules-and-policies/platform-manipulation [Accessed 12 October 2020].

8 For example, Google Threat Analysis Group, "TAG Bulletin: Q2 2020" (August 2020), https://blog.google/threat-analysis-group/tag-bulletin-q2-2020/ [Accessed 12 October 2020]. Google owns YouTube.

9 S. Bradshaw, H. Bailey, P. Howard, "Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation" (Oxford Internet Institute, 2021)

https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/127/2021/01/CyberTroop-Report20-FINALv.3.pdf [Accessed 30 January 2021].

10 S. Bradshaw and P. Howard, "The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation" (Oxford Internet Institute, 2019) https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf [Accessed 12 October 2020]

11 Blackbird.AI assessed that 27% of Twitter posts between 2 and 20 May 2020 were manipulated, with particularly high proportions of manipulation around specific right-wing, conspiracy and public health themes. Blackbird.AI, "COVID-19 (Coronavirus) Disinformation Report — Volume 3.0" (June 2020) https://www.blackbird.ai/blog/2020/06/10/covid-19-disinformation-report-vol-3-0/ [Accessed 14 December 2020]

12 *S. Woolley, "Bots and Computational Propaganda: Automation for Communication and Control", in N. Persily and J. Tucker, Social Media and Democracy: The State of the Field, Prospects for Reform (Cambridge: Cambridge University Press, 2020)*.

13 UN Human Rights Council Resolutions on "The promotion, protection and enjoyment of human rights on the Internet", A/HRC/RES/20/8 (2012), A/HRC/RES/26/13 (2014), A/HRC/RES/32/13 (2016), A/HRC/RES/38/7 (2018).

14 UN Office of the High Commissioner on Human Rights, "Guiding Principles on Business and Human Rights", adopted unanimously by the UN Human Rights Council in its Resolution on "Human rights and transnational corporations and other business enterprises", 6 July 2011, A/HRC/RES/17/4. The UN Guiding Principles are not a treaty. The duties for states set out in the UN Guiding Principles derive from the obligations of States Parties to the UN human rights treaties. The UN Guiding Principles' elaboration of corporate

responsibilities is not grounded in international legal obligation, hence the language of "responsibility" rather than "duty" or "obligation".

15 For example, the Human Rights Committee has referred to the responsibilities of business enterprises to respect human rights in its recent General Comments, such as General Comment No. 37 on Article 21 (Right of peaceful assembly), 23 July 2020, CCPR/C/GC/37, para.31; General Comment No. 36 on Article 6 (The right to life), 2 November 2018, CCPR/C/GC/36, para.22.

16 Other corporate actors, such as campaigning organisations, also have human rights responsibilities; these are outside the scope of this paper.

17 *Cyprus v Turkey (2002) 35 E.H.R.R. 30*.

18 UN Human Rights Committee General Comment No. 31 on The Nature of the General Legal Obligation Imposed on States Parties to the Covenant, 26 May 2004, CCPR/C/21/Rev.1/Add.13, para.10. The regional human rights tribunals have adopted a similar approach.

19 Observations of the United States of America on the Human Rights Committee's Draft General Comment No. 36 on Article 6 — Right to Life, 6 October 2017 https://www.ohchr.org/Documents/HRBodies/CCPR/GCArticle6/UnitedStatesofAmerica.docx [Accessed 12 October 2020] paras.13-15.

20 UN Human Rights Committee General Comment No. 36 on Article 6 (The right to life), 2 November 2018, CCPR/C/GC/36, para. 63. The Committee's General Comments and decisions are not legally binding (save arguably for any direction it issues on interim or provisional measures: UN Human Rights Committee, General Comment No. 33 on Obligations of States parties under the Optional Protocol to the International Covenant on Civil and Political Rights, 25 June 2009, CCPR/C/GC/33, para.19).

21 *M. Milanovic, Extraterritorial Application of Human Rights Treaties (Oxford: Oxford University Press, 2011)*.

22 UN Human Rights Committee, Concluding Observations on the fourth periodic report of the United States of America, 23 April 2014, CCPR/C/USA/CO/4, para.22.

23 UN Human Rights Council Resolution on "The right to privacy in the digital age", 26 September 2019, A/HRC/RES/42/15, preambular para.26.

24 *Human Rights Watch Inc & Others v Secretary of State for Foreign and Commonwealth Affairs and Others [2016] UKIPTrib 15_165-CH*, para.58. The Tribunal observed that it is obliged by domestic law not to do more than to keep pace with the jurisprudence of the European Court of Human Rights (para.60). The European Court of Human Rights has not considered this point.

25 *M. Schmitt (ed.), Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations (Cambridge: Cambridge University Press, 2017), p.185*.

26 Report of the Office of the UN High Commissioner for Human Rights, "The Right to Privacy in the Digital Age", 30 June 2014, A/HRC/27/37, para. 34.

27 B. Sander, "Democracy Under The Influence: Paradigms of State Responsibility for Cyber Influence Operations on Elections" (2019) 18 Chinese Journal of International Law 1, 45-49.

28 For a discussion of potential violations of public international law, see M. Milanovic and M. Schmitt, "Cyber Attacks and Cyber (Mis)information Operations during a Pandemic" (2020) 11 Journal of National Security Law & Policy 247 https://jnslp.com/wp-content/uploads/2020/12/Cyber-Attacks-and-Cyber-Misinformation-Operations-During-a-Pandemic_2.pdf [Accessed 31 January 2021].

29 J. Ohlin, "Did Russian Cyber Interference in the 2016 Election Violate International Law?" (2017) 95 Texas Law Review 1579, 1595–1598; B. Sander, "Democracy Under The Influence: Paradigms of State Responsibility for Cyber Influence Operations on Elections" (2019) 18 Chinese

Journal of International Law 1, 43–45.

30 B. Sander, "Democracy Under The Influence: Paradigms of State Responsibility for Cyber Influence Operations on Elections" (2019) 18 Chinese Journal of International Law 1, 44–45; M. Schmitt, "Virtual Disenfranchisement: Cyber Election Meddling in the Grey Zones of International Law" (2018) 19 Chicago Journal of International Law 30, 55–56.

31 *UN Guiding Principles on Business and Human Rights (2011), Section I: The State Duty to Protect Human Rights* reflecting states' positive obligations under the UN human rights treaties to protect individuals against acts by private entities that would impair individual enjoyment of their rights. See for example UN Human Rights Committee, General Comment No. 31 on The Nature of the General Legal Obligation Imposed on States Parties to the Covenant, 26 May 2004, CCPR/C/21/Rev.1/Add.13, para.8.

32 UN Guiding Principles on Business and Human Rights (2011), para.2: "States should set out clearly the expectation that all business enterprises domiciled in their territory and/or jurisdiction respect human rights throughout their operations," and para.3: "In meeting their duty to protect, States should: … (d) encourage, and where appropriate require, business enterprises to communicate how they address their human rights impacts."

33 UN Guiding Principles on Business and Human Rights (2011), para.3: "In meeting their duty to protect, States should … (c) provide effective guidance to business enterprises on how to respect human rights throughout their operations".

34 https://www.theyworkforyou.com/wms/?id=2019-07-22.HCWS1772.h [Accessed 14 December 2020]

35 European Commission, "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the European Democracy

Action Plan" (3 December 2020), COM(2020)790 final.

36 Singapore Protection from Online Falsehoods and Manipulation Act 2019 (No. 18 of 2019) https://sso.agc.gov.sg/Acts-Supp/18-2019/Published/20190625?DocDate=20190625 [Accessed 14 December 2020].

37 *M. Isaac and K. Browning, "Fact-checked on Facebook and Twitter, Conservatives Switch Their Apps" (11 November 2020), New York Times* https://www.nytimes.com/2020/11/11/technology/parler-rumble-newsmax.html [Accessed 14 December 2020].

38 See generally B. Sander, "Freedom of Expression in the Age of Online Platforms: The Promises and Pitfalls of a Human Rights-Based Approach to Content Moderation" (2020) 43 Fordham International Law Journal 939.

39 UN Guiding Principles on Business and Human Rights (2011), para. 25: "As part of their duty to protect against business-related human rights abuse, States must take appropriate steps to ensure, through judicial, administrative, legislative or other appropriate means, that when such abuses occur within their territory and/or jurisdiction those affected have access to effective remedy."

40 Most notably in the United States: Section 230 Communications Decency Act, 47 U.S.C. §230.

41 UN Human Rights Committee General Comment No. 36 on Article 6 (The right to life), 2 November 2018, CCPR/C/GC/36, para.22. While this General Comment only discusses the right to life, there is no suggestion that the Human Rights Committee intends to limit its view to this right.

42 See, for example, Observations of the United States of America and Comments of the Netherlands on draft UN Human Rights Committee General Comment No. 36, 2017 https://www.ohchr.org/EN/HRBodies/CCPR/Pages/GC36-Article6Righttolife.aspx [Accessed 12 October 2020].

43 UN Guiding Principles on Business and Human Rights (2011), Section II: The Corporate Responsibility to Respect Human Rights. See fn.13.

44 The International Bill of Rights comprises the Universal Declaration of Human Rights, the International Covenant on Economic, Social and Cultural Rights and the International Covenant on Civil and Political Rights with its two Optional Protocols: https://www.ohchr.org/documents/publications/compilation1.1en.pdf [Accessed 23 November 2020].

45 UN Guiding Principles on Business and Human Rights (2011), para.15. See also paras.16-20.

46 UN Guiding Principles on Business and Human Rights (2011), para.13.

47 UN Guiding Principles on Business and Human Rights (2011), para.21.

48 The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, and the Organization of American States (OAS) Special Rapporteur on Freedom of Expression, "Joint Declaration on Freedom of Expression and Elections in the Digital Age" (April 2020) https://www.article19.org/wp-content/uploads/2020/04/Joint-Declaration-Final.pdf [Accessed 12 October 2020] para.2(a)(i).

49 Recommendation of the Committee of Ministers to Member States on a Guide to human rights for Internet users, 16 April 2014, CM/Rec (2014)6 https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016804d5b31 [Accessed 25 November 2020] para.5.5.

50 UN Human Rights Council Resolution on "The right to privacy in the digital age", 26 September 2019, A/HRC/RES/42/15, para.8(a).

51 https://globalnetworkinitiative.org/ [Accessed 12 October 2020].

52 For example, *N. Nyabola, Digital Democracy, Analogue*

Politics: How the Internet Era is Transforming Politics in Kenya (London: ZedBooks, 2018).

53 *C Silverman et al., "I have blood on my hands: A whistleblower says Facebook ignored global political manipulation" (14 September 2020) BuzzFeed News* https://www.buzzfeednews.com/article/craigsilverman/facebook-ignore-political-manipulation-whistleblower-memo [Accessed 14 December 2020].

54 This article focuses on the ICCPR because it is a civil and political rights treaty of global application, with 173 States Parties (December 2020). The assumption that it reflects customary international law ("CIL") is not intended to be an argument as to the extent to which its various norms do reflect CIL.

55 As allegedly occurred in the 2016 US Presidential campaign, on the part of both the Internet Research Agency ("IRA") and Donald Trump's campaign team. Re IRA: P. Howard et al., "The IRA, Social Media and Political Polarization in the United States, 2012-2018" (Oxford Internet Institute/Graphika, 2018) https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/12/The-IRA-Social-Media-and-Political-Polarization.pdf [Accessed 14 December 2020] pp.32–34. *Re Trump campaign team: Channel 4 News Investigations Team, "Revealed: Trump campaign strategy to deter millions of Black Americans from voting in 2016" (28 September 2020), Channel4.com* https://www.channel4.com/news/revealed-trump-campaign-strategy-to-deter-millions-of-black-americans-from-voting-in-2016 [Accessed 14 December 2020].

56 *J. Williams, Stand out of our Light: Freedom and Resistance in the Attention Economy (Cambridge: Cambridge University Press, 2018).*

57 Documentary film directed by J. Orlowski, produced by Exposure Labs, 2020 https://www.thesocialdilemma.com/ [Accessed 14 December 2020].

58 "The right to freedom of thought, conscience and religion… in Article 18.1 is far-reaching and profound; it encompasses freedom of thoughts on all matters… this provision cannot be derogated from…" UN Human Rights Committee General Comment No.22 on The right to freedom of thought, conscience and religion (art.18), 30 July 1993, CCPR/C/21/Rev.1/Add.4, para.1. Freedom of opinion is "a right to which the Covenant permits no exception or restriction" and "Any form of effort to coerce the holding or not holding of any opinion is prohibited." UN Human Rights Committee General Comment No. 34 on Article 19: Freedoms of opinion and expression, 12 September 2011, CCPR/C/GC/34, paras.9-10.

59 Indeed, this conversation is important not only in the political context but also to determine the limits of acceptable behaviours in emerging markets of surveillance capitalism. Surveillance capitalism relies on 'behavioural futures markets' which involve developing a profile of each individual, predicting their reactions and sometimes their futures, and anticipating and nudging their needs and their views. *S. Zuboff, The Age of Surveillance Capitalism (New York: Public Affairs (Hachette Book Group), 2019)*.

60 *M. Nowak, UN Covenant on Civil and Political Rights: CCPR Commentary (Germany: NP Engel, 2005), p.442*.

61 E. Aswad, "Losing the Freedom to Be Human" (2020) 52(1) Columbia Human Rights Law Review 306.

62 Facebook, "July 2020 Coordinated Inauthentic Behaviour Report" (August 2020) https://about.fb.com/news/2020/08/july-2020-cib-report/ [Accessed 12 October 2020].

63 UN Office of the High Commissioner for Human Rights, "The Right to Privacy in the Digital Age", 3 August 2018, A/HRC/39/29.

64 At present this is a developing feature of all political campaigning, not just manipulation. See for example Information Commissioner's

Office, "Democracy Disrupted? Personal Information and Political Influence" (2018) https://ico.org.uk/media/action-weve-taken/2259369/democracy-disrupted-110718.pdf [Accessed 14 December 2020].

65 Article 19(2) ICCPR.

66 French Constitutional Council, Decision No. 2018-773 DC of 20 December 2018.

67 LOI n° 2018–1201 du 22 décembre 2018 [Law No 2018-1201 of 22 December 2018] (France) JO, 23 December 2018, 1; LOI n° 2018–1202 du 22 décembre 2018 [Law No 2018-1202 of 22 December 2018] (France) JO, 23 December 2018, 2.

68 M. Milanovic and M. Schmitt, "Cyber Attacks and Cyber (Mis)information Operations during a Pandemic" (2020) 11 Journal of National Security Law & Policy 247 https://jnslp.com/wp-content/uploads/2020/12/Cyber-Attacks-and-Cyber-Misinformation-Operations-During-a-Pandemic_2.pdf [Accessed 31 January 2021]

69 J. Bayer et al., "Disinformation and propaganda — impact on the functioning of the rule of law in the EU and its Member States" (European Parliament's Policy Department for Citizens' Rights and Constitutional Affairs, 2019) https://www.europarl.europa.eu/RegData/etudes/STUD/2019/608864/IPOL_STU(2019)608864_EN.pdf [Accessed 12 October 2020] pp.77–79.

