DEMAGOGUERY, TECHNOLOGY, AND COGNITION: ADDRESSING THE THREATS TO DEMOCRACY

*Stephan Lewandowsky**

▼

Numerous indicators suggest that democracy is under threat,[1] including in Europe. In 2020, *The Economist's* democracy index determined that one EU member state, Hungary, was no longer a democracy.[2] Throughout Europe, populist movements –mainly, but not exclusively, on the political right– have pitted "the people" against a presumed "elite" that is variously constructed as including mainstream media, politicians, experts, scientists, and academics.[3] The COVID-19 pandemic has put further pressure on societies by requiring restrictions on social behaviours to control the pandemic that are unprecedented in democracies and that may facilitate autocratization.

Although symptoms and causes of these trends are intertwined and difficult to tease apart, there is little doubt that wilful disregard of evidence and expertise,[4] accompanied by a flood of misinformation –on social media, in hyperpartisan news sites, and in political discourse– are at the heart of the challenge to democracies.[5] Misinformation matters: Exposure has been shown to make a

* The author acknowledges financial support from the European Research Council (ERC Advanced Grant 101020961 PRODEMINFO), the Humboldt Foundation through a research award, the Volkswagen Foundation (grant "Reclaiming individual autonomy and democratic discourse online: How to rebalance human and algorithmic decision making"), and the John Templeton Foundation (through the Honesty program awarded to Wake Forest University). The author also receives funding from Jigsaw (a technology incubator created by Google) and from UK Research and Innovation (through the Centre of Excellence, REPHRAIN).

1. *Freedom House, Freedom in the World 2020: A leaderless struggle for democracy* (Tech. Rep. 2020), at https://freedomhouse.org/sites/default/files/2020-02/FIW_2020_REPORT_BOOKLET_Final.pdf, accessed 7 March 2023; Anna Lührmann, Seraphine F. Maerz, Sandra Grahn, Nazifa Alizada, Lisa Gastaldi, Sebastian Hellmeier, Garry Hindle and Staffan I. Lindberg, *Autocratization Surges – Resistance Grows. Democracy Report 2020* (Tech. Rep.) (Gothenburg: V-Dem Institute, 2020).

2. Economist Intelligence Unit, *Democracy Index 2019: A Year of Democratic Setbacks and Popular Protest* (Tech. Rep.) (The Economist, 2020).

3. Silvio Waisbord, "The Elective Affinity between Post-truth Communication and Populist Politics", *Communication Research and Practice* 4 (2018): 17-34 [doi: 10.1080/22041451.2018.1428928].

4. Taner Edis, "A Revolt against Expertise: Pseudoscience, Right-wing Populism, and Post-truth Politics", *Disputatio* 9 (2020).

5. Stephan Lewandowsky, "Wilful Construction of Ignorance: A Tale of Two Ontologies", in

causal contribution to populist voting in Italy,[6] to triggering ethnic hate crimes in Germany and Russia,[7] and it has been shown to set political agendas.[8] Misinformation is particularly problematic because it has longer-term consequences: false information lingers in memory even if people acknowledge, believe, and try to adhere to a correction.[9] Lingering misinformation, in turn, can be politically consequential, for example when corrections of politicians' falsehoods do not affect people's feeling about the politician or their voting intention.[10]

Misinformation, however, does not exist in a vacuum: misinformation is disseminated (sometimes intentionally, in which case it is best referred to as *dis*-information) and it is consumed and shared by the public.[11] To understand the effects of misinformation on democracy thus requires an understanding of the processes of dissemination and consumption. In this chapter, I focus on two important drivers of misinformation spread and how they interact with human

Ralph Hertwig and Christoph Engel (eds.), *Deliberate Ignorance: Choosing Not To Know* (Cambridge, MA: MIT Press, 2020), 101-117; Silvio Waisbord, "Why Populism is Troubling for Democratic Communication", *Communication, Culture and Critique* 11 (2018): 21-34 [doi: 10.1093/ccc/tcx005].

6. Michele Cantarella, Nicolò Fraccaroli and Roberto Geno Volpe, "Does Fake News Affect Voting Behaviour?", *SSRN Electronic Journal* (2020) [doi: 10.2139/ssrn.3629666].

7. Leonardo Bursztyn, Georgy Egorov, Ruben Enikolopov & Maria Petrova, *Social Media and Xenophobia: Evidence from Russia* (National Bureau of Economic Research, 2019); Karsten Müller and Carlo Schwarz, "Fanning the Flames of Hate: Social Media and Hate Crime", *SSRN Electronic Journal* (2019) [doi: 10.2139/ssrn.3082972].

8. Chris J. Vargo, Lei Guo, and Michelle A. Amazeen, "The Agenda-setting Power of Fake News: A Big Data Analysis of the Online Media Landscape from 2014 to 2016", *New Media & Society* 20/5 (2018), 2028-2049 [doi:10.1177/1461444817712086].

9. Ullrich K. H. Ecker, Briony Swire and Stephan Lewandowsky, "Correcting Misinformation: A Challenge for Education and Cognitive Science", in David N. Rapp and Jason Braasch (eds.), *Processing Inaccurate Information: Theoretical and Applied Perspectives from Cognitive Science and the Educational Sciences* (Cambridge, MA: MIT Press, 2014), 13-38; Stephan Lewandowsky, Ullrich K. H. Ecker, Colleen M. Seifert, Norbert Schwarz, and John Cook, "Misinformation and its Correction: Continued Influence and Successful Debiasing", *Psychological Science in the Public Interest* 13 (2012): 106-131 [doi: 10.1177/1529100612451018]; Stephan Lewandowsky, Ullrich K. H. Ecker, and John Cook, "Beyond Misinformation: Understanding and Coping with the 'Post-truth' Era", *Journal of Applied Research in Memory and Cognition* 6 (2017): 353-369 [doi: 10.1016/j.jarmac.2017.07.008].

10. Briony Swire, Adam J. Berinsky, Sthephan Lewandowsky, and Ullrich K. H. Ecker, "Processing Political Misinformation: Comprehending the Trump Phenomenon", *Royal Society Open Science* 4/3 (2017): 160802 [doi: 10.1098/rsos.160802]; Briony Swire-Thompson, Ullrich K. H. Ecker, Stephan Lewandowsky, and Adam J. Berinsky, "They Might Be a Liar but They're My Liar: Source Evaluation and the Prevalence of Misinformation", *Political Psychology* 41/1 (2020): 21-34 [doi: 10.1111/pops.12586].

11. Stephan Lewandowsky, "Fake News and Participatory Propaganda", in Rüdiger F. Pohl (ed.), *Cognitive Illusions: Intriguing Phenomena in Thinking, Judgment, and Memory* (London: Routledge, 2022), 324-340 [doi: 10.4324/9781003154730-23].

cognition: First, I examine the role of demagogues; that is, political leaders who rely on false claims and promises, and emotive exploitation of people's prejudices, in order to gain power. How do they exercise power, and why do people accept demagoguery? Second, I examine the role of social media, focusing in particular on the role of algorithms. How does social media capture human attention? Why do people participate in sharing of misinformation?

21st century demagogues: divide and divert using social media

Demagoguery has been a looming threat to democracy since ancient Greece. Demagogues exploit a fundamental weakness of democracy: because ultimate power is held by voters, ruthless politicians can appeal to voters not through reason, as is the idealized democratic norm, but through emotion and simplistic appeals to the "people" and against a presumed "elite".[12] Some of the most horrific events in human history –such as the Nazi genocide– have resulted from the mobilization of large segments of the public by demagogues in pursuit of violent conflict.[13] The social-media technology available in the 21st century has given demagogues powerful new tools with which to reach the public and set the agenda on a scale never seen before.

This can be a positive development: Leaders can explain their actions and policy proposals, and they can engage in meaningful ways with the public. During the early stages of the pandemic, many political leaders used social media to keep the public informed and up-to-date about COVID-19-related developments and restrictions.[14] However, leaders have also used social media for less benevolent purposes. For example, former US president Donald Trump has used Twitter to spread disinformation and to divide American society by insulting nearly 500 people, places, and things within two years of taking office.[15]

12. Marnie Lawler McDonough, "The Evolution of Demagoguery: An Updated Understanding of Demagogic Rhetoric as Interactive and Ongoing", *Communication Quarterly* 66/2 (2018): 138-156 [doi: 10.1080/01463373.2018.1438486].

13. Michael Bang Petersen, "The Evolutionary Psychology of Mass Mobilization: How Disinformation and Demagogues Coordinate Rather Than Manipulate", *Current Opinion in Psychology* 35 (2020) [doi: 10.1016/j.copsyc.2020.02.003].

14. Michael Haman, "The Use of Twitter by State Leaders and its Impact on the Public during the COVID-19 Pandemic", *Heliyon* 6 (2020), e05540 [doi: 10.1016/j.heliyon.2020.e05540]; M. Reveilhac, "The Deployment of Social Media by Political Authorities and Health Experts to Enhance Public Information during the COVID-19 Pandemic", *SSM - Population Health* 19 (2022), 101165 [doi: 10.1016/j.ssmph.2022.101165].

15. Jasmine C. Lee and Kevin Quealy, "The 598 People, Places and Things Donald Trump Has

Donald Trump has also demonstrably used Twitter to affect political agenda setting and to divert public attention from issues that were politically harmful to him. To illustrate, when the cast of the "Hamilton" Broadway play pleaded for a diverse America at the end of a performance attended by Vice-President elect Pence in late 2016, Donald Trump tweeted vigorously and critically and demanded an apology from the actors. The Twitter activity coincided with the publication of a $25 million settlement of a lawsuit involving the defunct "Trump University", which included a $1 million penalty payment to the State of New York.[16] The politically damaging news about the settlement appeared to be largely drowned out by the Hamilton controversy. A Google Trends analysis revealed that the court settlement was of considerably less interest to the public than the Twitter event arising from Hamilton.[17]

The Hamilton affair is merely anecdotal. Systematic empirical evidence that Donald Trump used social media to divert attention from politically-inconvenient issues was provided by Lewandowsky, Jetter, and Ecker.[18] They explicitly tested the hypothesis that President Trump's tweets diverted media attention away from news that can be assumed to be politically harmful to him. Politically-harmful news was operationalized as coverage in the main media (New York Times [NYT] and ABC News) of the Mueller investigation into potential collusion between the Trump campaign with Russia during the 2016 election. Lewandowsky and colleagues hypothesized that the more the ABC and NYT reported on the Mueller investigation, the more Trump's tweets would mention keywords such as "jobs" or "China" that represented his political strengths. If that diversion to different issues were successful, then subsequent coverage of the Mueller investigation by ABC and NYT should be reduced. This patterns is precisely what was found by Lewandowsky and colleagues. Each additional ABC headline relating to the Mueller investigation was associated with 0.2 additional mentions of one of the keywords in Trump's tweets. In turn, each ad-

Insulted on Twitter: a Complete List" (2019), at https://www.nytimes.com/interactive/2016/01/28/upshot/donald-trump-twitter-insults.html, accessed 15 April 2023.

16. May Bulman, "Donald Trump 'Using Hamilton Controversy to Distract from $25m Fraud Settlement and Other Scandals'", at http://www.independent.co.uk/news/world/americas/donald-trump-hamilton-settlement-university-fraud-mike-pence-scandals-a7429316.html, accessed 4 April 2023.

17. Lewandowsky et. al., "Beyond Misinformation".

18. Stephan Lewandowsky, Michael Jetter, and Ullrich K. H. Ecker, "Using the President's Tweets to Understand Political Diversion in the Age of Social Media", *Nature Communications* 11 (2020): 5764 [doi: 10.1038/s41467-020-19644-6].

ditional mention of one of the keywords in a Trump tweet was associated with 0.4 fewer occurrences of the Mueller investigation in the following day's New York Times. This pattern did not emerge with placebo topics that presented no threat to the president, for example non-political issues such as football or gardening or other political topics such as Brexit.

Lewandowsky and colleagues thus presented empirical evidence in support of the hypothesis that President Trump's used Twitter to systematically divert attention away from a topic that is potentially harmful to him, which in turn appeared to suppress media coverage of that topic. It remains unclear whether Trump engaged in this behaviour intentionally or whether it reflected an intuition. It is clear, however, that Donald Trump was able to set the political agenda, contrary to the conventional wisdom that it is primarily the media, not politicians, that determine the agenda of public discourse in liberal democracies.[19]

Beyond affecting media coverage, Trump's misleading or false tweets, also tended to trigger supportive information cascades on social media propagated by his millions of followers. During the 2016 election campaign, Trump's tweets on average elicited three times as many retweets and likes as those by his opponent, Hillary Clinton.[20] Trump's ability to leverage social media in his support culminated in the violent insurrection on 6 January 2021. The armed assault on the Capitol was motivated by Trump's fabricated claim that his reelection had been "stolen" from him. Although this claim was shown to be false by virtually all mainstream media in the US and thoroughly dismissed by the courts, it was able to gather pace on social media.[21] In the 5 months following the 6 January insurrection, across 23 surveys, an average of 78% of Trump voters denied that President Biden was the legitimate winner of the election.[22]

19. Gary King, Benjamin Schneer and Ariel White, "How the News Media Activate Public Expression and Influence National Agendas", *Science* 358 (2017): 776-780 [doi: 10.1126/science.aao1100]; Maxwell McCombs, "A Look at Agenda-setting: Past, Present and Future", *Journalism Studies* 6 (2005): 543-557 [doi: 10.1080/14616700500250438].

20. Jayeon Lee and Weiai Xu, "The More Attacks, the More Retweets: Trump's and Clinton's Agenda Setting on Twitter", *Public Relations Review* 44 (2018): 201-213 [doi: 10.1016/j.pubrev.2017.10.002].

21. Rita Kirk and Dan Schill, "Sophisticated Hate Stratagems: Unpacking the Era of Distrust", *American Behavioral Scientist* (2021) [doi: 10.1177/00027642211005002]; Luke Munn, "More than a Mob: Parler as Preparatory Media for the U.S. Capitol Storming", *First Monday* 26/3 (2021) [doi: 10.5210/fm.v26i3.11574].

22. Gary C. Jacobson, "Driven to Extremes: Donald Trump's Extraordinary Impact on the 2020 Elections", *Presidential Studies Quarterly* 51 (2021): 492-521 [doi: 10.1111/psq.12724].

Donald Trump is not the only politician to use social media to his advantage. A recent analysis of millions of tweets by members of both houses of the US Congress revealed a striking political asymmetry.[23] Republicans were found to share links to untrustworthy websites on Twitter 9 times more often than Democrats between January 2016 and March 2022. Superimposed on that absolute difference is a temporal trend of increasingly greater divergence between Republicans and Democrats. Whereas information quality shared by Democrats has remained stable (and very high), the proportion of untrustworthy sites shared by Republicans doubled between 2016-2018 and 2020-2022. This behavior of the political leadership may help explain why several big-data analyses of the American public's news diets have found Republicans (especially extreme conservatives) to be far more exposed to misinformation and far more willing to share false information on social media.[24] The behaviour of the political leadership can contribute to the observed asymmetry among the public in at least two ways: first, by directly providing misinformation to Republican partisans and, second, by legitimizing the sharing of untrustworthy information more generally.[25]

Politicians clearly exercise considerable power through social media. But politicians' social media behavior constitutes only one side of the equation: supplying diversion, divisive information, and disinformation can only be effective and politically useful if there are consumers who are willing to accept and, ideally, share the information. Why, then, do people willfully consume disinformation? Or are people simply being duped, and they are passive victims of politicians' misinformation? It turns out that there is evidence for both of those processes.

Consider first partisans' willingness to accept information as true that is unequivocally and visibly false. Within 24 hours of Donald Trump taking office, White House officials falsely claimed that more people attended Trump's

23. Jana Lasser, Segun Taofeek Aroyehun, Almog Simchon, Fabio Carrella, David Garcia and Stephan Lewandowsky, "Social Media Sharing of Low Quality News Sources by Political Elites", *PNAS Nexus* 1 (2022), pgac186 [doi: 10.1093/pnasnexus/pgac186].

24. Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson and David Lazer, "Fake News on Twitter during the 2016 U.S. Presidential Election", *Science* 363 (2019): 374-378 [doi: 10.1126/science.aau2706]; Andrew M. Guess, Jonathan Nagler and Joshua Tucker, "Less Than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook", *Science Advances* 5 (2019), eaau4586 [doi: 10.1126/sciadv.aau4586]; Andrew M. Guess, Brendan Nyhan and Jason Reifler, "Exposure to Untrustworthy Websites in the 2016 U.S. election", *Nature Human Behavior* 4 (2020): 472-480 [doi: 10.1038/s41562-020-0833-x].

25. Lasser, Aroyehun et. al. "Social Media Sharing of Low Quality News Sources by Political Elites".

inauguration than any other previously. This claim was readily falsifiable by a range of evidence, including public transport data (Metro ridership) and photographs of the crowds on the National Mall during the inauguration. The false claim by the White House almost immediately became a prominent and polarizing issue. Schaffner and Luks conducted a study within two days of the controversy erupting that explored the impact of the administration's claim.[26] Participants were presented with two side-by-side photographs of the inaugurations of Barack Obama and Donald Trump, and had to identify the photo with more people in it. The difference in crowd size was so unambiguous that it was virtually impossible for good-faith responses to be incorrect. Indeed, only 3% of non-voters chose the incorrect picture. Among Trump voters, by contrast, this proportion was 15%. Given that the photos were unequivocal and the task trivial, Schaffner and Luks interpreted these results as revealing "expressive responding" of partisans. Instead of genuinely believing a misconception, partisans effectively chose to set aside unambiguous perceptual evidence and instead promulgated a politically-concordant falsehood –even if in this instance the "audience" was only an unknown experimenter. The proportion of people who were willing to do this meshes well with the proportion of people who have been observed to knowingly share false headlines.[27]

However, not all consumers of disinformation are willing participants in propaganda. Many people are exposed to disinformation and misinformation without actively seeking it out, but because content-curation algorithms are forcing the content on users.

Social media: attention and algorithms

Journalists have long known that "if it bleeds, it leads." People seek out news that is predominantly negative[28] or awe inspiring.[29] Online, users tend to share

26. Brian F. Schaffner and Samantha Luks, "Misinformation or Expressive Responding? What an Inauguration Crowd Can Tell Us about the Source of Political Misinformation in Surveys", *Public Opinion Quarterly* 82/1 (2018): 135-147 [doi: 10.1093/poq/nfx042].

27. Gordon Pennycook, Ziv Epstein, Mohsen Mosleh, Antonio A. Arechar, Dean Eckles and David G. Rand, , "Shifting Attention to Accuracy Can Reduce Misinformation Online", *Nature* 592 (2021): 590-595 [doi: 10.1038/s41586-021-03344-2].

28. Stuart Soroka, Patrick Fournier, and Lilach Nir, "Cross-national Evidence of a Negativity Bias in Psychophysiological Reactions to News", *Proceedings of the National Academy of Sciences of the United States of America* 116 (2019): 18888-18892 [doi: 10.1073/pnas.1908369116].

29. Jonah Berger and Katherine L. Milkman, "What Makes Online Content Viral?", *Journal of Marketing Research* 49 (2012): 192-205 [doi: 10.1509/jmr.10.0353].

messages that are couched in moral-emotional language.[30] By their very nature, digital media seem to amplify the role of emotion: the degree of moral outrage that is elicited by content online is considerably greater than for encounters in person or content in conventional media.[31]

This attentional bias is leveraged by social media platforms which exist only because our attention online has been commodified.[32] As a rule of thumb, when you use a "free" product online, *you* are the product. The more time users spend watching YouTube videos or checking their Facebook newsfeeds, the more advertising revenue is generated for the platforms. For the platforms, dwell time is the one and only currency that matters because it directly translates into advertising revenue. Platforms will seek to enhance dwell time by any means possible short of actually paying people to hang around.

It is unsurprising, therefore, that "fake news" and misinformation has become so prevalent online because false content –which by definition is freed from factual constraints– can exploit the human propensity to consume emotive and outrage-provoking content: misinformation on Facebook during the 2016 US presidential campaign was particularly likely to provoke voter outrage[33] and fake news titles have been found to be substantially more negative in tone, and display more negative emotions such as disgust and anger, than real news titles.[34] The platform's algorithms are trained to be sensitive to negative emotions: a former Facebook employee and whistleblower, Frances Haugen, revealed to the public in 2021 how the newsfeed curation algorithm favoured material that made people angry over material that elicited a "like" by a factor of 5.[35] Facebook thus "systematically amped up some of the worst of its platform, making it more prominent in users' feeds and spreading it to a much wider audience".[36]

30. William J. Brady, Julian A. Wills, John T. Jost, Joshua A. Tucker, Jay J. Van Bavel, "Emotion Shapes the Diffusion of Moralized Content in Social Networks", *Proceedings of the National Academy of Sciences of the United States of America* 114 (2017): 7313-7318 [doi: 10.1073/pnas.1618923114].

31. Molly J. Crockett, "Moral Outrage in the Digital Age", *Nature Human Behaviour* 1 (2017): 769-771 [doi: 10.1038/s41562-017-0213-3].

32. Tim Wu, *The Attention Merchants* (London: Atlantic Books, 2017).

33. Vian Bakir and Andrew McStay, "Fake News and the Economy of Emotions", *Digital Journalism* 6 (2018): 154-175 [doi: 10.1080/21670811.2017.1345645].

34. Jeannette Paschen, "Investigating the Emotional Appeal of Fake News Using Artificial Intelligence and Human Contributions", *Journal of Product & Brand Management* 29/2 (2020): 223-233 [doi: 10.1108/jpbm-12-2018-2179].

35. Pekka Kallioniemi, "Facebook's Dark Pattern Design, Public Relations and Internal Work Culture", *Journal of Digital Media & Interaction* 5 (2022): 38-54 [doi: 10.34624/JDMI.V5I12. 28378].

36. Jeremy B. Merrill and Will Oremus, *Five Points for Anger, One for a "Like": How Facebook's*

On YouTube, the recommender system is particularly important because by default, YouTube continues to play videos and present them to the user without an explicit request. There is now evidence suggesting that YouTube algorithms have actively contributed to the rise and consolidation of right-wing extremists in the US[37] and Germany.[38] A recent systematic review revealed that 14 out of 23 eligible studies implicated the YouTube recommender system in facilitating access to problematic content (e.g., extremist material), 7 produced mixed results, and only two did not implicate the recommender system.[39]

An over-arching difficulty in understanding algorithms and their effect on democracy is the lack of transparency and accountability. The delegation of choice from humans to algorithms under conditions of opacity and complexity raises questions about responsibility and accountability.[40] Who is responsible for a misinformation cascade? The human being who triggers it or the algorithm that is amplifying it in pursuit of user dwell time? This question is difficult to resolve unambiguously because the manufacturer or designer of the algorithm cannot predict its future behaviour in all circumstances. A designer may choose to weight anger during preceding engagements 5 times more than "likes" but that does not mean the designer knowingly facilitated misinformation cascades. It is therefore easy to claim that designers cannot be held morally or legally liable for the behaviour of their algorithms. This diffuse link between designers' intention and the actual behaviour of an algorithm creates a "responsibility gap" that is difficult to bridge with traditional notions of responsibility.[41]

The responsibility gap is amplified by the lack of transparency: Algorithms make decisions without public oversight, regulation, or a widespread under-

Formula Fostered Rage and Misinformation (2021), at https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/, accessed 4 April 2023.

37. Jonas Kaiser and Adrian Rauchfleisch, "Unite the Right? How YouTube's Recommendation Algorithm Connects the U.S. Far-right" (2018), at https://medium.com/@MediaManipulation/unite-the-right-how-youtubes-recommendation-algorithm-connects-the-u-s-far-right-9f1387ccfabd, accessed 4 April 2023 .

38. Adrian Rauchfleisch and Jonas Kaiser, "YouTubes Algorithmen sorgen dafür, dass AfD-Fans unter sich bleiben" (2017), at https://www.vice.com/de/article/59d98n/youtubes-algorithmen-sorgen-dafur-dass-afd-fans-unter-sich-bleiben, accessed 4 April 2023.

39. Muhsin Yesilada and Stephan Lewandowsky, "Systematic Review: YouTube Recommendations and Problematic Content", *Internet Policy Review* 11 (2022) [doi: 10.14763/2022.1.1652].

40. Nicholas Diakopoulos, "Algorithmic Accountability", *Digital Journalism* 3/3 (2015): 398-415 [doi: 10.1080/21670811.2014.976411].

41. Andreas Matthias, "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata", *Ethics and Information Technology* 6 (2004): 175-183 [doi: 10.1007/s10676-004-3422-1].

standing of the mechanisms underlying the resulting decisions. Facebook's reliance on anger over likes would never have become public knowledge without a whistleblower. At present, algorithms are considered proprietary trade secrets and operate as black boxes where neither individual users nor society in general know why information in search engines or social media feeds is ordered in a particular way.[42] The problem is compounded by the inherent opacity and complexity of machine-learning algorithms,[43] such that even creators or owners of algorithms may not be fully aware of their functioning. For example, YouTube's recommender system learns approximately one billion parameters and is trained on hundreds of billions of cases.[44] Predicting the response of the system in any particular situation is thus far beyond human capacity.

At present, knowledge about an algorithm can only be obtained by "reverse engineering";[45] that is, by seeking to infer an algorithm's design based upon its observable behaviour. Reverse engineering can range from the relatively simple (e.g., examining which words are excluded from auto-correct on the iPhone)[46] to the highly complex (e.g., an analysis of how political ads are delivered on Facebook).[47] Reverse engineering has uncovered several problematic aspects of algorithms, such as discriminatory advertising practices and stereotypical representations of Black Americans in Google Search,[48] and in the autocomplete suggestions that Google provides when entering search terms.[49]

42. Frank Pasquale, *The Black Box Society* (Cambridge, MA: Harvard University Press, 2015).

43. Pau B. de Laat, "Algorithmic Decision-making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?", *Philosophy & Technology* 31 (2018): 525-541 [doi: 10.1007/s13347-017-0293-z].

44. Paul Covington, Jay Adams and Emre Sargin, "Deep Neural Networks for YouTube Recommendations", in *Proceedings of the 10th ACM conference on recommender systems - RecSys '16* (2016): 191-198 [doi: 10.1145/2959100.2959190].

45. Diakopoulos, "Algorithmic Accountability".

46. Michael Keller, "The Apple 'Kill List': What Your iPhone Doesn't Want you to Type" (2013), at https://www.thedailybeast.com/the-apple-kill-list-what-your-iphone-doesnt-want-you-to-type, accessed 4 April 2023.

47. Muhammad Ali, Piotr Sapiezynski, Aleksandra Korolova, Alan Mislove and Aaron Rieke, *Ad Delivery Algorithms: The Hidden Arbiters of Political Messaging* (Tech. Rep. 2019), at https://arxiv.org/pdf/1912.04255.pdf, accessed 4 April 2023.

48. Latanya Sweeney, "Discrimination in Online Ad Delivery", *Queue* 11 (2013): 1-19 [doi: 10.1145/2460276.2460278]; Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: New York University Press, 2018).

49. Paul Baker and Amanda Potts, "'Why Do White People Have Thin Lips?' Google and the Perpetuation of Stereotypes via Auto-complete Search Forms", *Critical Discourse Studies* 10/2 (2013): 187-204 [doi: 10.1080/17405904.2012.744320].

In summary, much of the content consumed by the public is foisted upon them by opaque algorithms that are not subject to public scrutiny or accountability. What little we know about algorithms was obtained through painstaking reverse engineering or resulted from whistleblowing by former employees. That limited knowledge, however, should give rise to considerable concern and should stimulate action towards greater accountability. One step in this direction is the European Union's recent Digital Services Act, which came into force in October 2022, and which, among many other measures, requires large platforms to make available data to independent researchers to permit assessment of the risks and possible harms brought about by the platform's systems and to examine the accuracy, functioning, and testing of algorithms.[50]

Concluding comments

There is little doubt that democracy worldwide is under threat. Even countries whose democracies had appeared stable for decades if not centuries, such as the United States, have experienced recent episodes of political upheaval with a distinctly undemocratic character. There are many reasons for these developments that are difficult to disentangle. Here I identified two contributing factors: first, the ability of political leaders to exploit social media to divert attention from politically-inconvenient events and to spread disinformation. Second, the pernicious interaction between human attention and content-curation algorithms employed by the platforms to maximize user engagement.

Identifying solutions to these trends is beyond the scope of this chapter, although it is not impossible to envisage an Internet that is compatible with democracy rather than at least partially antagonistic to it. At the scholarly level, Lewandowsky and Pomerantsev provide a sketch of what that Internet for democracy might look like and how it might empower users rather than exploit them through a web of opaque algorithms.[51] At the policy level, the EU's recent Digital Services Act provides a pointer towards the regulation necessary to rein in the toxic power currently held by democratically unaccountable platforms.

50. Brandie Nonnecke and Camille Carlton, "EU and US Legislation Seek to Open Up Digital Platform Data", *Science* 375 (2022): 610-612 [doi: 10.1126/science.abl8537].

51. Stephan Lewandowsky and Peter Pomerantsev, "Technology and Democracy: A Paradox Wrapped in a Contradiction Inside an Irony", *Memory, Mind & Media* 1 (2022) [doi: 10.1017/mem.2021.7].