I will consider the approach taken in a number of different jurisdictions to the use of restrictions, including criminal law prohibitions, to prevent, deter or suppress the making of false online claims designed to influence the course of political events, especially (but not solely) elections. I will explain, first, how a commitment to freedom of expression ought to make state measures to deter the dissemination online of what I call false political 'viewpoint' information seem problematic. Secondly, I will show how a number of jurisdictions, including some liberal democracies, have wrongly been drawn into an authoritarian response to the spread of such false information: a response that tries to set the terms on which people form opinions and seek to make contributions to politics. Coercion may be employed (i) to secure equality conditions at the polls, when these are threatened by the dissemination of false 'participation' information, and (ii) to a limited extent, to support transparency about authorship; but coercion should not be used to deter the propagation of political viewpoint content, simply on the grounds of its falsity.

.

## 1. False Political Claims and the Ethics of Criminalisation

Many jurisdictions around the world use the criminal law, alongside a variety of regulatory and preventative or other 'discouraging' measures, to deter and prevent the publication or dissemination of false or misleading political information, especially when such 'information' is designed to influence elections or referendums.[1] Such (false) political information is of two basic kinds. First, there is information concerning the equality conditions under which democratic elections ought to be conducted: 'participation' information. This is information, for example, concerning where, when and how to vote. Secondly, there is 'viewpoint' information, with which I will be principally concerned. This is information about substantive political matters relating to candidates and policies ('Candidate X was

[1] Daniel Funke and Daniela Flamini, *A Guide to Anti-Misinformation Actions Around the World* (2018), https://www.poynter.org/ifcn/anti-misinformation-actions/.

found guilty of bribery'; 'Party Y's policies have been endorsed by the Pope'). For reasons explored in sections 4 and 5 below, state action to deter or prevent the dissemination of the latter kind of false 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 recently responded to (perceived) new threats from online activity to the integrity of 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.[2] However, we will see that in liberal democracies, if a 'militant democratic' approach to the dissemination of false political information is adopted,[3] such an approach can also justify a coercive response to the dissemination of viewpoint information purely on the grounds of its falsity, albeit a response narrower in scope than one characteristic of an authoritarian regime.

[2] Gábor Attila Tóth, 'Authoritarianism' (2017), https://oxcon.ouplaw.com/view/10.1093/law-mpeccol/law-mpeccol-e205, para 34.

[3] See Jan-Werner Muller, 'Militant Democracy', in Michel Rosenfeld and András Sajó (eds), *The Oxford Handbook of Comparative Constitutional Law* (Oxford: Oxford University Press, 2012), ch 59.

In modern liberal-democratic states, the rise of authoritarian-populist forms of criminalisation commonly reflects a perception amongst 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.[4] These liberal values are thought to have lost legitimacy, because they are considered not so much universal as 'elite' values, with implications for their role and influence in public life:[5]

> 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'.[6]

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.[7] By ironic contrast, though, in shaping the state's response to political misinformation, 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.[8]

---

[4] See e.e. Tim Newburn, '"Tough on Crime": Penal Policy in England and Wales' (2007) 36 Crime and Justice 425.

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

[6] Jennifer Hochschild, 'Dimensions of Liberal Self-Satisfaction: Civil Liberties, Liberal Theory, and Elite Mass Differences' (1986) 96 Ethics 386, 399.

[7] See e.g. 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.

[8] See Jan-Werner Muller, n. 3 above.

'Militant' democratic politics, in a narrow sense, involves 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.[9] 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 necessity and proportionality, such a commitment may appear to justify authoritarian restrictions on rights such as the right to advocate non-democratic forms of government, or to exercise free speech in other ways that appear to challenge or undermine democracy itself.[10] Key examples are at least some instances in which freedom of speech is used for the dissemination of false political information online (discussed in due course).[11] Yet, the taking of such militant democratic steps is in turn vulnerable to challenge on populist grounds, even though in general the public supports the suppression of 'criminal' content on social media networks.[12] Speaking of attempts to ensure verification of contentious claims or supposed 'fake news', Romain Badouard has expressed the view that:

[9] Jan-Werner Muller, n. 3 above, 1123-26.
[10] 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), https://www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf., 2.
[11] Jan-Werner Muller, n.3 above, at 1127.
[12] Tworek reports that 87% of German voters supported the controversial Internet Enforcement Act 2017, discussed in section below: Heidi Tworek, n. 10 above 2. However, as we will see, voters tend to support suppression of misinformation only when it is deployed by their political opponents: see text at n. below.

> 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]?[13]

What Badouard says about the political credibility of elite verification will *a fortiori* be true of elite attempts to use coercion to suppress political 'misinformation', even in the name of democracy itself. The challenges posed by populism – which in its modern form is itself, in part, a product of the online world[14] - to militant democratic attempts to deter the dissemination of such misinformation, whilst varied in form, come about in the following way. Populism typically rejects orthodox liberal ways of distinguishing between truth and fiction, or between fact and opinion or value, and in some instances seeks to politicise, for its own avowedly 'anti-establishment' ends, commitments to freedom of speech or the due process of law.[15]

This suggests that the use of coercion to suppress the foundations of the populist rejection of liberal politics would be illegitimate and self-defeating, and that a genuinely liberal, free speech approach to political misinformation needs to be nuanced. As the French Institute for Strategic Research (Ministry for the Armed

[13] Romain Badouard, *Le Désenchantement de l'Internet. Désinformation, Rumeur et Propagande (*FYP Editions, 2017), 48

[14] See Sven Engesser, Nayla Fawzi, and Anders Olof Larsson, 'Populist Online Communication: Introduction' (2017) 20 Information, Communication and Society 1279.

[15] The latter trend being aptly termed by Mudde, 'pathological normalcy', the radical interpretation of mainstream values: Cas Mudde, 'The Populist Radical Right: A Pathological Normalcy', [2008] *Willy Brandt Series of Working Papers in International Migration and Ethnic Relations 3./07*, https://muep.mau.se/bitstream/handle/2043/6127/WB%203_07%20MUEP.pdf?sequence%3D1. See also, Moyn, n. 7 above, discussing modern 'Lochnerism' (*Lochner v New York* 198 US 45 (1905)).

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.[16]

I argue that an equal right to politically unfettered free speech entails (i) protecting what I will call the equality conditions for political participation, (ii) rejecting, as a basis for state intervention, restrictions on political viewpoint content solely on the grounds that it is false or misleading, and (iii) giving only a cautious welcome to source-based justifications for making the exercise of free political speech more burdensome.

## 2. *The Pressure for Law Reform in Context*

A number of factors have conspired to generate a wide-scale legislative interest in reforming the law governing false or misleading political statements, particularly online statements, designed to shape election thinking and elections outcomes. In 1985, UK voters gave as their most important source of political information: television (63%), newspapers (29%), and radio (4%).[17] Only a tiny fraction claimed that their political and world news more broadly came from contact with other people, a position radically changed – if one includes virtual contact - since the

[16] Jean-Baptiste Jeangène Vilmer et al, *Information Manipulation: A Challenge for Our Democracies* (2018), https://www.diplomatie.gouv.fr/IMG/pdf/information_manipulation_rvb_cle838736.pdf,
[17] Ralph Negrine, *Politics and the Mass Media*, 2nd edition (London: Routledge, 1994), 2.

advent of the internet age. Correspondingly, in second half of the twentieth century, the need for reform of the law governing false political statements was perhaps kept adequately at bay (i) informally, by reliance on the perceived need to maintain tolerable working relations between politicians and journalists,[18] and (ii) more formally, by some regulation[19] – 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.[20] In the USA, for example, 22% of adults say that they use the Twitter platform. On that platform, some 6% of those on it who have public accounts on it are responsible for some 73% of tweets that mention national politics.[21] Significantly, over 55% of this group describe themselves as 'very' liberal or 'very' conservative, and 64% of the group view those with opposing political views in a very poor light.[22] The low 'knowhow' and cost barriers to entry associated with such technology have given such groups - through the use of social media platforms – communicative power without responsibility undreamt of by earlier generations of

[18] Aeron Davis, ' Journalist-Source Relations, Mediated Reflexivity and the Politics of Politics' (2009) 10 Journalism Studies 204-19.

[19] See e.g. https://www.ofcom.org.uk/tv-radio-and-on-demand/broadcast-codes/broadcast-code/section-six-elections-referendums.

[20] House of Commons, Digital, Culture, Media and Sport Committee, *Disinformation and Fake News: Final Report* (2019), https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf, para 202.

[21] https://www.pewresearch.org/fact-tank/2019/10/23/a-small-group-of-prolific-users-account-for-a-majority-of-political-tweets-sent-by-u-s-adults/

[22] Ibid. Although, in that regard, the belief that online activity exist in a so-called echo chamber in which they seek and hear only affirmation of their own opinions has been doubted: see Elizabeth Dubois and Grant Blank', 'The Myth of the Echo Chamber' (2018), https://theconversation.com/the-myth-of-the-echo-chamber-92544.

equally politically motivated but more ethically constrained journalists.[23] As 'responsible' journalism retreats behind its paywalls, social media users exploit their ability to reach out to 100s of millions of people, and to do so free of charge and – if they wish, for example, to post highly controversial, false or abusive content – anonymously.[24]

This is significant in a political context because – for example - during an election campaign there is likely to be an uptick in the generation by parties and their supporters of negative information and evaluation about their rivals' policies and personalities.[25] An election campaign is obviously the key opportunity for parties not merely to present themselves as fit for government, but also – perfectly legitimately - to explain why other parties are unfit for government.[26] Research shows that individual voters commonly give more weight to negative than to positive political information, and negative information is treated as more newsworthy by the media.[27] 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% more likely to be re-tweeted than true stories.[28]

[23] Kimberly Meltzer, 'Journalistic Concern about Uncivil Political Talk in Digital News Media: Responsibility, Credibility, and Academic Influence' (2015) 20 International Journal of Press and Politics 85-107.
[24] See, e.g., Ian Rowe, 'Civility 2.0: 'A Comparative Analysis of Incivility in Online Political Discussion' (2015) 18 Information, Communication and Society 121-38.
[25] See Martin Haselmayer, 'Negative campaigning and its consequences: a review and a look ahead' (2019) 17 French Politics 355; Samantha Bradshaw and Phillip Howard, 'The Global Organisation of Social Media Disinformation Campaigns' (2018) 71 Journal of International Affairs 23.
[26] William G Mayer, 'In Defence of Negative Campaigning' (1996) 111 Political Science Quarterly 437.
[27] Martin Haselmayer, n. 25 above, at 361 and 364; Michael Karanicolas, 'Subverting Democracy to Save Democracy: Canada's Extra-Constitutional Approaches to Battling "Fake News"' (2019) 17 Canadian Journal of Law and Technology 210, at 205.
[28] Soroush Vosoughi, Deb Roy, and Sinan Aral, 'The Spread of True and False News Online' (2018) Science 359 1146, at 1148.

So, attacks – using false, misleading or otherwise sensationalised claims - on the policies of parties and personalities of politicians, are likely to be more frequent and more systematic in the run-up to an election than at other times.[29] In that regard, so far as the generation of false information is concerned, it has been estimated that in the 2016 US Presidential election, false stories relating to the two candidates were shared at least 38 million times on Facebook, with just over half of Americans who recalled seeing these statements believing them.[30] Partly instrumental in this respect has been the use of (semi)automated bots - 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.[31] Even in liberal democracies, political parties do not hesitate to use bots, which are capable of increasing numbers of supposed followers on social media and enhance the chance that stories or hashtags they wish to highlight start trending.[32]

However, it has been the use of such techniques by foreign states or organisations that has attracted the most critical attention of liberal democratic politicians. For example, from 2014 onwards, the St Petersburg-based Internet

[29] Decisions to 'go negative' may, of course, also reflect more particular factors, such as the campaign environment, the dynamic interplay that develops between candidates during a campaign, and the nature of the issues in the campaign agenda: David F Demore, 'Candidate Strategy and the Decision to Go Negative' (2002) 55 Political Research Quarterly 669.

[30] See Hunt Allcot and Matthew Gentzkow, 'Social Media and Fake News in the 2016 Election' (2017) 31 Journal of Economic Perspectives 211.

[31] David Murray, 'Protecting Our Elections: Examining Shell Companies and Virtual Currencies as Avenues for Foreign Interference', Financial Integrity Network, June 26, 2018, https://www.judiciary.senate.gov/imo/media/doc/06-26-18%20Murray%20Testimony.pdf.

[32] Samantha Bradshaw and Philip Howard, n. 25 above, 28.

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 US citizens had some interaction, and 470 Facebook accounts that reached at least 126 million Americans.[33] 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.[34] In the dark arts of foreign 'interference' in elections and referendums, of course, it is Western countries themselves who are historically the masters, with the US 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.[35] Even so, many in the West have raised and continue to raise the bogey of foreign (Chinese?; Russian?) interference distorting opinions and outcomes in national elections,[36] not least in relation to the campaign supporting a UK exit from the European Union.[37]

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 netbots, and so on – is not so much the spread of the would-be influencer's

[33] Jean-Baptiste Jeangène Vilmer et al, n. 16 above, 85.
[34] David Wolchover and Amanda Robinson, 'Is Brexit a Russia-backed Coup?', New Law Journal, 24th January 2020.
[35] New York Times, Feb 17th 2018, 'Russian isn't the only one Meddling in Elections; we do it too'; Dov Levin, 'A Vote for Freedom? The Effects of Partisan Electoral Interventions on Regime Type' (2019) 63(4) Journal of Conflict Resolution 839.
[36] Jens David Ohlin, 'Did Russian Cyber Interference in the 2016 Election Violate International Law?' (2017) 95 Texas Law Review 1579.
[37] See e.g. House of Commons, Digital, Culture, Media and Sport Committee, n. 12 above, part 6; The Guardian, 4th November 2017, https://www.theguardian.com/politics/2017/nov/04/brexit-ministers-spy-russia-uk-brexit; Ewan McGaughey, 'The extent of Russian-backed fraud means the referendum is invalid' (2018), https://blogs.lse.ac.uk/brexit/2018/11/14/the-extent-of-russian-backed-fraud-means-the-referendum-is-invalid/; Ewan McGaughey, 'Could Brexit be Void?' (2018) 29(3) King's Law Journal 331, 334-36.

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.[38] Quite apart from the question whether any election or referendum in the West has been *decisively* influenced by foreign interference, it is clear that such interference can have a corrupting effect on democratic politics.[39] For example, research has shown that American voters take a highly partisan approach to evidence of foreign interference, being much more likely to condemn it, and to profess a loss of faith in the democratic process, when it appears helpful to their political opponents than when it helps their own party of choice.[40]

**3.Source-based and Content-based Restrictions, and 'Foreign' Falsehoods**

In analysing responses to these new threats, 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.[41] 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,[42] 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

[38] Jean-Baptiste Jeangène Vilmer et al, n. above at 77.
[39] House of Commons, Digital, Culture, Media and Sport Committee, n. 20 above, part 6.
[40] Michael Toms and Jessica LP Weeks, 'Public Opinion and Foreign Electoral Intervention' (2019), https://web.stanford.edu/~tomz/working/TomzWeeks-ElectoralIntervention-2019-08-13i.pdf.
[41] https://www.mtsu.edu/first-amendment/article/937/content-neutral.
[42] See Protection from Online Falsehoods and Manipulation Act 2019 (Singapore), s.8, discussed in section 6 below.

'partisan' or advertising messages.[43] Likewise, a ban on the publication of political messaging coming from outside a jurisdiction would be a source-based measure. 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.[44]

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.[45] Content-based prohibitions aimed at false or misleading political statements are widely employed in many jurisdictions, and we will consider some examples in due course. 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 6 below, 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).[46] More frequently,

[43] Canada Elections Act 2000 (as amended), s.325(2) & (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.

[44] See e.g. Electoral Commission, *Digital Campaigning: Increasing Transparency for Voters* (Electoral Commission, London 2018), para 26.

[45] See e.g. Liberty, https://www.libertyhumanrights.org.uk/human-rights/free-speech-and-protest/speech-offences.

[46] In Singapore 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 section 6 below.

measures that are partly source-based and partly content-based are aimed at false claims that are considered to be part of an attempt by one jurisdiction to influence political developments in another jurisdiction. Here are some examples.

In France, a 2018 amendment to earlier legislation[47] 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 they are judged, 'to harm the fundamental interests of the nation…particularly by disseminating false information'.[48] Under this militant democratic provision,[49] permissible sanctions include an order to stop broadcasting, fines up to 3% of the broadcaster's revenue, and in some cases withdrawal of authorisation to broadcast.[50] Another example is provided by section 282.4 of the Canada Elections Act 2000 (as amended in 2018-19), 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, such as the making of false claims about an election candidate, party leader or official (contrary to section 91 of the 2000 Act). Legislation in England and Wales

[47] 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.

[48] N. 47 above, Art 42-6 (as amended); https://www.conseil-constitutionnel.fr/en/decision/2018/2018773DC.htm. See https://www.loc.gov/law/help/freedom-expression/france.php#_ftn51. 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.

[49] On militant democracy, in the broad sense employed here, see section 1.

[50] N. 47 above Art. 42-1.

prohibits collusion in overseas influence on elections (whether or not involving false statements). Section 92 of the Representation of the People Act 1983, applicable to both elections and referendums, prohibits (subject to some exceptions) any person from bringing to bear on domestic campaigns foreign broadcasting content relating to the campaign, with the intention of influencing people to cast or refrain from casting their votes.[51]

There are reasons to treat such developments with caution. 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 that a government more respectful of international law and human rights will come to power in the targeted state.[52] Secondly, even if an overseas maker of a false statement is not him or herself by right protected in making it under a jurisdiction's constitution, there might still be a duty on a liberal-democratic legislator to tolerate (at least, in law[53]) the dissemination of the statement. 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

[51] The Canadian Elections Act 2000, s.330(1) has a parallel provision.
[52] Cecile Fabre, 'The Case for Foreign Electoral Subversion' (2018) 32(3) Ethics and International Affairs 283.
[53] I say, 'in law', because of course there is every reason for those who disagree with a statement to vigorously contest it through argument, and in that respect not to tolerate it at all.

restrictions apply only to false or misleading political claims, if 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*[54]:

> 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.[55]

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

> (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

[54] (2014) 45 F Supp 3d 765.

[55] *Susan B Anthony List v Ohio Elections Commission* (2014) 45 F Supp 3d 765, 769.

in doing so, if no contravention of subsection 330(1) or (2) is involved in the transmission.[56]

This exception is significant, in that it concedes the wrongfulness of trying – other than in a restricted range of cases - 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.[57] An irony about that is that evidence indicates that Russian attempts to influence the outcome of the 2016 Presidential election, for example, 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').[58]

More generally, though, 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 a usually a factual finding: for example, that someone putting up, further distributing

[56] On section 330, see n. 51 above.
[57] See e.g. Jeffrey L Kirchmeier, 'The Illusion of the Fact-Opinion Distinction in Defamation Law' (1988) 39 Case Western Reserve Law Review 867.
[58] Michael Jensen, 'Russian Trolls and Fake News: Information or Identity Logics?' (2018) 71 Journal of International Affairs 115, 122.

or sponsoring a post is based outside the jurisdiction, or that they have failed to reveal their identity or location.[59] So, the intervention of courts or regulators to police source-based restrictions may be 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'.[60] 'Trolling', of course, involves threats, hate speech, aggressive belittling, or other harms often justifiably made subject to criminal sanctions on content-based grounds.[61] So, if compelling online platforms to reveal the identity of 'trolls' (a source-based measure) serves to deter the dissemination of content-based wrongs, then so much the better.[62] 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. Perhaps political organisations, in particular, should be forced to be more honest about their use of bots.[63] 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

[59] As in the case of English law governing printed election material: see n. 43 above.
[60] Kate Jones, 'Online Disinformation and Political Discourse: Applying a Human Rights Framework' (2019), https://www.chathamhouse.org/sites/default/files/2019-11-05-Online-Disinformation-Human-Rights.pdf, 45 & 53.
[61] See Maeve Duggan, 'Online Harassment' (2014), https://www.pewresearch.org/internet/2014/10/22/online-harassment/.
[62] See e.g. the Harmful Digital Communications Act 2015 (New Zealand), s.22.
[63] Kate Jones, n. 60 above,

through a bot.[64] However, a bot may be acting simply as a kind of virtual 'amplifier,' something that may be used to spread valuable political (or *e.g.* health) information, as well as false claims.[65]

Further, so far as compulsory identity-disclosure is concerned, it ought also to be kept in mind that 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.[66] Compelling would-be speakers to reveal their identity, in such circumstances, may thus unacceptably confront them with a free speech dilemma (further considered below). In its 2019 White Paper,[67] the UK Government suggested that any Code of Conduct governing the taking down of content by social media companies must include steps to deter, 'users who deliberately misrepresent their identity to spread and strengthen disinformation'.[68] In the case specifically of what I am calling political viewpoint (dis)information, it is hard to see how such a recommendation could be squared with recognition that social media users may both reject official understandings of 'disinformation,' and also - in part for that reason - wish to maintain their anonymity. Giving priority to freedom of speech means accepting

[64] See the Bolstering Online Transparency Act 2018, SB 1001. For criticism, see https://www.wired.com/story/law-makes-bots-identify-themselves/.
[65] 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.
[66] For a defence of online anonymity, see Article 19, *Right to Online Anonymity* (2015), https://www.article19.org/data/files/medialibrary/38006/Anonymity_and_encryption_report_A5_final-web.pdf.
[67] HM Government, *Online Harms, White Paper*, CP 57 April 2019.
[68] HM Government, n. above, para 7.28.

that there is only a public interest in securing the identification of online agents in a limited range of cases: perhaps, cases where there is a clear threat of physical or mental harm (including harm threatened through the spread of some kinds of disinformation, such as dangerously inaccurate health advice).[69] I return to the suggestion that the extent of tolerance ought to reflect the nature and seriousness of the harm in section 9.

## 4. Restraint in Criminalising Misinformation: The Scepticism Principle.

I now turn my attention to the ethics of restrictions on substantive political content, content influencing substantive political beliefs. 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.[70]

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 judgement of public officials when exercising existing or creating new powers. However, it is convenient to speak in

[69] For discussion, see Carissa Véliz, 'Online Masquerade: Redesigning the Internet for Free Speech Through the Use of Pseudonyms' (2019) 36 Journal of Applied Philosophy 643.
[70] Frederick Shauer, *Free Speech: A Philosophical Enquiry* (Cambridge: Cambridge University Press, 1982), 86.

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. As the Washington Supreme Court has said, in declaring unconstitutional a prohibition on the malicious publication of a false statement of material fact about a candidate for public office, the statute was unconstitutional because it caused, 'the government, rather than the people, [to] be the final arbiter of truth in political debate'.[71] 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[72]). 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 judgements 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

[71] *Rickert v Public Disclosure Commission* (2007) 168 P.3d 826, 827.
[72] https://www.ipsos.com/ipsos-mori/en-uk/politicians-remain-least-trusted-profession-britain.

a letter criticising the proposal to transfer jurisdiction to hear complaints about false claims in 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.[73]

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. Whilst, 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.[74] The involvement of the state in threatening penal sanctions, in cases where 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

[73] Cited by Thomas LCJ, in *R (on the application of Woolas) v The Parliamentary Election Court* [2010] EWHC 3169 (Admin), para 23. For further discussion of this point, see e.g. 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.

[74] See e.g. Council of Europe, *Defamation and Freedom of Expression*, H/ATCM (2003) 1 (Strasbourg: Directorate of Human Rights, 2003).

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.[75]

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, just because (as Shauer rightly remarks) political leaders are fallible, does not mean government should not strive to provide access to political information for citizens, and to promote as public goods widespread political debate and participation.[76] Further, the use of criminal prohibitions – even 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 prevent the spread of false claims in relation to the voting process: what I am calling participation (mis)information.[77] 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.

[75] See e.g. *Oberschlick v. Austria No.1* (Application No. 11662/85), of 23 May 1991.
[76] See, generally, Eric Barendt, *Freedom of Speech*, 2nd ed (Oxford: Oxford University Press, 2007), 30-8.
[77] See e.g. the UK's Representation of the People Act 1983, s.115.

## 5. **Restraint in Criminalising Misinformation: The Autonomy Principle.**

The principle of scepticism – more properly, a sceptical attitude - is connected to another principle: the principle of political autonomy.[78] One aspect of the autonomy principle is that, in Eric Barendt's words (drawing on Thomas Scanlon[79]), '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'.[80] In itself, though, this statement of the principle does not take us far. That is because it is focused primarily on the autonomy of the potential audience 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, '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'.[81] Even so neither First Amendment jurisprudence (not further considered here),[82] not ECHR

[78] Thomas Scanlon, 'A Theory of Freedom of Expression' (1972) 1 Philosophy and Public Affairs 204.
[79] N. 78 above.
[80] Eric Barendt, n 76 above, 16.
[81] *Commonwealth v Lucas* (2015) 34 NE 3d 1242.
[82] 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.

case law,[83] give 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 socio-political 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 equality-based dimension to political engagement. In a voting context, what this entails is that, if I am freely to exercise the franchise, there must be accessible places for all to vote equally free from intimidation and undue influence. Everyone must have basic participation information about how, where and when to vote, and enough about candidates so that they can distinguish adequately between them. The counting of all the votes must be accurate and impartial. There must not be artificial (and in practice, biased) barriers to voting, such as disproportionately onerous identification requirements,[84] and so on. Guaranteeing that these equality conditions for free political engagement exist nationwide requires disinterested, collective - state – action, given the significant co-ordination 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 equality conditions may well be a perfectly proportionate step. False statements knowingly made about these

[83] Discussed briefly in section 6 below.

[84] 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.

participation conditions fall squarely within the scope of legitimate criminalisation, thus understood.

Can this argument about 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 Article 25 of the International Covenant on Civil and Political Rights[85] (and hence with the equality conditions just described).[86] 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 itself – and others - to freedom of information and to the promotion of honesty in politics through, for example, the introduction of codes of electoral conduct, or by providing tax breaks for independent fact-checking organisations, and so on. Further, 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

[85] '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 Article 3 of the First Protocol to the ECHR.
[86] See Kate Jones, n 60 above, 48.

help people to lead the lives they want to have by satisfying their false desires'.[87] However, in the case of political belief formation, the scepticism and autonomy principles insist that the role of the state is confined to, for example, non-coercively encouraging political engagement and fostering a culture of truthfulness and honesty in politics.

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.[88] 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*:[89]

---

[87] Joseph Raz, *The Morality of Freedom* (Oxford: Clarendon Press, 1986), 144.

[88] Worries about paternalism 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, many governments historically owe their election to their ability to persuade sufficient numbers of people to behave in just such a way.

[89] *Whitney v. California* (No. 3) 74 U.S. 357 (1927).

> 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.'[90]

The EU defines misinformation 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',[91] and 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 misinformation. The four pillars are:

> (i) improving the capabilities of Union institutions to detect, analyse and expose disinformation;
>
> (ii) strengthening coordinated and joint responses to disinformation;
>
> (iii) mobilising private sector to tackle disinformation;
>
> (iv) raising awareness and improving societal resilience.

In relation to (ii) 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'.[92] Similarly, in

---

[90] *Whitney v. California* (No. 3) 74 U.S. 357 (1927), at 377 (Brandeis J).
[91] EU High Representative of the Union for Foreign Affairs and Security Policy, *Action Plan Against Disinformation* (JOIN 2018, 36 Final), 1.
[92] Ibid., 8.

France, a Report from the Policy Planning Staff (Ministry for Europe and Foreign Affairs) and Institute for Strategic Research (Ministry for the Armed Forces)[93] defines political misinformation as, 'the intentional and massive dissemination of false or biased news for hostile political purposes',[94] 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'.[95]

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.[96]

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.[97] Such restrictions give rise to what can be called political falsehood dilemmas for speakers. For example, first, someone might know 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

[93] J-B Jeangène Vilmer n. 16 above.
[94] Ibid., 12.
[95] Ibid., 13.
[96] Ibid., 167. See also, NATO STRATCOM, *Internet Trolling as a Tool of Hybrid Warfare: The Case of Latvia* (2016), https://www.stratcomcoe.org.
[97] See *e.g*. Michael Karanicolas, 'Subverting Democracy to Save Democracy: Canada's Extra-Constitutional Approaches to Battling "Fake News"' (2019) 17 CJTL 202, at 223.

believed. Secondly, such a dilemma may arise when what the speaker intends to say is known by them to be untrue or misleading, but they nonetheless believe that there is sufficient reason to make the statement: in other words, when they may not be acting 'dishonestly' in a broad sense of that term.[98] In both instances, the existence of the political falsehood dilemma involves the worry that would-be contributors to political debate may be wrongly deterred by the prospect of sanctions from engaging in what is simply provocative, combative or attention-grabbing political speech.

The problem of political falsehood dilemmas is exacerbated by the fact that, in relation to political discourse, a significant element of value judgement must enter into the question whether or not a claim is 'misleading', or even whether or not it involves a 'false' claim.[99] Legal prohibitions on misleading statements require administrators and courts to grapple with the problem of disinformation and '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.[100] When it is known that legal uncertainties of these kinds may arise, people may find themselves facing a political falsehood dilemma, and could be deterred from making political claims that are simply provocative or controversial, to the detriment of a free speech polity.

---

[98] Jeremy Horder, *Excusing Crime* (Oxford: Oxford University Press, 2004), 48–50.

[99] See text at n 14 above.

[100] https://www.theguardian.com/technology/2019/nov/12/far-right-use-russian-style-propaganda-to-spread-misinformation; Kate Jones, n 60 above.

## 6.The Authoritarian Approach to Political Misinformation: Singapore.

As Bradshaw and Howard rightly observe,[101] authoritarian jurisdictions use the internet and social-media technologies to enhance their control over information, and their use of legislation to counter 'misinformation' may be an important dimension to that. What differentiates, in that regard, authoritarian misinformation strategies from the strategies at work in democracies? In democracies, so far as the spread of misinformation is concerned, concerted disinformation campaigns tend to arise mainly at election times,[102] and will not involve official state action to propagate such disinformation. By contrast, in authoritarian regimes, political viewpoint misinformation (and sometimes participation misinformation) is likely to be employed by state officials themselves on an ongoing basis: often, to target individuals such as dissidents or journalists. In democracies, targeting of this latter kind may be (routinely) engaged in by political parties, and 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 misinformation, we will see in due course there will be 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. Two key points of overlap are the extension of the criminal law to cover defamation of public officials or election candidates, and the

[101] Samantha Bradshaw and Philip Howard, n. 25 above, 25.
[102] Samantha Bradshaw and Philip Howard, n. 25 above, 28-29.

grant of powers – backed by the threat of 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 reject the use of the criminal law- or other forms of coercion - to protect officials or candidates from viewpoint misinformation, whilst vigorously countering (through coercion if need be) participation misinformation.

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

(a) The strengthening and broadening of coercive powers to deter and punish the dissemination of misinformation, including powers over social media platforms on which such information appears, without any attempt to prefer source-based measures to content-based measures;

(b) 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,[103] both in and beyond election times.

(c) Making a focus (in some instances, an exclusive focus), in law or in enforcement policy, alleged misinformation concerning government actions or the actions of other state officials, whether or not in relation to elections.

(d) Making high, often disproportionate levels of punishment available;

[103]Samantha Bradshaw and Philip Howard, n. 25 above, 26-27.

(e) The use of state-sponsored 'trolling,' targeting dissidents and journalists.[104]

It is important to note that at least the first of these features, and to some extent the second two, will also be found in some liberal legal systems influenced by militant democratic thinking, discussed in sections 7-9 below, 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 Singapore has introduced measures to counter misinformation that have a number of authoritarian characteristics, 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 (Article 14).[105] Under section 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 false statement of fact', and that communication of the statement is likely to have a range of effects, including:

(i) be prejudicial to the security of Singapore or any part of Singapore;

(ii) be prejudicial to public health, public safety, public tranquillity or public finances;

[104] Samantha Bradshaw and Philip Howard, n. 25 above, 28, speaking of Bahrain, Turkey, Azerbaijan and Russia.

[105] Asean Law Association, *The Singapore Legal System* (2018), https://www.aseanlawassociation.org/wp-content/uploads/2019/11/ALA-SG-legal-system-Part-2.pdf, ch 2. 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': Gábor Attila Tóth, n 2 above, para 19.

(iii) 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;

(iv) incite feelings of enmity, hatred or ill-will between different groups of persons; or

(v) 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 (a) to (d) of authoritarian responses listed above.[106]

An important point, first, is that it applies not only to content known to contain a false statement of fact, but 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 Article 10 of the ECHR.[107] Secondly, the communication of a false statement is

[106] For example, an individual found guilty under section 7 can be imprisoned for up to five years: section 7(2).
[107] *R (on the application of Woolas) v The Parliamentary Election Court* [2010] EWHC 3169 (Admin).

not, as such, prohibited,[108] unless it is likely to have one of the effects mentioned in the criteria listed in (i) to (v) 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. Thirdly, 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 misinformation widely or to a identified target audience. The need to satisfy one of conditions (i) to (v) might appear to rule out low-level informal communications, but that is far from clear. More liberal jurisdictions have sought to confine draconian laws of this type – such as Germany's Network Enforcement Act 2017 (considered below) - to platforms with millions of registered users.

Fourthly, Part 3 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 section 7. As the names imply, such directions require the person subject to them either to stop communicating the allegedly false information, or to declare a given communication by that person to be false and to publish a 'correction' in a form specified.[109] Extraordinarily, as well as applying to persons outside as well as within Singapore (so long as the

[108] Unless, by virtue of section 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.
[109] The full extent of what may be required is set out in sections 11 and 12.

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 their knew or had reason to know that the information was false.[110] In that regard, we should note the broad-ranging definition of 'false statement of fact' given in the definition section of the 2019 Act:

> (a) 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
>
> (b) 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, the section 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

[110] S.11(4); s.12(4).

as on anything done by the speaker. Finally, it is a Government Minister – not an independent body - who, by virtue of section 10(1) may require that a competent authority[111] issues a stop communication or a correction direction, politicising the process in an authoritarian way.[112] 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.

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 section 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.[113] Secondly, there is the fact that the right to free speech under Article 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 (Article 14(2)).[114] Liberal-democratic countries do not regard expediency as a legitimate ground for restricting freedom of speech in pursuit of such ends.

[111] A Statutory Board appointed by the Minister under section 6, and required to give effect to his or her instructions.

[112] Although s.17 permits an appeal to the High Court against the issuing of a direction, 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.

[113] 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 article 246 of the Chinese Criminal Code.

[114] For criticism, see Michael Hor, 'Freedom of Speech and Defamation: *Jeyaretnam Joshua Benjamin v. Lee Kuan Yew*' (1992) Singapore Journal of Legal Studies 542, 547–48.

Secondly, 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[115] expressly rules out the application of qualified privilege to allegedly defamatory statements made by or one behalf of election or presidential candidates. Further, 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.[116]

By way of contrast, in countries governed by Article 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,*[117] in considering false claims made about politicians, the European Court observed that, 'freedom of political debate is at the very core of the concept of a democratic society which prevails throughout the Convention'.[118] In *Otegi Mondragon v Spain*,[119] 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

[115] 1957, revised in 2014.
[116] *Jeyaratnam Joshua Benjamin v. Lee Kuan Yew* (1992) 1SLR(R) 791 (CA).
[117] *Lingens v. Austria*, 8 July 1986, https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-57523%22]}.
[118] *Ibid.,* para 42.
[119] *Otegi Mondragon v Spain* 15th March 2011 (Application no. 2034/07), para 50. See further, *Castells v. Spain*, judgment of 23 April 1992, Series A No. 236.

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 Article 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.[120]

Authoritarian regimes will typically reject such an approach, because it threatens the element of political domination at the heart of such regimes, 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.[121] For authoritarian regimes, control over 'misinformation' is a dimension to their control over the power narratives for that jurisdiction, such as control of what political issues are on the agenda,[122] which are not to be determined through free citizen engagement.[123]

In section 1, I said that one source of justification for a coercive response to the spread of false political viewpoint information, especially at election times, are particular liberal – be it considered universal, or elite – concerns. These are militant

[120] *Otegi Mondragon v Spain* 15th March 2011 (Application no. 2034/07), para 50. See further, *Castells v. Spain*, judgment of 23 April 1992, Series A No. 236, para 50.
[121] See further, European Audiovisual Observatory, *Disinformation in the Media under Russian Law* (2019), https://rm.coe.int/disinformation-in-the-media-under-russian-law/1680967369.
[122] See Steven Lukes, *Power: A Radical View* (London: MacMillan, 1974).
[123] See Gábor Attila Tóth, n 2 above.

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, has 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'.[124] So, for example, in the UK, 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.[125] 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'.[126] Nonetheless, in some jurisdictions, a militant democratic path to coercion has proved attractive

## 7. Authoritarian Liberalism: Germany.

[124] Jennifer Hochschild, n. 6 above, 386-87, summarising the main claim of Herbert McClosky and Alida Brill, *Dimensions of Tolerance: What Americans Believe about Civil Liberties* (New York: Russell Sage Foundation, 1
[125] https://reformpoliticaladvertising.org/blog/.
[126] Advertising Standards Authority, Annual Report 1999, https://www.asa.org.uk/asset/40D37406-BDD0-48A4-BF46C68DE81C1E39/, 2.

. 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 misinformation.[127] The 2017 Act builds on obligations first imposed by section 10 of the Telemedia Act (Germany) 2007:

> 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 (*i.e.* 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 section 3 of the 2017 Act requires that networks provide, 'effective and transparent procedure

[127] See e.g. Matthias Kettermann, 'Follow-up to the Comparative Study on Blocking, Filtering and Takedown of Illegal Internet Content' (2019), https://rm.coe.int/dgi-2019-update-chapter-germany-study-on-blocking-and-filtering/168097ac51.

[128] See the discussion in Article 19, *Germany: The Act to Improve Enforcement of the Law in Social Networks* (2017), https://www.article19.org/wp-content/uploads/2017/09/170901-Legal-Analysis-German-NetzDG-Act.pdf, 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 Heidi Tworek, n. 10 above, 3.

[129] These terms are not defined in the 2017 Act. For criticism, see Article 19, n 127 above, 12-13.

for handling complaints about unlawful content'. In the case of 'clearly violating' content, it must be removed within twenty four hours, and in the case of 'violating' content, within seven days; but 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. A failure to comply in an individual case will not lead to a fine, but 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 section 3(3) of the 2017 Act, a network platform can itself avoid the seven day deadline to act, if (i) there needs to be an investigation in to the truth or falsity of a claim, or (ii) 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 has not set up a self-regulatory body to meet the regulatory challenge, has been fined 2M euros under the 2017 Act for providing incomplete transparency reports.[131]

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

[130] See Network Enforcement Act: Regulatory Fining Guidelines (2018), https://www.bundesjustizamt.de/DE/SharedDocs/Publikationen/NetzDG/Leitlinien_Geldbussen_en.pdf?__blob=publicationFile&v=2; Stefan Theil, n. 127 above, 48. See further, the discussion of the approach in England and Wales in section 8 below.

[131] Stefan Theil, n. 127 above, 48; Ben Wagner et al, 'Regulating Transparency? Facebook, Twitter and the German Network Enforcement Act' (2020), https://www.researchgate.net/publication/338802975_Regulating_Transparency_Facebook_Twitter_and_the_German_Network_Enforcement_Act/link/5e2b217292851c3aadd7b7f5/download.

relating to content removal? The difficulty here is 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 (Article185), malicious gossip (Article 186), criminal defamation (Article 187), and treasonous forgery (Article 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. 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,

[132] Studies show that IT companies remove 70% of all hate speech notified to them within 24 hours: Matthias Kettermann, n. 127 above, 2.
[133] Although criminal defamation of the President of the Federation (Section 90); and criminal defamation of the state and its symbols (Section 90a) were excluded.
[134] Stefan Theil, n. 127 above, 44. See *Hoffer and Annen v Germany*, 397/07; 2322/07, Court, 13 January 2011.

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.[135] Faced with the prospect of sanctions for under-blocking, the rational network platform must inevitably incline towards over-blocking, impinging on Article 10 rights.[136]

These platforms are now the modern equivalent of the 'political police' envisaged in the late 1930s by Karl Loewenstein (who coined the term, and defended, 'militant democracy') whose task it was to be to defend democratic ideals in Germany, 'even at the risk and cost of violating fundamental principles'.[137] Inevitably, this kind of indirectly enforced self-censorship will lead to the taking down of what is merely challenging or provocative political content,[138] even though section 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).[139] Under the 2017 Act, not only is there is no prior determination by a court of what counts as 'violating' content, but

[135] Article 19, n 127 above, 16 & 20. In the case of an allegedly false claim, section 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.
[136] Wolfgang Schultz, 'Regulating Intermediaries to Protect Privacy Online – the Case of the German NetzDG' (2018), https://www.hiig.de/publication/regulating-intermediaries-to-protect-privacy-online-the-case-of-the-german-netzdg/.
[137] Karl Loewenstein, 'Militant Democracy and Fundamental Rights II' (1937) 31 American Political Science Review 638, 656-57, cited by Jan Muller, n. 3 above, 1120.
[138] Heidi Tworek, n 10 above, 3. However, for a contrary view, see Stefan Theil, n. 127 above, 49.
[139] See Heidi Tworek, n 10 above. A social network receiving more than 100 complaints in a year must publish its internal moderation practices.

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.[140] A 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.

In 2018, Twitter received 264,818 complaints in Germany regarding content, blocking 10% of content in consequence, YouTube received around 215,000 complaints, blocking 27% of content in consequence, and Facebook 1,704 complaints, leading to 20% of content being blocked.[141] This raises questions about the criteria used in each instance as a justification for taking the decision to remove content.[142] In its disaggregation of take-down statistics, Google indicated that 8695 complaints involving defamation or insult led to 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).[143] Theil's research reveals that in 2018-19, Facebook logged 500 NetzDG reports, with deletion of content following in 159 cases (roughly 33.8% of cases), and Twitter logged 256,462 reports of which 23,165 led to the deletion of content, roughly 9% of cases.[144] Defamation and

[140] Heidi Tworek, n 10 above, 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, n 127 above, 15.
[141] Matthias Kettermann, n. 131 above, 3. However, most of these complaints involved a breach of the platforms' community guidelines rather than the 2017 Act: Heidi Tworek, n 10 above, 5.
[142] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3300636; Heidi Tworek, n 10 above, 5.
[143] Tworek describes the 2017 Act, in consequence, as a 'community guidelines enforcement law': Heidi Tworek, n 10 above, 7. Insult-related offences under the 2017 Act were the most common kind offence 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%.
[144] Stefan Theil, n. 127, 50.

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.[145]

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.[146]

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'.[147] By contrast, I suggest that such a pious hope does not do full justice to the arguments against militant democratic justifications for employing coercive pressure, in cases involving false substantive political content. In that regard, as I will argue below when analysing the proposals for the UK, 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 platforms to the same governance structures, and risks of sanctions, in respect of the toleration of false or misleading political speech, as in the case of toleration of child pornography or incitements to engage terrorist violence? The answer we give is an

[145] Heidi Tworek, n 10 above, 5-6.
[146] Stefan Theil, n. 127 above, 50
[147] Stefan Theil, n. 127 above, 50.

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.[148] So, the concern remains that provocative and challenging political content is being unwittingly suppressed, along with genuinely harmful content.[149]

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'[150]) and, ironically, enhance its credibility amongst those disposed to support the speaker.[151] It is significant, in that regard, that there have been several high-profile deletions of content posted by German politicians.[152] 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.[153] counselled against Government heavy handedness, on the grounds that:

[148] 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: Heidi Tworek, n 10 above, 7.
[149] Article 19, n 127 above.
[150] J-B Jeangène Vilmer, A. Escorcia, M. Guillaume, J. Herrera, n 16 above, 169.
[151] See Heidi Tworek, n 10 above, 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.'
[152] See Stefan Theil, n. 127 above, 42.
[153] See J-B Jeangène Vilmer, n. 16 above.

> 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.[154]

Accordingly, only one of their 50 recommendations (recommendation 12) is concerned with the creation of prohibitions on the dissemination of false information.[155]

## 8. Authoritarian Liberalism: The United Kingdom.

One danger in the authoritarian-liberal response to digitally propagated misinformation of Germany and France 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 authoritarian regimes whose 'fake news' practices liberal regimes are, ironically, seeking to counter.[156] There is another danger. By giving government agencies, or even courts, the power to prohibit and punish the dissemination of political misinformation, 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.[157]

---

[154] J-B Jeangène Vilmer, n. 16 above, 172.
[155] J-B Jeangène Vilmer, n. 16 above, 173.
[156] See European Audiovisual Observatory, n 121 above.
[157] José Luis Martí and Philip Pettit, *A Political Philosophy in Public Life* (Princeton: Princeton University Press, 2010), at 38.

In that regard, it is noteworthy that significantly more people world-wide support the idea that journalists (75%) and online platforms (71%) should take the responsibility for weeding out misinformation over which they have influence, than support the notion that this is the responsibility of Government (61%).[158] How much better, if at all, do UK proposals for regulation on online speech fare?

In its 2019 White Paper on online harms,[159] 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.[160] Under the proposed UK approach, social media companies large and small 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.[161] The discharge of this duty of care, in accordance with a code of practice, will be overseen by a regulator, with powers (amongst other things) to fine those deemed to be in breach.[162] 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

[158] Reuters Institute, *Digital News Report 2018*, https://reutersinstitute.politics.ox.ac.uk/sites/default/files/digital-news-report-2018.pdf, 19.
[159] See HM Government, n 67 above.
[160] For a penetrating critique, see Victoria Nash, 'Revise and Resubmit? Reviewing the 2019 Online Harms White Paper' (2019) 11 Journal of Media Law 18.
[161] HM Government, n 67 above, 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.
[162] HM Government, n 67 above, paras 16-28.

speaking, as in Germany) on whether there was a breach of the duty of care in the way that the content came to be disseminated.[163] In that regard, as Victoria Nash suggests, the White Paper can be understood as a, 'manifesto for government-led platform governance': [164] soft regulation.[165] 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.[166]

In relation to disinformation, a key aspect of the White Paper approach is described thus:

> 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.[167]

The source-based aim described in the final words of this passage is an important one. There is likely to be a 'news quality obligation' imposed on companies, to ensure that companies, 'improve how their users understand the origin of a news article and the trustworthiness of its source'.[168]

---

[163] HM Government, n 67 above, para 19.
[164] Victoria Nash, n. 169 above, 23.
[165] See e.g. OECD, 'Soft Law', https://www.oecd.org/gov/regulatory-policy/irc10.htm.
[166] HM Government, n 67 above, 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, for example, Social Media Councils is to be preferred, see Article 19, Social Media Councils: Consultation (2019), https://www.article19.org/resources/social-media-councils-consultation/.
[167] HM Government, n 67 above, para 7.27.
[168] HM Government, n 67 above, para 7.29, citing *The Cairncross Review: A Sustainable Future for Journalism* (2019), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/779882/021919_DCMS_Cairncross_Review_.pdf.

So far as content-based measures are concerned, the White Paper anticipates that the Code of Practice will include, 'steps companies should take in their terms of service to make clear what constitutes disinformation'.[169] This claim equivocates over the right answer to the crucial question whether or not it is for individual companies, or for the regulator, to decide what amounts to 'disinformation.' So, it is unclear how or to what extent the Article 10 rights of those posting content will be protected.[170] More generally, on content-based restrictions, the White Paper says:

> Importantly, the code of practice that addresses disinformation will ensure the focus is on protecting users from harm, not judging what is true or not. There will be difficult judgement calls associated with this. The government and the future regulator will engage extensively with civil society, industry and other groups to ensure action is as effective as possible, and does not detract from freedom of speech online.[171]

There is a confusion about aim in the first sentence of this passage. The code of practice must, surely, be concerned with, 'what is true or not', if it is to be concerned with, 'protecting users from harm', because – for example – the potential harmfulness of advice about health products may depend on the truthfulness of what is being said by way of advice. By contrast, there is a respectable case for saying that no one is subjected to a risk of 'harm' (although

[169] HM Government, n 67 above, para 7.28.
[170] Victoria Nash, n. 169 above, 23.
[171] HM Government, n 67 above, para 7.31.

they may be inconvenienced, perhaps seriously) when encountering false political statements aimed at influencing substantive voting decisions.[172] In that regard, one of the, 'difficult judgement calls' referred to concerns the treatment of public figures. The general offence of criminal defamation was abolished in England and Wales, in all but exceptional circumstances, in 2009:[173] remedies for defamation must be sought through resort to private law, and to court-based, open justice. What, then, if social media companies come under pressure to restrict allegedly defamatory content when people – including public figures - complain to them and to the regulator about it? In that regard, the White Paper envisages social media companies having to take action only when abusive content, 'goes beyond free speech and impedes individuals' rights to participate'.[174] That is a proper and defensible distinction, in that it is essential – for example - that there is protection for the rights of those who are, say, systematically intimidated into silence and non-participation.[175] However, it is not an easy distinction for a company or regulator to turn into policy in a fair, open and consistent manner.[176]

In so far as the White Paper treats the dissemination of disinformation or misinformation about substantive political content in the same manner as it treats

[172] I exclude from this claim, of course, false statements that violate equality conditions, as discussed in section 5 above.
[173] Coroners and Justice Act 2009, s. 73. For the exception, see the Representation of the People Act 1983, s.106.
[174] HM Government, n 67 above, para 7.36.
[175] Damian Tambini, 'The Differentiated Duty of Care: a Response to the Online Harms White Paper' (2019) 11 Journal of Media Law 28, 30.
[176] The White Paper requires online platforms to make it easier for users to block content they do not wish to see (Ibid., para 7.24) A robust system of that kind ought to make it possible to set the bar higher when deciding to go further and take content down, block users or disable accounts. Companies will also be required to provide mechanisms for appealing against decisions to block content or users.

other online wrongdoing, such as incitement to engage in terrorism or to commit suicide, or the grooming of children for sexual activity, then it can be regarded as a militant democratic (liberal-authoritarian) set of proposals. As both Victoria Nash[177] and Damian Tambini [178] have argued, there ought to have been an unequivocal and clear commitment to distinguish, in terms of the appropriate regulatory response, between illegal harms and legal-but-unwanted harms and risks. In the White Paper, as it stands:

> 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.[179]

As Tambini suggests, the proportionate regulatory response to the legal-but-unwanted harm, or other wrongdoing such as the dissemination of false or misleading political speech, ought to be not the threat of sanctions, but 'monitoring, advice and transparency', with a focus on, 'consumer information, competition and switching'.[180]

## 9. Conclusion

A free speech culture is a culture that avoids militant democratic tendencies in its response to the problem of false or misleading claims regarding substantive politics. Online platforms are not themselves broadcasters and so, in relation to political content, they cannot reasonably be expected to ensure that such content

[177] Victoria Nash, n. 169 above.
[178] Damian Tambini, n. 184 above.
[179] Damian Tambini, n. 184 above.
[180] Damian Tambinin, n. 184 above, 29.

and coverage is balanced overall.[181] In a free speech culture, the key question is how to devise a modern interpretation of Justice Brandeis 'more speech' solution,[182] and put it at the forefront of any response to political misinformation online.[183] Answers to that question are not in short supply, as we have seen,[184] 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'.[185] A deterrent and punitive approach to false or misleading substantive political content, to the use of bots, or to the maintenance of anonymity, should be discarded in favour of a regulatory solution in which, 'States design a choice architecture without enforcing a particular choice'.[186] The UK White Paper proposes ensuring that platforms adopt the following steps (amongst others), to promote a flourishing free speech culture online:[187]

(i) Making content which has been disputed by reputable fact-checking services less visible to users;

(ii) Using fact-checking services, particularly during election periods;

(iii) Promoting authoritative news sources;

(iv) Promoting diverse news content;

[181] See, by contrast, the obligations enforced on broadcasters by the UK regulator, Ofcom: *The Ofcom Broadcasting Code* (2009), https://www.ofcom.org.uk/__data/assets/pdf_file/0027/19287/bcode09.pdf.
[182] Text at n 89 above.
[183] Reserving deterrence, suppression, and enforced transparency to cases where a specified and limited range of harms are threatened (whilst accepting there will always be debate about these). See section 9 above.
[184] Text following n 90 above.
[185] Jean-Baptiste Jeangène Vilmer, n 16 above, 170.
[186] Jean-Baptiste Jeangène Vilmer, n 16 above, 170.
[187] HM Government, n. 67 above, para 7.28.

(v) Ensuring that it is clear to users when they are dealing with automated accounts;

(vi) [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.

Unfortunately, because the White Paper must deal with online *harms*, and not just with substantive political misinformation, these positive, non-coercive steps to contribute to a better politics are wrongly co-mingled with steps (rightly) aimed at the blocking and suppression of genuinely harm-risking misinformation, such as false information about health.[188] Nonetheless, a better path forward is discernible here.

[188]HM Government, para 7.25.