ONLINE FREE SPEECH AND THE SUPPRESSION OF FALSE POLITICAL CLAIMS

Jeremy Horder*

Abstract: This article will consider the approach taken in different jurisdictions to the use of restrictions, including criminal law prohibitions, to prevent, deter or suppress the making of false political claims online, especially (but not solely) during election periods. The article will explain, first, how a commitment to freedom of expression ought to make state measures to deter the online dissemination of what the author calls false political "viewpoint" information seem problematic. Second, the article will show how a number of jurisdictions, including some liberal democracies, have wrongly been drawn to provide an authoritarian response to the spread of such false information: a response that tries to set the terms on which people come to give and form political viewpoints. Coercion may legitimately be employed (1) to secure due process at the polls, when this is threatened by the dissemination of what the author calls false "participation" information and (2) to support transparency about authorship to a limited extent; but coercion should not be used to deter the propagation of political viewpoint content, simply on the grounds of its falsity.

Keywords: *free speech; internet platforms; militant democracy; false political claims; disinformation; intentionally disseminated false information; coercion; electoral influence*

I. Introduction

Many jurisdictions around the world use the criminal law, alongside a variety of regulatory, preventative and other discouraging measures, to deter and prevent the publication or dissemination of false or misleading political information. Such measures are especially prominent when such disinformation is designed to influence elections or referendums.[1] In order to understand what may (with a greater or

* FBA, LLD, Professor of Criminal Law, Department of Law, London School of Economics, Houghton Street, London WC2A 2AE, UK. J.horder@lse.ac.uk. Formerly a Law Commissioner for England and Wales. His research has an emphasis on corporate criminal law, legal history, and law and politics. Most recently, he was the author of *Criminal Misconduct in Office: Law and Politics* (Oxford University Press, Oxford, 2018).

1 Daniel Funke and Daniela Flamini, *A Guide to Anti-disinformation Actions around the World* (2018), available at https://www.poynter.org/ifcn/anti-disinformation-actions/ (visited on 8 March 2021). It is important to note that my focus will be political disinformation, not hate speech or analogous wrongful harms. Hate speech is justifiably prohibited as a political silencing technique that is, in that respect, similar—as we will see—to election participation disinformation.

lesser degree of justification) be targeted, in that regard, it is helpful to distinguish two basic kinds of false or misleading political information relevant to the electoral context.[2] First, there is information concerning the procedures through which democratic elections are conducted, known as "participation" information. This is information, for example, concerning where, when and how to vote. Second, there is political "viewpoint" information with which I will be principally concerned (whether or not it is disseminated during an election or referendum period). This is fact-based information (not simply an opinion) about substantive political matters relating to candidates and policies, such as claims that "candidate X broke all her promises when she was last in office" or a claim that "party Y's policies have been endorsed by the Pope". For reasons explored in due course, state action to deter or prevent the dissemination of the latter kind of false information—political viewpoint information—ought to face a much higher bar, in terms of justification, if it can be justified at all (in my view, it cannot).

This justificatory task has taken on added significance because of the way in which some jurisdictions have responded to (perceived) new threats from online activity to the integrity of political information-provision at election times, and more broadly. For authoritarian regimes, a wide-ranging coercive and deterrent response to the online propagation of allegedly false political claims—whether those claims relate to participation or viewpoint information—is relatively easy to justify. A hallmark of such regimes is that they claim the right to set the key terms on which citizens engage (if at all) in politics, including in political debate, and can be expected to shape those terms—including what is to count as a "false" political claim that may be suppressed—to serve their own ends.[3] However, we will see that in liberal democracies, a "militant" democratic approach to the dissemination of false political information can also justify a coercive response to the dissemination of viewpoint (as well as participation) disinformation, purely on the grounds of its falsity, albeit a response narrower in scope than one characteristic of an authoritarian regime.[4]

In modern liberal-democratic states, the rise of authoritarian–populist forms of criminalisation commonly reflects a perception among legislators that many of the traditional liberal values, values that shaped a less authoritarian criminal justice system in the post-war era, have lost legitimacy.[5] These liberal values are thought

2 See Jeremy Horder, "Criminal Law at the Limit: Countering False Claims in Elections and Referendums" (2021) 84 *Modern Law Review* 429–455; Joshua S Sellers, "Legislating against Lying in Campaigns and Elections" (2018) 71 *Oklahoma LR* 141. The main focus of this article is so-called "disinformation", information that is known or suspected to be false. I will not be considering the difference in approach that may be appropriate for "misinformation", false information that is innocently shared. Accordingly, references throughout to "false viewpoint information" are in general reference to disinformation.

3 Gábor Attila Tóth, "Authoritarianism" (Oxford Constitutional Law, 2017) para.34, available at https://oxcon.ouplaw.com/view/10.1093/law-mpeccol/law-mpeccol-e205 (visited on 8 March 2021).

4 See Jan-Werner Muller, "Militant Democracy" in Michel Rosenfeld and András Sajó (eds), *The Oxford Handbook of Comparative Constitutional Law* (Oxford: OUP, 2012) Ch.59.

5 See, eg, Tim Newburn, "'Tough on Crime': Penal Policy in England and Wales" (2007) 36 *Crime and Justice* 425; Nicola Lacey, "Populism and the Rule of Law" (2019) 15 *Annual Review of Law and Social Science* 79.

to have lost legitimacy, because they are considered to be not so much universal as "elite" values, with implications for their role and influence in public life:[6] "If, for example, [liberal] elites simply hold different . . . views than the mass public, then the preservation of liberalism does not necessarily require restraints on rule by the [less liberal] masses".[7] If they are more elite than they are universal, liberal values will be politically vulnerable when confronted by, for example, the demands of a more vengeful and authoritarian (security-based) populist criminal justice agenda.[8] By ironic contrast, though, in shaping the state's response to political disinformation, authoritarian steps themselves may seem justified by some species of liberal—be it considered universal, or elite—concerns about the declining or decaying ethical character and quality of political debate online. One could think of this development as an outworking of so-called militant democratic thinking.[9]

Militant democratic politics, in a narrow sense, involve the restriction of collective rights deemed to be a direct threat to democratic values, such as the right to form political parties committed to the overthrow of democracy.[10] However, a commitment to militant democratic politics may also involve targeting individual rights, rather than just the collective rights of organisations and parties. Subject to contentious questions about proportionality, such a commitment may appear to justify authoritarian restrictions on rights such as the right to advocate non-democratic forms of government or the right to exercise free speech in other ways that appear to challenge or undermine democracy itself.[11] Key examples are some instances in which freedom of speech is used for the dissemination of false political information online.[12] For example in 2016, the Venice Commission (advising the European Commission) suggested that:

> not all ideas deserve to be circulated. Since the exercise of freedom of expression carries duties and responsibilities, it is legitimate to expect from every member of a democratic society to avoid as far as possible

6 For a penetrating early examination of this theme, see Sheldon Wolin, "The People's Two Bodies" (1981) 1 *Democracy* 9.

7 Jennifer Hochschild, "Dimensions of Liberal Self-satisfaction: Civil Liberties, Liberal Theory, and Elite Mass Differences" (1986) 96 *Ethics* 386, 399.

8 See, eg, Gerry Johnstone, "Penal Policy-Making: Elitist, Populist or Participatory?" (2000) 2 *Punishment and Society* 161–180. Of course, the so-called populist agenda may itself be the product of elite political groups. See, more broadly, Samuel Moyn, "On Human Rights and Majority Politics" (2019) 52 *Vanderbilt Journal of Transnational Law* 1135.

9 See JW Muller, "Militant Democracy" (n.4).

10 *Ibid*., 1123–1126.

11 Heidi Tworek, "An Analysis of Germany's NetzDG Law" (Working Paper, Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression, 2019) 2, available at https://www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf (visited on 8 March 2021); Ian Cram, "Keeping the Demos out of Liberal Democracy? Participatory Politics, 'Fake News', and the Online Speaker" (2019) 11 *Journal of Media Law* 113. For analysis of the more "militant democratic" decisions of the Supreme Court in the immediate aftermath of WWII, see Zachary S Price, "Our Imperilled Absolutist First Amendment" (2018) 20 *Journal of Constitutional Law* 817.

12 See JW Muller, "Militant Democracy" (n.4), 1127; I Cram, "Keeping the Demos" (n.11).

> expressions that express scorn or are gratuitously offensive to others and infringe their rights.[13]

Yet, the taking of such militant democratic steps by political guardians of the liberal-democratic tradition is vulnerable to challenge on populist grounds, even though in general, the public supports the suppression of "criminal" content on social media networks.[14] In populist thinking:

> [t]he elite is assumed to betray the people and to deprive it of its legitimate right to exercise power. The populist actor pitches himself or herself as a challenger of the elites and as an advocate of the people . . . The elite is also assumed to engage in a complicity with "the others" and to favour them over the people.[15]

Modern populism transforms established parties and their candidates from characterisation—in idealist republican terms—as essentially altruistic and dedicated (would-be) public servants into self-interested upholders of elite interests. With the immense power of online distribution behind it,[16] one important effect of that transformation is that populist political narratives lead increasingly to partisan polarisation and distrust.[17] Different sides adopt more intolerant and repressive attitudes and engage in what Zachary Price aptly calls "tit-for-tat degradation", with implications for what is regarded as "fair comment".[18] In spite of the work of the traditional media, of regulators and of fact-checking organisations, attempts to establish "the truth" in political debate are likely casualties of such developments. Online populist politics is not only an important driver behind distrust of and contempt for politicians from established parties seeking (as well as those already "in") office, but also behind challenges to what is to count as a "true", a "false" or a "misleading" political claim. Populists typically reject orthodox liberal ways of distinguishing between truth and fiction and between fact and opinion or value, commonly

---

13 European Commission for Democracy through Law, *Compilation of Venice Commission Opinions and Reports concerning Freedom of Expression and Media* (2016), available at www.venice.coe.int/webforms/documents/?pdf=CDL-PI(2016)011-e cited by I Cram, "Keeping the Demos" (n.11), 123.

14 Tworek reports that 87 per cent of German voters supported the controversial Internet Enforcement Act 2017, discussed in VI. Authoritarian Liberalism: Germany: H Tworek, "An Analysis" (n.11), 2. However, as we will see, voters tend to support suppression of disinformation only when it is deployed by their political opponents, see (n.41).

15 Sven Engesser, Nayla Fawsi and Anders Olof Larsson, "Populist Online Communication" (2017) 20 *Information, Communication and Society* 1279, 1281.

16 *Ibid*. See also Mark Littler and Matthew Feldman, "Social Media and the Cordon Sanitaire: Populist Politics, the Online Space, and a Relationship That Just Isn't There" (2017) 16 *Journal of Language and Politics* 510.

17 House of Commons, Digital, Culture, Media and Sport Committee, *Disinformation and Fake News: Final Report* (2019) part 6, available at https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf (visited on 8 March 2021).

18 See Z Price, "Our Imperilled Absolutist" (n.11).

replacing such "timeless" values and distinctions with a preference for perceived sincerity and authenticity in political messaging.[19] In that regard, speaking of attempts to ensure verification of contentious claims or supposed "fake news", Romain Badouard has expressed the view that: As fake news is the manifestation of popular distrust of the political and intellectual elite, how could verification by those same elites possibly convince those propagating it [the fake news]?[20] What Badouard says about the political credibility of elite verification will *a fortiori* be true of elite attempts to use coercion to suppress political "disinformation", even in the name of democracy itself.

This suggests that the use of coercion—or indeed of any other formal measures of discouragement—to suppress the foundations of the populist rejection of traditional (and especially militant) liberal-democratic politics would be illegitimate and self-defeating. As the French Institute for Strategic Research (Ministry for the Armed Forces) and Policy Planning Staff (Ministry for Europe and Foreign Affairs) puts it:

> Overregulation is a real danger, and even a trap set by our adversaries: far from being bothered with overzealous regulations, they will actually benefit from the controversy and divisions that it will create. We must be mindful of the risk of our actions having such unintended effects.[21]

Consequently, a genuinely liberal free-speech approach to political disinformation needs to be nuanced and largely non-coercive in character. I argue that the right approach to false political speech involves (1) protecting what I will call the equality conditions for political participation, (2) rejecting, as a basis for state intervention, restrictions on political viewpoint content solely on the grounds that it is false or misleading[22] and (3) giving only limited scope to source-based justifications for making the exercise of free political speech more burdensome.

---

19 See, eg, Martin Montgomery, "Post-Truth Politics? Authenticity, Populism and the Electoral Discourses of Donald Trump" (2017) 16 *Journal of Language and Politics* 619.

20 Romain Badouard, *Le Désenchantement de l'Internet. Désinformation, Rumeur et Propagande* (FYP Editions 2017) 48. There is evidence that "fact-checking", while supposedly politically neutral, is in fact more highly prized by people who are broadly liberal politically than by people whose thinking is more conservative: see Craig T Robertson and others, "Who Uses Fact-Checking Sites? The Impact of Demographics, Political Antecedents, and Media Use on Fact-Checking Site Awareness, Attitudes, and Behaviour" [2020] 5 *The International Journal of Press Politics* 217.

21 Jean-Baptiste Jeangène Vilmer *et al*., *Information Manipulation: A Challenge for Our Democracies* (2018) 176, available at https://www.diplomatie.gouv.fr/IMG/pdf/information_manipulation_rvb_cle838736.pdf (visited on 8 March 2021).

22 In broad terms, thus, I agree with Irina Katsirea on the need to prioritise the protection of free speech in the face of the threat posed by "fake (political) news": see Irina Katsirea, "'Fake News': Reconsidering the Value of Untruthful Expression in the Face of Regulatory Uncertainty" (2019) 10 *Journal of Media Law* 159. See also I Cram, "Keeping the Demos" (n.11). The rejection of state restrictions on false political viewpoint information should be understood to extend to mandatory requirements to post "fake news" alerts alongside supposedly false political statements. While free to take such action if they

II. Source and Content-Based Restrictions, and "Foreign" Falsehoods

A number of factors have conspired to generate a wide-scale legislative interest in reforming the law governing false or misleading political statements, especially those which are designed to shape election thinking and elections outcomes. In 1985, UK voters gave their preference to television (63 per cent), followed by newspapers (29 per cent) and radio (4 per cent) as their most important source of political information.[23] Only a tiny fraction claimed that their political and world news more broadly came from contact with other people, a position radically changed—when one includes virtual contact—since the advent of the internet age. Correspondingly, for much of second half of the twentieth century, the need for reform of the law governing false political statements was perhaps kept adequately at bay (1) informally, by reliance on the perceived need to maintain tolerable working relations between politicians and journalists[24] and (2) more formally, by some regulation[25]—albeit typically "light touch"—and journalistic codes of conduct, applicable to the traditional press. However, these restraints have become less significant, as those governed by them have been by-passed in terms of influence by social media users whose conduct falls outside their scope. In the United States, for example, while the Federal Election Commission noted in 2006 that only 18 per cent of all Americans cited the internet as their leading source of news about the 2004 Presidential election, by contrast, the Pew Research Center found that 65 per cent of Americans identified an internet-based source as their leading source of information for the 2016 election.[26] In that regard, a recent study of 126,000 news stories distributed on Twitter between 2006 and 2017 revealed that falsehoods (as established by fact-checking organisations) were 70 per cent more likely to be re-tweeted than true stories.[27] As Hannah Arendt remarked, "[l]ies are often much more plausible, more appealing to reason, than reality, since the liar has the great advantage of knowing beforehand what the audience wishes or expects to hear".[28] Partly instrumental, in this respect, has been the use of (semi)automated bots,

---

wish, internet platforms should not be *required* to classify political speech as reliable or unreliable for the purposes of issuing such alerts.

23 Ralph Negrine, *Politics and the Mass Media* (London: Routledge, 2nd ed., 1994) p 2.

24 Aeron Davis, "Journalist-Source Relations, Mediated Reflexivity and the Politics of Politics" (2009) 10 *Journalism Studies* 204–219.

25 See, eg, "Section Six: Elections and Referendums" (Ofcom, 6 January 2021), available at https://www.ofcom.org.uk/tv-radio-and-on-demand/broadcast-codes/broadcast-code/section-six-elections-referendums (visited on 8 March 2021).

26 Honest Ads Act 2019 (USA) (Congress.gov, 2019–2020), para.12, available at https://www.congress.gov/bill/116th-congress/senate-bill/1356/text (visited on 8 March 2021); Michael Erbschloe, *Extremist Propaganda in Social Media: A Threat to Homeland Security* (BRC Press: Boca Raton Florida, 2019) para.2.5.

27 Soroush Vosoughi, Deb Roy, and Sinan Aral, "The Spread of True and False News Online" (2018) *Science* 359 1146, 1148.

28 Hannah Arendt, *Crises of the Republic* (San Diego: Harvest Books, 1972) p 6.

fake Twitter or Facebook accounts that enable the spread of false claims though re-tweets and "likes" and non-state agencies organising internet "trolling" on a wide scale.[29]

In analysing responses to the perceived threats posed by online political falsehoods, we need to apply to this context a version of a First Amendment distinction between what the Supreme Court calls "content-neutral" (but I will call "source-based") and content-based grounds for intervention.[30] A ban on a post which fails a transparency test, as when it is anonymous or when it is not revealed that it was generated by a bot,[31] is a source-based measure. The same is true of the imposition of a requirement on social media platforms to maintain a register of those who disseminate election-related "partisan" or advertising messages.[32] Likewise, a ban on the publication of political messaging coming from outside a jurisdiction is a source-based measure.[33] In that regard, a common, source-focused theme in modern proposals to exercise control over the flow of information online is the supposed need to ensure that social media users are aware of the foreign origins and sponsors of posts, especially posts intended to affect election outcomes.[34] An example of why this theme has emerged is the finding that, from 2014 onwards, the St Petersburg-based Internet Research Agency at one point is reported to have controlled 3,814 human accounts and 50,258 bots on Twitter (with which nearly 1.5 million US citizens had some interaction) and 470 Facebook accounts that reached at least 126 million Americans.[35] Similarly, in the lead-up to the UK referendum on leaving the European Union, it was widely reported that over 150,000 Twitter accounts sourced in Russia posted content on Brexit.[36]

---

29 David Murray, "Protecting Our Elections: Examining Shell Companies and Virtual Currencies as Avenues for Foreign Interference" (Financial Integrity Network, 26 June 2018), available at https://www.judiciary.senate.gov/imo/media/doc/06-26-18%20Murray%20Testimony.pdf (visited on 8 March 2021).

30 David L Hudson Jr, "Other Articles in Legal Terms and Concepts Related to Speech, Press, Assembly, or Petition" (The First Amendment Encyclopedia), available at https://www.mtsu.edu/first-amendment/article/937/content-neutral (visited on 8 March 2021).

31 See Protection from Online Falsehoods and Manipulation Act 2019 (Singapore) s.8, discussed in V. Authoritarianism and Political Disinformation: Singapore.

32 Canada Elections Act 2000 (as amended) s.325(2) and (3). In English law, the name and address of the printer and promoter must be included on printed election material: Representation of the People Act 1983 s.110; Political Parties, Elections and Referendums Act 2000 s.143.

33 See, eg, the ban proposed on foreign political advertising via the internet, in the Honest Ads Act (USA) 2019 (n.26). See also the Representation of the People Act 1983 s.92.

34 See, eg, Electoral Commission, *Digital Campaigning: Increasing Transparency for Voters* (London: Electoral Commission, 2018) para.26.

35 JBJ Vilmer *et al*., *Information Manipulation* (n.21), 85. America is yet to adopt measures to require transparency in relation to internet political advertising by overseas entities: see Tim Lau, "The Honest Ads Act Explained" (Brennan Center for Justice, 17 January 2020), available at https://www.brennancenter.org/our-work/research-reports/honest-ads-act-explained (accessed 9 March 2021).

36 David Wolchover and Amanda Robinson, "Is Brexit a Russia-backed Coup?" *New Law Journal*, 24 January 2020. See, eg, House of Commons, Digital, Culture, Media and Sport Committee, *Disinformation* (n.17), part 6; Carole Cadwalladr, "Brexit, the Ministers, the Professor and the Spy: How Russia Pulls Strings in UK" *The Guardian* (4 November 2017), available at https://www.theguardian.

The regrettable rise in the use of such techniques by foreign states or organisations has attracted a great deal of critical attention from liberal democratic politicians, although in the dark arts of foreign "interference" in elections and referendums it is Western countries themselves, of course, who are historically the masters. The United States is estimated to have interfered overtly or covertly in no less than 81 foreign elections between 1946 and 2000, far more than any other major power.[37] Even so, many in the West have raised and continue to raise the bogey of foreign (Chinese or Russian?) interference, influencing opinions and outcomes in national elections.[38] A key difference from older examples of cold war meddling is that the modern aim of such interference—through fake online accounts, the use of net bots and so on—is not so much the spread of the would-be influencer's ideology, as the undermining of the political and social fabric of the target jurisdiction. Accordingly, the would-be influencer may intentionally generate contradictory messages: for example using opposing forms of content to encourage white supremacists and racial minorities to target one another.[39] Quite apart from the question of whether any election or referendum in the West has been *decisively* influenced by foreign interference, it is clear that such interference is potentially destabilising for the traditional model of liberal-democratic politics.[40] For example research has shown that American voters take a highly partisan approach to evidence of foreign interference. They are much more likely to condemn such interference and to profess a loss of faith in the democratic process, when the interference appears helpful to their political opponents than when it helps their own party of choice.[41]

By contrast, as the term makes clear, a content-based legal measure is focused on the substance of what someone has said. We find such measures in law when their target is defamatory statements, holocaust denial, threatening, insulting and hate-filled speech or incitements to engage in violence or terrorism as well as false political statements.[42] Content-based prohibitions capable of covering the last of these—false or misleading political statements—are widely employed in many jurisdictions. Here is an example. Although, in France, freedom of expression is an

---

com/politics/2017/nov/04/brexit-ministers-spy-russia-uk-brexit (visited on 9 March 2021); Ewan McGaughey, "The Extent of Russian-backed Fraud Means the Referendum Is Invalid" (2018), available at https://blogs.lse.ac.uk/brexit/2018/11/14/the-extent-of-russian-backed-fraud-means-the-referendum-is-invalid/ (visited on 9 March 2021); Ewan McGaughey, "Could Brexit Be Void?" (2018) 29(3) *King's Law Journal* 331, 334–336.

37 New York Times, "Russian Isn't the Only One Meddling in Elections: We Do It Too" (17 February 2018); Dov Levin, "A Vote for Freedom? The Effects of Partisan Electoral Interventions on Regime Type" (2019) 63(4) *Journal of Conflict Resolution* 839.

38 Jens David Ohlin, "Did Russian Cyber Interference in the 2016 Election Violate International Law?" (2017) 95 *Texas Law Review* 1579.

39 See JBJ Vilmer *et al*., *Information Manipulation* (n.21), 77.

40 House of Commons, Digital, Culture, Media and Sport Committee, *Disinformation* (n.17), part 6.

41 Michael Toms and Jessica LP Weeks, "Public Opinion and Foreign Electoral Intervention" (2019), available at https://web.stanford.edu/~tomz/working/TomzWeeks-ElectoralIntervention-2019-08-13i.pdf (visited on 9 March 2021).

42 See, eg, Liberty, "Your Human Rights", available at https://www.libertyhumanrights.org.uk/human-rights/free-speech-and-protest/speech-offences (visited on 9 March 2021).

"essential freedom", protected by arts.10 and 11 of the Declaration of Human and Civil Rights 1789, the freedom may be restricted, as when the form of expression is "tantamount to the abuse of this liberty in the cases determined by Law" (art.11). Under art.27 of the Republic Act 1881, the dissemination of disinformation is, in specified circumstances, made subject to criminal penalties:

> The malicious publication, dissemination and reproduction, by whatever means, of false news and documents which have been fabricated or falsified or mendaciously attributed to third parties, when this has disturbed the peace, or was capable of disturbing it, will be subject to a fine of 45,000 euros.[43]

In some cases, a legal measure aimed at false political statements is in part source-based and in part content-based. For example recent legislation in Singapore, discussed further in Section V, makes it an offence to generate a false claim (the content-based element), if the claim is disseminated by a bot (the source-based element).[44] Usually—an issue mentioned earlier—measures that are partly source-based and partly content-based are aimed at false claims considered to be part of an attempt by one jurisdiction to influence political developments in another jurisdiction.[45] Here are some examples. In France, a 2018 amendment to earlier legislation[46] made it possible for the independent broadcasting authority, the Conseil supérieur de l'audiovisuel (CSA), to prevent, suspend or prohibit TV or radio broadcasts controlled by a foreign state (following an initial warning), if the broadcast were judged "to harm the fundamental interests of the nation . . . particularly by disseminating false information".[47] Under this provision, permissible sanctions include an order to stop broadcasting, fines up to 3 per cent of the broadcaster's revenue and in some cases withdrawal of authorisation to broadcast.[48] Another example is provided by s.282.4 of the Canada Elections Act 2000 (as amended in 2018–2019)

---

43 The disturbance of the peace must be more than merely trivial: see Sandra Colliver *et al*. (eds), *Secrecy and Liberty: National Security, Freedom of Expression and Access to Information* (The Hague: Martinus Nijhoff, 1999) p 263.

44 In Singaporean law, the significance of the false claim having been generated by a bot is that, to justify insisting on its suppression, it will not be necessary—as it would in other cases—to show any likelihood of harm resulting therefrom: see V. Authoritarianism and Political Disinformation: Singapore.

45 For a defence of such measures, see JS Sellers, "Legislating" (n.2).

46 French Law No. 2018–1202, on the "fight against the manipulation of information", amending the Freedom of Communication Act No. 86–1067, September 1986. For a detailed analysis, see Rachel Craufurd-Smith, "Fake News, French Law and Democratic Legitimacy: Lessons for the United Kingdom?" (2019) 11 *Journal of Media Law* 52.

47 *Ibid*., arts.42–46 (as amended); "Decision no. 2018–773 DC of 20 December 2018" (Conseil Constitutionnel), available at https://www.conseil-constitutionnel.fr/en/decision/2018/2018773DC.htm (visited on 9 March 2021). See "Limits on Freedom of Expression: France" (Library of Congress), available at https://www.loc.gov/law/help/freedom-expression/france.php#_ftn51 (visited on 9 March 2021). There is an exceptional procedure, when the false information is distributed during an election campaign, that allows the CSA to suspend distribution of a broadcasting service during an election.

48 French Law No. 2018–1202 (n.46) art.42–1.

which prohibits, for the purposes of elections, "undue influence by foreigners". Influence engaged in by a foreigner will be "undue" in a content-based sense, if it involves influencing an elector through the commission of an offence contrary to Canadian law. In broad terms, such a prohibition might seem unexceptionable; but the prohibition extends to some controversial cases such as the offence of making false claims about an election candidate, party leader or official (contrary to the amended s.91 of the 2000 Act).[49]

How should we interpret these developments? There are reasons to treat them with caution. There can, subject to what I will say later about anonymity, ordinarily be no objection to a requirement simply to disclose that a source of political viewpoint disinformation is overseas (clumsy and sometimes arbitrary though such requirements may be). Measures to tackle disinformation that do no more than require "more speech" are easier to justify that those that directly restrict speech.[50] However, much more problematic is the question whether it is legitimate to restrict access to, or ban, a source of political viewpoint disinformation merely because it has an overseas source or connection. To begin with, it is not clear that it is necessarily unethical for one nation state (or group or party within that state) to seek to influence an election result in another nation state. An example of ethical influence might be where a foreign state acts in order to further the chances that a government more respectful of democratic principles, or international law and human rights, will come to power in the targeted state.[51] Moreover, a failure of legal toleration can lead to unfairness and arbitrariness. Is it right, for example, that Canadian law now treats political influence by former Canadian citizens who have exchanged Canadian for, say, French citizenship (and are hence "foreigners") in the same way as Moscow-based operators of bots set up to secure such influence? It is no real answer to this question to say that restrictions apply only to false or misleading political claims. That is because what is to be regarded as "false or misleading" is itself to be determined by the targeted state, with no opportunity (legislation almost never provides for this) for the maker of the claim to defend it. As Justice Black said in *Susan B Anthony List v Ohio Elections Commission*[52]: "We do not want the government . . . deciding what is political truth—for fear that the government might persecute those who criticize it. Instead, in a democracy, the voters should decide".[53]

A partial acknowledgement of the ethical problems that may arise is to be found in an important exception to the reformed s.282.4 of the Canada Elections Act 2000 (just mentioned). Subsection (3) exempts from the scope of the ban:

---

49 See further, Michael Karanicolas, "Subverting Democracy to Save Democracy: Canada's Extra-Constitutional Approaches to Battling 'Fake News'" (2019) 17 *Canadian Journal of Law and Technology* 201.

50 See the discussion of "more speech" at (n.89).

51 Cécile Fabre, "The Case for Foreign Electoral Subversion" (2018) 32(3) *Ethics and International Affairs* 283.

52 45 F Supp 3d 765 (2014).

53 *Ibid*., 769.

(a) an expression of . . . opinion about the outcome or desired outcome of the election;
(b) a statement . . . that encourages the elector to vote or refrain from voting for any candidate or registered party in the election; or
(c) the transmission to the public through broadcasting, or through electronic or print media, of an editorial, a debate, a speech, an interview, a column, a letter, a commentary or news, regardless of the expense incurred in doing so, if no contravention of subsection 330(1) or (2) is involved in the transmission.[54]

This exception is significant in that it concedes the controversiality of trying to exercise punitive control over the substantive content of political statements intended to influence elections, even when the statements are generated in foreign jurisdictions. The exception is, thus, a nod towards the importance of protecting free political speech. It is, nonetheless, controversial in its terms in that it makes criminal liability turn, in part, on the notoriously difficult distinction between statements of fact and statements of opinion.[55] An irony about that approach is that evidence indicates that, for example, Russian attempts to influence the outcome of the 2016 Presidential election were concentrated not so much on false information as on "opinion making" in the form of the creation of a narrative of political identity ("Trump is with us, and against the establishment").[56]

More generally, in policing political statements, is it easier to justify source-based regulation, as when coercing those posting on the net into revealing their identity? Source-based controlling measures are likely to be in one respect less controversial than content-based ones. The basis of a purely source-based ban is usually a factual finding: for example that someone putting up, further distributing or sponsoring a post is based outside the jurisdiction or that they have failed to reveal their identity or location.[57] So, the intervention of courts or regulators to police source-based restrictions may seem less politically charged and controversial than when such bodies must police content-based restrictions. Going beyond this point about process, Kate Jones has argued rightly that, "the right to freedom of expression does not entail that techniques for the manipulation of attention, such as use of bots and trolls, must be free of restriction".[58] "Trolling", of course, may involve threats, hate speech or other harms justifiably made subject to criminal sanctions on

54 On s.330, see M Karanicolas, "Subverting Democracy" (n.49).
55 See, eg, Jeffrey L Kirchmeier, "The Illusion of the Fact-Opinion Distinction in Defamation Law" (1988) 39 *Case Western Reserve Law Review* 867.
56 Michael Jensen, "Russian Trolls and Fake News: Information or Identity Logics?" (2018) 71 *Journal of International Affairs* 115, 122.
57 As in the case of English law governing printed election material: see Representation of the People Act 1983; Political Parties, Elections and Referendums Act 2000 (n.32).
58 Kate Jones, "Online Disinformation and Political Discourse: Applying a Human Rights Framework" (2019) 45 and 53, available at https://www.chathamhouse.org/sites/default/files/2019-11-05-Online-Disinformation-Human-Rights.pdf (visited on 9 March 2021).

content-based grounds.[59] So, if compelling online platforms to reveal the identity of "trolls" (a source-based measure) serves to deter the dissemination of material that is harmful in terms of its content, then so much the better.[60] However, it seems controversial to put the use of bots into the same category as online trolling, because their use raises a different set of issues. Political organisations, in particular, should certainly be forced to be honest about their use of bots.[61] In California, it is now a form of unlawful competitive behaviour to use a bot *via* a major online platform to communicate or interact, with a view to influencing an election vote, without disclosing (in a manner that is "clear, conspicuous, and reasonably designed") that the communication or interaction is through a bot.[62] However, it must be kept in mind that a bot may be acting simply as a kind of virtual "amplifier", something that may be used to spread valuable political (or eg health) information, as well as false claims.[63]

Further, so far as compulsory identity-disclosure is concerned, it ought also to be kept in mind that some individuals and political organisations—whether domestic or based overseas—may perfectly understandably wish to conceal their identity when articulating political claims. They may have well-justified fears of government censorship and prohibition or of retaliation from violent groups and individuals. As the UN Special Rapporteur has argued, anonymity provides individuals with:

> a means to protect their privacy, empowering them to browse, read, develop and share opinions and information without interference and enabling journalists, civil society organizations, members of ethnic or religious groups, those persecuted because of their sexual orientation or gender identity, activists, scholars, artists and others to exercise the rights to freedom of opinion and expression.[64]

Compelling would-be speakers to reveal their identity, in such circumstances, may thus unacceptably confront them with a free speech dilemma (further considered

59 See Maeve Duggan, "Online Harassment" (Pew Research Center: Internet & Technology, 2014), available at https://www.pewresearch.org/internet/2014/10/22/online-harassment/ (visited on 9 March 2021).

60 See, eg, the Harmful Digital Communications Act 2015 (New Zealand) s.22.

61 K Jones, "Online Disinformation and Political Discourse" (n.58).

62 See the Bolstering Online Transparency Act 2018, SB 1001. For criticism, see Renee Direstaideas, "A New Law Makes Bots Identify Themselves—That's the Problem" (Wired, 24 July 2019), available at https://www.wired.com/story/law-makes-bots-identify-themselves/ (visited on 9 March 2021).

63 For a defence of the use of bots, under the First Amendment, see Madeline Lamo and Ryan Calo, "Regulating Bot Speech" (2019) 66 *UCLA L Rev* 988.

64 3 Report of the UN Special Rapporteur, *Freedom of Expression* (2015), UN Doc A/HRC/29/32, cited by Privacy International (2019), *Submission to the House of Lords Select Committee on Democracy and Digital Technologies,* para.43, available at https://privacyinternational.org/sites/default/files/2020-03/19.09.20%20HoL%20Democracy%20and%20Digital%20Tech%20submission_final-1.pdf (visited on 9 March 2021). See also *Right to Online Anonymity* (2015) art.19, available at https://www.article19.org/data/files/medialibrary/38006/Anonymity_and_encryption_report_A5_final-web.pdf (visited on 9 March 2021).

later in this article). Social media users may also reject official understandings of "disinformation" and, perhaps in part for that reason, wish to maintain their anonymity.

The House of Lords Communications Committee suggested a "middle way" on the anonymity question, in relation to the application of criminal offences, in which someone would have to identify themselves in order to register with an internet platform but would be permitted to post anonymously thereafter. Nonetheless, the Committee recognised that even this is highly controversial.[65] Giving priority to freedom of political viewpoint speech may mean accepting that there is only a public interest in securing the identification of online agents posting political viewpoint disinformation in a limited range of cases. The right of an individual to seek the identification of a speaker might cover cases which there has been an invasion of private rights, so long as the law that makes the statement right-violating is itself consistent with human rights jurisprudence. The right of the platform to insist on identity disclose *ex ante* might cover cases where the source of the alleged disinformation is an account associated with an official or political party.

That may seem to involve an unduly tolerant attitude to the irresponsible spread of false or misleading political viewpoint information under the cover of anonymity. However, it should be kept in mind that protecting such a freedom is consistent with a platform policy of insisting that, for example, a "potentially unreliable information" warning is displayed prominently in all anonymous posts. In seeking to promote a flourishing free-speech culture, solutions encouraging critical engagement on the part of the audience are generally to be preferred to solutions inhibiting the freedom of speakers to engage with an audience. It is a very different matter, though, where someone has deliberately misrepresented their identity in order to spread disinformation. In any context, including in relation to political viewpoint information, intentionally to pretend to be another person in order to gain an audience or to gain credibility with an audience is to forfeit one's right to connect with the audience by such means on the platform in question. In that regard, it is noteworthy that, in a political context, in August 2018, Twitter removed approximately 50 accounts misrepresenting themselves as members of various state Republican parties.[66]

---

65 HL Select Committee on Communications, *Social Media and Criminal Offences* (2014–2015) para.54, available at https://publications.parliament.uk/pa/ld201415/ldselect/ldcomuni/37/3704.htm#a3 (visited on 9 March 2021). See also HM Government, *Online Harms, White Paper*, CP 57 April 2019, para.7.28. Department for Digital, Culture, Media and Sport, and Home Office, *Online Harms White Paper: Full government Response to the White Paper* (2020) para.2.35 (box 8), available at https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response#fnref:21 (visited on 9 March 2021). In the latter paper, the Government drew back from making any concrete proposals on the issue of anonymity, other than to say firms must address it in their terms and conditions.

66 Del Harvey and Yoel Roth, "An Update on Our Elections Integrity Work" (Twitter, 1 October 2018), available at https://blog.twitter.com/en_us/topics/company/2018/an-update-on-our-elections-integrity-work.html (visited on 9 March 2021).

III. Restraint in Criminalisation: The Scepticism Principle

I now turn my attention to the ethics of restrictions on false viewpoint political content, content embodying substantive political claims with a (false or misleading) factual element. One important principle of restraint, in this context, is what might be called the principle of scepticism. It is well articulated by Frederick Shauer:

> Freedom of speech is based in large part on a distrust of the ability of government to make the necessary distinctions, a distrust of governmental determinations of truth and falsity, an appreciation of the fallibility of political leaders, and a somewhat deeper distrust of government power in a more general sense.[67]

Thus described, the "principle" of scepticism is not so much a principle as a (healthy) public attitude, an attitude diametrically opposed to blind faith—encouraged by authoritarian regimes—in the judgment of public officials when exercising existing or creating new powers. However, it is convenient to speak in terms of the "principle" of scepticism, a principle which may have a number of manifestations.

An important aspect of Shauer's account of the principle's application is his claim that it involves "distrust of governmental determinations of truth and falsity". In that respect, the principle is to be found at work, for example, in a refusal—or least, a deep reluctance—to give government agencies or private bodies (such as social media platforms) when directly coerced by such agencies, the responsibility for determining the (in)appropriateness of false political viewpoint content for dissemination. In declaring unconstitutional a prohibition on the malicious publication of a false statement of material fact about a candidate for public office, the Washington Supreme Court said the statute was unconstitutional because it caused "the government, rather than the people, [to] be the final arbiter of truth in political debate".[68] Knowing that distrust and suspicion may arise in such circumstances, it can be tempting to turn the responsibility for policing political discourse over to courts (usually, more highly trusted by ordinary people[69]). The acceptability of doing this will depend on the legal culture in which such move is made, but there are significant risks associated with it. These risks are linked to the principle of scepticism in that judges called on to make the relevant judgments will be open to accusations of (elite) politically motivated censorship. This point was raised by the Lord Chief Justice for England and Wales as long ago as 1868 in a letter criticising the proposal to transfer jurisdiction to hear complaints about corruption in

67 Frederick Shauer, *Free Speech: A Philosophical Enquiry* (Cambridge: Cambridge University Press, 1982) p 86.

68 *Rickert v Public Disclosure Commission* 168 P 3d 826, 827 (2007).

69 "Politicians Remain the Least Trusted Profession in Britain" (Ipsos, 30 November 2017), available at https://www.ipsos.com/ipsos-mori/en-uk/politicians-remain-least-trusted-profession-britain (visited on 9 March 2021).

elections from Parliament to the courts. Chief Justice Cockburn complained to the Lord Chancellor that:

> The decision of the Judge given under such circumstances will too often fail to secure the respect which judicial decisions command on other occasions. Angry and excited partisans will not be unlikely to question the motives which have led to the judgment. Their sentiments may be echoed by the press. Such is the influence of party conflict, that it is apt to inspire distrust and dislike of whatever interferes with party objects and party triumphs.[70]

It is also important to tease out another aspect of the principle of scepticism. This is that a sceptical attitude ought to lead to an asymmetrical approach to the criminalisation of certain kinds of wrong. While, for example, it normally ought to be possible for any individual to take steps using private law to protect their reputation, a sceptical attitude ought to be taken to the use of the criminal law to protect, in particular, the reputation or dignity of public officials or election candidates.[71] The involvement of the state in threatening penal sanctions, in cases where officials or would-be officials believe that they have been defamed or insulted, is by its nature compromised and open to abuse, when analysed from the viewpoint of scepticism. Officials may call for a police investigation into a rival's supposed insult or allegedly defamatory remark for political reasons (to damage their rival's political credibility). Even if the call for an investigation is *bona fide*, if it is made during an election period, then if the police say that the investigation must be delayed until after the election, that decision will inevitably appear as much open to political motivation as a decision to launch an investigation straight away would be.[72]

The principle of scepticism should not, though, be regarded as inconsistent with any kind of government activity in the area of free speech and free expression, even in relation to elections and political participation. For example, although (as Shauer rightly remarks) political leaders are fallible, governments should nonetheless strive to provide access to political information for citizens and to promote as public goods widespread political debate and participation.[73] Further, the use of criminal prohibitions on some forms of speech ought to be regarded as perfectly consistent with adherence to the principle of scepticism. For example it is uncontroversial that many jurisdictions use the criminal law to deter the spread of false

70 Cited by Thomas LCJ, in *R (Woolas) v Parliamentary Election Court* [2012] QB 1, [23]. For further discussion of this point, see, eg, Alan Renwick and Michela Palese, *Doing Democracy Better: How and Information and Discourse in Election and Referendum Campaigns in the UK Be Improved?* (London: University College London Constitution Unit, 2019) Ch.2.

71 See, eg, Council of Europe, *Defamation and Freedom of Expression*, H/ATCM (2003) 1 (Strasbourg: Directorate of Human Rights, 2003).

72 See, eg, *Oberschlick v Austria* (1995) 19 EHRR 389.

73 See, generally, Eric Barendt, *Freedom of Speech* (Oxford: OUP, 2nd ed., 2007) pp 30–38.

claims in relation to the voting process: what I call participation disinformation.[74] There is no special reason to look askance at government efforts, say, to prevent or punish online campaigns to deceive people into thinking that an election or referendum is to take place a day later than the official poll. This point warrants further exploration, in relation to the next principle of restraint, the autonomy principle.

IV. Restraint in Criminalisation: The Autonomy Principle

The principle of scepticism—more properly, a sceptical attitude—is connected to another principle: the principle of political autonomy.[75] One aspect of the autonomy principle is that, in Eric Barendt's words (drawing on Thomas Scanlon[76]), "a person is only autonomous if he is free to weigh for himself the arguments for various courses of action that others wish to put before him".[77] In itself, though, this statement of the principle does not take us far enough. That is because it is focused primarily on the autonomy of the potential audience members and not on the autonomy of speakers. The freedom to weigh arguments for oneself will be worthless without access to those arguments. So, an important—perhaps more important—dimension to the principle of autonomy is that speakers must be free to put forward—and undeterred in putting forward—arguments, so that they may be weighed. Does that include the freedom to put forward political arguments that are (intended to be) false or misleading? In striking down a statute creating an offence of knowingly or recklessly making a false statement about political candidates or ballot initiatives, the Massachusetts Supreme Court answered with a "yes" to this question. The Court suggested that that such an offence was likely to chill "the very exchange of ideas that gives meaning to our electoral system".[78] Even so neither First Amendment jurisprudence (not further considered here)[79] nor ECHR case law gives an unequivocally affirmative or negative answer to the question. So, what is the right approach?

One way to search for the answer is to set the autonomy principle in a proper context. My political well-being, my flourishing as a political citizen, depends substantially on government action to defend and promote what might broadly be called the freedom and equality-based dimensions to political participation. Where voting is concerned, what this entails is that, for example, there must be accessible places for all to vote, without intimidation or undue influence (eg through

74 See, eg, the UK's Representation of the People Act 1983 s.115; J Horder, "Criminal Law" (n.2); JS Sellers, "Legislating" (n.2).

75 Thomas Scanlon, "A Theory of Freedom of Expression" (1972) 1 *Philosophy and Public Affairs* 204.

76 *Ibid*.

77 E Barendt, *Freedom of Speech* (n.73), 16.

78 *Commonwealth v Lucas* 34 NE 3d 1242 (2015).

79 The relevant part of the First Amendment reads "Congress shall make no law . . . abridging the freedom of speech, or of the press". See further, Simon Rodell, "*False Statement v Free Debate*: Is the First Amendment a License to Lie in Elections?" (2008) 60 *Florida Law Review* 947.

secret balloting). There must also be sufficiently well-publicised details about how, where and when to vote, and there must be enough information about candidates such that voters can distinguish adequately between them. Further, there must not be artificial (and in practice, biased) barriers to voting, such as disproportionately onerous identification requirements,[80] the counting of all the votes must be accurate and impartial and so on. Guaranteeing that these equality and freedom conditions for political engagement exist nationwide requires disinterested, collective (state) action, given the significant coordination problems to be solved. As is the case in many other instances in which the law must step in to solve such problems, the use of the criminal law to deter and punish breaches of, or attacks on, the freedom and equality conditions may well be a perfectly proportionate as well as necessary step.[81] False statements knowingly made about these participation conditions fall squarely within the scope of legitimate criminalisation.

Can this argument about freedom and equality be taken a step further to justify prohibitions on the manipulation or distortion of politics through the dissemination of false viewpoint information? For example it has been suggested that the toleration of the use of algorithms that manipulate what voters see or which distort the impression they get of how public debate is going may be inconsistent with the "right to vote" under art.25 of the International Covenant on Civil and Political Rights[82] (and hence with the freedom or equality conditions just described).[83] Even if that—militant democratic claim—is true, it would not necessarily provide a proportionate justification for using coercion to restrict such (misuse of) free speech. To begin with, a state can legitimately commit a jurisdiction to freedom of information and to the promotion of honesty in politics through non-coercive measures. Examples involving state action would be the introduction of codes of electoral conduct promoting transparency and honesty or the provision of tax breaks for independent fact-checking organisations. Further, in the case of platforms, action can be taken to promote or highlight a range of sources of political viewpoint information and to ensure that persistently unreliable sources find their posts accompanied by unreliability warnings and links to alternative sources of information.[84]

The scepticism and autonomy principles combine to condemn as unethical the use of coercion to deter practices aimed at the formation of substantive political beliefs and intentions, even when the result is that the beliefs and intentions may be formed on a false basis. To be sure, as Raz puts it "one does not help people

80 William D Hicks *et al*., "A Principle or a Strategy? Voter Identification Laws and Partisan Competition in the American States" (2015) 68 *Political Research Quarterly* 18.

81 See, eg, J Horder, "Criminal Law" (n.2); JS Sellers, "Legislating against Lying" (n.2).

82 "Every citizen shall have the right and the opportunity . . . without unreasonable restrictions (a) To take part in the conduct of public affairs, directly or through freely chosen representatives; (b) To vote and to be elected at genuine periodic elections". See also art.3 of the First Protocol to the ECHR.

83 See K Jones, "Online Disinformation" (n.58), 48.

84 See Department for Digital, Culture, Media and Sport, and Home Office, *Online Harms* (n.65), paras.7.27–7.31.

to lead the lives they want to have by satisfying their false desires".[85] Under the principle of scepticism, the taking by the state of coercive steps to deter and punish the dissemination of false political viewpoint information is illegitimate because officials are fallible. They will too often be tempted to breach the trust placed in them by serving self-interested purposes. Under the autonomy principle, the taking of such steps is equally to be ruled out but primarily for speaker-focused reasons.[86] Consequently, more important to the autonomy principle are the rights of speakers, the disseminators of (false) political viewpoint information, to be free from the threat of coercion in creating their own political narrative and disseminating it in a manner of their own choosing. For adherents of the autonomy principle, the use of coercion to deter disseminators of false political viewpoint information is regarded as unnecessary, because a flourishing political culture can adequately address falsehood through the republican remedy of so-called "more speech" solutions. As Justice Brandeis famously put it in *Whitney v California*:[87]

> To courageous, self-reliant men [*sic*], with confidence in the power of free and fearless reasoning applied through the processes of popular government . . . [i]f there be time to expose through discussion the falsehood and fallacies, to avert the evil by the processes of education, the remedy to be applied is more speech, not enforced silence.[88]

The EU defines disinformation as "verifiably false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public, and may cause public harm".[89] It has called for urgent and immediate action to protect the Union, its institutions and its citizens against disinformation. However, its proposed "four pillars" response involves not the threat of coercion but a "more speech" solution, focused on quickly identifying and calling out disinformation. The four pillars are:

(1) improving the capabilities of Union institutions to detect, analyse and expose disinformation;
(2) strengthening coordinated and joint responses to disinformation;
(3) mobilising private sector to tackle disinformation and
(4) raising awareness and improving societal resilience.

---

85 Joseph Raz, *The Morality of Freedom* (Oxford: Clarendon Press, 1986) p 144.

86 There are worries about paternalism undermining dog the audience-focused argument that what matters is how one addresses the risk that voters will encounter false political speech and hence take a "wrong" turning in terms of their political beliefs and intentions. Indeed, the cynic might add that many governments historically owe their election to their ability to persuade sufficient numbers of people to behave in just such a way.

87 74 US 357 (1927).

88 *Ibid*., 377 (Brandeis J).

89 EU High Representative of the Union for Foreign Affairs and Security Policy, *Action Plan against Disinformation* (JOIN 2018, 36 Final) 1.

In relation to (2), for example, what is proposed is "prompt reaction via fact-based and effective communication . . . to counter and deter disinformation, including in cases of disinformation concerning Union matters and policies".[90] Similarly, in France, a Report from the Policy Planning Staff (Ministry for Europe and Foreign Affairs) and Institute for Strategic Research (Ministry for the Armed Forces) defines political disinformation as "the intentional and massive dissemination of false or biased news for hostile political purposes"[91] and says that "Governments can and should come to the aid of civil society . . . [because they] cannot afford to ignore a threat that undermines the foundations of democracy and national security".[92] Yet, once again, the primary solution proposed is not the use of coercion but a "more speech" solution. In preference to a top-down approach, the Report recommends "horizontal, collaborative approaches, relying on the participation of civil society", accompanied by the establishment of a national entity responsible for the detection and countering of information manipulation.[93]

As well as being unnecessary, content-based restrictions aimed at false substantive political content are also a disproportionate infringement of the autonomy of speakers, when they are backed by coercive threats in virtue of their unacceptably chilling effect.[94] Such restrictions give rise to what can be called political falsehood dilemmas for speakers. To begin with, legal prohibitions on false statements require officials to grapple with the problem of "half-truths": content that contains an element of truth, but which is packaged in accompanying mistruths or selective reporting of factual information that creates a skewed view.[95] Is a statement comprising an untrue and a true element a "false" statement or perhaps merely a "misleading" statement? When it is known that legal uncertainties of this kind may arise, people may find themselves facing a political dilemma over whether to speak or not and could be deterred from making political claims that are simply provocative or controversial to the detriment of a free speech polity. Further, a political falsehood dilemma arises when someone knows that if they make a political statement that turns out to be false, their claim that they honestly believed in its veracity when the statement was made is unlikely to be believed. Such a dilemma may also arise when what the speaker intends to say is known by them to be untrue or misleading, but they nonetheless believe that there is a compelling moral reason to make the statement: in other words, when they may not be acting "dishonestly" in a broad

90 EU High Representative of the Union for Foreign Affairs and Security Policy, *Action Plan* (n.89), 8.

91 JBJ Vilmer *et al*., *Information Manipulation* (n.21), 12.

92 *Ibid*., 13.

93 *Ibid*., 167. See also, NATO STRATCOM, *Internet Trolling as a Tool of Hybrid Warfare: The Case of Latvia* (2016), available at https://www.stratcomcoe.org (visited on 9 March 2021).

94 See, eg, M Karanicolas, "Subverting Democracy" (n.49), 223.

95 Alex Hern, "Far Right 'Use Russian-style Propaganda to Spread Misinformation'" *The Guardian* (12 November 2019), available at https://www.theguardian.com/technology/2019/nov/12/far-right-use-russian-style-propaganda-to-spread-disinformation (visited on 9 March 2021); K Jones, "Online Disinformation" (n.58).

sense of that term.[96] In these two instances, the political falsehood dilemma arises even though the would-be speaker may share a conception of political truth with those who will stand in judgment over his or her utterances. However, it is obvious that such a dilemma also arises when someone disseminates or shares political information they believe to be true or defensible, but which they appreciate is regarded by those who will stand in judgment as false or indefensible (eg "President Obama was not born in the United States"[97]). In such cases, the existence of the political falsehood dilemma involves the worry that would-be contributors to political debate may be wrongly deterred from engaging in what is simply combative or attention-grabbing political speech by what they see as the prospect of sanctions imposed by those (including the courts) they regard as political enemies.[98]

V. Authoritarianism and Political Disinformation: Singapore

As Bradshaw and Howard rightly observe,[99] authoritarian jurisdictions use the internet and social-media technologies to enhance their control over information, and their use of legislation to counter "disinformation" may be an important dimension to that. What differentiates, in that regard, authoritarian disinformation strategies from the strategies at work in democracies? In democracies, so far as the spread of disinformation is concerned, concerted disinformation campaigns tend to arise mainly at election times[100] and will not involve official state action to propagate such disinformation. By contrast, in authoritarian regimes, political viewpoint disinformation (and sometimes participation disinformation) is likely to be employed by state officials themselves on an ongoing basis, often to target individuals such as dissidents or journalists as well as citizens at large. In democracies, targeting of this latter kind may be (routinely) engaged in by political parties, regime-friendly media outlets, bloggers and so on but will nonetheless be outside the scope of permissible state action, either directly or through agents. Turning to strategies to counter political disinformation, we will see in due course that there will be an overlap in important respects in terms of what is legally proscribed or permissible, between authoritarian regimes and militant democratic regimes which employ coercion in the service of elite interests. In that regard, two key points of overlap are the extension of the criminal law to cover defamation of public officials or election candidates and the grant of powers—backed by the threat of

96 Jeremy Horder, *Excusing Crime* (Oxford: OUP, 2004) pp 48–50.

97 Anthony Zurcher, "The Birth of the Obama 'Birther' Conspiracy" (16 September 2016), available at https://www.bbc.co.uk/news/election-us-2016-37391652 (visited on 9 March 2021).

98 For these reasons, in *R v Zundel* [1992] 2 SCR 731, the Supreme Court of Canada struck down a criminal prohibition on "wilful publication of false statement or news that person knows is false and that is likely to cause injury or mischief to a public interest" as unconstitutional.

99 Samantha Bradshaw and Phillip Howard, "The Global Organisation of Social Media Disinformation Campaigns" (2018) 71 *Journal of International Affairs* 23, 25.

100 *Ibid*., 28–29.

coercion—to require the suppression of allegedly false political viewpoint information on the grounds of content-based objections. By contrast, democratic states inspired by the autonomy and scepticism principles to reject militant forms of democracy will generally not use the criminal law—or other forms of coercion—to protect officials or candidates from viewpoint disinformation, while vigorously countering (through coercion if need be) electoral participation disinformation.

An authoritarian response to political viewpoint disinformation, whether or not confined to election periods, will typically be characterised by a number of features, including one or more of:

(1) The strengthening and broadening of coercive powers to deter and punish the dissemination of disinformation, including powers over social media platforms on which such information appears, without any attempt to prefer source-based measures to content-based measures;
(2) Giving power to the executive to invoke procedures and/or to determine what information has false or misleading content or even confining such powers to the executive,[101] both in and beyond election times;
(3) Making a focus (an exclusive focus in some instances) in law or in enforcement policy, alleged disinformation concerning government actions or the actions of other state officials, whether or not in relation to elections;
(4) Making high, often disproportionate levels of punishment available;
(5) The use of state-sponsored "trolling", targeting dissidents and journalists.[102]

It is important to note that at least the first of these features (1) and to some extent the second and third features (2 and 3) will also be found in some liberal legal systems influenced by militant democratic thinking, discussed later on, even though liberal regimes will generally eschew the fourth and fifth features.

I will take as exemplar of an authoritarian approach a controversial recent set of legal changes in Singapore. I focus on Singapore, because in 2019, Singapore introduced measures to counter disinformation that have a number of authoritarian characteristics. This is in spite of Singapore's status as a constitutional republic with a Westminster style of representative, democratic government based on the separation of powers, with a constitutional guarantee of freedom of speech (art.14).[103] Under s.7 of Singapore's Protection from Online Falsehoods and Manipulation Act 2019, "a person must not do any act in or outside Singapore in order to communicate in Singapore a statement knowing or having reason to believe that . . . it is a

101 *Ibid*., 26–27.
102 *Ibid*., 28, speaking of Bahrain, Turkey, Azerbaijan and Russia.
103 Asean Law Association, *The Singapore Legal System* (2018) Ch 2, available at https://www.aseanlawassociation.org/wp-content/uploads/2019/11/ALA-SG-legal-system-Part-2.pdf (visited on 9 March 2021). On a constitutional–theoretical map, Singapore could perhaps be located between a genuine Westminster-style democracy and a fully authoritarian regime that has only a "facade constitution": GA Tóth, "Authoritarianism" (n.3), para.19.

false statement of fact", when that communication of the statement is likely to have a range of effects, including:

(1) be prejudicial to the security of Singapore or any part of Singapore;
(2) be prejudicial to public health, public safety, public tranquillity or public finances;
(3) influence the outcome of an election to the office of President, a general election of Members of Parliament, a by-election of a Member of Parliament, or a referendum;
(4) incite feelings of enmity, hatred or ill-will between different groups of persons; or
(5) diminish public confidence in the performance of any duty or function of, or in the exercise of any power by, the Government, an Organ of State, a statutory board, or a part of the Government, an Organ of State or a statutory board.

There are a number of points to note about this provision, which in essence contains all the first four features (1)–(4) of authoritarian responses listed before.[104]

First, an important point is that the provision applies not only to content known to contain a false statement of fact, but also to cases in which the maker has "reason to believe" that his or her communication involves a statement that is false. Clearly, this greatly extends the scope of the provision, giving both individuals and media outlets far less control over and knowledge of the extent of their potential liability. In England and Wales, similar laws extending liability to "reason to believe" cases have been held to be in essence inconsistent with the right to free speech protected by art.10 of the ECHR.[105] Second, the communication of a false statement is not, as such, prohibited[106] unless it is likely to have one of the effects mentioned in the criteria listed in (1)–(5) above; but these criteria are both broad and vague. In truth, the criteria should be seen less as a restraint on the scope of the prohibition and more as an encouragement to the prosecution to target certain kinds of conduct. Third, the prohibition draws no distinction between cases in which (say) one person communicates—perhaps relatively informally—to only one other person or to a small number of other people and cases in which a more wide-ranging and systematic attempt is made to spread disinformation widely or to an identified target audience. The need to satisfy one of conditions (1)–(5) might appear to rule out low-level informal communications, but that is far from clear. More liberal jurisdictions have sought to confine draconian online disinformation laws—such as Germany's

104 For example, an individual found guilty under s.7 can be imprisoned for up to five years: s.7(2). The maximum sentence rises to 10 years' imprisonment in a case where an "inauthentic online account or bot" is used in breaching s.7: s.7(2)(c).

105 *R (Woolas) v Parliamentary Election Court* [2012] QB 1.

106 Unless, by virtue of s.8, it stems from a "bot" in which case communication of the false information—or enabling another person to communicate it—is itself sufficient to amount to an offence.

Network Enforcement Act 2017 (considered later in this article)—to platforms with millions of registered users.

Fourth, Pt.3 of the 2019 Act provides for an alternative to prosecution, namely the issuing of a "correction direction" or a "stop communication direction", measures extending even further than s.7. As the names imply, such directions require the person subject to them either to stop communicating the allegedly false information or to provide a declaration that the given communication is false and to publish a "correction" in a form specified.[107] Extraordinarily, as well as applying to persons outside as well as within Singapore (so long as the communication is in Singapore), the "stop communication" and "correction" directions can be issued to and will bind the person who communicates the false statement of fact *irrespective* of whether they knew or had reason to know that the information was false.[108] In that regard, we should note the broad-ranging definition of "false statement of fact" given in the definition section of the 2019 Act:

(1) a statement of fact is a statement which a reasonable person seeing, hearing or otherwise perceiving it would consider to be a representation of fact; and
(2) a statement is false if it is false or misleading, whether wholly or in part, and whether on its own or in the context in which it appears.

There is little or no scope here for the courts to rein in the operation of the 2019 Act by interpreting "statement of fact" in a restrictive way.

Particularly notable, in this respect, are the evaluative elements to the determination of whether a statement is false, namely whether it is "misleading" and whether it is false or misleading "in context". This definition deprives the determination of whether a statement of fact is indeed "false" of what would otherwise be an important element of objectivity, making it hard for speakers to determine whether they have "reason to know" that a statement is false. By this means that the s.7 offence is turned into one of, in effect, strict liability and (what is more) liability dependent as much on how a Minister evaluates a statement as on anything done by the speaker. Finally, it is a Government Minister who, by virtue of s.10(1) may require that a competent authority—a Statutory Board, appointed by the Minister under s.6 and required to give effect to his or her instructions—issue a stop communication or a correction direction, politicising the process in an authoritarian way.[109] Almost needless to say, breach of a stop communication or correction direction, without reasonable excuse, is a criminal offence, punishable by up to 12 months' imprisonment.

107 The full extent of what may be required is set out in ss.11 and 12.

108 "Protection from Online Falsehoods and Manipulation Act 2019", s.11(4); s.12(4).

109 Although s.17 permits an appeal to the High Court against the issuing of a direction, it is possible if and only if the person affected has first asked the Minister to withdraw the direction and he or she has refused. No time scale is given for the Minister to respond, making the issuing of a direction a particularly powerful weapon at election times.

In that regard, we should note that, in virtue of a Code of Practice issued under the 2019 Act, prescribed internet intermediaries are duty-bound to put in place reasonable due diligence measures, among other things, to ensure: "Prioritising relevant and authoritative information and increasing the visibility of such information where appropriate in automatically ranked distribution channels, or reducing the visibility of material that is the subject of a Part 3 or Part 4 Direction".[110] By this means that the newly created Protection from Online Falsehoods and Information Act Office can seek to co-opt internet platform into pursuing the government's online political information agenda.

These developments must be set alongside three already well-known features of Singaporean law that make it more authoritarian than democratic in this area of law. First, defamation is a criminal offence under s.499 of the Singapore Criminal Code, if D makes or publishes any imputation concerning any person, intending to harm, or knowing or having reason to believe that such imputation will harm the reputation of such person.[111] Second, there is the fact that the right to free speech under art.14 of the Constitution may be overridden not only when it is "necessary" but also when it is "expedient" to do so, in the interests of security, public order, friendly relations with other countries or morality (art.14(2)).[112] Liberal-democratic countries do not regard expediency as a legitimate ground for restricting freedom of speech in pursuit of such ends. Third, there is the fact that no greater licence is given to criticise public figures at elections times, either in criminal law or in the civil law of defamation. Section 14 of Singapore's Defamation Act[113] expressly rules out the application of qualified privilege to allegedly defamatory statements made by or on behalf of election or presidential candidates. The courts have held that there is no defence of qualified privilege in defamation law applicable to criticism of public figures respecting the performance of their duties, even when the offending statements are made at a political rally.[114]

By way of contrast, in countries governed by art.10 of the ECHR, although (controversially) offences involving defamation of public officials have been held to be consistent in general terms with rights to freedom of expression, the European Court has found such consistency hard to establish when political speech is involved. In *Lingens v Austria*,[115] in considering false claims made about politicians, the European Court observed that "freedom of political debate is at the very

110 Protection from Online Falsehoods and Information Act Office, "Code of Practice for Giving Prominence to Credible Online Sources of Information", para.5a, available at https://www.pofmaoffice.gov.sg/documents/Prominence%20Code.pdf (visited on 9 March 2021).

111 Although there is an exception for good faith expression of opinions respecting the conduct and character of another touching a "public question". See, for comparison, the definition of defamation under art.246 of the Chinese Criminal Code.

112 For criticism, see Michael Hor, "Freedom of Speech and Defamation: *Jeyaretnam Joshua Benjamin v Lee Kuan Yew*" (1992) *Singapore Journal of Legal Studies* 542, 547–548.

113 1957, revised in 2014.

114 *Jeyaratnam Joshua Benjamin v Lee Kuan Yew* [1992] 1 SLR(R) 791 (CA).

115 (1986) 8 EHRR 407.

core of the concept of a democratic society which prevails throughout the Convention".[116] In *Otegi Mondragon v Spain*,[117] the Court was concerned with a false allegation by an elected representative and spokesperson for a Parliamentary group, who said at a press conference that the Spanish King "protects torture and imposes his monarchical regime on our people through torture and violence". The elected representative was convicted of insulting the Spanish King, but the conviction was held to be inconsistent with art.10. The Court held that:

> There is little scope under Article 10 for restrictions on freedom of expression in the area of political speech or debate—where freedom of expression is of the utmost importance—or in matters of public interest. While freedom of expression is important for everybody, it is especially so for an elected representative of the people.[118]

Authoritarian regimes will typically reject such an approach, because it threatens the element of political domination at the heart of such regimes in which citizens are in law and practice subject to the will of those in authority and both legally and politically insecure in so far as they seek to challenge that subjection.[119] For authoritarian regimes, control over "disinformation" is a dimension to their control over the power narratives for that jurisdiction, such as control of what political issues are on the agenda,[120] which are not to be determined through free citizen engagement.[121]

In s.1, I said that one source of justification for a coercive response to the spread of false political viewpoint information, especially at election times, is particular liberal—be it considered universal or elite—concern. These are militant democratic concerns about the declining or decaying ethical character (devaluing or attacking democracy) and ethical quality (failing to distinguish fact from fiction) of political debate. The rise of dishonest internet electioneering and its impact on—as well the participation in it, of—the public at large have re-awakened militant democratic ("establishment") fears that, as some American sociologists sought to demonstrate in the 1980s, "political elites are more tolerant than the masses and . . . the mass public is dangerously intolerant in at least some areas".[122] So, for example, in the

116 *Ibid.,* [42].

117 *Otegi Mondragon v Spain,* 15 March 2011 (Application No. 2034/07), [50]. See further, *Castells v Spain*, judgment of 23 April 1992, Series A No 236.

118 *Ibid*..

119 See further, European Audiovisual Observatory, *Disinformation in the Media under Russian Law* (2019), available at https://rm.coe.int/disinformation-in-the-media-under-russian-law/1680967369 (visited on 9 March 2020).

120 See Steven Lukes, *Power: A Radical View* (London: MacMillan, 1974).

121 See GA Tóth, "Authoritarianism" (n.3). See, eg, the new law governing disinformation in Russia: Federal Law of March 18, 2019 No 30-ФЗ, "On Amending the Federal Law: On Information, Information Technologies and the Protection of Information".

122 See J Hochschild, "Dimensions of Liberal Self-satisfaction" (n.7), 386–387, summarising the main claim of Herbert McClosky and Alida Brill, *Dimensions of Tolerance: What Americans Believe about Civil Liberties* (New York: Russell Sage Foundation, 1983) p 1.

United Kingdom, the Coalition for the Reform of Political Advertising has called for the introduction of a requirement that what they call, "objective factual claims" in political advertisements must be substantiated and that a body (possibly the Electoral Commission, or perhaps a new body) should have the power to ensure compliance with this requirement.[123] It is perhaps telling, though, that one of the bodies the Coalition suggests might take on such a role—the UK Advertising Standards Authority—has in the past specifically rejected it. In 1999, the then Chair of the Authority said: "The free flow of argument in the cut and thrust of open debate is the best antidote to political advertising that misleads or offends".[124] Nonetheless, in some jurisdictions, a militant democratic path to coercion has proved attractive.

VI. Authoritarian Liberalism: Germany

In Germany, the Netzwerkdurchsetzungsgesetz (NetzDG) 2017, the Network Enforcement Act 2017, has sought to establish an uneasy compromise between a commitment to free speech and an authoritarian drive to suppress disinformation.[125] Freedom of expression in Germany is protected by art.5(1) of the Basic Law:

> Everyone shall have the right freely to express and disseminate his opinion by speech, writing, and pictures and freely to inform himself from generally accessible sources. Freedom of the press and freedom of reporting by means of broadcasts and films are guaranteed. There shall be no censorship.

However, in interpreting art.5(1), it has been held that the propagation of "incorrect information" is not an interest that merits protection,[126] although the threat of sanctions must not be used in such a way as to chill freedom of expression.[127] In that regard, the 2017 Act builds on an implied permission to make network platforms responsible for controlling content in certain circumstances—where there is knowledge of the content and an absence of diligence in responding appropriately to that knowledge—first created by s.10 of the Telemedia Act (Germany) 2007:

123 "Blog Posts about the Coalition for Reform in Political Advertising" (The Coalition for Reform in Political Advertising), available at https://reformpoliticaladvertising.org/blog/ (visited on 9 March 2021). For a similar suggestion, at an official level, see *Digital Technology and the Resurrection of Trust* (Report of Session 2019–2021, HL Paper 77, 2020).

124 Advertising Standards Authority, Annual Report (1999), 2, available at https://www.asa.org.uk/asset/40D37406-BDD0-48A4-BF46C68DE81C1E39/ (visited on 9 March 2021).

125 See, eg, Matthias Kettermann, "Follow-up to the Comparative Study on Blocking, Filtering and Takedown of Illegal Internet Content" (2019), available at https://rm.coe.int/dgi-2019-update-chapter-germany-study-on-blocking-and-filtering/168097ac51 (visited on 9 March 2021).

126 BVerfGE 85, 1 (1991) (Critical shareholders' case), discussed in I Katsirea, "Fake News" (n.22), 179.

127 BVerfGE 54, 208 (*Böll*), discussed in I Katsirea, "Fake News" (n.22), 179.

> Service providers shall not be responsible for the information of third parties which they store for a recipient of a service, as long as 1. they have no knowledge of the illegal activity or the information and, as regards claims for damages, are not aware of any facts or circumstances from which the illegal activity or the information is apparent, or 2. upon obtaining such knowledge, have acted expeditiously to remove the information or to disable access to it.[128]

The 2017 Act imposes on major social networks (those with over 2 million registered users in Germany)—other than those providing essentially "journalistic or editorial" content[129]—an obligation to remove (ie delete or block) a range of prohibited content. Knowledge of such content (triggering the applicability of the provisions) is liable to come from users' complaints, respecting which s.3 of the 2017 Act requires that networks provide "effective and transparent procedure for handling complaints about unlawful content". In the case of "clearly violating" content, it must be removed within 24 hours and in the case of "violating" content, within seven days. However, there is no indication of how the line between these categories is to be drawn and no defence that the network platform was acting in good faith in drawing it in a particular place, subsequently found to be inappropriate. Having said that, a failure to comply in an individual case will not lead to a fine. Instead, systematic and persistent management failures that have permitted violating content to appear may lead to an administrative penalty of up to 50 million euros (in the case of corporate bodies and legal persons).[130] Under s.3(3) of the 2017 Act, a network platform can itself avoid the seven-day deadline to act, if (1) there needs to be an investigation into the truth or falsity of a claim, or (2) a procedure has been put in place whereby an independent self-regulatory body, accredited by the Ministry of Justice, has been established by a platform to make a binding decision on whether to delete or block content. Facebook, which had not at that point set up an independent body to meet the regulatory challenge, was fined 2 million euros under the 2017 Act for providing incomplete transparency reports.[131]

---

128 See the discussion in Article 19, *Germany: The Act to Improve Enforcement of the Law in Social Networks* (2017), available at https://www.article19.org/wp-content/uploads/2017/09/170901-Legal-Analysis-German-NetzDG-Act.pdf (visited on 9 March 2021); and in Stefan Theil, "The Online Harms White Paper: Comparing the UK and German Approaches to Regulation" (2019) 11 *Journal of Media Law* 41, 46. On the inspiration of militant democratic thinking behind the Act, see H Tworek, "An Analysis" (n.11), 3.

129 These terms are not defined in the 2017 Act. For criticism, see Article 19, *Germany* (n.128), 12–13.

130 See Network Enforcement Act: Regulatory Fining Guidelines (2018), available at https://www.bundesjustizamt.de/DE/SharedDocs/Publikationen/NetzDG/Leitlinien_Geldbussen_en.pdf?__blob=publicationFile&v=2 (visited on 9 March 2021); S Theil, "The Online Harms" (n.128), 48. See further, the discussion of the approach in England and Wales in VIII. Conclusion.

131 S Theil, "The Online Harms" (n.128), 48; Ben Wagner *et al*., "Regulating Transparency? Facebook, Twitter and the German Network Enforcement Act" (2020), available at https://www.researchgate.net/publication/338802975_Regulating_Transparency_Facebook_Twitter_and_the_German_Network_Enforcement_Act/link/5e2b217292851c3aadd7b7f5/download (visited on 9 March 2021). Facebook

How effective might such a self-regulatory body (or a platform acting on its own behalf) be in drawing up a clear and robust human rights jurisprudence relating to content removal? The difficulty here is not so much the nature of the procedure just described but the controversial nature of the content covered. "Violating" content includes content amounting to or involving the incitement of offences under some 22 statutes in the German Criminal Code. Included are such offences as hate crime, child pornography and other content harmful in analogous ways;[132] but "violating" content under the 2017 Act goes further. It covers content amounting to such offences under the German Criminal Code as criminal insult (art.185), malicious gossip (art.186), criminal defamation (art.187) and treasonous forgery (art.100a).[133] These are controversial extensions. For example as Stefan Theil observes:

> [O]ne might attract criminal liability for defamation when describing a specific abortion doctor's work as "babycaust" even though this does not fall under most definitions of hate speech: it is not based on attributes such as race, religion, ethnic origin, sexual orientation, disability, or gender.[134]

The 2017 Act expects network platforms to implement political censorship policies in relation to such content and more broadly in relation to supposedly false political claims falling under the headings of criminal insult, gossip or defamation.[135] In spite of the inherently controversial character of such censorship, there is to be no prior *official* determination of whether the content in question in fact contravenes one of the 22 statutes to which the 2017 Act applies. So, for example, network platforms will themselves be required to give consideration, in appropriate cases, to whether the person posting allegedly violating content possessed the relevant fault element for the offence and—in a case where the complaint is about content containing an allegedly false claim—to whether there is an appropriate factual basis to the content.[136] Faced with the prospect of sanctions for under-blocking, the

---

has now set up an independent Oversight Board to which complaints about content can be taken if the matter is not resolved satisfactorily internally. Facebook will regard itself as bound by the Board's decisions. The Oversight Board can request that its decisions in individual cases are applied to similar cases and may also issue policy recommendations to Facebook, available at https://about.fb.com/news/2019/09/oversight-board-structure/ (visited on 9 March 2021).

132 Studies show that IT companies remove 70 per cent of all hate speech notified to them within 24 hours: M Kettermann, "Follow-up to the Comparative Study" (n.125), 2.

133 Although criminal defamation of the President of the Federation (Section 90) and criminal defamation of the state and its symbols (s.90a) were excluded.

134 S Theil, "The Online Harms" (n.128), 44. See *Hoffer and Annen v Germany*, 397/07; 2322/07, Court, 13 January 2011.

135 The European Court regards the development of such measures as consistent with art.10: *Delphi v Estonia,* App No 64569/09 (Judgment June 16, 2015).

136 Article 19, *Germany* (n.128), 16 and 20. In the case of an allegedly false claim, s.3(2)(3)(a) of the 2017 Act requires the person who posted the content to respond to the complaint before the network platform takes its decision.

rational network platform must inevitably incline towards overblocking, impinging on art.10 rights.[137]

Platforms are now the modern equivalent of the "political police" envisaged in the late 1930s by Karl Loewenstein (who coined the term "militant democracy" and defended it) whose task it was to be to defend democratic ideals in Germany, "even at the risk and cost of violating fundamental principles".[138] Inevitably, this kind of indirectly enforced self-censorship will lead to the taking down of what is merely challenging or provocative political content,[139] even though s.3(1) provides that social networks with over 2 million registered users must provide a complaints procedure for those whose content is taken down (as well as for those who wish to complain about content).[140] Under the 2017 Act, not only is there is no prior determination by a court of what counts as "violating" content, but it is also left to network platforms to decide how to balance the right to free speech against the risk that content will be found to be violating.[141] As Ian Cram observes, for commercial platforms, the priority for shareholders will inevitably be the avoidance of legal penalties.[142] In that respect, an additional concern about how that balance is likely to be struck is that, under the 2017 Act, there is no penalty for taking down too much *inoffensive* content. As the House of Lords Select Committee on Democracy and Technology has observed:

> The evidence we have received suggests that ensuring free expression online is not the priority for content moderation processes . . . the main concern behind platforms' content moderation activities was brand management . . . [and] . . . moderating inexpensively rather than accurately and that this is reflected in the working conditions of some of the human moderators.[143]

137 Wolfgang Schultz, "Regulating Intermediaries to Protect Privacy Online—The Case of the German NetzDG" (2018), available at https://www.hiig.de/publication/regulating-intermediaries-to-protect-privacy-online-the-case-of-the-german-netzdg/ (visited on 9 March 2021).

138 Karl Loewenstein, "Militant Democracy and Fundamental Rights II" (1937) 31 *American Political Science Review* 638, 656–657, cited by JW Muller, "Militant Democracy" (n.4), 1120.

139 See H Tworek, "An Analysis" (n.11), 3. However, for a contrary view, see S Theil, "The Online Harms" (n.128), 49.

140 H Tworek, "An Analysis" (n.11). A social network receiving more than 100 complaints in a year must publish its internal moderation practices. For Facebook's development of its complaints' procedure, see "Establishing Structure and Governance for an Independent Oversight Board" (Facebook, 17 September 2019), available at https://about.fb.com/news/2019/09/oversight-board-structure/ (visited on 9 March 2021).

141 H Tworek, "An Analysis" (n.11), 3: "The NetzDG process does not require a court order prior to content takedowns nor does it provide a clear appeals mechanism for victims to seek independent redress". See also Article 19, *Germany* (n.128), 15.

142 See I Cram, "Keeping the Demos" (n.11), 125.

143 HL Select Committee on Democracy and Digital Technologies, *Digital Technology and the Resurrection of Trust* (HL Paper 77, 16 June 2020) paras.114–115.

In 2018, Twitter received 264,818 complaints in Germany regarding content, blocking 10 per cent of content in consequence; YouTube received around 215,000 complaints, blocking 27 per cent of content in consequence and Facebook got 1,704 complaints, leading to 20 per cent of content being blocked.[144] This raises questions about the criteria used in each instance as a justification for taking the decision to remove content.[145] In its disaggregation of take-down statistics, Google indicated that 8695 complaints involving defamation or insult led to the content taken down because of breach of Google's community guidelines, and 3206 complaints led to defamatory or insulting content being taken down because the content was adjudged to breach the 2017 Act (but not the guidelines).[146] Theil's research reveals that in 2018–2019, Facebook logged 500 NetzDG reports, with deletion of content following in 159 cases (roughly 33.8 per cent of cases), and Twitter logged 256,462 reports of which 23,165 led to the deletion of content, roughly 9 per cent of cases.[147] Defamation and insult have been, along with hate speech, the most common kind of complaint, with Google receiving 45,190 complaints about defamation and insult in all compared to, for example, 27,308 complaints about sexual content.[148] Nonetheless, Theil argues that: "the vast majority of reports do not result in deletion of content. This is problematic for critics of online regulation as a cornerstone of their argument relies on demonstrating that overblocking is more than a theoretical possibility".[149]

In his view, sophisticated systems of online regulation can not only address online harms, but may also, "have civilising influence on online expression instead of foreshadowing the end of a free and open discourse".[150] The risk is that such a pious hope does not do full justice to the arguments against militant democratic justifications for employing coercive pressure, in cases specifically involving false political viewpoint content.[151] In that regard, as I will argue below when analysing the proposals for the United Kingdom, there is a need for a "tiered" approach that tailors legal-regulatory steps to the nature of the harms or risks in issue. To put it in terms of a simple question: is there really a justification for subjecting network

144 M Kettermann, "Follow-up to the Comparative Study" (n.125), 3. However, most of these complaints involved a breach of the platforms' community guidelines rather than the 2017 Act: H Tworek, "An Analysis" (n.11), 5.

145 William Echikson and Olivia Knodt, "Germany's NetzDG: A Key Test for Combatting Online Hate" (CEPS Policy Insight, 2018), available at https://ssrn.com/abstract=3300636 (visited on 9 March 2021); H Tworek, "An Analysis" (n.11), 5.

146 Tworek describes the 2017 Act, in consequence, as a "community guidelines enforcement law": H Tworek, "An Analysis" (n.11), 7. Insult-related offences under the 2017 Act were the most common kind of offences complained about to Facebook (460 complaints), with Twitter reporting 75,925 insult-related complaints. By 2019, complaints had dropped in overall numbers by around 15 per cent.

147 S Theil, "The Online Harms" (n.128), 50.

148 H Tworek, "An Analysis" (n.11), 5–6.

149 S Theil, "The Online Harms" (n.128), 50.

150 *Ibid*., 50.

151 See also I Cram, "Keeping the Demos" (n.11), 138, discussing the potential for suppression of controversial political speech in online platforms' in-house or "community" guidelines themselves.

platforms to the same governance structures and risks of sanctions in respect of the toleration of false or misleading political viewpoint speech, as in the case of toleration of child pornography or incitements to engage terrorist violence? The answer we give is an important one, because the figures show that the threat of sanctions is playing a role in Germany, albeit a minor one, in influencing major social media platforms in the way they seek to disassociate themselves from the dissemination of false political information.[152] So, the concern remains that provocative and challenging political content is being unwittingly suppressed, along with genuinely harmful content.[153]

Moreover, making public the use of the relevant provisions to suppress speech in controversial cases may have the opposite effect to that intended: usage may highlight the content to a wider audience ("to refute is also to reiterate"[154]) and, ironically, enhance its credibility among those disposed to support the speaker.[155] It is significant, in that regard, that there have been several high-profile deletions of content posted by German politicians.[156] With this problem in mind, France's Institute for Strategic Research (Ministry for the Armed Forces) and Policy Planning Staff (Ministry for Europe and Foreign Affairs)[157] counselled against Government heavy handedness on the grounds that: "As one of the roots of the problem is distrust of elites, any "top down" approach is inherently limited. It is preferable to champion horizontal, collaborative approaches, relying on the participation of civil society".[158] Accordingly, only 1 of their 50 recommendations (recommendation 12) is concerned with the creation of prohibitions on the dissemination of false information.[159]

VII. A New Dawn? The Proposals for England and Wales

One danger in an authoritarian–liberal response to digitally propagated disinformation is that it becomes, in effect, almost indistinguishable in form (even if intended to operate differently in practice) from the response of the more fully

152 Although, as Heidi Tworek points out, it is unclear to what extent a take-down leads to suppression of the false information, insult or other offending content or simply brings greater attention to it through its publication elsewhere: H Tworek, "An Analysis" (n.11), 7.

153 Article 19, *Germany* (n.130). See also I Cram, "Keeping the Demos" (n.11), 140, "by excluding all but the rational, consensus-seeking, civilised speaker, deliberative accounts of political pluralism impede the creative power of the demos. In consequence, very little consecratory discourse is let in".

154 JBJ Vilmer *et al*., *Information Manipulation* (n.21), 169.

155 See H Tworek, "An Analysis" (n.11), 4: "A few days after NetzDG came into force in January 2018, prominent AfD (Alternative für Deutschland) politician Beatrix von Storch saw one of her social media posts removed from Twitter and Facebook under the law. Widespread media coverage of this incident, including the post's content and its potential illegality, seemed to confirm fears of the Streisand effect, or what one journalist dubbed the 'Storch effect'".

156 See S Theil, "The Online Harms" (n.128), 42.

157 See JBJ Vilmer *et al*., *Information Manipulation* (n.21).

158 *Ibid*., 172.

159 *Ibid*., 173.

authoritarian regimes whose "fake news" practice liberal regimes are, ironically, seeking to counter.[160] There is another danger. By giving government agencies, or even courts, the power to prohibit and punish the dissemination of political viewpoint disinformation, states will diminish what ought to be a "republican" commitment to nurture citizens who, rather than being cowed by or automatically deferential to authority, "can speak their minds, walk tall among their fellows, and look others squarely in the eye. They can command respect from those with whom they deal, not being subject to their arbitrary interference"[161] In that regard, it is noteworthy that significantly more people world-wide support the idea that journalists (75 per cent) and online platforms (71 per cent) should take the responsibility for weeding out disinformation over which they have influence than support the notion that this is the responsibility of Government (61 per cent).[162] In this respect, how do UK proposals for regulation on online speech fare?

In its 2019 White Paper on online harms,[163] the UK government sought to carve out a middle way between, on the one hand, a libertarian guarantee of free speech, and on the other hand, an authoritarian system permitting the suppression (*ex post facto*) of substantive content deemed by a government regulator or court to be false or misleading.[164] Under the proposed UK approach, large and small social media companies will be bound by a statutory duty of care to protect users from a range of harms and to minimise the commission or encouragement of such harms by doing what is "reasonably practicable" towards those ends.[165] However, only the companies with the largest reach and enabling the widest range of material to be uploaded (dubbed "Category 1" service providers) will have a duty of care that covers legal-but-potentially-harmful material, such as disinformation.[166]

The discharge of this duty of care, in accordance with a code of practice, will be overseen by a regulator, with powers (among other things) to fine those deemed to be in breach.[167] However, in a case where unlawful or otherwise inappropriate content has appeared on a platform, the regulator's powers—for example to impose fines—will bear not on the nature of the content as such, but (broadly speaking, as in Germany) on whether there was a breach of the duty of care in the way that the

160 See European Audiovisual Observatory, *Disinformation* (n.119).

161 José Luis Martí and Philip Pettit, *A Political Philosophy in Public Life* (Princeton University Press 2010) p 38.

162 Reuters Institute, *Digital News Report 2018*, 19, available at https://reutersinstitute.politics.ox.ac.uk/sites/default/files/digital-news-report-2018.pdf (visited on 9 March 2021).

163 See HM Government, *Online Harms* (n.65).

164 For a penetrating critique, see Victoria Nash, "Revise and Resubmit? Reviewing the 2019 Online Harms White Paper" (2019) 11 *Journal of Media Law* 18.

165 HM Government, *Online Harms* (n.65), para.35. Relevant harms include hate speech, encouragement to engage in terrorism, harassment, offences against children, children accessing inappropriate content and encouragement to commit suicide or self-harm.

166 Department for Digital, Culture, Media and Sport, and Home Office, *Online Harms* (n.65).

167 HM Government, *Online Harms* (n.65), paras.16–28. The Government has also not ruled out introducing criminal liability for individual directors of firms that are in serious breach.

content came to be disseminated.[168] As Victoria Nash suggests, the White Paper can be understood as a "manifesto for government-led platform governance":[169] soft regulation. Regulation is to be soft, in that it will, for example, be up to companies to decide exactly how they fulfil their obligations under the duty of care, a decision that can even include whether to fulfil the obligations by adopting means other than following the Code of Practice.[170] An important distinction from the approach in Germany is that the possibility of over-blocking or inappropriate take-down is directly addressed. Companies will be obliged to take reasonable steps to monitor and evaluate the effectiveness of their systems in preventing the incorrect removal of legitimate content, and the Regulator will be expected (in what is, in any event, normal practice) to employ an escalating system of enforcement measures "to avoid incentivising content takedown, with judicial oversight to safeguard the most severe sanctions like blocking".[171]

What is the approach specifically to disinformation? An initial aspect of the White Paper approach concerns the presentation of "news" on platforms:

> Companies will need to take proportionate and proactive measures to help users understand the nature and reliability of the information they are receiving, to minimise the spread of misleading and harmful disinformation and to increase the accessibility of trustworthy and varied news content.[172]

The source-based aim described in this passage is an important one. There is, in effect, likely to be a "news quality obligation" imposed on companies to ensure that they "improve how their users understand the origin of a news article and the trustworthiness of its source" (a proposal derived from the Cairncross Review of Journalism[173]).[174] The White Paper approach is, in this respect, quite similar to the Code of Practice promulgated with the same end in view in Singapore (as we saw in s.5). The idea appears to be that the platform companies must, first, step away from

168 HM Government, *Online Harms* (n.65), para.19.

169 V Nash, "Revise and Resubmit?" (n.164), 23.

170 HM Government, *Online Harms* (n.65), para.20. For the view that even soft State regulation is liable to go too far in impeding freedom of speech, and a system of guidance provided by, eg, Social Media Councils is to be preferred, see art.19, *Social Media Councils: Consultation* (2019), available at https://www.article19.org/resources/social-media-councils-consultation/ (visited on 9 March 2021).

171 Department for Digital, Culture, Media and Sport, and Home Office, *Online Harms* (n.65), para.2.25 and Annex A.

172 HM Government, *Online Harms* (n.65), para.7.27. In its 2020 response to consultation on the White Paper, the Government speaks of the need to provide "support for platforms to help users better identify the reliability and trustworthiness of news sources and the introduction of a 'news quality obligation' on platforms": Department for Digital, Culture, Media and Sport, and Home Office, *Online Harms* (n.65), para.6.3 (Box 22).

173 *The Cairncross Review: A Sustainable Future for Journalism* (2019), available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/779882/021919_DCMS_Cairncross_Review_.pdf (visited on 9 March 2021).

174 HM Government, *Online Harms* (n.65), para.7.29.

political neutrality by making judgments about the reliability of political viewpoint claims (along with the reliability of other material that is or may be harmful, such as fake health advice) and, second, ensure that free speech entails "free reach" only in relation to trustworthy news content. In other words, algorithms that determine who sees what must minimise the spread of "untrustworthy" news content, rather than being driven solely by the needs of advertisers to reach more individuals (as a result of which, there may be an incentive to encourage the spread of sensationalist claims).[175]

A number of questions arise in relation to this strategy, so far as political viewpoint content is concerned. Must the maintenance of "varied" news sources involve internet platforms in actively balancing access granted to differing political viewpoints, and, if so, does that entail—say—treating white supremacist and inclusive content as equal in terms of the coverage that they are to receive? Balancing might be considered a curious obligation to place on platform companies. Why, for example, cannot they openly declare themselves to be especially welcoming or even exclusively concerned with a narrow range of political viewpoint content? Contrariwise, an internet site hosted on a platform might act purely as a news aggregator, tagging all stories with a word or phrase in them and leaving the reader to sort the wheat from the chaff. Must internet platforms really concern themselves with the extent to which such sites conform to notions of "trustworthiness", even if one can sufficiently clearly articulate what "trustworthiness" is to involve, when determining whether to grant access? This approach leaves a number of matters unclear. What will new entrants to news reporting have to do to prove their trustworthiness, and what will outlets deemed untrustworthy have to do to show that they are now trustworthy? There will be a need for consistency of judgments in relation to trustworthiness, reliability and variety, but that will also have to be weighed against the desire to accommodate changing perceptions of these matters. It is noticeable that in proposing to oblige platforms to take "proportionate and proactive measures" to understand the sources and reliability of news, the White Paper is gold plating. It goes beyond the recommendations of the Cairncross Review, which envisaged simply an obligation on platforms to report on the steps they have themselves been taking to help readers understand the source and trustworthiness of news.[176]

So far as content-based measures are concerned, the White Paper was criticised by both Victoria Nash[177] and Damian Tambini[178] for failing to give an unequivocal and clear commitment to distinguish, in terms of the appropriate regulatory response, between illegal harms and risks and legal-but-unwanted harms and risks (such as the risks posed by many kinds of disinformation). In fairness to the White Paper, it gave a commitment—vague though it might have been—that "the

175 HL Select Committee on Democracy and Digital Technologies, *Digital Technology* (n.143), para.100.
176 *Ibid*.
177 V Nash, "Revise and Resubmit?" (n.164).
178 Damian Tambini, "The Differentiated Duty of Care: A Response to the Online Harms White Paper" (2019) 11 *Journal of Media Law* 28.

expectations placed on firms, as set out in codes of practice, will vary according to the category of harm";[179] and in its response to consultation on the White Paper, the Government spells this commitment out in a little more detail. First, the key obligations binding on platforms, irrespective of size or reach, are to block or remove content when it involves illegality, particularly the most serious kinds of illegality such as child abuse and terrorist material.[180] In that regard, it is worth noting that in a single month during a lockdown in 2020, the Internet Watch Foundation and its partners blocked at least 8.8 million attempts by UK internet users to access videos and images of children suffering sexual abuse.[181]

Obligations to block or remove content that is legal-but-potentially-harmful are (as indicated above) imposed only on larger platforms with fewer restrictions on the nature of content that can be uploaded. Such platforms will be under an obligation:

> to take proportionate and proactive measures to help users understand the nature and reliability of the information they are receiving, to minimise the spread of misleading and harmful disinformation and to increase the accessibility of trustworthy and varied news content.[182]

It is troubling that "more speech" obligations, such as the obligation to assist users with critical engagement, are simply bundled together in this passage with obligations to restrict the spread of disinformation. For example the response goes on to suggest that platforms must address the issue of "making content which has been disputed by reputable fact-checking services less visible to users";[183] but the mere fact that there has been such a challenge from a fact-checker cannot by itself justify a restriction. From a free-speech perspective, it ought to have been made clearer that restrictions must be clearly justified by a risk of significant harm, whereas political viewpoint disinformation, in and of itself, is best met by a "more speech" solution. Having said that, the White Paper consultation response does at least give the appearance of linking disinformation with significant harm. It says that "this type of activity can range from online bullying and abuse, to advocacy of self-harm, to spreading disinformation and misinformation"[184] and then goes on to say that platforms will have to deal with "other types of harm that spread online—from dangerous misinformation spreading lies about vaccines to destructive pro-anorexia content".[185] However, in the final analysis, it will be left to sec-

179 HM Government, *Online Harms* (n.65), para.6.6.
180 Department for Digital, Culture, Media and Sport, and Home Office, *Online Harms* (n.65), paras.8–9.
181 Internet Watch Foundation, "Millions of Attempts to Access Child Sexual Abuse Online during Lockdown" (2020), available at https://www.iwf.org.uk/news/millions-of-attempts-to-access-child-sexual-abuse-online-during-lockdown (visited on 9 March 2021).
182 Department for Digital, Culture, Media and Sport, and Home Office, *Online Harms* (n.65), para.7.27.
183 *Ibid*., para.7.28.
184 *Ibid*., para.10.
185 *Ibid*., foreword.

ondary legislation to determine what kinds of disinformation and misinformation must be addressed by platforms in their terms and conditions.[186]

In analysing the White Paper, Tambini suggested that "The extension of pseudo-liability for harmful but not illegal content is problematic and, together with harsh sanctions could lead to significant chilling of freedom of expression".[187] By contrast, Tambini suggests that the proportionate regulatory response to legal-but-unwanted harm ought to be not the threat of sanctions but "monitoring, advice and transparency", with a focus on "consumer information, competition and switching".[188] Has the response to consultation allayed these concerns? The response speaks of identifying "priority harm" that platforms must address in their terms and conditions,[189] and we have just seen that the examples given are the ones in which the concern is either risks of physical harm or public disorder or disinformation liable to spark or entrench racism and hate.[190] A contemporary example given, in relation to the latter, is of a false suggestion that certain groups were not following social distancing rules during the pandemic.[191]

What is disappointing, though, is the failure to mention the approach that is to be taken to political kinds of disinformation or the failure of any special measures that are appropriate at election times, beyond a single, unanalysed suggestion that platforms ought to use fact-checking services, "particularly during election periods".[192] There should, for example, have been some mention of Twitter's elections-specific support portal, which is designed to provide faster responses to alleged abuses.[193] Electoral participation disinformation ought to be a "priority harm" and is so treated by some major platforms. During the 2018 US midterm elections, Twitter took enforcement action under its civic integrity guidelines on nearly 6,000 Tweets identified as voter suppression attempts (many of which originated in the United States itself). Facebook's policy on the removal of content includes the view that "there are some extreme forms of disinformation that we remove from our platform, including disinformation that leads to offline violence and disinformation that leads to voter suppression".[194] It is absolutely correct to make this equation between hate speech and electoral participation disinformation,

186 *Ibid*., foreword.
187 D Tambini, "The Differentiated Duty" (n.178).
188 *Ibid*., 29.
189 Department for Digital, Culture, Media and Sport, and Home Office, *Online Harms* (n.65), para.2.83.
190 *Ibid*., foreword.
191 *Ibid*., 2.77; Nazia Parveen, "Police Investigate UK Far-right Groups over Anti-Muslim Coronavirus Claims" *The Guardian* (5 April 2020), available at https://www.theguardian.com/world/2020/apr/05/police-investigate-uk-far-right-groups-over-anti-muslim-coronavirus-claims (visited on 9 March 2021).
192 Department for Digital, Culture, Media and Sport, and Home Office, *Online Harms* (n.65), para.7.28.
193 D Harvey and Y Roth, "An update" (n.66).
194 "Expanding Our Policies on Voter Suppression" (Facebook, 15 October 2018), available at https://about.fb.com/news/2018/10/voter-suppression-policies/ (visited on 9 March 2021); "Annual Self-assessment Reports of Signatories to the Code of Practice on Disinformation 2019" (European Commission, 29 October 2019), available at https://ec.europa.eu/digital-single-market/en/news/annual-self-assessment-reports-signatories-code-practice-disinformation-2019 (visited on 9 March 2021).

and it is proportionate to regard take-down (or at the very least, severe restriction) as the appropriate remedy.[195]

VIII. Conclusion

A robust, liberal-republican free speech culture is a culture that avoids militant democratic tendencies in its response to the problem of false or misleading claims regarding viewpoint politics. Online platforms are not themselves traditional broadcasters and so, in relation to political content, they cannot reasonably be expected to ensure that such content and coverage is accurate and balanced overall.[196] In a culture committed to free speech, the key question is how to devise a modern interpretation of Justice Brandeis "more speech" solution,[197] a solution encouraging critical user engagement with disputed content, and put it at the forefront of any response to political viewpoint disinformation online.[198] Answers to that question are not in short supply, as we have seen,[199] but few states have fully embraced them, being unwilling to face the fact that "any purely governmental response . . . is bound to be regarded as biased and propagandist".[200] A deterrent and punitive approach to false or misleading political viewpoint content to the use of bots or to the maintenance of anonymity should be discarded.[201]

The UK White Paper proposed ensuring that platforms adopt the following steps (among others) to promote a flourishing free-speech culture online:[202]

(1) making content which has been disputed by reputable fact-checking services less visible to users;
(2) using fact-checking services, particularly during election periods;
(3) promoting authoritative news sources;
(4) promoting diverse news content;
(5) ensuring that it is clear to users when they are dealing with automated accounts; and
(6) [Having] reporting processes which companies should put in place to ensure that users can easily flag content that they suspect or know to be false, and which enable users to understand what actions have been taken and why.

195 Indeed, as Facebook itself points out, some hate speech is directed at voter suppression, such as a claim that someone of a particular race or religion is unfit to vote or hold office: *ibid*.

196 See, by contrast, the obligations enforced on broadcasters by the UK regulator, Ofcom: *The Ofcom Broadcasting Code* (Ofcom, 2009), available at https://www.ofcom.org.uk/__data/assets/pdf_file/0027/19287/bcode09.pdf (visited on 9 March 2021).

197 See EU High Representative of the Union for Foreign Affairs and Security Policy, *Action Plan* (n.90).

198 Reserving deterrence, suppression and enforced transparency to cases where a specified and limited range of harms are threatened (while accepting there will always be debate about these).

199 Text following (n.89).

200 See JBJ Vilmer *et al*., *Information Manipulation* (n.21), 170.

201 *Ibid*., 170.

202 HM Government, *Online Harms* (n.65), para.7.28.

Whatever the virtues of this approach in relation to online *harms* such as hate speech, electoral participation information or false and dangerous information about health,[203] step (1) in particular is not appropriate in the case of political viewpoint disinformation. It is not the role of the state—or its agents such as electoral commissions, internet ombudsmen and the like—to seek to constrain the spread of wrongheaded political viewpoints or to coerce private sector agents such as internet platforms into engaging in such constraint. Instead, political activists and sympathetic private organisations must strive to create and disseminate powerful counter-narratives to what they see as bad politics:

> Instead of trying the design the institutions which, through supposedly "impartial" procedures would reconcile all conflicting interests and values, the task for democratic theorists and politicians should be to envisage the creation of a vibrant "agonistic" public sphere of contestation where different hegemonic political projects can be confronted.[204]

203 *Ibid*., para.7.25.

204 Chantal Mouffe, *On the Political* (London: Routledge, 2005) p 3.